The University Of Sheffield.

Investigating online user groups for digital cultural heritage: a study of National Museums Liverpool

By:

David Walsh

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Sheffield

Faculty of Social Sciences

Information School

June 2024

Abstract

This study seeks to fill a gap in the field of virtual museum visitors by providing a more comprehensive understanding of different user groups and exploring the creation of a tool for future researchers and museum professionals to quickly gather and analyse user groups in their own studies. This study aimed to identify the information-seeking characteristics of different user groups for museum websites through the investigation of seven research questions. The first research question focused on the identification of the main user group categories present in the Digital Cultural Heritage literature. The literature review revealed a lack of a comprehensive approach to the identification and use of a group label. The second research question explored the similarities, shared characteristics, and differences of the identified groups, revealing that attempting to define a label through literature via its characteristics proved difficult. The third and fourth research questions examined how user groups are defined based on survey responses and how they compare to those in the literature, finding that the "general public" and "non-professional" visitors make up nearly 80% of all visitors. The fifth research question compared user groups identified in the transaction logs to those in the literature and survey results, suggesting that focusing on labelled user groups may not be the most effective approach. The sixth research question identified user groups from survey responses using the PAM clustering method, resulting in a set of more distinguishable user groups. The final research question focused on how comprehensive user group definitions can be generated from the cluster results, producing more detailed and actionable user group definitions.

Acknowledgements

I would like to express my deepest gratitude to my supervisory team, Paul Clough, Mark Hall, Frank Hopfgartner and Jonathan Foster for the guidance I have received during this process. I am proud of what I have accomplished, and a great deal of that is due to the four of you.

I would also like to thank my family and friends for being there for me through this amazing journey. In particular, I would like to thank my parents, partner and kids, who all just believed in me, kept me grounded during the highs of this process and supported me during the lows, but also to love and encourage me through the process. Words cannot express how grateful I am to have had such a supportive family. Collette and Mark, thank you for being my best friends and stopping me from taking things too seriously.

Finally, a huge thank you to National Museums Liverpool for allowing me to perform this study.

Declaration

I, the author, confirm that the Thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means (www.sheffield.ac.uk/ssid/unfair-means). This work has not been previously been presented for an award at this, or any other, university.

A list of publications relating to this work can be found in Chapter 1.5.

Table of Contents

List of Tables

List of Figures

Chapter 1

INTRODUCTION

1.1 Research Context

According to Sullivan (2016), digital cultural heritage (DCH) is the use of technology in the service of preserving cultural or natural heritage. However, with the widespread availability and accessibility of technology, identifying the characteristics of those who interact with DCH services can only serve to enhance the overall user interaction experience and the reputation of the DCH provider.

Cultural heritage (CH) institutions are attempting to reach the widest audience possible. The advancement of internet technology has removed the barriers of attracting visitors to the physical buildings to observe the artefacts/artworks. Many institutions have, or are in the process of, undertaking large digitisation projects and hosting most, if not all, of their collections in online catalogues. The goal of this work is to attract more visitors to the websites and physical museums or galleries.

However, the institutional websites are seeing large numbers of website visitors arriving at the site and leaving instantly (bounce rate); which is opposite of the intended goal of engaging visitors. Therefore, there is clearly a need to develop an understanding of website users and their contexts for visiting the website in order for the museum development team to provide content and user interface systems that are better suited for and support users.

1.2 Research Motivation

The classical roles of the museum are collecting, preserving, research and basic interpretation. However, museums, especially if publicly funded, are seen as being obliged to give things back to society to "justify their existence", (Fleming, 2007), emphasising the significance of the relationship between the users and the museum's collections. To assist in the enhancement of these relationships, the introduction of technology, in the form of an accessible visual medium, usually a website, plays a significant role. Filippini-Fantoni and Bowen (2007); Fisher and Twiss-Garrity (2007); Durbin (2008) established consensus that an online space is very important in providing the pre- and post-museum visit experience, but too little is known about the users of the online sites, why the users go to these sites and how museums could adapt the sites to their needs (Farber and Radensky, 2008; Roberto, 2008; Salgado, 2008; Samis, 2008).

Currently, almost all the published literature focuses on DCH website user groups such as experts, professionals, non-professionals (hobbyists), students, and academics. These labels were identified from the physical environment while also inferring that these groups are the majority of visitors to a DCH website. However, the large bounce rates being experienced by these websites may suggest otherwise.

Furthermore, those undertaking research work in DCH have based their work on only one or two groups and characteristics, such as motivation and task, in a single study. Although such studies may provide useful insights into the behaviour of these specific groups, they might not accurately capture the full range of user information-seeking behaviours and preferences and as such not be truly representative of the wider user base. Moreover, attempting to merge characteristics from various studies to create user groups could result in incomplete or inaccurate definitions that do not adequately capture the nuances of user behaviour. The UK Government service manual highlights this by stating, "always start by learning about the people who will use it. If you do not understand who they are or what they need from your service, you cannot build the right thing" (UK, 2017).

However, it was found that the participant users being used as representatives of the user groups for the studies consisted of people that were at hand and were not

truly representative of the end users. Rather, they seem to dismiss anyone outside of their focus group or label those outside users as the "general public". The use of the term "general public" as a catch-all user label is misleading. The Cambridge Dictionary (2020) defines the general public as "ordinary people, especially all the people who are not members of a particular organisation, or who do not have any special type of knowledge".

However, it is crucial to recognise that the general public, who likely constitute the majority of museum visitors, are not a homogeneous group. This diversity needs to be acknowledged and catered for in the design of digital environments. Ignoring the varied characteristics and needs of such a broad user base could jeopardise a DCH institution's relevance and survival in a technology-driven world.

Therefore, it is important for researchers and practitioners in DCH to be able to develop a more nuanced understanding of user information-seeking behaviours and preferences of those who engage with the digital environment that go beyond simplistic labels or incomplete definitions, in order to provide a meaningful experience.

1.3 Aim, Objectives and Research Questions

The aim, objectives and research questions grew out of the key gaps identified in the review of relevant literature described in Chapter 2.

1.3.1 Aim

This study seeks to categorise and describe the user groups that typically engage with DCH websites, focusing on their specific information needs, motivations, technical and knowledge levels, and data consumption preferences. The aim is to develop actionable user group descriptions that can support UX professionals and website developers, together with informing future research.

1.3.2 Objectives

The objectives of this research are designed to systematically develop a deep understanding of user groups within DCH contexts, as follows:

OB1: To systematically review and categorise existing user groups within the DCH literature, identifying gaps and inconsistencies in current user group definitions.

OB2: To conduct a comprehensive survey among DCH website users to identify characteristics and behaviours of distinct user groups.

OB3: To analyse user interaction data from museum website server logs to understand real-world user behaviours, further refining user group categories based on digital engagement patterns.

OB4: To apply cluster analysis techniques to survey data, with the aim of identifying data-driven user groups based on their characteristics and behaviours.

OB5: To develop a tool for DCH, UX, and web development professionals that generates comprehensive user group profiles from clustering data.

1.3.3 Research Questions

To guide the investigation, the following research questions have been formulated and will be systematically addressed throughout this thesis:

RQ1: What are the main user group categories present in the Digital Cultural Heritage literature?

RQ2: What are the identified groups' similarities, shared characteristics, and differences?

RQ3: How are user groups defined based on the survey responses?

RQ4: How do the aspects of the user groups defined in the survey compare to those in the literature?

RQ5: How do user groups identified in the transaction logs compare to those identified in the literature and survey results?

RQ6: What are the patterns and characteristics of user groups as determined by cluster analysis of survey data?

RQ7: How can comprehensive and usable user group profiles be generated from the clustering data?

To better understand those users who interact with a DCH environment, National Museums Liverpool (NML) allowed their website users to be the focus of this research. NML has a very large heterogeneous collection of artefacts across its museums and galleries, which could result in a single person being an expert in one area/collection presented on the website but having no knowledge of items/subject in another and so belonging to the general public user group in this less knowledgable part of the collection. If domain knowledge (or job in the organisation) is the distinguishing factor in this categorisation, then what are the propositions of the user groups accessing the NML site?

To garner a picture of a DCH audience, what could be considered a historical identification of user labels and characteristics was obtained through a literature review, while the primary data was gathered through an online survey via NML that captured individual responses to specific questions in addition to the gathering of unobtrusive data through keystroke actions. This allowed for the decomposition of the overall research questions into themes that were captured in a data set, which was then examined through various forms of analysis, e.g., log analysis and cluster analysis.

However, during the course of this part-time PhD study, which spanned several years, the direction of NML changed. The key contacts at the museum, initially engaged in the project, left the organisation. As a result, the direct collaboration with NML could not be sustained, and the contact was eventually lost. This impacted the final evaluation stage of the thesis, which could no longer be conducted directly with NML's staff and users. Consequently, the evaluation was adapted to include a broader spectrum of UX and web design/development experts across various fields, extending yet still including the DCH domain. Fortunately, some of the former staff from NML were also included in this revised evaluation process.

1.4 Organisation of Thesis

Following this introduction, Chapter 2 presents a literature review. Chapter 3 describes the methodology crafted to address the objectives and answer the research questions, and Chapters 4 to 8 report on the phases of work and their findings that form the body of research work for this project, with each Phase building from the last culminating in the proposed outcome in Chapter 8. Chapter 9 summarises the discussions from each Phase whilst also identifying the limitations, contributions and future research agenda and presenting the recommendations and conclusions.

1.5 Thesis Material

The research described in the following publications forms the basis of the material presented in the chapters of this thesis:

- **Chapter 4:** Walsh, D., Clough, P., and Foster, J. (2016). User categories for digital cultural heritage. In Clough, P., Goodale, P., Agosti, M., and Lawless, S., editors, *Proceedings of the First International Workshop on Accessing Cultural Heritage at Scale co-located with Joint Conference on Digital Libraries 2016 (JCDL 2016)*, volume 1611 of *CEUR Workshop Proceedings*. CEUR-WS.org

- **Chapter 5:** Walsh, D., Hall, M., Clough, P., and Foster, J. (2017). The ghost in the museum website: Investigating the general public's interactions with museum websites. In Kamps, J., Tsakonas, G., Manolopoulos, Y., Iliadis, L., and Karydis, I., editors, *Research and Advanced Technology for Digital Libraries*, pages 434–445, Cham. Springer, Springer International Publishing Walsh, D., Hall, Mark M.and Clough, P., and Foster, J. (2020). Characterising online museum users: a study of the national museums liverpool museum website. *International Journal on Digital Libraries*, 21(1):75–87

- **Chapter 6:** Walsh, D., Clough, P., Hall, M. M., Hopfgartner, F., Foster, J., and Kontonatsios, G. (2019). Analysis of transaction logs from national museums liverpool. In Doucet, A., Isaac, A., Golub, K., Aalberg, T., and

Jatowt, A., editors, *Digital Libraries for Open Knowledge*, pages 84–98, Cham. Springer International Publishing

- **Chapter 7:** Walsh, D., Clough, P., Hall, M. M., Hopfgartner, F., and Foster, J. (2021). Clustering and classifying users from the national museums liverpool website. In Berget, G., Hall, M. M., Brenn, D., and Kumpulainen, S., editors, *Linking Theory and Practice of Digital Libraries*, pages 202–214, Cham. Springer International Publishing

Any material re-use in this thesis is permitted by the publishers.

Springer: covering (Walsh et al., 2016, 2017, 2019, 2020, 2021) - "Authors have the right to reuse their article's Version of Record, in whole or in part, in their own thesis. Additionally, they may reproduce and make available their thesis, including Springer Nature content, as required by their awarding academic institution." [1]

ACM: covering (Walsh and Hall, 2015) - "Authors can include partial or complete papers of their own (and no fee is expected) in a dissertation as long as citations and DOI pointers to the Versions of Record in the ACM Digital Library are included." [2]

1.5.1 Additional Publications

Additional relevant publications by the author are:

- Walsh, D. and Hall, M. (2015). Just looking around: Supporting casual users initial encounters with digital cultural heritage. In Gade, M., Hall, M., Huurdeman, H., Kamps, J., Koolen, M., Skov, M., Toms, E., and Walsh, D., editors, *Proceedings of the First International Workshop on Supporting Complex Search Tasks co-located with the 37th European Conference on Information Retrieval (ECIR 2015)*, volume 1338 of *CEUR Workshop Proceedings*. CEUR-WS.org

- Speakman, R., Hall, M. M., and Walsh, D. (2018). User engagement with generous interfaces for digital cultural heritage. In *Digital Libraries for Open Knowledge: 22nd International Conference on Theory and Practice of Digital*

[1] `https://www.springer.com/gp/rights-permissions/obtaining-permissions/882`

[2] `https://authors.acm.org/author-resources/author-rights`

Libraries, TPDL 2018, Porto, Portugal, September 10–13, 2018, Proceedings 22, pages 186–191. Springer

- Hall, M. and Walsh, D. (2021). Exploring digital cultural heritage through browsing. In Golub, K. and Liu, Y.-H., editors, *Information Organization in Digital Humanities: Global Perspectives*, Digital Research in the Arts and Humanities, pages 261–284, United Kingdom. Routledge

Chapter 2

LITERATURE REVIEW

Since 2005, rapid technological advancements have led to the exponential growth of internet usage, becoming integral to daily work and social activities. According to Watters (2022), the World Wide Web (WWW) has revolutionised the world, with users spending significant time engaging in various online activities. With over 5.1 billion active internet users, 4.28 billion accessing the web via mobile, and 4.2 billion on social media, the presence and potential monetisation opportunities are immense. Therefore, businesses and organisations across all sectors must prioritise user experience to remain competitive. For commercial entities, embracing user experience ensures business continuity, with heavy investments in market research to attract and retain their demographic through personalised online experiences.

This is not the case for cultural heritage (CH) organisations such as museums, that may fall under a charity status, especially those designated as National museums and are directly funded by their respective central governments. CH organisations have traditionally emphasised in-person experiences, offering free entry to their significant collections. However, prolonged national lock-downs have severely affected this model and financial sustainability. As a result, adapting to the growing digital economy requires more than replicating physical experiences online; it necessitates understanding and engaging with digital users to create meaningful experiences that cater to their needs and interests.

This chapter examines the literature related to those who interact with Digital Cultural Heritage (DCH) through their information-seeking behaviour, how they engage with the digital collections/content, what support mechanisms are provided

and the subsequent categorisation of user groups based on current literature. The remainder of this chapter is organised into the following sections; Section 2.1 Cultural Heritage, how the physical environment engaged its visitors; Section 2.2 Digital Cultural Heritage (DCH), the inclusion of technology; Section 2.3 Users, who they are and how they have been studied and categories; Section 2.4 Information behaviour and Information seeking, the models that have been proposed; Section 2.5 Interfaces and Interactions, what interfaces are offered and also proposed as solutions; Section 2.6 How users are studied in the web industry; focussing on the UX and UCD procedures and techniques; Section 2.7 provides a summary of the literature review.

2.1 Cultural Heritage

According to UNESCO (2009) Institute for Statistics, "Cultural heritage includes artefacts, monuments, a group of buildings and sites, museums that have a diversity of values including symbolic, historical, artistic, aesthetic, ethnological or anthropological, scientific and social significance. It includes tangible heritage (movable, immobile and underwater), and intangible cultural heritage (ICH) embedded into cultural and natural heritage artefacts, sites or monuments. The definition excludes ICH related to other cultural domains such as festivals, celebrations etc."

Originally, museums served as private collections for the wealthy, with limited access for the "respectable" and "elite" classes. Early "public" museums mainly catered to the middle and upper classes. For example, when the British Museum opened in 1759, in order to control crowds and protect artefacts, visitors applied in writing to gain entry. Only in the 19th century were all social classes and age groups able to visit the museum, especially on public holidays (Goldgar, 2000).

Today, CH institutions, such as museums and galleries, regularly accommodate a wide array of visitors, ranging from those with minimal knowledge or interest to academic experts (Roppola, 2013). Upon arrival, non-expert visitors may simply aim to spend time in the museum or gallery (Hennes, 2002; Medić et al., 2015), while those with a bit more understanding may have specific goals, such as viewing particular items or collections.

To facilitate engagement, CH institutions have developed a number of successful

strategies for supporting visitors at all levels, especially those who arrive with little or no knowledge, by providing them with the following:

1. **Floor plans:**, assist visitors in obtaining an overview of the entire collection and navigating the physical building(s), improving their visit experience (Talbot et al., 1993). These are particularly beneficial for families with young children, providing them with the freedom to move swiftly through the collections.

2. **Fliers:** promote specific items or collections, benefiting all users in the physical setting and providing returning visitors with previews of upcoming attractions.

3. **Guide-books:** offer curated contextual information about collections and artefacts, enriching the visitor's learning experience. These are ideal for sharing with companions, especially for infrequent visitors or family groups (Hood, 2004). Additionally, guidebooks serve as post-visit learning aids or memorabilia.

4. **Audio guides:** offer an immersive alternative to guidebooks, allowing visitors to engage more deeply with artefacts. These also serve as accessibility aids for those with hearing or attention disabilities and cater to non-native language speakers with multilingual options (Martins, 2012; Straughan, 2019). According to Woodruff et al. (2001), non-personal audio guides eliminate competing stimuli, enhancing the visitor experience. Short audio clips, preferred by visitors, provide control and integrate seamlessly into conversations.

5. There are also **guided tours** for the visitors who like the opportunity to not only learn about the collection and certain items but also to have the ability to ask questions from the guide as they go (Straughan, 2019). Guided tours can also offer a behind the scenes and give an idea of what working in a museum is really like (Straughan, 2019).

Nevertheless CH institutions face ongoing challenges, ranging from funding cuts to issues related to artefact maintenance, restitution, and visitor engagement. Challenges such as funding cuts have led to reduced budgets and staff layoffs, impacting exhibit quality and object preservation, as well as visitor experiences (Newman and

Tourle, 2011; Hunt, 2019; Rex, 2020). Moreover, the current energy price crisis adds further uncertainty to funding availability, raising questions about how museums will cope (Bailey, 2022).

More recently, museums have received criticism for their narrow and biased historical focus, prompting a need for more inclusive exhibits. Changing visitor expectations emphasise interactive and engaging experiences, compelling museums to adapt digitally and offer innovative exhibits to remain relevant (Ahmed et al., 2020; Komarac et al., 2020). For instance, these exhibits may include showcasing previously-stored artefacts unseen due to space constraints (Baskas, 2013); or the preservation of artefacts that are naturally decaying, like Aleppo's great mosque (Fangi, 2019; News, 2013); as well as war-damaged artefacts such as The Temple of Bel at Palmyra levelled in 2015 (Shaheen, 2015).

However, all these issues were exasperated during the COVID-19 pandemic. In March 2020, approximately 90% of museums worldwide closed their physical sites (UNESCO, 2020), resulting in significant financial losses for institutions reliant on visitor revenue (Nemo, 2020). To mitigate these losses and maintain community engagement, many museums shifted focus to online content (King et al., 2021). The Art Fund report (Hadley, 2020) emphasises the importance of expanding online collections to reach audiences during lockdowns and beyond. Nevertheless, despite efforts to attract its core visitors back through digital services, many institutions have yet to regain pre-pandemic visitor levels three years later (Bailey, 2022). Consequently, understanding and engaging with online users is now paramount (Hadley, 2020).

2.2 Digital Cultural Heritage

Digital Cultural Heritage (DCH) involves preserving and presenting cultural objects through digitisation (Vane, 2020), including both tangible and intangible heritage. It offers detailed representations of objects and customs that may be vulnerable to destruction. Digitisation of artefacts also presents the opportunity for learning experiences of these objects and diverse cultures through various online media e.g., mobile apps, virtual reality, and social media, fostering accessibility, appreciation, and broader audience engagement for CH institutions.

2.2.1 Digitisation

Traditionally, CH institutions primarily exhibit artefacts in physical spaces, limiting the number on display due to space and curation constraints. As a result, many artefacts remain in storage rather than being showcased. To address this, institutions have undertaken large-scale digitisation efforts, making their collections accessible online (Nauta and van den Heuvel, 2015). However, websites typically provide general visitor information, such as opening hours and facility details, or an overview of what can be found in the physical buildings. According to Hughes (2004) with digitisation, institutions now offer online databases showcasing digitised artefacts, enhancing user experience through creative digital presentation methods.

The process of digitisation involves creating a digital version, usually a database of the institution's catalogue or index (Vane, 2020). The database entry is often a simple replication of the meta-data entries from the museum's paper documentation about the artefacts that describe the item and possibly an image/low-resolution photograph or two. Although, it could include hi-resolution photographs, 3D scans, videos, audio of the artefacts or pieces of art that try to fully digitise the item (Vane, 2020).

However, catalogues or indexes were initially designed to furnish museum staff and researchers (Eklund, 2011; Vane, 2020) with artefact metadata that included provenance, descriptive details, and organisational information, e.g., dates, condition, material, style, rights, acquisition, genre, etc. These records are typically generated upon object acquisition, but for items in long-term storage that predated structured curation processes, metadata may be minimal or incomplete. This is not a unique problem as different institutions may have different cataloguing practices, which again could lead to varying levels of detail in the metadata structure and content (Vane, 2020). Agirre et al. (2013) show that the metadata of records are often limited and incomplete. During this digitisation process, the data is sometimes cleaned and standardised in an attempt to resolve issues of individual archivists' unique documenting methods and to also possibly include any missing data.

Despite the existence of standards and guidelines for digitising collections, the standards and the methods of capture are continually evolving as technology evolves. The technological advancements over the last decade means that even existing dig-

ital collections are sometimes outdated because of the method or approach used at the time of digitisation, for example, paper documents would previously have been scanned in low-resolution black and white but would today be photographed in colour and in high-resolution (Lourdi and Nikolaidou, 2009).

2.2.2 Measuring Website Traffic and Bounce Rates

Establishing an online presence, through the creation of a website, is a common strategy for businesses and organisations to provide digital accessibility and engage with users through technology. Website traffic (Enge et al., 2012), akin to physical footfall, measures the number of visitors accessing a website over a specified period, indicating its effectiveness in attracting and engaging users. Measuring website traffic can be assessed through tools such as Google Analytics, and provides data on visitor numbers, page views, session duration, and more (Google, 2023). Other key metrics include unique visitors, page views, bounce rate, and session duration, all of which offering insights into website effectiveness and user experience.

However, according to, Sarraf (1999); Drivas et al. (2021); Hall and Walsh (2021); Rodríguez-Vázquez et al. (2023) there has always been an issue with DCH websites, both pre- and post-pandemic, regarding high bounce rates. A bounce rate, as defined by Google (2023), is calculated when a user opens a single page on the site and exits without making any further requests to the Analytics server during that session. This typically indicates a single-page visit with only one request to the server. For example, Corona (2021) evaluated the website traffic and bounce rates for the Louvre Museum pre-, during and post-pandemic closures. While a drop in website users over this period was observed, bounce rates remained constant at almost 55% which is far from ideal as it indicates that one out of two users abandons the website.

The implications of a high bounce rate may suggest that users are finding the information they need on the first page of content they visit, but this is unlikely. More commonly, a high bounce rate indicates that users are not finding the desired information or that the content quality is poor, prompting them to return to the search engine results (Drivas et al., 2021). Another possibility is that the pages lack an engaging user experience, failing to encourage further exploration of the site's content, which is typically a goal for all businesses and organisations (Drivas et al.,

2021; Sarraf, 1999; Voorbij, 2010).

Examining the bounce rate of an exemplar DCH museum website (NML) is explored in Chapters 5 and 6 where the data indicates the high bounce rate claims made by others and reveals that these problems could be linked to not understanding the audience(s) or the way the content should be presented to or experienced by the audience.

Additionally, while this thesis initially focused on exploring user experience, it became evident that there was a significant knowledge gap concerning the understanding of DCH website users within the field of study. Addressing this preliminary knowledge gap is essential before tackling the broader knowledge gap related to user experience in DCH websites. As a result, the thesis topic was pivoted to focus on studying the users of DCH museum websites.

2.3 Users

A user in the context of CH and DCH can refer to anyone who engages with cultural heritage resources, whether it be physical sites, artefacts, or digital content. This can include visitors to physical museums or heritage sites, researchers, students, educators, or anyone interested in exploring and learning about cultural heritage. In the DCH context, users access information via online databases, virtual exhibits, and other digital projects (Terras, 2008).

2.3.1 User Visitor Models in Cultural Heritage

One of the first studies undertaken regarding visitors to physical museums was conducted in 1884 at the Liverpool Museum. Four groups of users: students, observers, loungers, and German and Scandinavian immigrants were identified (Hein, 1998). Subsequent studies have investigated various museum visitors in various contexts, including their motivations (Falk, 2009), who they visited with (Dierking and Falk, 1994; Spellerberg et al., 2016), the role adopted during their visit (Falk, 2009), and their engagement with the museum (Templeton, 2011).

Falk's (2009) model based on five museum visitor identities, focused on users' needs and assumed all visits to cultural institutions are perceived in similar ways

regardless of ethnicity, age, class background or personal history and could be captured through the following labels:

1. **Explorers** – motivated by personal curiosity,
2. **Facilitators** – motivated by other people and their needs (i.e. a parent bringing a child),
3. **Experience-Seekers** – motivated by the desire to see and experience a place (i.e. tourists),
4. **Professional/Hobbyists** – motivated by specific knowledge-related goals (i.e. a scholar researching a specific topic),
5. **Rechargers** – motivated by a desire for a contemplative or restorative experience Falk (2009).

Falk (2009) used qualitative data to show that demographic characteristics and factors, such as time of year, were not enough to truly understand and predict visitor behaviour. Rather, the reasons for the visit /engagement were the main motivating factors. Motivation has been defined as an internal (intrinsic) state that arouses, directs and maintains behaviour (Woolfolk, 2001). A user's motivations to visit a collection are sometimes internal: enjoyment, curiosity, or personal needs and interests and require no pressure or incentive as the task is the reward in itself. Others are external (extrinsic): work, study, or assisting someone else to find or explore the collection. Both intrinsic and extrinsic motivations can produce the same amount of effort and eventual end result (Schaller and Goldman, 2004).

Fantoni et al. (2012) described five reasons for users engaging with the Indianapolis Museum of Art (IMA) website: (i) plan a visit to the museum; (ii) find specific information for research or professional purposes; (iii) find specific information for personal interest; (iv) engage in casual browsing without looking for something specific; and (v) make a transaction on the website. Ham (2013) found similar results but was able to extend the definitions to include captive and non-captive audiences when studying learners (see Table 2.1).

Table 2.1: User motivations and DCH

Fantoni et al.'s (2012) motivations	**Label**
Find specific information for personal interest	Curiosity
Find specific information for research or professional purposes	Work
Plan a visit to the museum	Plan visit
Engage in casual browsing without looking for something specific	Pleasure
Ham's (2013) motivations	
Find for educational (school) reasons,	Learn captive
Find for personal reasons	Learn non-captive

While some groups are sometimes identified and referred to in terms as abstract as novice or expert (Johnson, 2008), more commonly, user groups have their identity based on a profession (e.g., curator, librarian, researcher, teacher, or student). Alternative groups have been based on user interest or motivations (e.g., tourist, explorer, general user), or age group (e.g., adult, child) (see Table 2.2). These archetypal users are often described by nothing more than a title, leaving the reader to imagine and fill in the blanks, as the analysis in Chapter 4 will show. Other times they are described by very simplistic one or two sentence descriptions, and occasionally they are described using personas (Rasmussen and Petersen, 2012).

Table 2.2: Groups and their differentiators Walsh et al. (2020)

User Group	**Motivation**	**Domain/CH Knowledge**	**Task**	**Location**	**Frequency of visit**
General Public	Personal Pass time	Low to Intermediate	Pre-Visit	Merseyside Northwest	1st or Yearly
Non-Professionals	Personal	Intermediate	Known Item	Merseyside NorthWest England Outside UK	1st or Yearly
Students	Study	Low	Pre-Visit Museum Overview Collection Overview	Merseyside NorthWest Outside UK	1st
Academics	Study Work	High to Very High	Known Item Known Collection	Outside UK	Varies
Teachers	Personal Work	Intermediate to High	Pre-Visit	Merseyside	Yearly
Museum Staff	Work	High to Very High	Pre-Visit Known Collection	Merseyside	Monthly Daily

However, aspects of users that are less likely to change over time include their psychological attributes, such as cognitive style, which constitute their personal context. Modelling the user's context allows systems to personalise and adapt to the user's situation and will also attribute to variations in their information seeking behaviour (Michalakis and Caridakis, 2022). Consideration should therefore be taken into account of the users' individual differences, as well as their group attributes, and broader contexts, e.g. geographical, social, culture when looking at categorising the user (Taylor, 1982).

2.3.2 Example User Groups

This section explores three examples of user groups and their characteristics. The first example is Europeana, where archetypal users are created using personas to aid in platform development and evaluation. The second example is the Library of Congress National Digital Library Program, where users are classified into categories based on their motivations, domain knowledge, library system knowledge, focus, and time allocated to the task.

2.3.2.1 Europeana

Chowdhury (2015) examines the European Commission's challenge to Europeana, Europe's biggest aggregated collection of cultural assets, to provide a platform for European citizens to see heritage pieces. She emphasises the diversity of user cultures, ages, academic levels, and behaviours in order to demonstrate the fallacy of attempting to standardise a European user. However, the development and evaluation of Europeana required the formation of paradigmatic users. Personas have been used to characterise people, with the idea that "each persona represents numerous users and a set of personas reflects a spectrum of the intended user groups" Rasmussen and Petersen (2012). A persona includes features such as: names, jobs, feelings, interests, goals they wish to fulfil, turning the abstract and very short descriptive role category into a 'real' person Roussou et al. (2013).

Personas can provide detailed knowledge on the users' domain knowledge, as they are always written with a scenario in mind where the user performs a task on the system they are being used for. Technical knowledge is often indicated by

nothing more than a few words "Tech-Savy" or "some web usage". To date the development of personas has undergone 3 revisions Bergstrom (2010) that are derived from various forms of data: user studies, transaction log analyses and demographic data. The Europeana Persona Catalogue v3 consists of 5 (previously 7) personas developed around the broad dimensions of search literacy (e.g., IT knowledge, task knowledge and language), and information-seeking behaviour (e.g., search strategy and personality). An example persona is as follows: *"Jukka: PhD in music and professor at a university. Very confident about technology and always on the look-out for new stuff and new ways of communicating, on the computer as well as on his iPhone. Very confident about searching and finding useful and relevant results."*

2.3.2.2 Library of Congress NDL

In the context of the Library of Congress (LC) National Digital Library (NDL) Program, Marchionini et al. (2003) describes users and their demands. The authors go into great length about user categories, which are derived from users' motives (associated with specific information-seeking tasks), domain expertise, library system knowledge, task focus (e.g. amount of information needed), and time allotted to the work. As a result, the following categories emerged:

1. **LC staff**: high motivation, medium domain knowledge, high library system knowledge, high focus, and limited time allocations.

2. **Hobbyists**: e.g., genealogy, Civil War, railroads, other examples), high motivation, typically high domain knowledge, a range of library system knowledge, high focus, and high time allocations.

3. **Scholars**: e.g., historians, sociologists, anthropologists, authors, high motivation, high domain knowledge, high library system knowledge, high focus, and high time allocations.

4. **Professional researchers**: e.g., picture researchers, high motivation, medium domain knowledge, average to high library system knowledge, very high focus, and medium time allocations.

5. **Rummagers (browsers)**: e.g., PhD students looking for topics; scholars looking for new directions, topics, high motivation, medium domain knowl-

edge, range of library system knowledge, low focus, and medium to high time allocations.

6. **Object seekers**: e.g., some authors, CD-ROM/ multimedia developers, TV/video producers, and instructional materials developers, high motivation, range of domain knowledge, low library system knowledge, high focus, and low to medium time allocations.

7. **Surfers** e.g., those who are curious, those who bump into the NDL, etc., low motivation, low domain knowledge, low library system knowledge (but may be high computing system knowledge), low focus, and very low time allocations.

8. **Teachers K-16**: medium motivation, medium to high domain knowledge, low to medium library system knowledge, medium focus, and low time allocations.

9. **Students K-16**: low to medium motivation, low domain knowledge, low library system knowledge, low to medium focus, and low to medium time allocations.

Consideration of personal attributes (e.g. age, gender and cognition), domain expertise and technical expertise is given in producing categories, along with consideration of information seeking task.

2.3.2.3 Casual Leisure Users

Ardissono et al. (2012, p.74) state that visitors to web based collections "are often first- and short-time visitors to an unknown place". This means that, they are either in constant need of help in finding relevant information, or that providing them with appropriate information is challenging because, their interests and needs are unknown.

Walsh and Hall (2015, p.01) describe a casual user as someone "who has just stumbled across [the digital] collection in the same way that they would wander into the CH institution's physical space". These casual users have no specific goal or information need in mind when they come to a digital collection (Villa et al., 2013), but instead they could simply be wishing to: waste time, rest or look for peace and relaxation, experience something engaging, or have a desire for something interesting

or challenging (Elsweiler et al., 2011). However, according to Mayr et al. (2016, p.01) and Pousman et al. (2007, p.1145) casual users of digital collections do not possess the experts' knowledge of the collection and how to navigate or search them, nor do they have the expertise in analytic thinking or in reading visualisations. Therefore, in order to assist this group with their information seeking in digital collections, especially their initial interactions; it is essential to understand these users and the features that could assist them.

2.3.3 Users in Digital Cultural Heritage

Often the most common reason for visiting a museum website is to plan an upcoming visit to the physical museum (Marty, 2007b; Villaespesa et al., 2015; Booth, 1998). While it is tempting to apply physical visitor models (see Section 2.3.1) to the digital world because of this reason, there is no certainty that the two entirely overlap (Cunliffe et al., 2001), and online visitors should, therefore, be studied in their own context (Peacock and Brownbill, 2007; Marty, 2008).

Previous studies have attempted to identify the diversity of users of DCH which has resulted in a strategy that simplifies the virtually unlimited possibilities of user profiles by creating generic groups or categories of users - 'stereotypes' (Allen, 1997) for example; motivation (Fantoni et al., 2012); engagement (Templeton, 2011); role adopted during the visit(Spellerberg et al., 2016); user expertise and profession (Vilar and Šauperl, 2014; Pantano, 2011); information needs (Booth, 1998; Marchionini et al., 2003); or the chances of wandering in (Walsh et al., 2017).

Whilst Fantoni et al.'s (2012) motivations aligned with Falk's (2009) framework, it is clear that there could be differences between the physical and the digital spaces. Fantoni et al. (2012) states that many museums and other cultural organisations have adopted Falk's (2009) CH user classifications as a means of segmenting online audiences, even though these classifications were devised for the physical museum. Adopting user classifications from physical museums for a DCH website may have limitations due to the different motivations and behaviours of online users, leading to potentially narrow and outdated understandings of user needs and preferences. The classifications may also be too broad or limited to capture the diverse range of

user experiences, and not account for the dynamic nature of online user behaviour. Therefore, it is important to critically evaluate and adapt these classifications to avoid misinterpretations of user needs and ineffective design decisions.

To avoid making this mistake, Villaespesa et al. (2015) uses audience-centric research to characterise online visitors to the Tate UK website and audience segmentation methodologies to categorise website visits based primarily on the incentives driving users to the site e.g., personal interest research, student research, professional research, inspiration, enjoyment, art news, repeat visit planning, first visit planning, and organisational information. As the segmentation is mostly dependent on motive, the same user may be in one type of visit mode on one visit and another on another.

2.3.4 Moving between Roles and Expertise

Many of the studies mentioned above, attempt to identify users without considering how users develop their knowledge over time. Instead, they focus on capturing users at a single point in time. However, some prior studies have shown how people can play multiple roles in relation to a single system (Van Hooland, 2006; Cifter and Dong, 2008) and that roles can change over time depending on age, personal/social circumstances and motivations, as well as users' relationship with technology (Kelly and Russo, 2008). This change in needs and expectations can also occur either because interest in the domain or domain knowledge has increased, or once the initial experience has been satisfied, users look for what else is on offer (Marty, 2007a). According to Cifter and Dong (2008), in the case of subject, domain knowledge or technical expertise, this may occur through learning (see Figure 2.1).

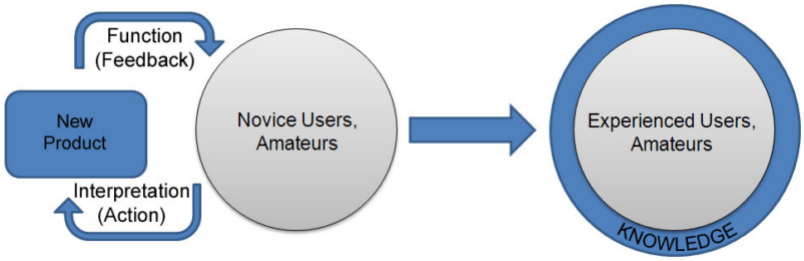


Figure 2.1: Cifter and Dong's (2008) process of gaining experience

By acquiring knowledge and/or experience, users can advance in one or multiple

dimensions, moving from being a novice to becoming an expert. Figure 2.2 includes an element of iterative feedback. For example, in the case of domain knowledge, the knowledge gained, in turn, leads to insight and action as users develop (and apply) their knowledge (Taylor, 1982). Note that a hobbyist can be either a novice hobbyist or an expert hobbyist or something inbetween.

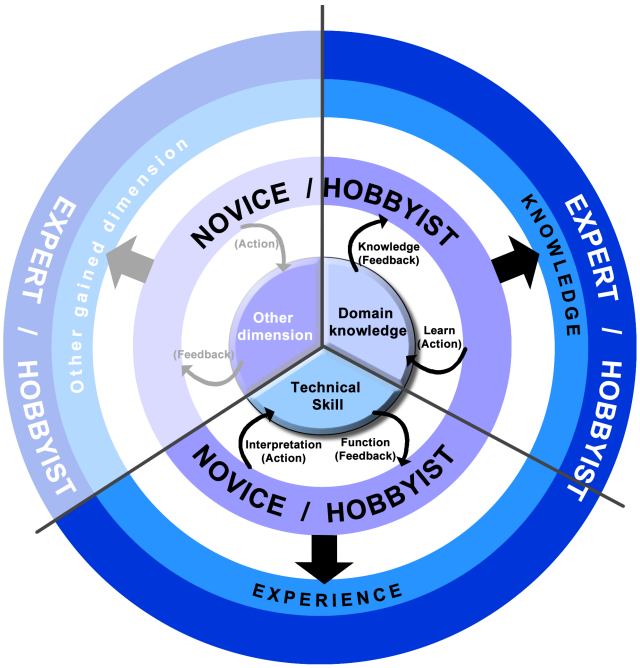


Figure 2.2: Adapted version of Cifter and Dong's (2008) process of gaining experience to accommodate for multiple dimensions

2.4 Information Behaviour and Information Seeking

Users have often been categorised and modelled in the fields of information-seeking and context-aware systems. Marchionini (1989) outlined the process as a special case of problem-solving, involving recognising the information problem, planning and conducting the search, evaluating results, and iterating if needed. Wilson (1999) elaborated on this, defining information behaviour as the overarching field, within which lies information-seeking behaviour, and within that, the more specific field of Information search behaviour (see Figure 2.3).

Figure 2.3: Wilson's (1999) nested model of information behaviour research

Wilson's (2000) framework defines information behaviour as a broad concept covering how individuals interact with information by encompassing needs, seeking behaviour, and use. Information-seeking behaviour is a subset of this, involving the intentional effort to find information that meets specific goals by utilising various information systems. Within this, "information search behaviour" includes directed searching with clear goals and use of specific search terms and browsing, which involves exploring without specific queries or goal.

Russell-Rose and Tate (2012) also considered the behavioural differences of search system users and focus on two dimensions: domain expertise and technical expertise. The inclusion of "expertise" can have significant effects on how people find and use information. Domain expertise reflects familiarity (or experience) with a subject, while technical expertise captures proficiency with using computers, search systems and the internet. Users are mapped to both dimensions as novices or experts, with the inclusion of how experience can move users from novices to experts over time.

2.4.1 Theoretical Models of Users Information Seeking and Behaviour

The study of information seeking and the creation of theoretical models designed to explain the phenomenon can be traced back to as early as the 1940s with the development of Shannon's (1948) Communication Model (see Figure 2.4). Since then, information-seeking models have been continuously refined over the years (Ellis et al., 1993).

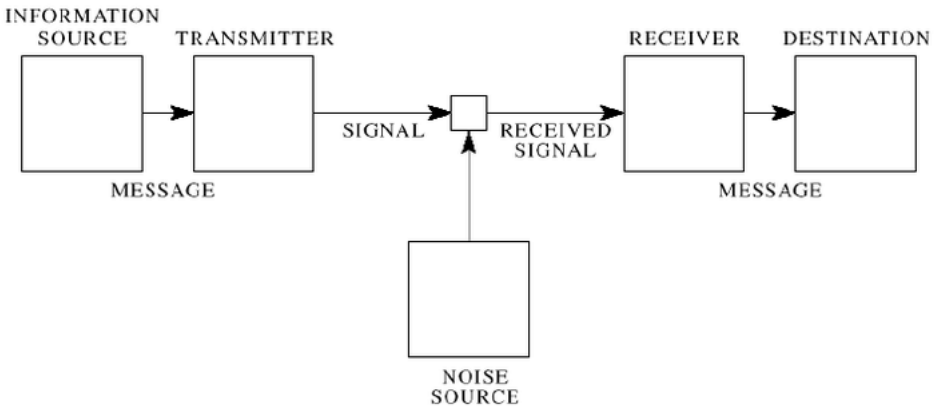


Figure 2.4: Schematic diagram of a general communication system (Shannon, 1948)

Originally designed to optimise telecommunication, Shannon (1948) model has become pivotal in information-seeking despite its broader focus (see Figure 2.4). It explains fundamental communication processes and addresses noise reduction's crucial role in accurate message transmission. In information-seeking, noise encompasses barriers hindering information acquisition. Acknowledging noise's significance enables seekers to strategise, minimising their impact and improving information acquisition (Lee, 1988).

Wilson's (1981) Model of Information Behaviour (see Figure 2.5) expanded upon Shannon's (1948) model, by focusing on how individuals actively seek out information to satisfy their needs and goals.

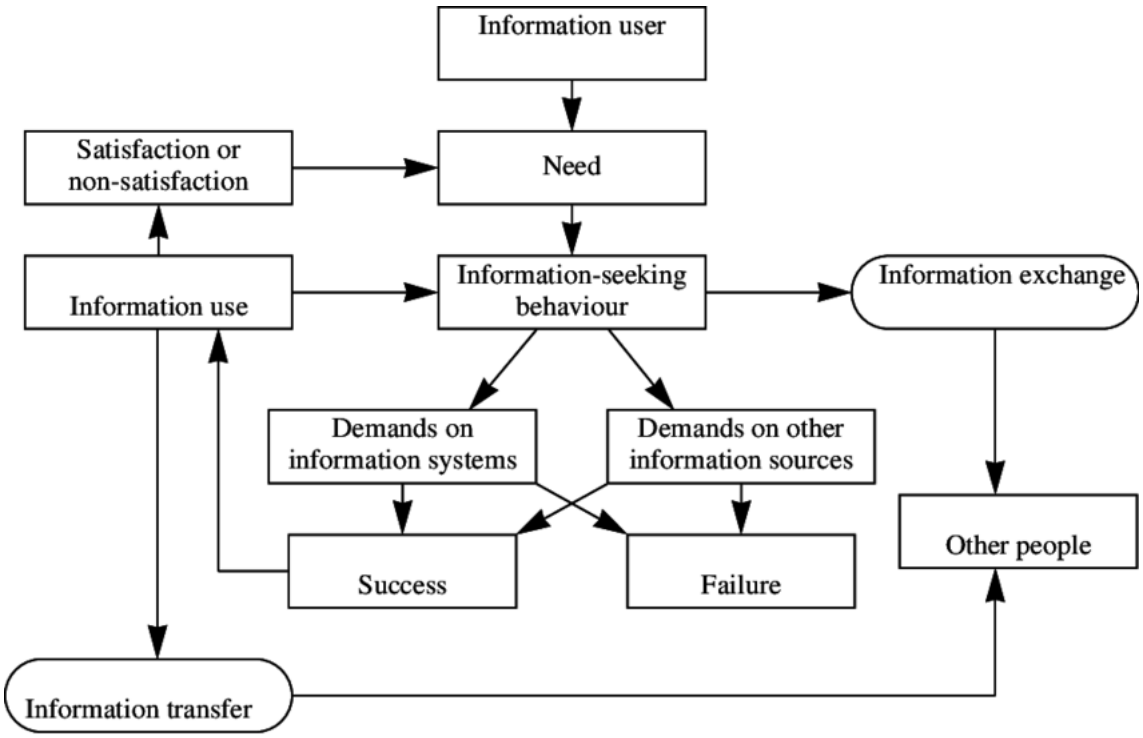


Figure 2.5: Model of Information Behaviour (Wilson, 1981)

Wilson's model exposes various facets of "information-seeking behaviour" beyond merely addressing "information needs". It delineates the process starting from the recognition of an information need, followed by the utilisation of diverse sources or services to locate pertinent information. Success leads to the fulfilment of partial or complete information needs, while failure prompts further search endeavours. Additionally, the model acknowledges information sharing among individuals as a vital component, akin to Shannon's communication model but in a more intricate form (see Figure 2.4).

However, information-seeking behaviour encompasses more than just transmitting information; it involves locating, evaluating, and utilising information. Belkin et al.'s (1982b) Anomalous State of Knowledge (ASK) framework delves into the cognitive aspect of individuals, addressing how they recognise gaps in their knowledge, leading to information needs. ASK extends Shannon's model by incorporating psychological factors underlying information needs, similar to Wilson's models, which emphasise the initiation of information-seeking behaviour by information needs. Therefore, while Wilson's (1981) nested model explores the perceived need by information users, Belkin's ASK model explains the origin of this need by con-

sidering the users' anomalous state of knowledge.

Establishing how users locate information was explored by Kuhlthau's (1989) through the Information Search Process (ISP) Model (see Figure 2.6) which depicts the users experiencing various stages of both task and emotion during the search process. These stages include Initiation, marked by uncertainty or apprehension; Selection, where users may feel optimism post-selection; Exploration, accompanied by confusion, anxiety, and frustration; Formulation, characterised by decreased anxiety and increased confidence; Collection, wherein confidence further increases; and finally, Action, culminating in a sense of achievement (Kuhlthau, 1991; Russell-Rose and Tate, 2012).

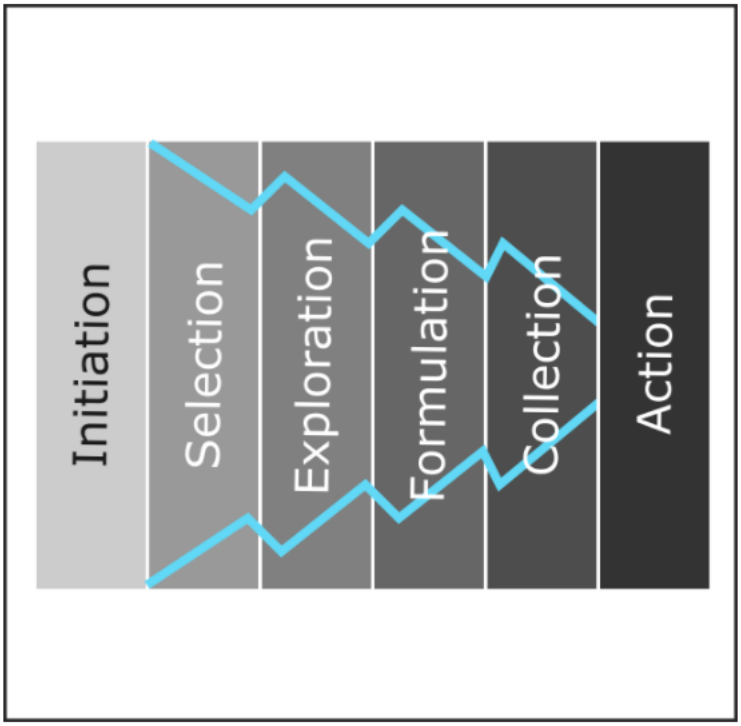


Figure 2.6: Information Search Process (ISP) Model (Kuhlthau, 1989)

However, Ellis's (1989) extended Kuhlthau's work further through the development of the "Behavioural Model of Information Seeking" (see Figure 2.7), by adding a cognitive perspective that underscores the users' mental model of the information system and intentions in the information-seeking process. This is achieved by providing a more detailed account of individuals' actions at various stages of information seeking.

- **Starting** (Ellis) = Initiation stage (Kuhlthau) - individuals recognise their

information needs and initiate their search by identifying potential sources and gathering preliminary information.

- **Chaining** (Ellis) = Exploration stage (Kuhlthau) - users follow references, citations, or links from one source to another to uncover relevant information.
- **Browsing** (Ellis) = Exploration and Collection stages (Kuhlthau) - users casually explore the information landscape or evaluate potential sources to gather information related to their topic.
- **Differentiating** (Ellis) = Formulation stage (Kuhlthau) - users start to differentiate between relevant and irrelevant information sources, focusing their search more narrowly.
- **Monitoring** (Ellis) = Collection stage (Kuhlthau) - users actively track specific sources or areas of interest to stay informed about updates or new developments in their field.
- **Extracting** (Ellis) = Collection and Presentation stages (Kuhlthau) - users gather relevant information from selected sources and prepare to present their findings.

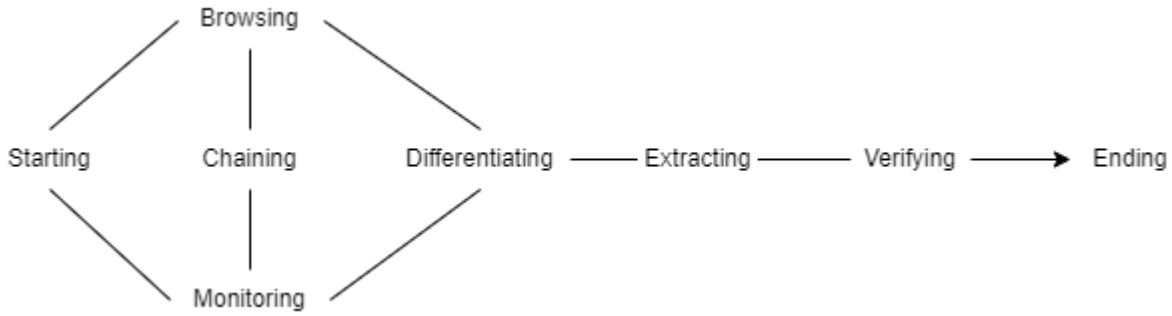


Figure 2.7: Information Search Process (ISP) Model (Ellis, 1989)

Ellis (1989) also argues that the social context should be considered in designing information systems that support users' information-seeking behaviour.

Bates's (1989) berry picking model (see Figure 2.8) accounted for the searchers' evolving information needs as they learned from the materials found in previous search results. The model emphasises the iterative and nonlinear nature of the

information-seeking process. Typically, users start with a narrow feature of a broader topic or a relevant reference and move through multiple sources. Each new piece of information they encounter leads to new ideas and directions, resulting in a new conception of the query. This ongoing process is known as an evolving search, where users continually modify the search terms and the query itself.

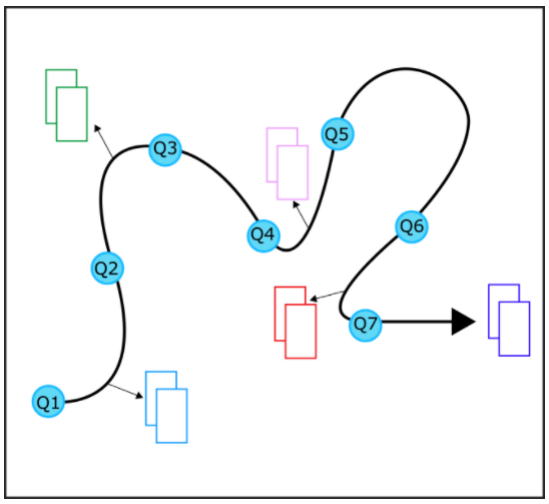


Figure 2.8: Bates Berry Picking model (Bates, 1989)

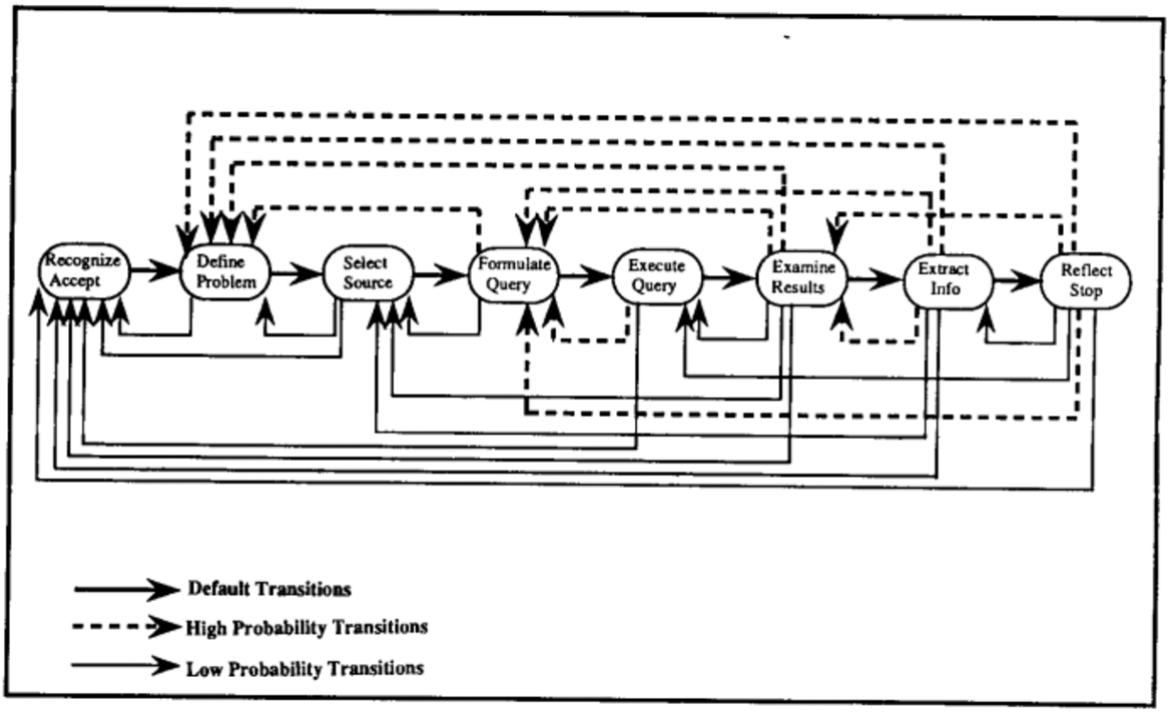


Figure 2.9: Marchioninis' Information Seeking, 7-step process of resolving an information need (Marchionini, 1995)

However, information-seeking is a complex process influenced by various personal and environmental factors. The information seeker's unique mental models, experiences, abilities, and preferences all play a role in information seeking (Bates, 1989; Ellis, 1989; Kuhlthau, 1991). The information-seeking process is both systematic and opportunistic involving strategic decision-making and the interaction of various factors such as the information seeker, the task, the search system, the domain, the setting, and the search outcomes (Marchionini, 1989, 1995). Figure 2.9 shows how the information-seeking process comprises several sub-processes that begin with recognising and accepting the problem and continue until the issue is resolved or abandoned. Understanding the problem depends on knowledge of the task domain and may also be influenced by the setting/environment.

The expansion of technology, however, gave rise to novel models for users seeking information (Pirolli and Card, 1999; Blandford and Attfield, 2010). Pirolli and Card's (2005) model draws from ecology and animal behaviour, such as foraging for food, to illustrate the decision-making process involved in information seeking. It emphasises how users adapt their search strategies based on the cost and benefit of information, ultimately leading to the concept of "satisficing," where users stop searching once they find information that meets their minimum requirements rather than seeking the best solution. This model was then expanded by incorporating cognitive processing and introducing the "sensemaking model for intelligence analysis." (see Figure 2.10) explains how individuals make sense of complex information, generate hypotheses, and develop mental models to interpret data. It also highlights the role of feedback loops in the sensemaking process, as users continually adjust their mental models based on new information.

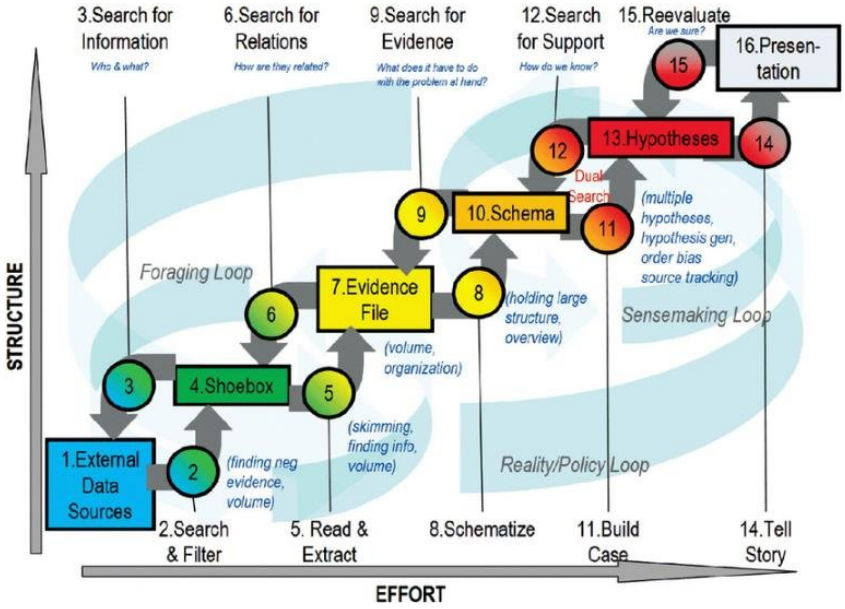


Figure 2.10: Pirolli and Card's notional model of sensemaking model for intelligence analysis (Pirolli and Card, 2005)

Blandford and Attfield (2010) proposed an information journey model (see Figure 2.11), was based on empirical studies of user behaviour (Ellis, 1989; Bates, 1989; Kuhlthau, 1989; Wilson, 1981). The framework consists of four main activities: recognising an information need, acquiring information, interpreting and validating information, and using information. This model differs from previous models as it requires users to validate, interpret and use the information to adjust their information needs, making it dynamic, it also acknowledges serendipity as a part of the information-seeking experience, as it considers the occasions when information is stumbled upon without being actively sought. For example, the journey may not begin with the individual recognising a need, as they may encounter information without prior recognition and not immediately use it; it also includes validating and interpreting information, highlighting the intrinsic problem of finding the right information at the right moment.

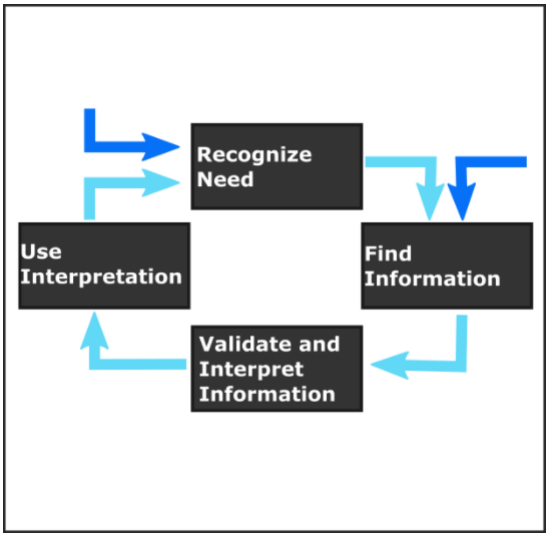


Figure 2.11: Blandfords information journey model (Blandford and Attfield, 2010)

Wilson (2017) identified that many information-seeking models have one common limitation: in defining the potential stages that users may experience, they discover that it is not straightforward to define the exact process the searcher is likely to experience. The models were then created in a linear structure that represents the ideal path, but one that a user rarely takes. Instead, the searchers move between a series of active, passive and reflective states in no set or predictable way, as illustrated in Marchionini's (1995) model (see Figure 2.9).

Intending to create a new model that captures the search process from the perspective of a searcher seeking information, Wilson (2017) presented the "Tetris model of resolving information needs within the information seeking process". Unlike existing models that categorise behaviours and stages, this new model focuses on the searcher's motivations as they transition between these behaviours and stages. The model takes a literal mapping approach to the Tetris game and identifies that the uncompleted horizontal lines on the game board are the Information need(s). As the searcher searches or comes across new information, each piece of new information is represented as a new block. Every block is a slightly different shape, and the searcher attempts to fit that new knowledge into the board wherever it will work, sometimes fitting perfectly in the line and completing the knowledge gap for the information need. Still, often the blocks do not fit entirely and only partially fill the information gap (partially resolving the information need). The model also allows for those blocks that do not fit the current need and so get stacked, creating

a new information need. The Tetris model alludes to the constant learning that occurs by indicating that any completed lines/information needs are removed from the board to leave room for new information needs. The main difference between this model to all the others is that the Tetris model does not perceive progress as moving through a series of stages but instead in resolving information needs, which is much potentially closer to how searchers would see the problem themselves (got a problem, solve it, move on). With this respect, the model links closely but expands on that of Bates (1989) berry picking concept. By taking the idea that multiple parts of information help to fulfil the need and that new information builds on past learnt knowledge and also enables better searches to be undertaken, but expanding on Bates' model by allowing for the information that does not add to the need but instead broadens the horizons and creates a new parallel information need.

Critiques of information-seeking models highlight their tendency to oversimplify the complex nature of information-seeking behaviour. Some researchers argue that traditional models such as the one proposed by Wilson (1981), and other stage models, overlook the social and cultural factors influencing information-seeking behaviours (Chatman, 1999). Moreover, there is a concern that these models may overlook the significance of emotional aspects in information-seeking behaviour, including feelings of worry, anxiety, frustration, and uncertainty (Case and Given, 2016). Furthermore, doubts emerge regarding the universality of these models for all types of information-seeking behaviour (Dervin, 1983). For example, research on information seeking in online environments, for instance, indicates that individuals may engage in diverse search behaviours depending on their tasks, the nature of information sought, and the functionalities of the information system utilised (Marchionini, 2006).

While most theories typically tie information encounters to tasks motivated by work or personal goals, Elsweiler et al. (2011) introduce an adapted version of Ingwersen and Järvelin's (2005) nested model of contexts, focusing on casual-leisure scenarios driven by hedonistic needs, where people engage in search behaviours for pleasure rather than specific information. Elsweiler et al. (2011) highlight key differences in information behaviour within casual-leisure contexts namely: 1. Tasks are often motivated by achieving a particular mood or state, often linked to the individual's quality of life and health; 2. The act of searching itself often takes precedence

over finding specific information; 3. Information needs are frequently ill-defined or absent, with transient needs possibly emerging to facilitate casual pursuits; 4. Success is not necessarily contingent on finding specific information or results.

However, while there are observable behavioural differences in casual leisure information behaviour, these contexts still adhere to established models of information seeking behaviour, involving searching for and consuming information from various sources e.g., text, video, image etc. (Elsweiler et al., 2011). Nevertheless, while the hedonistic pursuit of pleasure through casual leisure information has been examined, whether the individual's emotional state at the outset of the information encounter influences their casual leisure information behaviour and the types of experiences sought remains unexplored.

Despite criticisms, information-seeking models remain a valuable resource for researchers and practitioners in the information field. These models provide a framework for understanding information-seeking behaviour and can inform the development of information systems and services.

2.5 Interfaces and Interactions

Many models discussed in Section 2.4 rely on search interfaces, as search has become synonymous with web use due to search engines like Google and Bing. Studies of experienced or professional users (see Section 2.3) often assume these users understand the data and can convert their needs into queries (Broder, 2002). However, this overlooks users with undefined needs or who struggle to translate their information needs into system-understandable queries (Prescott and Erway, 2011; Speakman et al., 2018).

CH institutions often contend with extensive collections, making it challenging to present all items simultaneously or in a browseable format for the digital environment. Search-style interfaces allow users to quickly and efficiently find specific items or narrow down their interests (Vane, 2020). According to Adams et al. (2005), the search interface provides a mechanism by which the user can articulate their information needs into a query and examine and evaluate the output. However, significant usability challenges with search interfaces have been identified, particularly for users unfamiliar with a collection that requires specific terms (Borgman, 1996) or those

with undefined information needs (Speakman et al., 2018; O'Brien and Toms, 2008). Similarly, a lack of domain knowledge or focused goals can make the search box a significant barrier for such users (Belkin et al., 1982a; Whitelaw, 2015). These challenges include the difficulty in forming effective queries resulting in the potential for disengagement due to inaccurate search results or non-intuitive interface designs (Norman, 2002; Lehmann et al., 2012). Consequently, many studies advocate for more exploratory, browser-based interfaces, allowing users to discover information serendipitously or in a structured way without precise search terms (Koch et al., 2006; Ruecker et al., 2011).

Therefore, while traditional search interfaces often fail to address the increasingly complex and diverse data needs of today's users, particularly given the vast amounts of multi-media data that defy simple keyword description (Lang et al., 2013), the emergence of "next-generation" search technologies aim to address these challenges by integrating advanced features such as query refinement, context-sensitive results, and meta-data-based filtering, thereby enhancing the user experience by supporting dynamic information retrieval (Morris and Teevan, 2012; White, 2016).

Generative AI represents this advancement in search technology, capable of providing direct responses based on deep learning models e.g., the GPT series of models (Pavlik, 2023). These AI systems enhance traditional search functionalities by improving search relevance through understanding language semantics and user intent (Ryan, 2023). They also offer innovative features such as query rephrasing, personalised search results (Cao et al., 2023; Pisoni et al., 2021), and the ability to process and analyse multimedia content for more intuitive and interactive searches (Lim et al., 2023; Zakraoui et al., 2019).

However, while incorporating generative AI into search systems may enhance the user search experience of the DCH collection, the uptake of such interfaces over standard search by the DCH institutions is still minimal. This may primarily be based on issues such as trust and apprehension surrounding the creation of false information by the AI's that could potentially result in misinformation being presented as fact from institutions which as seen by users as places/sources of knowledge and truth (Kusnick et al., 2024; Spennemann, 2024).

Accordingly, the digitisation of CH collections has made them more accessible through web interfaces to a broader audience (Skov et al., 2004; Skov and Ingwersen,

2008). This expanded audience necessitates a different approach to that of the standard search box (Walsh et al., 2017). According to Koch et al. (2006) and Ruecker et al. (2011), most users prefer to explore navigation structures (browsing) rather than search for information.

Browsing is the process of exploring content in a more unstructured way, often by following links in the body or navigation bar from one web-page to another or by scrolling through social media feeds or other content streams. Browsing does not involve a specific query or goal, instead, the aim is to explore rather than to find specific pieces of information and selecting those that appear interesting or relevant (Bates, 1989).

Browse interfaces often use visual metaphors and structured navigation that can lead to more engaging experiences. They are especially beneficial for less experienced users who might feel overwhelmed by the blank search box of typical museum search interfaces (Whitelaw, 2015; Coburn, 2016). Moreover, these browsable interfaces are suggested to be more inclusive, enabling serendipitous discovery and exploration of collections (Bates, 1989; Hall and Walsh, 2021).

In parallel to the evolving search technologies, technological development in browse interfaces aims to make digital collections more accessible and engaging. Termed either "Rich Prospect Browsing" or "Generous Interfaces", these concepts offer a more inclusive approach by providing a comprehensive overview of collections upon entry, allowing users to navigate and explore without requiring specific queries (Ruecker et al., 2011; Whitelaw, 2015). These interfaces are designed to be intuitive, allowing users to zoom and filter through a meaningful depiction of every item in the collection, supporting serendipitous discovery and deeper engagement with the content (Foo, 2016; Coburn, 2016).

Such interfaces challenge the traditional "ungenerous" search-based systems that hide information behind queries, proposing instead a user-centric design that reveals more information upfront. This approach aligns with the principles of Schneiderman and Plaisant's (2010) "overview first, zoom and filter, then details-on-demand," enabling users to interact with the collection in a more fluid and engaging manner. The development of these interfaces represents a significant shift towards accommodating diverse user needs, including those who may not know exactly what they are looking for (Whitelaw, 2015; Coburn, 2016).

However, irrespective of these technological advancements and possibilities, almost all museums offer a standard search interface as the main means of allowing users to explore their curated collections and sub-collections, reverting to their traditional stance of expecting their users to know what to search for.

This research study therefore, does not delve into the specific interface designs used in DCH but instead focuses on the broader categories of 'search or browse' (Walsh and Hall, 2015; Prescott and Erway, 2011) and 'desktop or mobile' as these distinctions are crucial for understanding user preferences in DCH settings. This approach, while informed by a detailed review of various models (Vane, 2020; Ruecker et al., 2011), aims to capture overarching trends that could influence future interface development that may cater for diverse user needs.

2.6 Studying Users in the Web/UX Industry

The experience that users have when engaging with museums, museum artefacts and museum websites has seen significant attention over recent years (Zahidi et al., 2013, 2014; Ismail, 2007; Meyerson et al., 2012; Beaudoin, 2012; Sweetnam et al., 2012, 2013). As user experience (UX) becomes more widely recognised by museums and museum management as an important factor in delivering services that the audiences want, the line between business strategy and the design of the user experience is fading (Kolko, 2015). According to Knemeyer (2015) "user experience (UX) has become a mission-critical consideration for companies in every industry, and of every shape and size". With more museums starting to engage with user experience design (UXD) practices, it is essential to understand the different approaches that can be adopted.

2.6.1 User Centred Design Process

User-Centered Design (UCD) originated from the work of Kling's (1977) in the 1970s and was further developed by Donald Norman's research laboratory at the University of California, San Diego in the 1980s. Building on earlier studies (Lucas Jr, 1971; Lucas, 1975; Hedberg and Mumford, 1975; Hedberg, 1980) Kling established a theoretical framework for designing systems with a user-centric focus. UCD gained

wider recognition (Norman and Draper, 1986; Norman, 1988) by encompassing the four basic user principles of; visibility, accessibility, legibility and language, at the centre of the design process.

An advancement of the UCD process is Human Centered Design (HCD), which emphasises considering users as individuals with names, feelings, and emotions. Unlike UCD, which may depersonalise users to mere objects (Cooley, 1989, 1996), HCD focuses on the human aspect of design. UCD/HCD is widely adopted by software and website UX/design teams today, employing an iterative process that prioritises user needs and involvement throughout the four main phases of design and development e.g., Phase 1) Researching the context of use, Phase 2) Specifying user requirements, Phase 3) Producing design solutions, and Phase 4) Evaluating designs against requirements. (see Figure 2.12) Users are typically engaged through investigative research methods e.g., such as surveys and interviews and design techniques e.g., brainstorming, storyboarding, and prototyping.

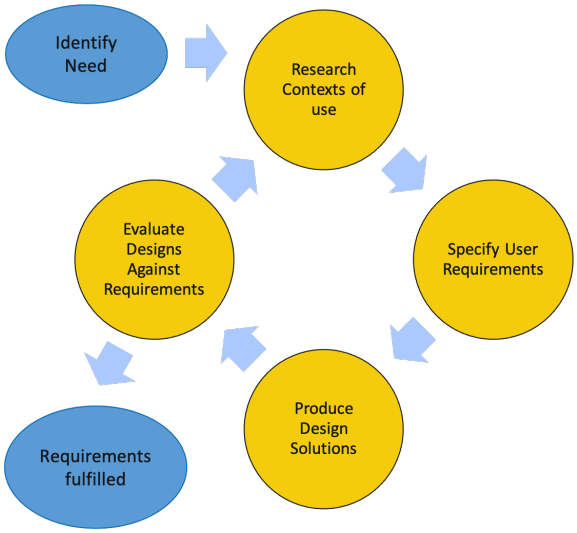


Figure 2.12: UCD Process

2.6.2 User Experience Design

User Experience (UX) Design has evolved from Norman (1988) and diversified into a number of methodologies that are practised within the design, web and software industry, for example, The British Design council's Double Diamond Framework; The IDEO's HCD ideology and Stanford University Design Schools design thinking process.

2.6.2.1 Double Diamond

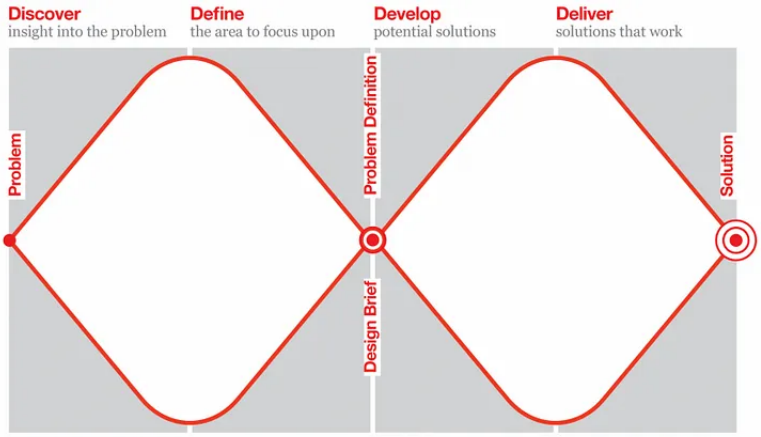


Figure 2.13: British Design Council's Double Diamond Framework

The British Design Council's Double Diamond (British Design Council, 2005) (see Figure 2.13) is a visual framework that represents the design process. It is commonly used as a way to structure the design process and ensure that designers consider all aspects of the problem at hand.

The Double Diamond model comprises four phases: Discover, Define, Develop, and Deliver. During Discover, designers research and gather information to understand the problem; Define, using the gathered information the problem is defined and user needs identified; Develop involves creating and testing potential solutions, while Deliver focuses on implementing and launching the solution.

2.6.2.1.1 IDEO HCD Ideology

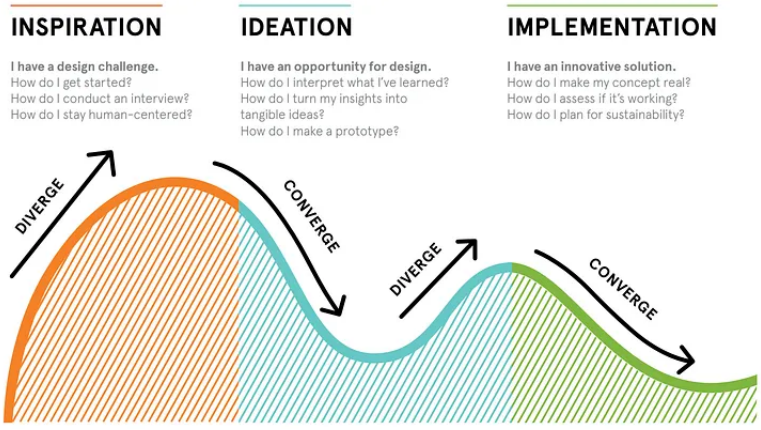


Figure 2.14: IDEO HCD Ideology

The IDEO HCD (Human-Centered Design) ideology and toolkit (IDEO, 2015) (see Figure 2.14) and User Experience Design (UXD) share similar principles and goals.

The IDEO HCD ideology comprises three phases: Inspiration, Ideation, and Implementation. During Inspiration, designers conduct user research to develop empathy and understand consumer demands; in Ideation, designers generate ideas, create prototypes, and iterate based on feedback. Finally, during Implementation, designers bring the final solution to market and continue to refine based on user feedback.

UXD, similar to IDEO HCD, encompasses understanding user needs, generating prototypes, and refining solutions based on user feedback. It involves user research, personas, journeys, wireframing, prototyping, and testing. While IDEO HCD is a broad framework applicable to any design challenge, UXD specifically targets digital products and services. However, both processes prioritise user understanding and tailored solutions.

2.6.2.2 Design Thinking

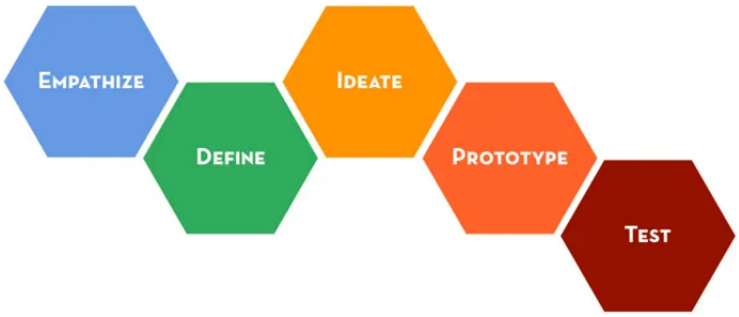


Figure 2.15: Stanford Design Schools Design thinking process

Design thinking is a iterative process used by teams to understand users, challenge assumptions, redefine problems, and create innovative solutions for prototyping and testing. The process comprises of five stages (see Figure 2.15): Empathize, Define, Ideate, Prototype, and Test. During the Empathize stage, designers conduct a form of ethnographic research to understand users' needs, an objective shared with this thesis. In the Define stage, designers identify the problem and user requirements based on research findings. In the Ideate stage, ideas are generated and selected. Then, in the Prototype stage, designers create prototypes, followed by testing and

refining them based on user feedback in the Test stage.

Commonalities can be found between Design Thinking and UXD, although the Design Thinking process is a broad framework applicable to any design challenge, while UXD focuses specifically on designing digital products and services. However, both approaches prioritise understanding the user and designing solutions to meet their needs.

Additionally, UXD, Design Thinking, and HCD are closely related approaches that aim to solve design problems. While they share similar principles and methods, there are also subtle differences in their focus and application, for example:

Similarities:

- All three approaches are centred around understanding the needs and behaviours of users, and designing solutions that meet those needs.
- All adopt a range of methods and techniques that help designers to understand user needs and behaviours, generate ideas and solutions, and test and refine their designs. Some of the most commonly used methods and techniques include:

    1. **User Research**: This involves gathering data about the needs, goals, and behaviours of users through techniques such as interviews, surveys, and observation.
    2. **User Personas**: These are fictional characters that represent the different types of users who will interact with the product or service. They help designers to understand the needs and behaviours of different user groups.
    3. **User Journeys**: These are visual representations of the users' experience with the product or service, from initial awareness to post-purchase. They help designers to identify pain points and areas for improvement.
    4. **Ideation Techniques**: These are methods for generating and selecting ideas, such as brainstorming, mind mapping, and SCAMPER.
    5. **Prototyping**: This involves creating low-fidelity or high-fidelity models of the design solution, which can be tested and refined based on user feedback.

6. **Usability Testing**: This involves observing users as they interact with the product or service and gathering feedback on its usability, effectiveness, and efficiency.
7. **Iterative Design**: This involves repeating the design process of creating prototypes and testing with users multiple times, using feedback from users and stakeholders to refine the solution.

- They all prioritise empathy and understanding, and encourage designers to put themselves in the shoes of their users.

Differences:

- UXD is specifically focused on designing digital products and services, while Design Thinking and HCD can be applied to a broader range of design challenges.
- Design Thinking is a more structured process that follows a specific set of stages, while HCD is a broader approach that encompasses multiple design methodologies.
- HCD places a greater emphasis on social impact and sustainability, while UXD and Design Thinking are more focused on creating effective and usable solutions for users.
- UXD is typically more focused on the final product or solution, while Design Thinking and HCD may also consider the larger context and systems in which the design problem exists.

Overall, UXD, Design Thinking, and HCD all embrace a user-centred approach and iterative design. Their distinctions lie in focus, application, and the methods they employ. Depending on the design challenge, any of these approaches could prove effective independently although designers may opt to blend aspects of each for a tailored process.

2.6.3 Benefits of Adopting UXD in DCH

Adopting a UXD stance provides a developer with the experiences a user may have when interacting with the digital service they use, for example:

- **Improved User Engagement**: UXD approaches can help to create digital cultural heritage experiences that are more engaging and enjoyable for users, leading to increased interest in cultural heritage and greater understanding and appreciation of it.

- **Increased Accessibility**: UXD approaches can help to make digital cultural heritage experiences more accessible to a wider range of users, including those with disabilities or language barriers.

- **Enhanced Learning**: UXD approaches can help to create digital cultural heritage experiences that are more informative and educational, leading to increased learning and knowledge retention among users.

- **Preservation of Cultural Heritage**: By creating digital cultural heritage experiences that are engaging and accessible, UXD approaches can help to preserve cultural heritage by making it more accessible to a wider audience.

- **Improved Collaboration**: UXD approaches can facilitate collaboration between designers, cultural heritage professionals, and end-users, leading to more effective and impactful digital cultural heritage experiences.

- **Increased Funding Opportunities**: By demonstrating the value and impact of digital cultural heritage experiences, UXD approaches can help to attract funding from a wider range of sources, including government agencies, private foundations, and corporate sponsors.

Several studies within DCH explore specific aspects of the UXD process. Some investigate user satisfaction (Zahidi et al., 2014), while others address issues like navigating digital collections (Ismail, 2007; Meyerson et al., 2012); or enhancing interface usability (Zahidi et al., 2013).

However, although numerous studies share similarities with these, certain emerging issues deviate from the UXD paradigm. Beaudoin (2012) suggested that the needs of users are one of the contexts that should be considered because it would influence users' positive experience and response of digital cultural collections. Different types of users have different user requirements and needs. While Sweetnam et al. (2012) points out that current digital platforms in cultural heritage often

cater to specific goals or user groups, resulting in interfaces that may be either too simplistic or overly complex for certain users, leading to varying degrees of user satisfaction.

There are many studies of individual user groups as discussed earlier in this literature review, but many were studied in isolation of the other groups, which would suggest that the UXD process could be followed closer by studying all users of the system at once, which could then lead to more satisfactory systems being created.

2.6.4 Personas

A persona is a method to model, summarise, and communicate research findings about observed or researched individuals. Cooper (2004) argued that by creating personas, designers could gain better insight into the needs and behaviours of their target audience, and create products that align with those needs.

Personas are depicted as a specific person that is a fictitious character profile that represents a real world/system user (Goltz, 2014). The persona is a document (usually a single side of a page) that details an informed, but made up typical user (it is never a full representation of a real user but is a made up representation of all users in that group). They serve as design tools by keeping system interface designers focused on the needs, goals, and frustrations of users (Cooper, 2004). Goltz (2014) claims that they also enable the designer to focus on a manageable and memorable cast of characters, instead of focusing on thousands of individuals. Personas can be used by the development team to help guide decisions about product features, navigation, interactions, and even visual design (Goodwin, 2005).

It encourages the team not to design solely for themselves or a single customer (who might be an outlier). Instead, they should consider questions such as: 'Would this persona prefer it this way or that?', 'Do they need it one way or another?', 'What would their expectations be?'. By understanding who the users are, their likes and dislikes, goals, skills, and personalities, a skilled developer/designer can then attempt to adopt the users' perspective. Personas also force the team into maintaining focus by designing/building to meet the personas goals. A good example of a persona, by Soloman (2021) for her Just Eat case study, can be seen in Figure 2.16. The best

personas are based on user data. This could be gathered from Google Analytics, on-site reviews, social media, ecommerce data, off-site reviews (yelp or google places), or even from customer estimation tools (quantcast, Alexa or google add planner).

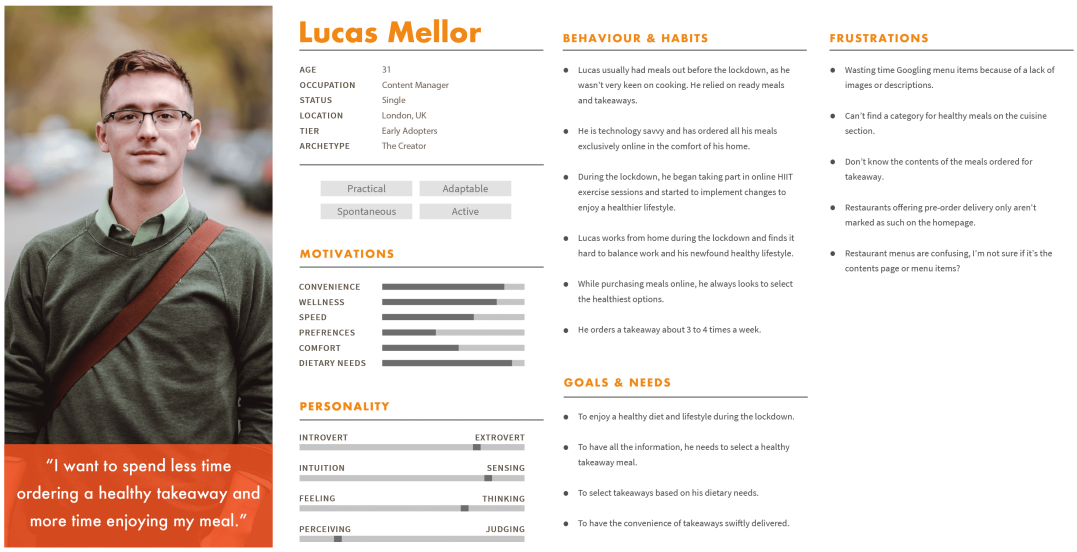


Figure 2.16: Soloman (2021) Example of a persona created for Just Eat.

Personas offer numerous advantages in product design (Miaskiewicz and Kozar, 2011), aiding in identifying user needs, formulating design requirements, and crafting user-friendly products. Although, Miaskiewicz and Kozar (2011) acknowledge challenges like potential bias and the necessity for continual maintenance and updates in persona creation.

Exploring how the perceptions and utilisation of personas among designers and user experience professionals was undertaken, Matthews et al. (2012) discovered that while most participants deemed personas effective in product design, their usage lacked consistency. Some employed personas as a means of communication to explain the target audience to stakeholders, whereas others utilised them as a design tool to shape product features and requirements. Additionally, the application of personas varied based on the type of product under design, with consumer products showing higher likelihood of persona usage compared to enterprise products, which cater to a broader array of users with diverse needs. Matthews et al. (2012) suggest that there is a need for more education and training on the use of personas to ensure that they are used effectively and consistently across different types of products and design teams.

The creation of Personas, according to Cooper (2004) and Goodwin (2008) should be derived from qualitative data gathered through methods like interviews or observations. Utilising three types of data combinations: qualitative personas, qualitative personas validated with quantitative data, and quantitative personas (i.e., QPC) assists in the formulation of authentic personas (Mulder and Yaar, 2006; Salminen et al., 2020). Additionally, mixed methods approaches have also been devised (Pruitt and Grudin, 2003; Salminen et al., 2018) to enhance the selection, enrichment, and refinement of personas through iterative data collection. While the mixed methods approach is considered optimal in industry, constraints such as time and budget often hinder development teams from acquiring the necessary quality of data. These limitations can impede researchers' ability to comprehend the audience fully and may compromise the accuracy of the resulting personas (Bagnall et al., 2005).

2.6.4.1 Issues with Personas

Traditional qualitative methods for creating personas have faced criticism due to high costs, limited scalability, and lengthy manual processes (Thoma and Williams, 2009) (Salminen et al., 2020). While personas are often employed to address internal issues stemming from a lack of consensus on user requirements (Rönkkö et al., 2004), they can suffer from subjectivity and lack of representative data, as small user bases are utilised to expedite the process (Chapman and Milham, 2006). (Matthews et al., 2012) study on UX practitioners revealed that although all participants understood personas and used them for communication, most did not use them to inform their own design work. This reluctance stemmed from poorly informed or vague Personas. However, participants with detailed and user-informed personas, supported at all levels of the business, experienced numerous benefits, including facilitating interaction design activities, defining product functionality, and assessing design effectiveness (Cooper, 2004; Pruitt and Grudin, 2003; Pruitt and Adlin, 2010).

Personas offer a generalised representation of user motivations, goals, and needs within a specific group, outlining universal experiences and expectations. While personas typically lack detailed information seeking behaviour descriptions, examples like the seven personas for Europeana by Rasmussen and Petersen (2012), each with

a defined search strategy, demonstrate their potential value in aiding design decisions. Nevertheless, Salminen et al. (2022a) raised concerns regarding the number of personas created for a project, with arguments favouring fewer personas due to reduced cognitive overload that may hinder the designers and developers decision-making. However, Salminen et al.'s (2022a) research suggests that having more than 10 Personas, particularly in interactive systems, enhances the representation of diverse populations and improves inclusivity in designs. Matthews et al.'s (2012) findings also indicate that practitioners heavily invested in personas often utilise an extensive set, though for a project for many development teams utilising between 3 and 5 personas avoids unwieldiness (U.S. General Services Administration (GSA) Technology Transformation Service, 2019).

While the main issues with personas are that:

1. personas when not user data led are guesses at best and can be seen by the team as not useful.

2. personas often lack in useful details.

Personas serve as a valuable tool for fostering empathy and providing a general user background for development teams. This is exemplified by their adoption in major tech companies like Spotify (Torres de Souza et al., 2019). However, they fall short in offering detailed insights into user preferences for system usage, content absorption, access methods, and task-specific knowledge levels.

2.6.4.2 Personas Based on User Research

According to (Pruitt and Grudin, 2003) personas should be based on empirical data to accurately depict users' needs, goals, and behaviours. They emphasise the importance of creating realistic and representative personas and using them throughout the design process to ensure alignment with user needs. This empirical data gathering process, termed "user research," involves various methods such as surveys, interviews, focus groups, observations, and user testing. The aim is to understand users' needs, behaviours, motivations, and pain points. Analysis of the collected data identifies patterns and themes, grouping users with similar characteristics and behaviours to form personas. For example, both Pretorius and Sangham (2016);

Chang et al. (2008) conducted user research to develop Personas. Pretorius and Sangham studied various user groups engaging with government online services, resulting in the creation of five personas which were used to develop the online services to better meet the needs of the users. While Chang et al. (2008) crafted four personas for the development of personal finance software, which led to an improved product and increased user satisfaction.

Empirical user research is crucial, but the type of data used can influence the results. Jansen et al. (2021) reviewed methods for persona creation, highlighting the strengths and weaknesses of qualitative, quantitative, and mixed methods. While they suggest that no single approach is superior, emphasis should be placed on the importance of aligning methods with specific needs, goals and available resources with a focus on how a digital approach may benefit practitioners who adopt any of the approaches. Mijač et al. (2018) notes that data-driven approaches can enhance accuracy but may face challenges regarding data reliability, interpretation, and bias in the development.

2.6.4.3 Data-Driven/Automated Personas

Creating personas through empirical user research is valuable for design, but it can be time-consuming and costly. As vast amounts of user data, captured on users as they interact with websites, are often underutilised, interest in automated persona generation is growing. This section of the literature review explores the conceptual basis, research agenda, and quantitative methods for clustering and creating personas.

Salminen et al. (2019) propose using data-driven automated personas to analyse large volumes of online user analytics data across different domains. Their system generates data-driven personas from social media and online analytics data while safeguarding user privacy. It prioritises user segments based on online behaviour, adds demographic information, and creates rich detailed persona profiles dynamically adding characteristics such as names, photos, and descriptive quotes. The system can handle colossal datasets from platforms like YouTube, Facebook, and Google Analytics. However, computational challenges associated with automatic persona generation were found along with data dis-aggregation, cross-platform mapping, and

content filtering. Addressing these issues will require interdisciplinary collaboration across various fields such as algorithmic computer science and user studies with an interpretative touch.

Alternatively, Nielsen et al. (2019) propose that automated persona generation can aid online content creators in crafting engaging content tailored to their target audience. They suggest a research agenda focused on developing algorithms to identify key user characteristics, selecting appropriate data sources, and evaluating the accuracy and utility of automated personas.

Work undertaken by An et al. (2016) investigated automated persona generation using social media data. They proposed a framework that combines social media, demographic, and psycho-graphic data to create personas automatically. By leveraging data from platforms such as Facebook, Twitter, and LinkedIn, and applying clustering algorithms, they suggest a more efficient and precise method for persona creation compared to traditional approaches.

However, while there is a growing interest in quantitative persona creation, there is a lack of standardisation. Salminen et al. (2020) advocate for developing guidelines and best practices for quantitative persona creation, it's essential to recognise the constraints of statistical methods. Quantitative data may overlook the intricate aspects of human behaviour, emotions, and motivations, while standardisation may result in oversimplified or homogeneous personas. For example Figure 2.17 (Jansen et al., 2005; Jansen, 2006; Jansen et al., 2007, 2017; Jansen, 2021; Salminen et al., 2018) contains significantly less textual information and relies more on statistical data compared to a standard persona example (see Figure 2.16).

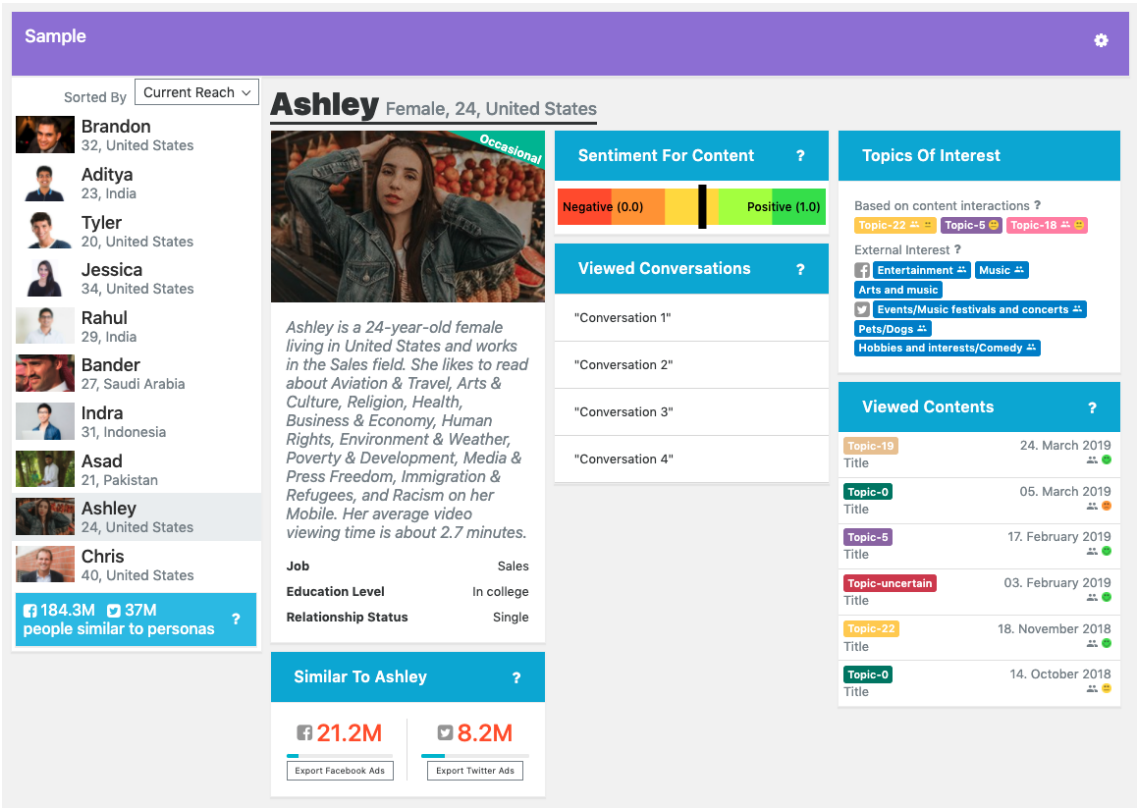


Figure 2.17: Example of an automated persona.

Therefore, to address this, qualitative methods like ethnographic studies and user interviews should be integrated to ensure a comprehensive understanding of diverse user groups(Salminen et al., 2020).

Creating personas that rely on multiple data sources presents a challenge for those lacking access to all required data. There has however been some limited work around creating a data-driven automated persona from restricted data sources. Several studies (Salminen et al., 2022b; Boyce et al., 2018; Holden et al., 2017; Kim and Wiggins, 2016; Lee et al., 2020; Schäfer et al., 2019) have attempted to create personas using survey data or semi-structured interviews, resulting in personas with limited usability and comparability to other studies (see Figure 2.18).

Figure 2.18: Salminen et al. (2022b) Example of a persona created from a survey.

Many of these studies focused on the health sector: Boyce et al. (2018) investigated natural medicine usage without specifying the clustering algorithm used; Schäfer et al. (2019) examined elderly people using German healthcare ICT devices and applications, employing the k-means clustering algorithm; Holden et al. (2017) studied older people with heart health issues using semi-structured interviews and hierarchical clustering, although the algorithm is unspecified; Lee et al. (2020) also employed the k-means algorithm and principal component analysis (PCA) to create personas for millennials' media and media-related device usage; Kim and Wiggins (2016) focused on the expectations and perceptions of library service users, utilising factor analysis and an unidentified clustering algorithm; Salminen et al. (2022b) demonstrated successful persona creation without a clustering algorithm, employing a statistical heuristic approach based on limited survey data, resulting in personas with limited content.

According to Gibbons (2018), Empathy maps extend personas to help developers/designers understand user problems, goals, perceptions, and interactions with a system. Empathy maps are collaborative visualisations, usually developed by the UX team, to articulate what is already known about this type of user/persona. This allows the development team to create a shared understanding of the users' needs and to feel more compassion about the issues, counter any negative impacts, and make improvements that will aid the user. Further insights into user behaviour can

be obtained by linking the empathy map with user engagement via understanding the user journey. According to Bell (2022), customer journey mapping identifies the sequence of experiences a customer has with an organisation, from recognising a need to fulfilling it. A crucial part of this mapping is understanding users' emotions along the way, revealing what aspects of the experience leave them satisfied, thrilled, or disappointed. While Bradley et al. (2021) notes that a good customer journey map relies heavily on quality personas and that by using customer journey maps with personas overlaid, designers can visualise how each persona interacts with various touchpoints in a service or product experience.

However, the creation of automated data-driven personas entail risks, with outdated data sources posing the greatest danger. According to Adlin and Pruitt (2010) who advises against using outdated data, as user needs may change over time, necessitating current data for accurate persona creation. An automated approach to persona generation will be presented in Chapter 8.

2.7 Summary

An online presence is now standard practice for businesses and organisations reliant on customers for profitability and survival. In the digital sphere, customers have become users who expect the same level of satisfaction, but quicker. To ensure satisfaction, it's crucial to understand how users experience and interact with a digital platform, identify who the users are, and understand their needs in order to create a meaningful digital experience.

Typically, CH organisations fall behind profit-driven industries in adopting new technologies due to limited funding. Focusing on DCH users, the literature review highlights broader issues that need addressing before tackling specific UX concerns like systems, interfaces, content presentation, and high museum bounce rates.

While the original goal of this thesis was to address this issue, the literature review revealed a knowledge gap in understanding DCH users. Specifically, the lack of holistic studies on DCH museum user groups. Despite frequent references to user groups, the identified groups lack comprehensive characterisations and are often examined in a narrow, varied manner, ignoring broader aspects. Furthermore, inconsistent labelling of user groups makes comparisons difficult.

The examination of the literature found numerous studies on physical and digital CH users, with most focusing on single characteristics, ignoring previous research, and oversimplifying the audience. Few studies offered a comprehensive description of museum audience groups, while others applied Falk's user motivation-based group definitions, designed for physical museum visitors, to the digital spaces without considering their appropriateness. Yet the online museum environment differs significantly from the physical setting, leading to additional motivations like social interaction and remote access to resources. Falk's model focuses solely on motivation, overlooking factors like technical expertise and evolving technology, therefore making its adoption to digital contexts challenging, especially when faced with an unaccounted diverse audience.

Establishing the digital CH audience proved difficult from the literature as it was discovered that many user groups were identified with limited detail, some overlapping and using different terminology for similar groups, hindering their usefulness and support. Studies often reported on groups like experts, academics, and students, but mismatches exist in their definitions and characteristics. A higher proportion of the literature focuses on professional users, possibly due to accessibility, making group comparisons challenging. Despite this, it became apparent that there appeared to be a consensus in many of the published works that DCH environments' main audiences are professional or academic users, which upon closer investigation may not be an accurate representation. Similarly, the term "General Public" was often used as a catch-all for user groups outside of the publications main purpose, which may also prove that these users are the most common visitors to museum websites and require closer attention.

Investigating why users interact with a website involved examining their information-seeking behaviour. Most existing models focus on professional or expert users, assuming they have a work-related task and a known information need. However, these models overlook users who visit for leisure. Elsweiler et al. (2011) model attempted to address this, although it was still assumed that the user has an information need and therefore does not capture the nuances of casual leisure information seeking in DCH, such as cultural and emotional factors. Furthermore, the existing models treat users as static entities, failing to consider their evolution, learning, or the impact of external factors on their behaviour (Bates, 1979).

The literature review also highlighted the systems and interfaces used in DCH and their impact on museum bounce rates and the underexposure of collection artefacts. Traditional search-based systems hide collections, favouring users with specific queries and subject expertise. This approach overlooks users without predefined needs or query formulation skills, excluding them from accessing lesser-known items. To address this, Rich Prospect browsing, and Generous interfaces aim to open collections for browsing. However, screen size limits Rich Prospect browsing, making items unrecognisable due to their small size, while Generous interfaces show a selection of items but risk making the collection appear disjointed or hidden from users.

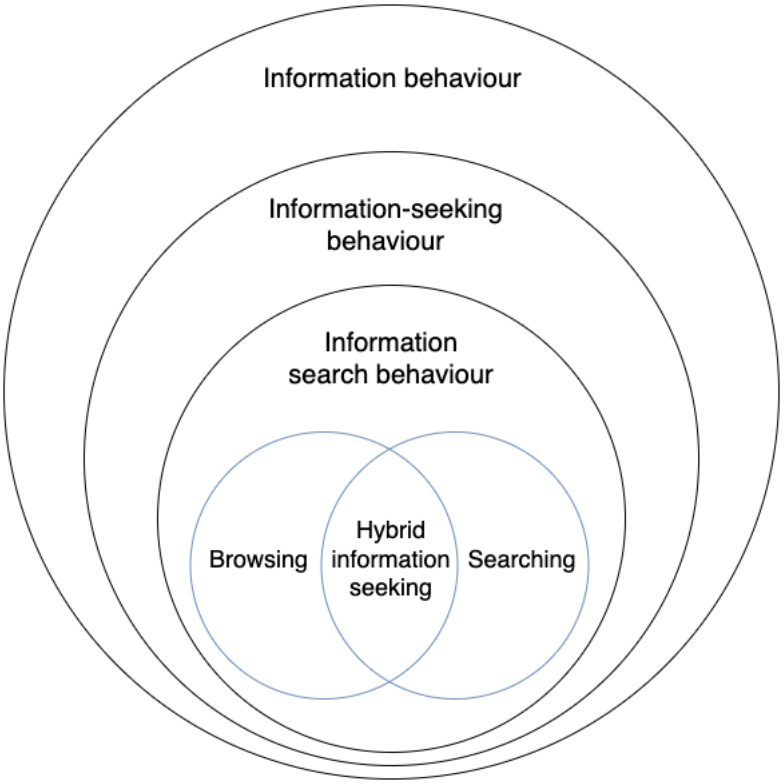


Figure 2.19: An adapted version of Wilson's (1999) nested model of information behaviour research that encompasses browsing and searching and hybrid information seeking

Figure 2.19 shows an adapted version of Wilson's (1999) nested model, the browsing and search circles are shown as subcategories within the information search behaviour circle, representing the specific strategies people employ during the information-seeking process. The overlap between browsing and searching represents a blended

information-seeking behaviour, where individuals use both strategies interchangeably and simultaneously to find information. This combined approach allows users to leverage the strengths of both browsing and searching, depending on their evolving information needs and context.

Recognising the link between user experience and success metrics e.g., lower bounce rates and increased engagement, is at the forefront of the web industry. Understanding users' wants, needs, abilities, and motivations enables UX teams to develop tailored solutions. However, many of the systems depicted in the literature lack this comprehensive understanding, resulting in a mismatch between what is required by those within the DCH environment and what is needed by others.

This thesis will investigate the users of DCH and produce a comprehensive view of the user groups capturing all characteristics/dimensions, which will then assist in developing a method that groups can be studied in an economical and quick yet comparable way so museums, as well as researchers, can create and report their own studies of groups.

Chapter 3

METHODOLOGY

3.1 Introduction

This chapter introduces the methodological approach adopted for this work and how it has guided data collection, analysis and development of theory. This involves presenting the fundamental beliefs and rationale driving the research process through an appropriate mechanism.

There are various strategies that can be used for investigating problems (Cohen et al., 2018), though each approach to research has its own qualities. According to Cohen et al. (2018), paradigms do not necessarily drive the research, but instead, research is driven by the purposes of the investigation. Hence, due to the continually evolving nature of society's demand for technological advancement, this research has been situated in real-world practice-oriented focus informed by a mixed-methods research approach.

The discussion will, therefore, concentrate upon the methodological challenges of undertaking a "real-world" problem through the philosophical lens of pragmatism (Yvonne Feilzer, 2010). A pragmatic world view (Creswell, 2014) allows the researcher to concentrate upon the research question(s) (Johnson and Onwuegbuzie, 2004), rather than be preoccupied with ontological and epistemological debates about reality and laws of nature (Cherryholmes, 1994).

The chapter will also discuss the essential background and fundamental guidelines common in the spectrum of approaches to a mixed methodology; the subsequent sections describe data collection and the phases for this study. The remainder of the

Chapter is organised as follows: Section 3.2 outlines the research paradigm; Section 3.3 describes the Pragmatic Mixed Methods Approach. Section 3.4 describes the research design used in this work based on an adapted version of the Mixed Methods approach. Section 3.5 discusses the ethics, and Section 3.6 provides a summary of the methodology.

3.2 Research Paradigm

According to Candy (1989) all theories can be grouped into three main taxonomies: Positivist, Interpretivist and Critical paradigms. The inclusion of a fourth, Pragmatism (Tashakkori and Teddlie, 2003) converges elements of the three to form what can be considered a merging of diametrically opposed positions (Gage, 1989).

Pragmatism is essentially practice driven (Denscombe, 2008) with truth and reality being in a state of flux, e.g. alternating between subjective and objective, scientific and humanistic; preferring utility, practical consequences, outcomes and heurism over the dogmatic rigid pursuit of an idealised single representation of "reality". Pragmatism is a straightforward, solution-seeking, practical approach to life and problems, with research being judged by whether it has found out what was needed to be known regardless of whether the data and methodologies are quantitative or qualitative (Yvonne Feilzer, 2010).

Establishing "what works" through practical "creation" and "artefacts" to solve problems assist in determining the world as opposed to a distinctive form of enquiry. Flexibility and adaptation of frames of reference are preferable to the assumptions that problems cannot change or mutate and once "solved" remain that way. No longer a "slave" to methodological loyalty, academic community or social context (Oakley, 1999) pragmatism adopts a methodologically diverse, pluralist approach to research drawing on multiple epistemologies while placing the emphasis on "fitness for purpose" and relevance regarding "reality" as both objective and socially constructed (Johnson and Onwuegbuzie, 2004).

According to Cohen et al. (2018), utilising a combined approach to investigation enhances the quality of the research (Suter, 2005). While Chatterji (2004) argues that in order to discover "what works" a mixed-methods approach is unavoidable. Pragmatism has its own standards of rigour and is not an unprincipled approach

in that the research must deliver useful, practicable, reliable and valid answers to questions put by the research (Cohen et al., 2018).

3.3 Pragmatic Mixed Methods Approaches

This study used a mixed-methods design (Tashakkori and Teddlie, 2003), which is a procedure for collecting, analysing and "mixing" both quantitative and qualitative data at some stage of the research process within a single study to develop a more complete understanding of a research problem (Creswell, 2003). The rationale for mixing is that neither quantitative nor qualitative methods are sufficient by themselves to capture the trends and details of the situation, such as the engagement levels of users with DCH. When used in combination, quantitative and qualitative methods complement each other and allow for more complete analysis (Greene et al., 1989; Tashakkori et al., 1998).

In quantitative research, there is a reliance on numerical data (Charles and Mertler, 2002). Post-positivist claims are used to generate knowledge in ways such as cause and effect reasoning, reducing variables to particular variables, hypotheses and questions, measurement and observation, and theory testing. To evaluate the degree and frequency of correlations, a researcher separates variables and causally connects them. Furthermore, the researcher will pick the factors to explore and tools that will generate extremely trustworthy and accurate results.

Alternatively, qualitative research is "an inquiry process of understanding" where the researcher develops a "complex, holistic picture, analyses words, reports detailed views of informants, and conducts the study in a natural setting" (Creswell, 1998). Knowledge assertions in this method are based on constructivist (Guba et al., 1994) or advocacy/participatory (Mertens, 2014) viewpoints. Data is obtained from persons engaged in the everyday life of the place in which the inquiry is structured in qualitative research. The values that these people perceive for their world are used to analyse data. Finally, it "produces an understanding of the problem based on multiple contextual factors" (Miller and Slater, 2000).

In a mixed-methods approach, researchers build knowledge on pragmatic grounds, (Creswell, 2003) asserting truth is "what works" (Cohen et al., 2018). Approaches are chosen, in addition to variables and units of analysis, that are more applicable

to finding an answer to the research question (Tashakkori et al., 1998). A major principle of pragmatism is that quantitative and qualitative methods are compatible and that both numerical and text data, collected sequentially or concurrently, can help better understand the research problem.

When developing a mixed-methods study three key considerations must be addressed: priority, implementation, and integration (Creswell, 2003). Priority involves determining whether quantitative or qualitative methods receive more emphasis in the study. Implementation pertains to the sequence of data collection and analysis—whether they occur sequentially, in stages, or concurrently. Integration refers to the phase of the research process where quantitative and qualitative data are combined or linked.

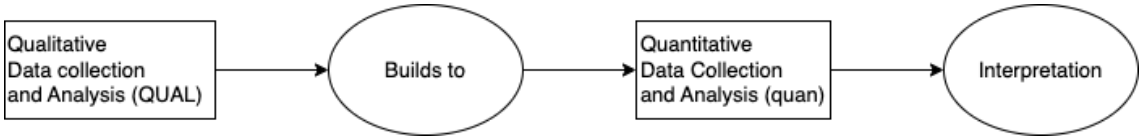


Figure 3.1: Exploratory Sequential Mixed Methods Process (Creswell, 2014)

An exploratory sequential mixed-methods design commences with an initial qualitative phase of data collection and analysis, and then uses the findings in a second quantitative data collection and analysis phase that builds on the initial phase. The purpose of this approach is to develop confidence with specific samples of populations and to see if the data from a few individuals (qualitative phase) can be generalised to a large sample of the population (quantitative phase) with a final integration phase that draws together the data from the separate strands (Creswell, 2014).

To undertake this research, the NML website was used in order to be able to conduct an in-depth exploration of those user groups that interact with the website and web collections. NML was studied in a cross-sectional manner to ensure the users were the natural audience of the website. Taking a snap-shot of the audience at a specific point facilitated a comprehensive comparison of the different dimensions/-variables that make up the characteristics of each user group, such as demographics, domain knowledge, general CH knowledge, location, goal, task, behaviour etc.

3.4 Research Design

This research used an adapted version (see Figure 3.2) of the traditional exploratory sequential mixed-methods design (see Figure 3.1), consisting of five distinct phases (Creswell, 2002, 2003).

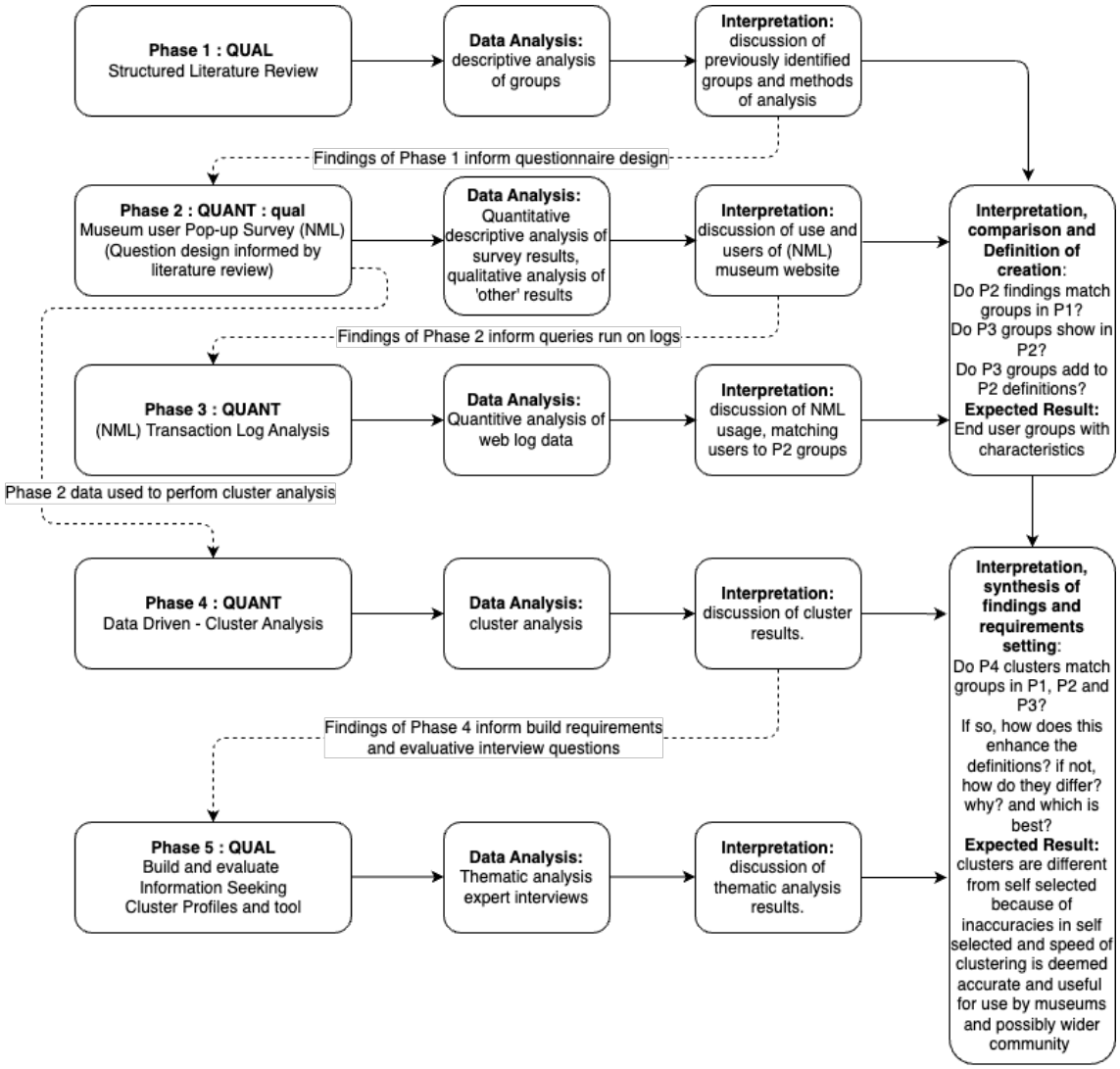


Figure 3.2: Exploratory Sequential 5 Phase Mixed Methods Research Design

3.4.1 Phase 1: Establishing current DCH User Groups through the Literature: The Systematic Literature Review

Developing an online resource for diverse audiences presents a significant challenge. To begin this process involves gaining a thorough understanding of these audiences,

given the extensive and varied research conducted in DCH and its audiences already. Therefore, exploring the literature to identify and understand user groups and their characteristics is a logical starting point. However, an initial review reveals a lack of comprehensive studies on the characteristics of DCH museum user groups. Specifically, there was no existing work to identify what user groups have been defined and used within the cultural heritage and wider GLAM literature, and what any identifying and distinguishing characteristics may be.

According to Tranfield et al. (2003), a systematic review process offers a reliable foundation for designing research, as it is based on a more comprehensive understanding of what is known about a subject. However, obtaining relevant publications from established literature can be a complex process. As suggested by Grant and Booth (2009) "the expansion in evidence-based practice has lead to an increasing variety of review types", and as such requires an appropriate strategy to identify what is relevant and what can be left out.

A systematic literature review (SLR) identifies, selects and critically appraises publications in order to answer a clearly formulated question (Ndayizigamiye, 2022) and should follow a clearly defined plan where the criteria are clearly stated before it is begun. An SLR should be a comprehensive, transparent search conducted over multiple databases. It should involve a well thought out search strategy that has a specific focus or answers defined questions. In addition, the SLR must include the type of information searched, critiqued and reported within known timeframes; search terms, search strategies (including database names, platforms, dates of search) and limits to ensure repeatability.

Phase 1: Objective:

OB1: To systematically review and categorise existing user groups within the digital cultural heritage literature, identifying gaps and inconsistencies in current user group definitions.

3.4.1.1 Systematic Literature Review: PRISMA

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), is an evidence-based minimum set of items aimed at helping scientific authors to report a wide array of systematic reviews and meta-analyses, primarily used to assess

the benefits and harms of a health care intervention. PRISMA focuses on ways in which authors can ensure a transparent and complete reporting of this type of research.

The PRISMA strategy offers the replicability of a systematic literature review through the adoption of a hierarchical process. Essentially researchers define research objectives aligned with the research question, specify keywords, and develop exclusion and inclusion criteria. During the review stage, relevant articles are identified, and irrelevant ones are removed. Articles are then analysed according to predetermined categories. The PRISMA statement consists of a 27 item checklist (Page et al., 2021) for assessing the quality of a systematic review, along with a four phase flow diagram (see Figure 3.3).

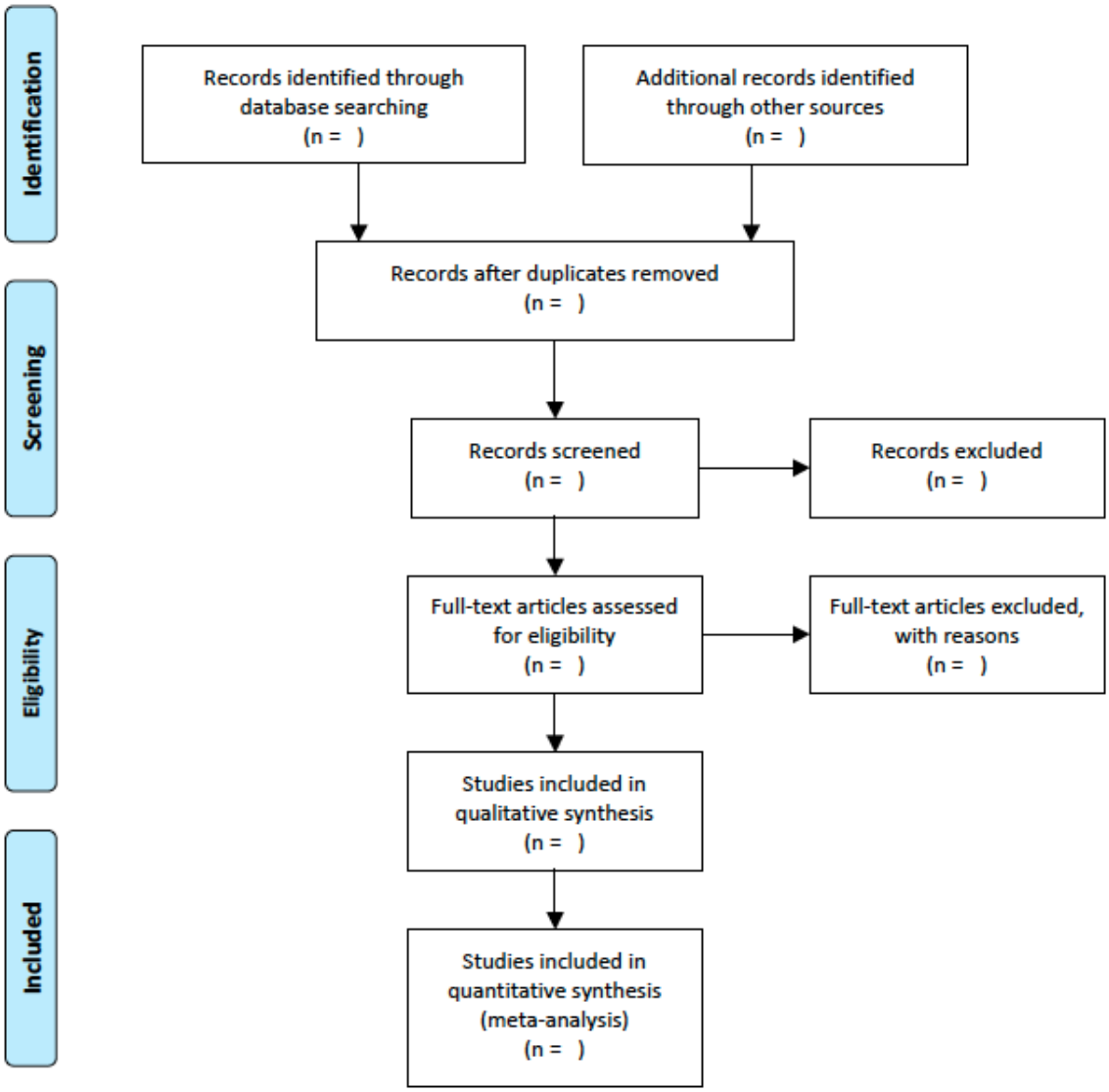


Figure 3.3: PRISMA 2009 diagram. Moher et al. (2009)

The phases consist of:

Stage 1 Identification - publications are identified from named databases through the incorporation of keywords. Keywords are words or short phrases that represent the main ideas of the research topic or question. Searching a computerised database allows for the combining of keywords using AND and OR. Using AND to combine keywords will retrieve publications that include both of the terms, while using OR will retrieve publications that include either of the terms;

Stage 2 Screening - the title and abstract of each publication is read to determine whether it contains material that would be relevant or helpful to the systematic review. This is usually a simple "yes/no" choice with reasons for exclusion noted;

Stage 3 Eligibility - requires the full publication to be read in order to determine if the publication is relevant to the field being looked at/and or the research question being posed. However, in order to ensure boundaries e.g., what is acceptable and what is not, for the SLR, Inclusion and Exclusion criteria are developed, usually after setting the research question but before the search is conducted;

Stage 4 Included - Finalising the list of publications to include in the review requires a decision regarding how many of these publications can be included in a quantitative synthesis, or meta-analysis, is made.

3.4.1.2 Analysis

In addition to Stage 4 of PRISMA, a temporal statistical analysis was conducted on the data to establish the existence of a timeline of popular user group labels used among authors. Temporal statistical analysis enables the examination and modelling of the behaviour of a variable in a data set over time (e.g., to determine whether and how concentrations are changing over time) (I.T.R.C., 2013). This type of analysis is used to examine patterns and trends in data that is collected at regular intervals, e.g., daily, weekly, monthly, or annually which can then help in

making predictions about future events and/or inform decision-making and support effective problem-solving.

3.4.1.3 Implementation of the research methodology: Phase 1 - Undertaking PRISMA

A review of existing studies of DCH was conducted in April 2016 utilising an adapted PRISMA (see Figure 3.4) (Moher et al., 2009) approach to identify and compare how users had been identified, categorised and grouped in the literature.

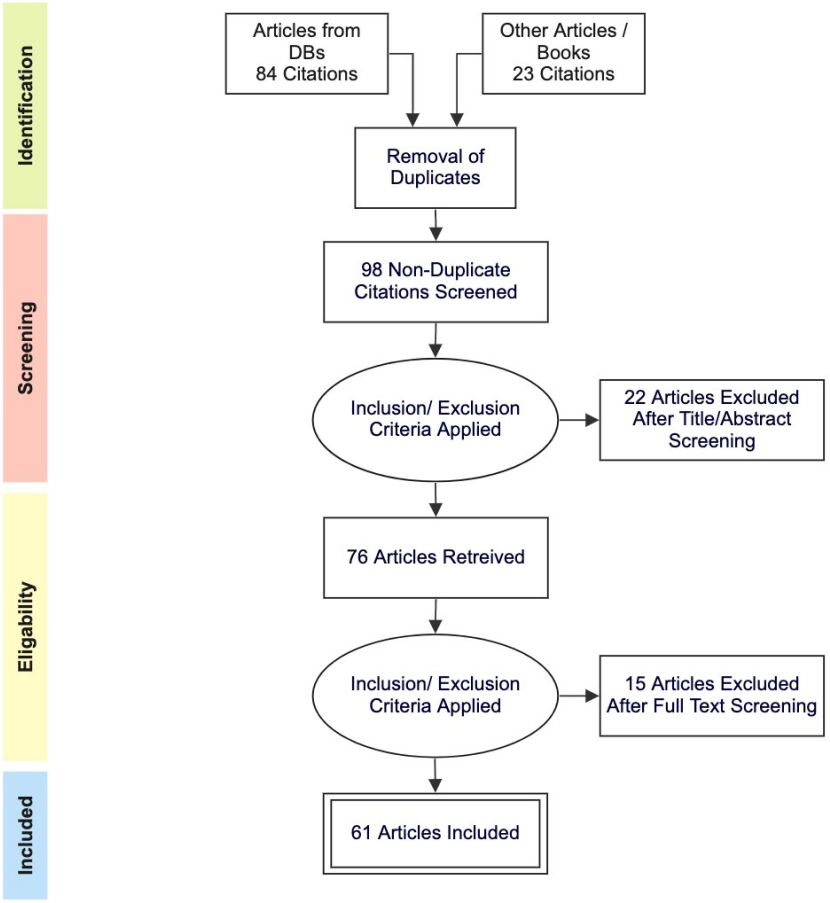


Figure 3.4: Phase 1: PRISMA Methodology Results

Step 1 - Identification: Relevant articles were identified using a search strategy based upon keyword search of computerised databases. Google Scholar was initially used, as this allowed for a number of databases to be searched in one search (Jean-François et al., 2013). An identical search was carried out for each database to ensure nothing was missed that was not listed on Google Scholar. The databases searched were: Web of Science, Wiley Online Library, IEEE Xplore, ACM Digital

Library, SpringerLink, Emerald Journals, and ScienceDirect (Elsevier).

Determining the keywords that were to be used required consideration of the thesis research objective 1, and particularly RQ1 and RQ2 and the questions these raise, such as: How are user groups characterised and identified in DCH literature? As database searches look for the exact words entered into the search box or the synonyms of the words, using these RQs in their entirety did not provide any meaningful results. Therefore, extracting words associated with the RQs and the questions surrounding them would generate a number of potential candidate publications. The keywords that were used in various combinations came from the following four groups: for web-enabled systems, the following keywords were used *web, website, virtual, online*; For Digital Cultural Heritage: *museum, DCH, digital cultural heritage*; For user identifying keywords: *users, visitors, user needs, information needs, user studies, personas* were used; and for interaction keywords *searching, exploring, browsing.* The shorter the length of the key phrase or a single keyword, the more generic the results generated in the search and the inclusion of irrelevant publications in the large volume that is likely to be returned. Using a longer-tailed keyphrase of multiple keywords returns a more precise and smaller more focused set of search results to filter through. In order to generate those more desirable papers, a minimum of 3 keywords and a maximum number of 4 keywords were used - 1 from each group. e.g. online - digital cultural heritage - user needs to a 4 word - online - DCH - users - browsing. Other examples include: 'virtual museum visitors', 'digital cultural heritage website users', 'museum website users', 'museum personas', 'digital cultural heritage user studies' and 'DCH website information needs'. All searches were carried out as exact searches by wrapping the key phrase in double quotes in all searches on all systems. Only papers that had a free downloadable PDF were accepted. Many of the keyword combinations returned no results in any database search. The exact number of results per keyword has been lost between the analysis taking place in April/May 2016 and 2024. However, Appendix A.3 provides a current search restricting the findings to 2016. This search shows that only 13 of the keyphrases return any results (2367 of the 2380 possible combinations produce nothing) and that the number with a PDF available is considerably less (25% of results).

The types of publications generated by the use of the keyphrases included studies

of museum websites and online digital collections, digital cultural heritage users and user behaviour, and search patterns and visitor journeys, resulting in eighty-four publications with a PDF being identified for review.

After conducting some snowball sampling based on the retrieved publications (Radjenović et al., 2013), an additional twenty-three publications were added. These were mostly white papers (often unpublished) from funded projects, such as Europeana, the PATHS[1] and CULTURA[2] projects as well as PhD thesis. The publications were then filtered for duplication (publications that appeared in more than one database), which resulted in a total list of ninety-eight publications to start screening. The full list of **98** publications is listed in Appendix A Table A.1.

Once the publications were identified and collected from the keyword search, additional steps were taken to ensure relevance to the research area through undertaking an initial publication overview to a full read of said publication.

Step 2 - Screening: A single manual screening approach was adopted, as this was an efficient use of time and the resources available, using the following inclusion and exclusion criteria:

1. Inclusion
    (a) A user category, role or group or label must be evident in the publication; or
    (b) User motivations were identified for groups of users accessing DCH websites.
2. Exclusion
    (a) Publications that did not meet said inclusion criteria, or
    (b) Studies identifying user groups only in physical cultural heritage settings; or
    (c) Publications cited in digital cultural heritage referring to prior categories, such as those used in healthcare (Cifter and Dong, 2008).

[1] `http://www.paths-project.eu/`
[2] `http://www.cultura-strep.eu/`

This resulted in the exclusion of 24 publications when screening the titles and abstracts (see Appendix A Table A.1).

Occasionally, the title or abstract of the publication provided little by way of identifying a user, group or role, which resulted in a need to review the full text. Full-text inclusion and exclusion criteria were created in order to facilitate the eligibility of the publication (see Step 3).

Step 3 - Eligibility:

Inclusion criteria were developed in relation to the following: (i) a user category, role or group or label must be evident in the publication, together with at least one user group's characteristics; or (ii) user categories, role or group must be evident in the publication even though motivations were less clear, but the boundaries of the groups were apparent; or (iii) user motivations were identified for groups of users accessing DCH websites.

Exclusion criteria consisted of the following: (i) publications that did not meet said inclusion criteria, or (ii) studies identifying user groups only in physical cultural heritage settings; or (iii) publications cited in digital cultural heritage referring to prior categories, such as those used in healthcare (Cifter and Dong, 2008).

After applying these criteria to the full texts a further 15 publications were excluded due to the eligibility rules. This final set of 59 publications consisted of 2 books (Stebbins, 2007; Falk, 2009), 1 thesis (Lin and Gregor, 2006), 1 report (Goodale et al., 2011) and 55 papers that were used for the extraction of the user groups and their characteristics (see Appendix A Table A.2)

Step 4 - Data Preparation: The data preparation stage required the identification of all potential user group labels in the publications to be established. Data were extracted in relation to the user group labels, the author's name, year of publishing, the paper's theme, the types of participants used for the study and the type of GLAM the publication focused on.

Once identified, each of the 59 publications were read in full. From the text, any user group label identified was extracted along with any descriptive analysis and or associated characteristics that the authors provided in terms of an explanation for

the label usage. Phrases were extracted from the publication that related wholly to the user group label being used. for example, Cifter and Dong (2008) identified that professional users "*have a well-defined goal for his/her activities, that they act within the formal part of the economy. They usually have good knowledge about the task that they perform with the product: they are trained, and they may have previous experience with the product or the relevant task. They also have gone through extensive training to achieve particular knowledge which is valuable in a social or economical context*" (A complete set of results can be found in Appendix B). Additionally, the removal of some user group labels was also conducted at this stage due to a user group label or its definition being overtly generic in that it could be applied to multiple user groups (e.g. visually impaired user, disabled or book lover, could apply to a user from any group), not meaningful (e.g. use of an Intermediate label). Where publications focused their study on a single user group label (e.g. experts) and then indicated that other potential user group labels would be combined as a single user group label (e.g. general public, users) the ill-defined nature of the user group label would exclude the generic label from the current study (see Appendix C.1).

3.4.1.4 Basic Data Set Cleaning

After establishing the primary data set of 169 user group labels from the 59 publications (see Appendix B) a basic data cleaning step was applied e.g., merging duplicate groups. Duplicate groups were defined as two (or more) groups that appear in a different sources and

- are plural/singular uses, such as "Researchers" / "Researcher";
- differ only in a generic suffix, such as "Student" / "Student group";
- have other minor language differences that do not imply any actual difference, such as "Lay user" / "Laymen"

Applying these definitions resulted in a 70 group labels being merged, resulting in a total of 95 user group labels (see Appendix C.2) that form the actual analysis.

3.4.2 Phase 2 Survey of DCH Users (NML)

Developing a system for user groups by merging characteristics from various publications found in the literature or using incomplete definitions and understanding of the groups could lead to numerous issues.

However, recognising the importance of understanding customers is vital for achieving success. By gathering information about customers, an organisation gains deeper insights into their behaviours and needs. Identifying and comprehending customers' preferences enables the delivery of exceptional service that meets and exceeds their expectations. This, in turn, fosters a positive organisational reputation, increases visibility, enhances conversion rates, and propels growth (Stickdorn and Schneider, 2011).

Therefore, to acquire insight into genuine DCH customers or users, a comprehensive survey will be conducted among users of the NML website.

Phase 2: Objective:

OB2: To conduct a comprehensive survey among DCH website users to identify characteristics and behaviours of distinct user groups.

3.4.2.1 National Museum Liverpool

National Museums Liverpool was established as a national museum in 1986 with origins dating back to 1851 when the 13th Earl of Derby donated his natural history collection to the town of Liverpool, founding the Liverpool Museum.

National Museums Liverpool is one of 12 national museums in England and Wales and the only national museums group created in its own right within an English regional city. Today, National Museums Liverpool consists of 8 museums and galleries: The World Museum, the Walker Art Gallery, Merseyside Maritime Museum, the International Slavery Museum, the Museum of Liverpool, Sudley House, Lady Lever Art Gallery and the HM Revenue and Custom National Museum. Each museum/gallery being housed in their own physical building located in or around Liverpool city centre, protecting an outstanding heterogeneous collections comprised of over 4 million objects related to art, history (spanning from Ancient Egypt to slavery) and science.

Visitor levels to the physical museum has seen some fluctuation in numbers. In the year 2018/19 the physical venues received over 3.9 million visitors. With the World Museum being the most visited museum outside of London, closely followed by the Museum of Liverpool and the Merseyside Maritime Museum. The year 2019/20 saw slightly fewer visitors (probably due to the closure of the museums as part of the covid restrictions) with just over 3 million visitors, 44.7 thousand of which were visitors from overseas.[3] [4] [5]

3.4.2.1.1 NML Website

The NML have a 2030 vision that includes a 'for everyone' part, which aims to have over 6 million visitors engaged each year through digital. Seven of the eight museums currently have an accompanying website, which saw more than 2.3 million visitors in both the years of 2018-19 and 2019-20.

The NML website (see Figure 3.5) is divided up so that each museum/gallery have their own area. There was also a dedicated kids area (removed in the website revamp in 2019), shop and some generic information pages about the group linked from the NML homepage.

3.4.2.2 Examining Museum (NML) Users

To gain an understanding of the users and their experience when interacting with the NML website, the main data collection method for this phase was an online survey. The survey was chosen in order to gain responses from as wide a range of the users as possible who are not reachable by face to face surveys or interviews in the physical settings, as it is assumed that the online users may be different users to those that visit in person. Even though the potential audience is international, the survey was presented only in English as the NML website exists in English only and as such it is reasonable to assume that potential participants would have sufficient English skills to respond to the questions accordingly.

The advancement of technology has seen the transformation of this data collection technique through digitisation and hence the ability to reach far ranging

[3] https://www.liverpoolmuseums.org.uk/about/index.aspx

[4] https://www.liverpoolmuseums.org.uk/wml/history/WML_150_years.pdf

[5] https://annualreview.liverpoolmuseums.org.uk/#highlights

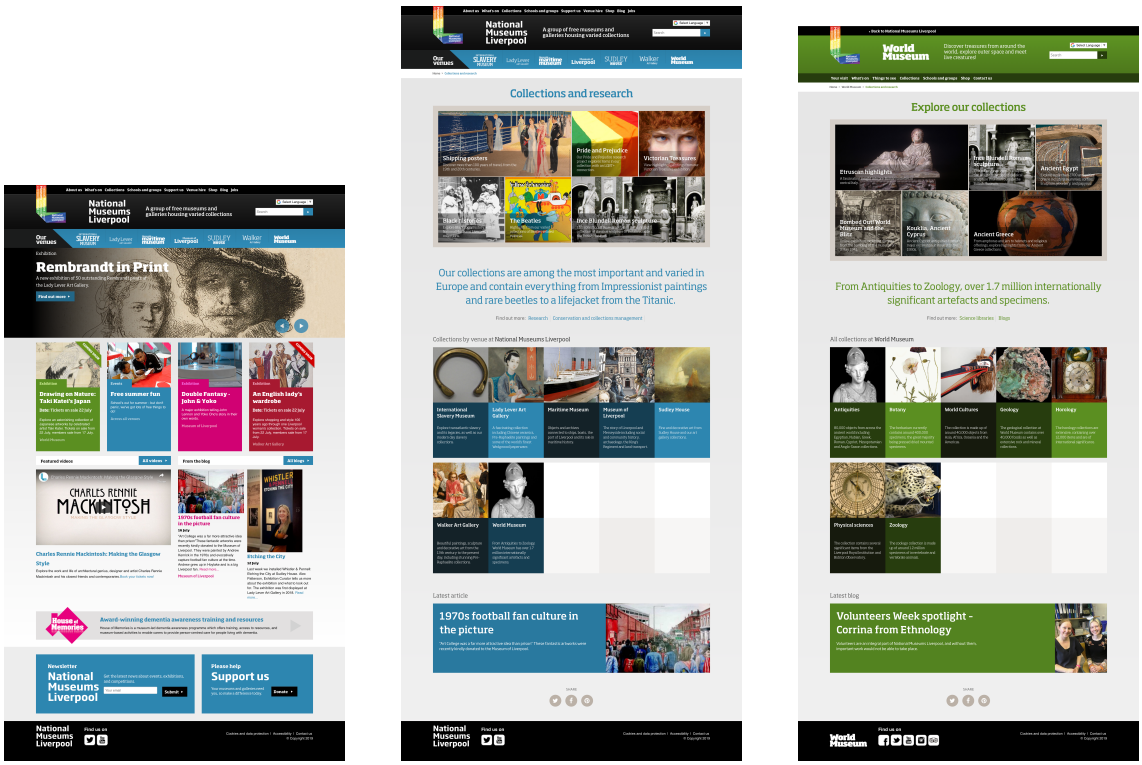


Figure 3.5: NML webpages - Homepage - Collections homepage - WML Collection page

audiences as well as having the potential to collect large amounts of data efficiently (i.e. with less error due to the lack of transferring written data on to a computer), economically (as it requires low human resource efforts while collecting or managing data), all within a relatively short time frames (Regmi et al., 2017). However, the disadvantages noted by Rivas (2006) e.g., participant/response/self-selection bias, high non-response and non-completion rates, and the potential for question miscomprehension by participants, are still relevant.

3.4.2.3 Survey Design

It was agreed in collaboration with NML that the survey would not exceed 25 questions due to historically low completion rates when exceeding this limit. Keeping the survey to 22 questions, therefore, almost ensured a good response rate and minimised any potential inconvenience to the participants.

The design of the questions and groupings (see Table 3.1) is centred around the six dimensions in which users have been previously studied individually (refer to Phase 1 - Motivation, Task, Engagement, Domain and Technical expertise, shared

experiences). Additional questions were incorporated to capture preferences for content types, perceived usage specific to the NML website (such as times and days), as well as general demographic information. Lastly, the users were asked to identify what group they would place themselves into for this visit to the NML site. This last question offered the users a series of labels taken from literature and also offered an "other" option where the users could provide their own label if they wished.

Table 3.1: Survey Questions

Question Group	**Questions**	**Sources**
Motivation	**Today I am visiting the NML website: (Select one)** : For personal reasons; For study (College, Further education, University); For work; For no specific reason or to pass time.	Spellerberg et al. (2016); Fantoni et al. (2012)
Task	**What is the primary purpose of your visit to the NML website today?**: Gain knowledge of a specific museum object (a certain clock (The Barry astronomical clock) or statue (Huskisson Statue) for example); Gain knowledge of a specific type of object (Egyptian burial objects); Gain general knowledge of a collection (our art collection, the Egyptian or horology collection); Gain knowledge of the collections that the museums exhibit; Prepare for a visit to the museum; To buy an item from the online gift shop; To keep up to date with the museums news via the blogs; I don't know.	europeana (2014); Skov (2009); Stiller (2014); Goldman et al. (2004)
Preferred Content	**Please rank the following content types in order of your preferred preference (e.g. move your most preferred to the top of the list, and your least to the bottom)**: Text; Pictures; Videos; Audio (speech or music); Interactive/hands-on content (quizzes, virtual tours).	Goldman et al. (2004); Pitkow and Recker (1994)
Shared Experience	**Are you sharing this website experience?**: No, I am on my own; I am looking with a child/children by my side; I am looking with another person remotely;	Goldman et al. (2004)
Engagement	**How frequently do you visit the NML website?:** This is the first time; A couple of times a year; Monthly; Weekly; Daily. **Roughly how many web pages have you viewed on the NML website during this visit?:** Just the one; 2 - 10; 11 - 20; 21+. **When on a web page about an object or collection, do you typically?:** Read the whole page from top to bottom in depth; Scan the page to find the relevant words you are looking for?; look for pictures first and then read;	europeana (2014); Russell-Rose and Tate (2012) Belkin et al. (2009); Russell-Rose and Tate (2012); Pitkow and Recker (1994) Taylor (2015); Pitkow and Recker (1994)
Domain Knowledge	**In the context of cultural heritage and your current visit to the NML website, please select the appropriate statement:** I have some experience and background knowledge; I'm highly experienced and have extensive background knowledge; I'm a novice with little knowledge within this area; I don't know **Rate your general Cultural Heritage knowledge (1 = low and 5 = high) using a Likert scale**	Skov (2009)
Technical expertise	**When seeking information on a website, which method do you prefer?:** The search box; Using content links or navigational buttons.	Taylor (2015); Pitkow and Recker (1994); Spink et al. (1999)

	When using the search box to search for content that you do not find in the first set of results, do you?: Give up; Search a further 2 or 3 times; Persist until you find the material/answer you were looking for.	Mansourian (2008); Spink et al. (1999, 2001)
.	**Please rate your proficiency with the Internet tools and services (only rate those that are relevant) 1 = low (used a few times) 5 = high (use it all the time):** Facebook; Twitter; Instagram; Professional networking (LinkedIn); Internet banking; Online shopping; Streaming music service (Spotify); Streaming Video (YouTube, BBC iPlayer); Document creation and reading (Google Docs).	Russell-Rose and Tate (2012)
Usage	**What time of day do you usually undertake the type of activity you are primarily using the NML website for today? (Please select all that apply):** Morning; Lunchtime; Afternoon; Evening.	Lapatovska et al. (2011)
	What type of device are you using for this visit to NML website?: Mobile Phone; Tablet PC (iPad, Amazon Fire, Android tablet etc); Desktop or Laptop PC	Pitkow and Recker (1994)
Demographics	**How old are you?:** Under 18; 18 - 34; 35 - 54; 55 - 64; 65 - 74; 75+	Pitkow and Recker (1994); Hall et al. (2012a)
	What gender best describes you?: Male; Female; Rather not say	Hall et al. (2012a); Pitkow and Recker (1994)
	Please select completed educational programs/-courses?: Secondary School (GCSE's, O-levels); Further education / college diploma ; Undergraduate Degree; Master Degree; Doctorate; Professional Certifications; None; Other	Hall et al. (2012a); Pitkow and Recker (1994)
	Please select in progress educational programs/-courses?: Secondary School (GCSE's, O-levels); Further education / college diploma ; Undergraduate Degree; Master Degree; Doctorate; Professional Certifications; None; Other	Hall et al. (2012a); Pitkow and Recker (1994)
	Where in the World are you at the moment?: Merseyside; Northwest England; Rest of England; Wales, Scotland or Ireland; Outside of the UK	Pitkow and Recker (1994)
	Which of the following categories best describes your employment status?: Employed, working full-time; Employed, working part-time; Student (Full time); Student (Part time); Not employed, looking for work; Not employed, NOT looking for work; Retired; Disabled, not able to work	Hall et al. (2012a)
Groups	**Which of the following groups would you place yourself in for this visit to the NML website? (Select as many as you think applicable):** Non-professional researcher (hobbyist, amateur historian); Professional researcher (historian, genealogist); Academic (scholar, post doc researcher, academic support); Teacher; Student (college, university, further education); Museum staff (curator, archivist); General public/user; Other	Not asked in previous literature.

Many of these questions are taken directly or are slight adaptations of questions found in and previously asked in the literature, the sources for which are indicated in the source's column of Table 3.1. The literature in Phase 1 and the literature review chapter focussed on the users of DCH and explored the characteristics often in isolation. In order to make the question set coherent, some of the questions required a slight change for example Fantoni et al. (2012) asked "What is your main reason for visiting the IMA website today?" which was changed to "Today I am visiting the NML website: (Select one): For personal reasons; For study (College, Further education, University); For work; For no specific reason or to pass time." changing the question from an open-ended question to a closed style with the options being the findings from the Fantoni et al. (2012) work.

Ordinarily, age bands are grouped in 10 year band widths. However, there are examples where alternative groupings have been adopted, for example: Dewing (2012) when studying social media usage used 16 to 34, 35 to 54, 55 to 64 and 65 and over. Rivas et al. (2020) used similar groupings in the research of developing classifier models and analysis of user demographics for medical social media posts. However, Rivas et al. (2020) used five groups: 0 to 17 years, 18 to 34 years, 35 to 44 years, 46 to 64 years, and older than 65 years. This type of grouping appears frequently within the medical user research and general user research fields (Mathis, 1999; on Longevity, 2016). Mathis (1999) adopted a slightly adapted grouping of 18 to 34, 35 to 54, 55 to 64, 65 to 74 and over 75's. This allowed for a more granular look at the older population. Based on the research within the DCH field of hobbyists and how these tend to be spread over the older generations, this seemed a good format to adopt for this research. The minimum age for the study was required to be 18 for both the ethics and the museum. Therefore the age categories settled on for this study were: 18 to 34, 35 to 54, 55 to 64, 65 to 74 and 75+. The 18 to 34, will capture individuals who are at the beginnings of their interests in DCH, research or careers. The 35 to 54, participants, usually enjoying the influential years of their careers. The 55 to 64 will capture the pre-retirement group, the 65 to 74 the newly retired group, and the 75+ will capture the established retired users. While this has some limitations, e.g., due to variation of the age of retirement; the scale aligned with previous work in this area.

3.4.2.4 Creation of the Popup Survey

The use of questionnaires permits the collection of both qualitative and quantitative data using closed response or scaling questions (Payne and Whittaker, 2006).

Following a descriptive design focusing on the areas of the survey will allow this study to count the population which have a certain opinion (Oppenheim, 1992), specifically the users' perception of their levels of various engagement components; cognition, emotion, participation and performance similar to other studies researching in the area of DCH.

Using self-completion questionnaires allows the individual user to provide information without feeling pressured in the presence of the researcher. This provides a fair chance of receiving responses that contain both truthful and accurate information, as the user had the time to thoroughly consider the information that they were providing (Walliman, 2006). For this research, an anonymous questionnaire (Appendix A) was distributed to the user of the NML website. The questionnaire was a list of questions in a format that enabled a range of answers to be obtained in a relatively structured manner, supporting an easier data analysis process (Matthews, 2010).

3.4.2.5 Pilot Survey Testing

In order to validate the proposed questions, an initial pilot survey was conducted with 8 participants, made up of NML Museum professionals and academic researchers. Each pilot user was asked to complete the survey and then to provide any additional comments they felt pertinent in relation to the survey format, delivery mechanisms, or any other comments they wanted to make in order to enhance the final version of the survey prior to its publication (see Appendix G)

Subsequent amendments were made based on the feedback received from the participants, for example, reducing the amount of information provided in the Consent form (Participant 1) and ensuring that the "do not wish to participate" link does not go to the Survey completed page (Participant 2). Question amendments were made in line to comments relating to simplification of wording (Participant 2 and Participant 4) (see Appendix G).

3.4.2.6 Ethical Issues Regarding Questionnaires

The questionnaire, whether traditional or digital, will always be seen as an intrusion into the life of the participant, be it in terms of time (survey completion), sensitivity or invasion of privacy (questions being asked).

Therefore, some important ethical issues to adhere to when conducting a survey are confidentiality, informed consent and the right to withdraw consent at any time. The respondent's right to confidentiality should always be respected, and any legal requirements on data protection should be adhered to. In the majority of surveys, the participant should be fully informed about the aims of the survey, and the participant's consent to participate in the survey must be obtained and recorded. Equally, it should be recognised that participants have the right to change their minds and, as such, withdraw consent, which again requires the provision of information regarding how to do this.

To address these ethical situations, the survey was designed to be as unobtrusive as possible. The user was invited to participate via a pop-up, which allowed them to ignore or contribute. If they ignored it, then they would not see the invite again, respecting their wishes not to contribute. For those users who did wish to respond then, they were shown details of how and where their response data would be used, who had access to it and also informed that they could exit at any time and that none of their personal details would be asked for or stored. The data was also cleaned initially, removing any accidentally added personal details by auto form fillers etc, and all IP addresses were anonymised. Any partially completed surveys were also removed.

3.4.2.7 Implementation of the Research Methodology: Phase 2 Survey Deployment

The survey was set up using Poll Daddy (now rebranded as Crowd Signal [6]) and hosted on the NML website, where participants were recruited via a small banner pop-up, which appeared after a 10-second delay of the page loading (see Figure 3.6). The Poll Daddy system was chosen on the advice of NML as they had used it on previous surveys they had run and knew that it both integrated well with the NML

[6] `https://crowdsignal.com/`

website and also that users were already familiar with how it looked. The decision to use an intercept pop-up over an instant or exit pop-up survey was primarily to capture a potentially wider audience during their initial interactions with the website (Andrews et al., 2003); this being based on the users' freshly aroused interest as opposed to an "instant pop-up" which can be perceived as aggressively interfering with the users valuable time and content consumption, or the "exit pop-up" which can often be skipped or seen as a nuisance by users who think they have left the site.

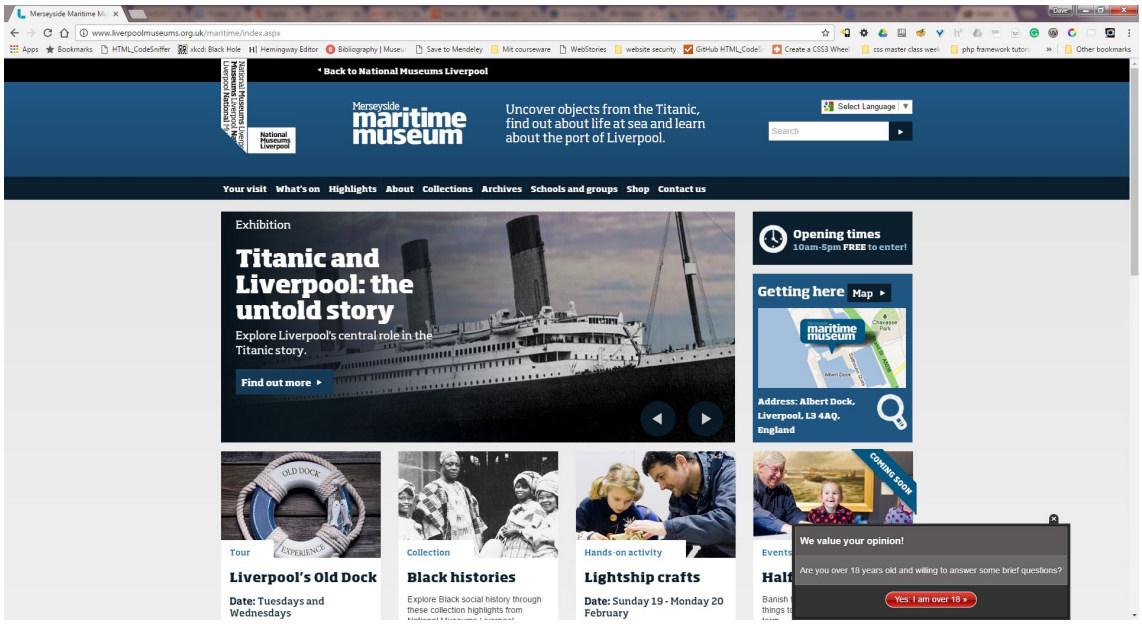


Figure 3.6: Survey pop-up

To ensure the full potential spectrum of users was captured, the survey was available for a two-week period (from 1/2/2017 to 14/2/2017) via the museum areas of the website: World Museum [7], International Slavery Museum [8], Sudley House [9], and the Maritime museum [10]. For a second two-week period (from 15/2/2017 to 1/3/2017) it was made available via the Galleries areas: Walker art gallery [11] and Lady Lever art gallery [12]. The reason they were not run simultaneously was because the Galleries had just run their own survey and there was a possibility of survey

[7] http://www.liverpoolmuseums.org.uk/wml
[8] http://www.liverpoolmuseums.org.uk/ism
[9] http://www.liverpoolmuseums.org.uk/sudley
[10] http://www.liverpoolmuseums.org.uk/maritime
[11] http://www.liverpoolmuseums.org.uk/walker
[12] http://www.liverpoolmuseums.org.uk/ladylever/

fatigue (Porter et al., 2004) impacting upon the potential response rate. The survey was also not conducted on the Museum of Liverpool site due to a previously planned and long running survey the NML were conducting on this one museum site. The survey was restricted to a single completion by an individual, which was enforced via use of cookies to maintain anonymity.

Although self-selection and response biases were recognised as potential issues, given the website's global audience, an online survey with a popup on the actual NML website was deemed the most feasible method to reach the broadest possible respondent base.

3.4.2.8 Data Preparation

During the 28 days the survey was available a total of, 1118 potential participants interacted with the survey pop-up; of which 663 completed survey attempts, prior to the data being cleaned, and 455 abandoned survey attempts were undertaken. This resulted in a 59% response rate.

In order to verify the level of completeness of the 663 surveys, a further examination was performed. The data was cleaned by removing those participant surveys where Question 1 (the consent agreement) was not accepted (90 participants) and where all 22 questions had not been fully answered (45 participants). Next, to ensure all participants were over 18 any participant who had selected the age group of 'under 18' were removed. Age had been specified in the consent form and on the button they click (Figure 3.6) but additional screening was included to capture under 18's who missed the first few mentions, this resulted in 9 responses being removed. The data was then filtered for any personal data and responses where the survey had been abused by non-authentic respondents, resulting in a further 5 removals and leaving a total of 514 (50.4%) response rate.

3.4.3 Phase 3: Transaction Log Analysis (NML)

Whenever users interact with online services, the communication between user and system creates a record of the interaction and kept on the servers owned by the organisation (Jansen, 2006). Although, as technology advances, servers can be configured to ignore the collection of this superfluous data or restrict the type and

time that it is held for. If available, however, these interactions can help to provide insights into users' search and navigation behaviours, such as the pages accessed, time spent on pages, and patterns of use Russell-Rose and Clough (2016) (e.g., sequential pages accessed (Kachhadiya and Patel, 2018) or search queries entered and reformulated (Lau and Horvitz, 1999)).

This phase will utilise the data from the NML museum website server logs in order to gain a more nuanced understanding of how users engage with DCH websites.

Phase 3: Objective:

OB3: To analyse user interaction data from museum website server logs to understand real-world user behaviours, further refining user group categories based on digital engagement patterns.

3.4.3.1 The TLA Process

According to Jansen (2006) the TLA process involves the collection, preparation and analysis of the automatically gathered data. Using transaction logs allows for the unobtrusive collection of robust data from real users while interacting with a website in real time. Establishing what information is being looked for and collected is therefore provided by the research questions. Data preparation or eliminating surplus data, allows for the intentional focus of the data to be aligned to the research questions through the utilisation of either analysis software or bespoke programming scripts.

A TLA focuses on three common levels of analysis; term, query, and session. A term, consists of a string of characters (text) separated by a delimiter, e.g., a comma, semi-colon, or quotes; the frequency it appears; and whether it is unique or of high usage. This is then extended through the use of a query (e.g., a collection of terms input by the user) and uses the query as a base metric e.g., establishing initial, identical (duplication by the user), modified and subsequent queries as well as its complexity (e.g., use of Boolean and or other operators) and structure. Further expansion of TLA is gained when examining a session. A session interaction occurs when there is an exchange between the user and the system during a period of time, usually between five and 120 minutes, and captures all the interrelationships between the user/system and therefore presents an opportunity for investigation of

user information seeing behaviour.

However, previous work on TLA typically relies on grouping requests into *sessions*: all concurrent requests from the same IP address and User-Agent, often within some cut-off period (e.g., 30 minutes). Jones et al. (2000) undertook a manual analysis of a digital library log and were able to identify user demographics and the searches and search patterns the users adopted. Chen and Cooper (2001) found a user session could be characterised, based on 47 variables from a library catalogue log, these were grouped into six clusters which characterised users' interactions (mostly reflecting domain knowledge and technical knowledge). Yu et al. (2005) used Decision Trees and features based on temporal, content or communication activities, to identify 'interesting' users from server logs (see Table 3.2).

Table 3.2: Yu et al. (2005) clustering features

Type	**Attribute**	**Description**
Temporal attributes	A1	Accessing between midnight and 7 a.m.
	A2	The total session time
	A3	Statistics such as the time a visitor accesses the site, the total time a visitor stays at the site, and the different amounts of time a visitor stays on various pages
Page attributes	A4	The total number of accessed pages during the whole session
	A5	The accessing width (the number of child pages accessed from a single page)
	A6	The accessing depth (the depth of the pages accessed from a single page)
	A7	The percentage of graphic files requested compared to the total number of accessed pages
	A8	Number of accesses to the robots.txt file
Communication attributes	A9	Access methods (such as Get, POST, and Head) that visitors use to interact with the site

Wang et al. (2016) used a similarity graph to identify clusters of similar users based on clickstream data. To visualise and understand user behaviours, iterative feature pruning was used to capture the natural hierarchy within user clusters. Additionally, an unsupervised learning method was used to build interaction behaviour models from the clickstream data. Dev and Liu (2017) found that clustering based on task sequence rather than click sequence is more robust as it eliminates noise (such as mistakes or unintended actions) and creates more reliable clusters. Zhang and Kamps (2010) identified searcher stereotypes of novice and expert users using

search logs. They manually analysed user query terms, session lengths, session duration, and repeat visits to establish shared traits among the user groups. They were able to identify additional groups that shared traits with the novice and expert groups.

However, Stenmark (2008) states: "When studying how ordinary Web users interact with Web search engines, researchers tend to either treat the users as a homogeneous group or group them according to search experience." Stenmark (2008) goes on to claim that in order to capture the variety in behaviour known to exist amongst searchers, an automatic clustering technique (based on self-organising maps and search log data) is required. They discovered that users can be usefully separated into distinguishable segments based on their actual search behaviour. Overall, they identified six clusters and condensed them into three groups: casual seekers (fact seekers - intranet users looking for quickly retrieved answers); a more holistic group (longer sessions and more reading time); and information-seeking-savvy employees who formulate longer queries and browse through more documents than other groups.

Therefore, as the focus of this phase is to further identify the information seeking behaviours of the NML website in an attempt to enhance the already established characteristics of user groups, Jansen's (2006) three stage model will be adopted and enhanced further with the inclusion of cluster analysis as determined by Stenmark (2008).

3.4.3.2 Implementation of the Research Methodology: Phase 3

The transaction log analysis conducted followed the standard TLA processes (Jansen et al., 2009) (see Section 3.4.3).

3.4.3.3 Data Collection

Initially, logs were gathered from the Microsoft IIS web server hosting the NML sites for the period of the user survey ± 1 month (1 Jan - 30 Mar 2017). Server logs for this phase comprised of 586,868 rows, each representing a page request from 'real' users and those from robots (also known as crawlers or spiders). IIS logs can store referrer details, however, this feature was not enabled for the NML website. The

format of logs is standard IIS format:

> *Log Format*: date time s-ip cs-method cs-uri-stem cs-uri-query s-port cs-username c-ip cs(User-Agent) sc-status sc-substatus sc-win32-status time-taken

Example log entry: 2015-07-08 00:00:00 172.17.101.21 GET /onlineshop/graphics/product/thumbnails/sewing-box-new-thumb.jpg - 80 - 66.249.64.146 Googlebot-Image/1.0 304 0 0 15

Each part of the IIS log entry contains specific information about the users' request. The 'date' and 'time' fields record when the request was made. The 's-ip' is the server IP address that handled the request, and 'cs-method' indicates the HTTP method used (e.g., GET, POST). The 'cs-uri-stem' details the resource path requested, while 'cs-uri-query' provides the query string if one was used in the request. The 's-port' is the server port number that received the request.

The 'cs-username' is the username of the authenticated user, though it is often represented by a hyphen if the request was made anonymously. The 'c-ip' field holds the client IP address that made the request. The 'cs(User-Agent)' field describes the browser or bot that made the request, which is important for distinguishing between 'real' users and robots. The 'sc-status' is the status code returned by the server, such as 200 for a successful request or 404 for a not found error.

'sc-substatus' and 'sc-win32-status' are sub-status and win32 status codes that provide additional error information, but these are typically less commonly used than the main status code. Lastly, 'time-taken' represents the amount of time the server took to process the request in milliseconds. It should be noted that for the NML website, some fields like 'cs-username' and 'cs-uri-query' were not utilised and are marked with a hyphen to indicate their absence in the logs. Table I.1 in Appendix I shows a sample of the NML log data broken down into columns by data.

3.4.3.4 Preparing the Data

The log data was gathered in a series of log files (one per day). These were concatenated into one log file for ease. This then allowed for the identification and removal of bots, crawlers and developer validation tools (e.g. link checkers etc). This was

done using both the user_agents package and also a series of manual checks of the user agents and known IP addresses of bots and cross-checking with DNS lookups.

To illustrate how the bots impacted the system, the following tables show the amount of activity during the survey period.

	Jan	**Feb**	**Mar**	**Total**
Uncleaned	147730	262339	176535	586604
Cleaned	140134	248615	163574	552323
Bots	7596	13724	12961	34281

Table 3.3: Number of bots accessing the NML site per month

	Mon	**Tue**	**Wed**	**Thur**	**Fri**	**Sat**	**Sun**	**Total**
Uncleaned	81437	100312	101252	97263	85549	54952	65839	586604
Cleaned	76950	94371	95791	91769	80419	50879	62144	552323
Bots	4487	5941	5461	5494	5130	4073	3695	34281

Table 3.4: Number of bots accessing the NML site per day

It is clear that February saw more bot activity than the other months (see Table 3.3), and that there is no particular day that stands out as getting more bot traffic, however there is a drop on Sundays (see Table 3.4). Chandler (2020) identifies that the introduction of non-human participants such as bots do not accurately represent the intended study population and therefore should be removed from the data set. Therefore, in order to avoid the damage to this research that bot activity may inflict upon the findings, manual scrutiny of the data logs was performed in an attempt to weed out as much bot activity as possible.

As the code was refined, the data was cleaned further and the following was removed: entries without a 200 HTTP status code (successfully displayed pages); entries that were not GET requests removed; all rows with potential user identifying data (e.g., names, emails, credit card numbers, etc.); request entries for background files (e.g., images; CSS, JavaScript, font and Adobe Flash files; XML data and linked PDF files); Rows for sitemap.txt and a developer's test.aspx page was also removed. This meant that the remaining log data contained successful page requests only. All the other rows mentioned earlier that were removed are for assets to these pages, and so for this analysis are essentially noise in the data.

At this stage the location (country, county and city levels) of requests was identified based on the IP address and using the `IP2location` library[13] and the Country, regions and city were added to the dataset. However, the version of IP2Location used was the free tier "lite" version, which contained limitations regarding some abstract locations not being available in the lite database. Any rows where the location data was not able to be identified were removed. These rows were removed to maintain the integrity of the spatial analysis, as imputing these values could introduce additional bias, potentially skewing the results.

Next, segmenting the entries, or *sessionisation*, into user sessions based on IP address, User-Agent and a 30 minutes time cut-off, ensured that the full activity for the session was captured (as outlined in He and Göker (2000)). This saw the removal of some bot entries that were not previously identified by the bot Agent in the library. These bots at this stage became very apparent as there were either a series of requests for numerous pages made all at the same time (far more than a human could ever manage to make within a second or minute), or a series of pages requested a couple of seconds apart as so to not get blocked by automatic bot stoppers, but these pages were all different and the request timings were perfectly called every few seconds, which is not enough time for a human to have the page loaded and to have engaged with the page. This data, therefore, had to be removed to ensure the log data for the rest of the analysis was as accurate as possible. With entries grouped into sessions, the log data was fully anonymised, with all IP addresses randomly hashed.

After pre-processing and sessionisation, the transaction log data was reduced to a dataset of 552,323 rows of page requests.

3.4.4 Phase 4: Cluster Analysis (NML)

Cluster analysis is a statistical machine-learning technique used to group similar objects or data points into clusters, based on their characteristics or attributes. Everitt et al. (2011) describes it as "essentially about discovering groups in data". Kassambara (2017b) explains this in a little more depth by identifying clustering as "one of the important data mining methods for discovering knowledge in multidimensional

[13]https://lite.ip2location.com/ip-address-ranges-by-country

data. The goal of clustering is to identify patterns or groups of similar objects within a data set of interest".

According to Harmouch (2021), Cluster analysis, clustering, or data segmentation, is an unsupervised machine learning technique that groups data samples into clusters based on similarities using predefined distance measures like the Euclidean distance. In simple terms, cluster analysis is where the analysts allow the computer to examine the data and determine how the data should be classified or grouped. These groupings can be made based on pre-defined groups of classifications using discrimination and assignment methods (supervised machine learning) or with no initial groupings (unsupervised machine learning).

Therefore, this phase will extend the analysis of the NML survey data undertaken in Phase 2 and Phase 3 by using a data-driven approach. In order to do so, the self-classification survey question (Q21) will be disregarded as this phase aims to assess the accuracy of self-classification.

Phase 4: Objective:

OB4: To apply cluster analysis techniques to survey data, with the aim of identifying data-driven user groups based on their characteristics and behaviours.

3.4.4.1 Clustering Types

There are five key unsupervised methods for clustering that could be potentially adopted;

Partitioning: Partitioning clustering is a method of grouping data points into clusters based on their similarity. The goal is to split the data into distinct, non-overlapping subgroups, where data points within a cluster are more similar to each other than they are to points in other clusters. The strengths of this approach lie in being easy to understand and implement, the ability to handle non-linearly separable data and can be scaled up for large data-sets. The weaknesses however revolve around the approach not producing an optimal number of clusters, while determining the similarity measure can be challenging in addition to its sensitive to initial conditions that can result in different solutions for the same data (Kassambara, 2017b).

Hierarchical: Hierarchical clustering is a method of grouping data points into

clusters based on their similarity and by creating a hierarchy of clusters. This is achieved by either building a dendrogram, where each branch represents a cluster, or by creating a set of nested clusters. The strengths of this approach are found in the ability to handle large data-sets, visualising the hierarchy of clusters in a dendrogram in addition to handling non-linearly separable data. Weaknesses again appear around the area of not being able to produce the optimal number of clusters or being able to determine the similarity measure easily. Computationally, this is an expensive process for large data-sets and again may produce different results for the same data due to the choice of linkage method (Kassambara, 2017b).

Fuzzy: Fuzzy clustering is a method of grouping data points into clusters based on their degree of membership in each cluster. Unlike traditional clustering methods that assign each data point to a single cluster, fuzzy clustering allows a data point to belong to multiple clusters with varying degrees of membership. While this approach can handle data points with ambiguous or uncertain cluster assignments, as well as being able to process non-linearly separable data and generate a soft classification of data points into multiple clusters weakness again lies with the inability to produce the optimal number of clusters, is computationally expensive, determining the appropriate membership function can be challenging and results can be sensitive to the choice of the membership function and the number of clusters (Kassambara, 2017b).

Density: Density-based clustering is a method of grouping data points into clusters based on the density of data points in a particular region. The algorithm works by identifying regions in the data where there is a higher density of data points compared to other regions, and then grouping these data points into a cluster. The strengths of this approach are its ability to manage data points with variable densities, identify clusters of arbitrary shapes and sizes and manipulate data sets with noise and outliers. Weaknesses are evident in relation to how computationally expensive it is for large data-sets, in its sensitivity to the choice of the density threshold, the neighbourhood size and with clusters of differing densities as well as encountering difficulty with high dimensional data (Kassambara, 2017b).

Model: Model-based clustering is a method of grouping data points into clusters based on a statistical model. The algorithm fits a parametric or non-parametric model to the data and then uses this model to assign data points to clusters. This

approach can manage non-linearly separable data while handling multi-modal data distributions as well as high-dimensional data. However, model selection can be challenging and may impact the results because of the approach being sensitive to the initialisation of model parameters in addition to data that does not fit the assumed model. This approach can also be computationally expensive for large data-sets (Kassambara, 2017b).

Establishing which algorithm to select proved problematic as there does not appear to be a preferred method suggested by literature, rather many previous papers that focused on clustering chose to only discuss a single algorithm approach, action it and report on the results (Dolnicar, 2002; Kansal et al., 2018; Kaufman and Rousseeuw, 2009; Krantz et al., 2009; Padmaja and Sheshasaayee, 2016; Kuncheva et al., 2006). This potentially created a number of issues: firstly, it meant that it was not feasible to provide any context for the results and because no comparison can be made to the results of other algorithms, there is a question over any generalisability of the research results, and secondly, future researchers cannot make educated decisions on which algorithm to choose for their own work.

Therefore, since the survey data is categorical (or nominal), the five previously discussed unsupervised clustering methods could potentially be adopted. However, not all of them were suitable depending on the specific characteristics of the data and the goals of the analysis (see Table 3.5).

During the clustering process, various algorithms were utilised to cluster the data, each producing unique results owing to their varied measurement calculations. After converting the data to a matrix using GOWER, all algorithms were evaluated for suitability. However, since the data is categorical (or nominal), only algorithms suitable for this type of data were explored. (e.g., Partitioning (PAM, CLARA); Hierarchical (HClust (agglomeration (bottom-up)), Diana (divisive (top-down))) and Fuzzy (fclust, fanny).

3.4.4.2 Implementation of the Research Methodology: Phase 4

This phase sought to initially discover if there are sub groups that self classified as the General Public; and also to understand the differences between the user clusters/groups of all the captured response data, this includes their information

needs, technical ability, general and specific CH subject knowledge.

The earlier manual analysis (Phases 2 and 3) of this data set revealed that performing a manual analysis was a very time consuming and potentially expensive process for a museum to undertake, as well as requiring the analyst to have good statistical and analytical knowledge, which is generally not available in-house. It also revealed that a large proportion of the survey respondents (49% n=253 from 514) considered themselves as General Public at the point of completing the survey. The results from this showed that the self-selected General Public user group mapped with most of the other user groups and so suggests that there are sub-groups within this homogeneous General Public group, or it could have suggested that when people self-assign to a group, they do so based on their role at that exact time, e.g. a museum staff member may class themselves as "museum staff" during the working day when undertaking work tasks, but at a weekend when undertaking a personal task they make self-classify as "general public". This user discretion hence brings into question the whole use of the self classification and requires a more in depth review.

Therefore, removing Q22 and performing a cluster analysis on all the General Public data would allow for the exploration of the hypothesised sub-groupings within this category; and allow the clustering algorithm to surface the natural groups that are in the whole data set, without what could be perceived as respondent bias being placed on the potentially untrustworthy self categorisation question (Q22).

3.4.4.3 Selection and Justification of PAM and Hclust K-mediods algorithms

Selecting the right clustering algorithm is crucial for uncovering meaningful insights from survey data. Given the various clustering techniques available, an empirical evaluation was conducted to determine the most suitable method(s) for the user survey dataset. This evaluation involved comparing several clustering algorithms, including Partitioning Around Medoids (PAM), Hierarchical clustering using Ward's method (Hclust), and other notable algorithms like CLARA (PAM for large datasets), DIANA (Model), and FANNY (Fuzzy). The comparison was based on two key performance metrics: the Silhouette score and the Dunn Index. The

Silhouette score measures how similar an object is to its own cluster compared to other clusters, indicating the quality of the clustering. The Dunn Index, on the other hand, assesses the compactness and separation of the clusters, with higher values denoting better clustering quality.

An empirical approach assesses each algorithm's ability to produce distinct and clear clusters within the survey data. Assessment ensures reliable clustering outcomes and understandable, actionable insights from the analysis. It forms the basis for selecting the best clustering methods to meet the research goals.

In the evaluation of clustering algorithms applied to the survey dataset, both Partitioning Around Medoids (PAM) and hierarchical clustering using Ward's method (Hclust) have shown particular strengths that align with the research objectives. PAM, with its relatively stable Silhouette and Dunn index scores and with its performance peak at $k = 3$, achieving a Silhouette score of 0.1537017 and a Dunn Index of 0.1148573, demonstrates a reasonably good structure with clusters that are compact and well-separated. This underscores PAM's utility in identifying distinct and interpretable clusters, which is crucial for detailed analysis of survey data.

Similarly, hierarchical clustering (Hclust) using Ward's method has also proven to be effective, particularly noted by its highest Silhouette score of 0.1807585 at $k = 3$. Ward's method, known for its ability to minimise the variance within each cluster, yields slightly better cohesion and separation for this specific number of clusters. This method's performance, especially its superior Dunn Index at higher k values (e.g., 0.2040201 at $k = 8$), suggests its usefulness in uncovering hierarchical structures within the data, which can offer additional insights into the dataset's natural groupings.

The decision to also consider Hclust alongside PAM is informed by the complementary perspectives these methods provide. While PAM offers a stable and nearly deterministic clustering suitable for straightforward interpretation and application, Hclust, particularly with Ward's algorithm, complements this by exposing the hierarchical relationships between clusters, which can be invaluable in exploring the data's underlying structure more deeply.

The selection of Ward's method over other variations of hierarchical clustering is justified by its consistent performance in achieving higher quality clusters, as evidenced by the empirical data. Ward's method effectively minimises the sum of

squared differences within all clusters, a criterion that aligns well with the aim to achieve high-quality, well-defined clusters in the survey dataset.

Table 3.5: Clustering Performance of User Survey Data (Silhouette Score / Dunn Index Score)

K	PAM	CLARA	FANNY	HCLUST_WARD	HCLUST_SINGLE	HCLUST_COMPLETE	DIANA	KMEANS
k=2	0.199 / 0.129	0.101 / 0.119	0.129 / 0.124	0.159 / 0.261	0.180 / **0.313**	**0.224** / 0.181	0.201 / 0.129	0.142 / 0.124
k=3	**0.183** / 0.131	0.122 / 0.119	0.086 / 0.000	0.151 / 0.210	0.050 / **0.303**	0.170 / 0.190	0.181 / 0.132	0.174 / 0.124
k=4	0.170 / 0.148	0.132 / 0.115	0.129 / 0.124	0.121 / 0.192	0.048 / **0.309**	0.169 / 0.206	**0.174** / 0.152	0.114 / 0.136
k=5	0.175 / 0.151	0.068 / 0.129	0.129 / 0.124	**0.183** / 0.220	-0.037 / **0.308**	0.162 / 0.213	0.177 / 0.152	0.131 / 0.142
k=6	**0.166** / 0.152	0.108 / 0.133	0.095 / 0.000	0.161 / 0.200	-0.049 / **0.308**	0.159 / 0.145	0.165 / 0.153	0.108 / 0.144
k=7	**0.180** / 0.160	0.113 / 0.136	0.129 / 0.124	0.103 / 0.197	-0.062 / **0.298**	0.138 / 0.154	0.179 / 0.161	0.098 / 0.141
k=8	0.161 / 0.162	0.091 / 0.136	0.129 / 0.124	**0.190** / 0.230	-0.089 / **0.296**	0.143 / 0.162	0.163 / 0.164	0.114 / 0.152
k=9	**0.164** / 0.164	0.119 / 0.142	0.129 / 0.124	0.100 / 0.151	-0.093 / **0.296**	0.142 / 0.165	0.164 / 0.167	0.131 / 0.172
k=10	0.133 / 0.167	0.075 / 0.140	0.129 / 0.124	0.104 / 0.180	-0.121 / **0.294**	**0.139** / 0.167	0.134 / 0.169	0.130 / 0.166

PAM was therefore selected, despite being time consuming and computer intensive (run time and memory are quadratic) due to it's intuitive, robustness to noise and outliers compared to k-means (due to the properties of distances being used), and its ability to produce a "typical individual" or "medioid" for each cluster (useful for interpretation) (Filaire, 2018).

3.4.4.4 Clustering Environment

For the cluster analysis, the Kaggle.com website was utilised. Kaggle is a subsidiary of Google and provides powerful servers that are accessible in a secure, private format. Kaggle hosts Jupyter-style notebooks that can run Python or R. For this work, an R environment was adopted running version 3.6.3 (29/02/2020). R was chosen over Python in order to pre-empt the creation of a tool that non-technical and non-scientific museum staff could easily use in the next chapter. R has a package called R-Shiny that allows for the creation of user interfaces, and this drove the technology choice.

In order to undertake the analysis of the survey response data-set the following libraries were installed and utilised: tidyverse, factoextra, fpc, NbClust, FactoMineR, readxl, dplyr, cluster, Rtsne, DataExplorer, ggsci, cba, corrplot, clustertend, seriation, vtree, KlaR, hbrthemes, reshapre2, mclust, nomclust, fclust, caret, randomForest, ranger, clValid.

Preparation of the Kaggle environment also required the original character data loaded into a data frame and converted into a special data structure called a factor. Similar to an Excel spreadsheet, R data frames are a type of data structure designed to hold such tabular data. A data frame consists of a number of rows and columns, with each column representing some variable or feature of the data and each row representing a record. A data frame is actually just a list where each object (column) is a vector with the same number of items.

Factors are intended to hold categorical data: variables that can take on one of several distinct values from a set, e.g., gender, visit reason, visit purpose and frequency of website visit, and assign each category an integer value. The number of factor categories or "levels" is equal to the number of unique elements in the vector used to make the factor, e.g., a factor representing biological gender would have two

levels: male and female. The ID field was dropped from the dataset as each row was assigned a unique number which would add bias for the higher numbers, but this field is not a factor or a response to any survey question. Rows with any "empty" or "unknown" fields also known as missing values were dropped from the dataset before analysing it because: the missing values can introduce bias into the analysis, which can skew the results. For example, if a large number of observations are missing for a specific variable, the mean or median of that variable will not be an accurate representation of the true value. Also, many statistical techniques, such as linear regression and many clustering algorithms, cannot handle missing values and will produce errors if they are present in the dataset. Lastly, removing the missing values improve the performance of algorithms by reducing the number of observations in the dataset.

3.4.4.5 Data Collection and Preparation

Utilising the survey from Phase 2, 21 survey questions (see Table 3.1) were considered as potential features/variables. The self-selected categorisation (Q22) was omitted from the feature analysis but was included in the dataset to allow for cross-referencing and analysis, which was analysed in isolation.

3.4.4.5.1 Feature Reduction and Selection for Clustering Analysis

The feature reduction process from the initial 22 questions/features to the final 8 features/variables involved an evaluation of the survey questions to identify the features that were most indicative of distinct user groupings. This evaluation process used both empirical data analysis with theoretical insights focusing on features' ability to capture significant differences among users in terms of their website engagement and cultural heritage interests.

Step 1: Initial Feature Evaluation: The initial step involved a thorough evaluation of each survey question for its relevance to the research questions and its potential to distinguish between user groups. This evaluative process, grounded in both theoretical considerations and exploratory data analysis, allowed for the preliminary identification of redundant or non-discriminatory features. The exploratory

data analysis included the production of a series of box-plot diagrams and entropy scores, one for each feature. The low entropy features were excluded due to their low variance, which would have a lack of impact on clustering outcomes. The features removed due to their scores being less close to zero (all < 0.22) were "gender, content types and education". All other features returned moderate to high levels of entropy (> 0.6), suggesting a meaningful variation that could potentially help in identifying distinct clusters.

The features removed based on their irrelevance to the defining factor of user groups were "Internet proficiency skills and Device type".

Step 2: Multiple Correspondence Analysis (MCA): To further refine the feature set, Multiple Correspondence Analysis (MCA) was employed, facilitating the examination of relationships between categorical variables. MCA served as a critical tool for visualising the data structure, enabling the identification of dimensions that capture significant variance within the dataset. Through this analysis, features contributing minimally to the variance were earmarked for exclusion.

Step 3: Cluster Tendency Assessment: An assessment of cluster tendency using the Hopkins statistic provided empirical evidence of the dataset's suitability for clustering. This step was pivotal in validating the feature selection process, ensuring that the reduced set of features retained the inherent clusterability of the data.

Stages two and three were undertaken iteratively with the lowest contributing features removed each time unit the Hopkins statistic returned a value of less than 0.5.

The 8 variables for use in the cluster analysis dataset that remained after the feature reduction steps were "visit reason, visit purpose, frequency of visit, domain knowledge, General CH knowledge, location, age and employment". From the analysis, these were deemed clusterable and important in profiling user groups and produced a Hopkins statistic of 0.1826, suggesting that the 8 variables are clusterable (Bezdek and Hathaway, 2002; Walsh et al., 2020).

3.4.4.5.2 Further Pre-processing

The selected categorical (nominal and ordinal) variables were as follows: visit reason (nom, 4 levels), visit purpose (nom, 9 levels), frequency of website visit (ord, 5 levels), level of domain knowledge (ord, 4 levels), level of general CH knowledge (ord, 5 levels), location (nom, 5 levels), age group (ord, 5 levels), employment status (nom, 8 levels). The self-assigned user group variable (user group (nom, 8 levels)) reflects how the survey participants defined themselves as in accordance with group labels that had been identified in the literature (see Appendix F.1): Academic (n=25), General Public (n=253), Museum Staff (n=10), Non-Professional (n=137), Other (n=26), Professional (n=5), Student (n=33), Teacher (n=25).

Further pre-processing included removing cases with 'unknown' responses (e.g. for levels of knowledge). These results were excluded to prevent the introduction of bias that could result from data imputation (Newman, 2014). The Categories with low counts were also merged to reduce the number of variable categories e.g., 'daily', 'weekly' and 'monthly' frequency of visit (see Table 5.15) was combined into a single 'regular' category, thereby reducing the data-set to 487 cases.

In order to successfully segment the data, the sample data set needed to be of a sufficient size (McGinn and Kotamraju, 2008; Dolnicar, 2002; Dolnicar et al., 2014), free of biases (Dolnicar et al., 2016), and the number of variables used cannot be too high for the sample (Dolnicar et al., 2016). Using Formann's (1984) (cited in Sarstedt et al. (2014)) rule of thumb, meant that the sample and cluster relation in this study is viable: $2^8 = 256$ and 487 / 256 providing 1.9 times the minimum sample size. Qiu and Joe (2006) suggest that the sample size should amount to a minimum of 10 times the number of clustering variables, which when applied to the data-set reveals a six times excess of the required 80. Adversely, using Dolnicar et al.'s (2014) approach of having 70 times the number of variables means that the sample of 487 is just short of sufficient, e.g., 70 x 8 = 560 with 73 under the minimum. However, exceeding two of the three approaches, provided confidence that the size of the sample would not lead to a sub-optimal segmentation solution, and thereby validated the data set as adequate for this phase.

3.4.4.6 Additional Data-set Preparation for General Public and Non-General Public Classification

In addition to the validation of the entire data set, examination of the General Public and Non-General Public group labels was also performed. Originally, the survey respondents were asked to self classify into one or more of 7 groups. Table 5.43 presents the most frequently selected groups, which resulted in a General Public sample size of 253. Applying Formann's (1984) 2^m rule of thumb resulted in the General Public only data failing by 3 results (0.98%) and obviously still failing the Dolnicar et al. (2014) 70 times requirement and yet passing the Qiu and Joe (2006) requirements.

Removing the General Public group left the Non-General Public group with a sample of 234. Applying Formann's rule of thumb in this instance resulted in an overall fail by 22 results (0.91%), still failed, Dolnicar et al.'s but also still passed Qiu and Joe's requirements.

Using the same data-set that was used for the manual descriptive analysis in Chapter 5, the fields were all converted to factors which enabled the further processing of said data with the PAM algorithm. This resulted in the ID field being dropped along with any empty or "unknown" fields. The fields of: visit reason, visit purpose, frequency, domain knowledge, ch knowledge, age, location, employment and aggregated group remained. At this point a series of clustering using various algorithms was undertaken and groups emerged, but these clusters were unstable, and the resulting clusters did not make much sense. Closer examination of the data revealed some fields should have been ordinal and not factors and so these were changed, and the fields of frequency, CH knowledge, age, visit purpose and employment were also re-organised into levels or named groups that meant there were less unique groups in each field (for example: in the age field there were very few participants selecting 65-74 and also very few selecting 75+ and so these were grouped together to form a single 65+ option).

3.4.5 Phase 5: Automatically Describing User Groups

Phases 1 to 4 have involved efforts to gather, identify, and collate the characteristics defining user groups. The insights gained from each of these phases have shaped the

progression to the next, thereby collectively contributing to the development of this phase.

Phase 4, was a continued attempt to extract further identifying information-seeking behaviours e.g., user groups and characteristics, from the Phase 2 data, via an unsupervised machine learning approach. This approach saw the emergence of clusters (see Table 7.14) that were recognisable in the literature (Phase 1) and the manual survey (Phase 2). The creation of these clusters revolved around the extracted compacted characteristics, or mediods, via PAM algorithm, that then offered up a potential descriptive narrative for each of the user groups (Phase 4) (see Chapter 7 Section 7.3.2).

However, while there are clear clusters emerging and descriptions being produced, the argument for this work has always been around the usability and comparability of shared descriptions. For example, in Phase 4 (Section 7.3.2), the results are short and only descriptive of the medoid. These descriptions are hiding lots of information about the composition of each factor in that cluster.

The lack of shareability is also highlighted in many of the published papers utilised in Phase 1. According to Graham et al. (2016), humanities scholars were aware of the potential benefits of machine learning, such as improving data analysis and discovering new patterns, but lacked the technical expertise to implement these technologies effectively. This suggests that a tool to undertake this work may aid in the use of more detailed (published) user group definitions.

Similarly, according to Parry et al. (2018), museum professionals do not possess the technical skills needed to implement machine learning projects. The study revealed that skills were just one challenge among many affecting technology use. Larger issues included inadequate funding to hire digital specialists, leading to ad-hoc outsourcing; projects managed by non-specialists with specialists brought in too late; and, in well-resourced museums, specialists being spread too thinly across numerous projects, limiting time and scope for each. Generally, the use of machine learning (ML) techniques in the context of CH is weak and that the application of these techniques is still limited Fiorucci et al. (2020).

However, while this indicates that both audiences would benefit from detailed audience information and a tool that simplifies complexity, eliminating the need for hiring/outsourcing and/or retraining staff for cluster analysis and interpretation,

it also allows digitally skilled museum staff to focus more on other projects. The detailed textual descriptions required by researchers may not be as practical for the museum practitioners responsible for the website. Therefore, a different representation of the same data that aligns more with the current practices of the web industry e.g., Personas, may be more suitable.

Therefore, this phase proposes the exploration of a pre-persona visual approach, based on the information-seeking cluster results, called an "information-seeking cluster profile" (ISCP) will be developed.

Phase 5: Objective:

OB5: To develop a tool that generates comprehensive user group profiles from clustering data, complete with detailed textual and visual presentations for early design use by researchers and DCH, UX, and web development professionals

3.4.5.1 Exploring Cluster Analysis via Heuristics

In cluster analysis, the medoid is a point within a cluster that is considered representative of the other points in the cluster. It is typically based on some measure of centrality, such as the point closest to the cluster's mean or median (Jin and Han, 2010). However, although the medoid is intended to be representative of the other points in the cluster, it is not a perfect representation, but rather one point within the cluster that may not capture the diversity of opinions or behaviours of others. Therefore, it is important to consider the medoid in the context of the entire cluster, as a way to highlight key characteristics (Han et al., 2011).

The previous clustering work carried out in Phase 4 produced clusters and presented these in the form of a set of medoids that showed one respondent's set of scores across all the factors (survey questions) for each cluster. The hierarchical clustering produces a different end result: it does not provide centroids or medoids. Instead, it outputs a hierarchical tree structure that shows how each of the members are grouped at multiple levels. By showing how close the branches are, it demonstrates how similar the clusters are to each other. To gain an understanding of the characteristics of the clusters produced in hierarchical clustering, a further process is required that examines the features or variables. This analysis allows for a detailed view of how all members of the cluster contribute.

However, it can be difficult to identify which variables are most important in defining the clusters. This is because hierarchical clustering takes into account all variables equally when determining cluster membership. As a result, it can be challenging to understand the underlying relationships between variables and how they contribute to the formation of the clusters. A further complication is that variables with similar profiles may be given different weights depending on how they are combined during the clustering process. This can make it difficult to understand the relative importance of different variables in defining the clusters. Therefore, for a detailed understanding of the survey data of users of a DCH website, it is important to understand the users' characteristics/skills/abilities/wants and needs; therefore, no matter which algorithm is used (PAM or Hierarchical), there needs to be a way to interpret these post clustering. In an approach to combat this potential misrepresentation of the clusters, it is proposed to have heuristics to evaluate each factor of the cluster results and provide a multi-valued breakdown.

3.4.5.1.1 Applied Heuristics:

In order to improve the descriptiveness of the user group definitions, the following heuristics are applied to each factor of the cluster results which generates a more refined and accurate definition of the user group, representing the majority of the cluster members as opposed to only the medoid.

1. Identify the attributes that make up 75% of each factor of the cluster data.
2. Limit the value to a max of 3 attributes in the output.

Adopting a 75% heuristic to identify the attributes that make up 75% of each factor of the cluster data will help to:

Simplify the analysis: Examining all attributes in a cluster analysis can be overwhelming and time-consuming. Focusing on the top 75% of attributes simplifies the analysis and provides a more concise picture of the factors driving clustering results.

Identify the most important attributes: By focusing on the top 75% of attributes in each factor, the most important attributes that are contributing

to the clustering results are identified. This can provide insights into the underlying patterns in the data and help to guide further analysis.

Aid interpretation: By identifying the top attributes in each factor, a better understanding of the characteristics of the clusters and how they differ from each other emerges. This will aid in interpretation and provide insights into the underlying structure of the data.

Support decision-making: Identifying the top attributes in each factor enables a more informed decision to be made about how to use the clustering results.

Ideally, the cluster results would have a clean majority of results for each factor, although this was rarely the case with having so many factors in the data set. Factors that cover numerous values creates a more specific and quantitative view of the cluster and adds depth and a sense of realism to the clusters/user groups.

Restricting the maximum number of values to 3 attributes in the output is to further simplify the analysis and provide a clearer picture of the most important attributes driving the clustering results. This can save time and resources that might otherwise be spent analysing irrelevant or unimportant attributes. Focusing on a smaller number of attributes also ensures that results are clear and easily understandable to a wide audience. This can be especially important in fields such as DCH, where the results of data analysis are often communicated to non-expert audiences who may not have a deep understanding of the process or the underlying data. Additionally, limiting the output to a maximum of 3 attributes can help to prioritise which variables to focus on for further analysis or intervention. For example, studying user engagement with a DCH website may consist of cluster analysis on group users based on their browsing behaviours. By limiting the analysis to the top 3 attributes that contribute to each cluster, patterns in user behaviour can be quickly identified and targeted interventions to improve engagement developed.

Data that is not situated within the top 3 attributes or the 75% makeup of the factor, are excluded from any of the presentations, but remains available in the script outputs should further exploration be desired.

Applying the heuristics to the cluster results will produce a very detailed description of the cluster. Additionally, including the range of values and the breakdown

of percentages can be useful in providing a more complete picture of the cluster. The range of values gives the idea of the diversity of the data within the cluster and also highlights any outliers. While the breakdown of percentages provides a more detailed view of the distribution of the data within the cluster, which will help to better understand the characteristics of the cluster and distinguish it from others. This level of detail not only allows for a comparative analysis of the clusters in the one study to be undertaken, but it also allows for the user groups defined in one study to be compared to findings from other studies.

3.4.5.2 Validation of ISCPs

To assess the accuracy and usefulness of the profiles created, the method of creation, and the proposed adaptation to the standard UX workflow, a qualitative analysis was conducted through a series of interviews with UX and web development professionals active in the industry. Thematic analysis, as per Braun and Clarke (2006), is a qualitative method to identify, analyse, and report patterns/themes in data. It's categorised into inductive (bottom-up) and deductive (top-down) approaches. Inductive analysis is a developmental strategy where codes/concepts emerge as the researcher reads, understands and becomes familiar with the data Braun and Clarke (2021); Byrne (2022). A deductive, or a priori, analysis is the application of predetermined codes that have been created from concepts drawn from the literature, theory, or hypothesis that the researcher has developed and applied to the data (Boyatzis, 1998; Joffe, 2012). Employing a hybrid (inductive/deductive) approach in thematic analysis enhances rigour by integrating the exploratory data-driven bottom-up inductive method with the top-down confirmatory approach to align codes to a pre-defined deductive code book (Guest et al., 2014; Swain, 2018; Proudfoot, 2023). The applied thematic analysis begins with an inductive phase, where the analysis is driven by the data itself without initial reliance on pre-established frameworks or categories. This phase is exploratory, aimed at uncovering themes and codes directly from the data. After the initial inductive coding and theme development, the approach incorporates a deductive phase where the emergent themes are matched with or related to a set of pre-defined codes that are derived from existing theories and frameworks. This phase is confirmatory, aimed at integrating and validating the findings within

the context of specific research questions. This hybrid "Inductive/Deductive Applied Thematic Analysis" approach leverages the strengths of both inductive and deductive methodologies, providing a comprehensive analysis that is both grounded in the data and relevant to the pre-existing theoretical or research frameworks. It embodies the flexibility and adaptability advocated in applied thematic analysis, as described by Guest et al. (2006), allowing researchers to explore new territories within the data while also making their findings relevant to specific research goals or theoretical discussions.

3.4.5.3 Qualitative Analysis

This research used a hybrid approach as it combines the strengths of both inductive and deductive methods, allowing a thorough exploration of the data while still being informed by the research questions and the theoretical framework established through the systematic literature review, survey, and clustering analysis. The inductive element pertaining to the interviewing of DCH professionals, museum web developers, and UX experts captured nuanced perspectives on the ISCP tool's utility, applicability, and implications. This generated new insights and themes directly from the interviews. The deductive component was necessary to ensure that the analysis remained focused and relevant to the research objectives. This involved seeking themes that specifically related to the effectiveness, usability, and impact of the proposed method and its comparison with traditional methods as outlined in the research questions.

3.4.5.3.1 Interview, Design and Deployment

In order to establish what practitioners think about a) the idea of ISCPs and data-driven personas and b) the ISCP tool (what they like, dislike, would change and if they would adopt) the main data collection method for this phase was an interview. Interviews allow for the gathering of rich and in-depth data that captures the complexity of participant's experiences and perspectives. Interviews offer several advantages, including the ability to probe and clarify participants' responses, establish a rapport with participants, and tailor the interview to the specific needs of the research project.

3.4.5.3.2 Question Planning

The interview questions were tailored to capture both the specific insights related to RQ7 and broader trends in UX practices and user profiles. The questions were focused on:

- outlining the participant's suitability for the study.
- gaining their opinions of personas in the UX process.
- gaining their understanding of data-driven personas.
- validating the ISCP profiles produced and shown to the participants.
- getting participants' open and honest opinions of the ISCP tool and the change to the UX process.
- identifying if the participants could see any weaknesses it improvements that could be made to the profiles or tool.
- seeing if they would adopt a tool like the ISCP.

Adopting a semi-structured interview process allows the pre-determined questions to be asked while also allowing flexibility for questions that are based on the participants responses (see questions schedule in Appendix J.2.1).

3.4.5.3.3 Pilot Study

The interview questions were devised and subjected to approval by the research/-supervisory team. Once approved, initial pilot interviews where performed with both an experienced service designer and a senior user researcher. This resulted in a number of changes to the wording of questions and also the order in which some questions were to be asked. Further pilot interviews then took place with eight third-year Web Design and Development B.Sc. students, who have all studied several modules around UX design and the web development process. This resulted in additional minor adjustments to some of the wording of a few of the questions to aid the explanation of certain concepts.

3.4.5.3.4 Recruiting Participants

Participants were initially recruited through known contacts in UX, web development, and the DCH community, as well as contacts identified from LinkedIn within the DCH UX industry. Initially, fifteen emails were sent out, yielding six positive responses (40% response rate). Subsequently, each interviewee was requested to recommend potential contacts who would be a good fit for this study (snowball/non-probability sampling approach) (Parker et al., 2019). Participants were asked to email contact details to the researcher rather than providing them during the interview for data protection purposes. An additional seven emails were sent out based on these recommendations, resulting in three positive responses and agreements to participate (43% response rate). The overall response rate was 41%.

3.4.5.3.5 Interview Process

The interviews were arranged and conducted via Microsoft Teams. Each session was recorded and automatically transcribed using the MS Teams transcription tool, while the interviewer also took notes (referred to as Initial Analytical Interest). Although the interviews began with typical demographic questions, their semi-structured format allowed for slight deviations to explore deeper or additional themes and stories.

3.4.5.3.6 Member Checking

In order to validate the reliability of the study findings, "member checking" was performed. Member checking is a qualitative research strategy employed in theme analysis that contributes to establishing the trustworthiness of qualitative research, validating findings and recognising participants as co-constructors of knowledge. This approach aids in mitigating researcher bias, thereby boosting the credibility and reliability of research outcomes (Birt et al., 2016).

There are two forms of member checking: 'real-time member checking', also known as 'in-situ member checking' (Zairul, 2021), and 'post-analysis member checking', commonly referred to simply as 'member checking' (Birt et al., 2016).

Real-time member checking involves the interviewer seeking immediate feedback from the interviewee on the accuracy of their understanding or interpretation of what the interviewee has shared through summarising, paraphrasing, or reflecting

back the participant's statements and asking for confirmation or clarification. This method can enhance the accuracy and reliability of the data collected by minimising any misunderstandings or misinterpretations early on. It also helps to build trust between the interviewer and the participant, as it demonstrates the interviewer's commitment to accurately understanding and representing the participant's experiences. Real-time member checking is a valuable tool for improving the quality of qualitative data and can complement traditional post-analysis member checking. It helps capture more accurate and nuanced understandings of participants' experiences and perspectives, thereby strengthening the overall credibility of the research findings (Zairul, 2021).

Post-analysis member checking entails communicating the study findings, interpretations, or particular analytical outputs e.g., thematic categories or codes, to the participants who provided the original data. The goal is to solicit comments on the accuracy of the researcher's interpretations of the data. Birt et al. (2016) identifies that there are several ways this validation can be conducted: Returning the transcript to the interviewee asking for confirmation; a member-check interview using the transcript or using the analysis of that participant's data; A member-check focus group with participants or a sample of others with similar experience or member-check interviews, both using the synthesised, analysed data and themes. Regardless of the approach taken, adjustments to theme analysis are made based on participant feedback. This may involve modifying topics, reinterpreting data, or incorporating new insights to ensure the analysis accurately reflects participants' opinions and experiences.

Although both methods are non-exclusive and can complement each other, this study specifically opted for real-time/in-situ member checking due to its immediacy and effectiveness in capturing participants' thoughts and reactions. The decision was driven by the study's focus on understanding nuanced perspectives where immediate clarification and interaction were deemed crucial. This approach facilitated a deeper engagement with participants during the interview process and aligned with the study's priorities of accuracy in data collection and building rapport and trust. The primary distinction between the two methods is their timing, with real-time/in-situ member checking conducted during the interviews and post-analysis member checking afterwards. Opting for the former was essential for immediate

clarification of participant responses and interpretations, eliminating the need for follow-up meetings or communications, which was preferred over post-analysis due to potential changes in participants' trains of thought.

3.4.5.4 Thematic Analysis Process

Thematic analysis is the process of identifying patterns or themes within qualitative data. Though some researchers (Braun and Clarke, 2006; Clarke and Braun, 2013) argue that thematic analysis is a methodology in its own right, others, such as Morse et al. (2021) and Boyatzis (1998), consider thematic analysis as a method, practice or a tool and as it is not tied to a particular epistemological or theoretical perspective, the approach provides a flexible and diverse approach to data analysis (Braun and Clarke, 2006; Clarke and Braun, 2013).

The method of thematic analysis used in this study was a hybrid approach that incorporated the two main contrasting philosophical methods of reasoning: a bottom-up, inductive, content-driven exploratory process and a top-down, deductive theoretical process. The inductive process generated a series a posteriori (post-empirical, or after the interviews) codes derived from data generated from the individual questions asked in the interviews (Glaser and Strauss, 2017; Charmaz, 2006; Boyatzis, 1998), while the deductive process produced a set of a priori (or pre-empirical) codes (similar to the work of Crabtree and Miller (1992)) developed from the research aim, the research question (RQ7). The inductive codes are aligned to the deductive codes as well as allowing for new codings of the data to be identified meaning that theory was both a precursor to, and an outcome of, the data analysis procedures (Braun and Clarke, 2006; Proudfoot, 2023) (see Table 3.6).

3.4.5.4.1 Stage 1 - Transcription Amendment

With Time stamps:

> *"0:5:5.650 –> 0:5:25.330*
>
> *Participant 4*
>
> *Yes and no. I think what I'll give you is a classic example of sort of a design phrase taking you know from like discovery to alpha/beta and taking through that. But sometimes we'll get projects that are midway*

Table 3.6: Theme generation and development process

Stage	Step	Description
1) Transcription Amendment	1	Checking and fixing auto-transcriptions
2) Familiarisation with the data	2	Reading transcripts thoroughly
3) Generating Deductive Codebook	3	Generate codebook using RQ and Interview Questions
4) Coding	4	Codes generated (inductively) in text
	5	Deductive coding and aligning inductive codes to the deductive codebook
	6	Codes checked for duplication and similarities (combined or removed)
	7	Codes grouped into sub-categories
	8	Group codes into themes
	9	Reviewing themes
	10	Defining and naming themes
5) Analysis	11	Themes and data interpreted and analysed rather than descriptive
	12	Critical friend employed to ensure themes and data sufficient
6) Writing	13	Method clearly outlined
	14	Themes provide a useful basis for discussion of ISCP validation in the text

through. Sometimes we'll get projects that have a very short deadline on them. So that will limit the things that we can do.

0:5:25.370 –> 0:5:35.770

Participant 4

In the ideal world, we would approach everything the same way. Discovery Alpha beta live, or that way, or try and get as much research beforehand, but that's not always the case." (Participant 4)

Without timestamps:

" Participant 4

Yes and no. I think what I'll give you is a classic example of sort of a design phrase taking you know from like discovery to alpha/beta and taking through that. But sometimes we'll get projects that are midway through. Sometimes we'll get projects that have a very short deadline on them. So that will limit the things that we can do. In the ideal world, we would approach everything the same way. Discovery Alpha beta live, or that way, or try and get as much research beforehand, but that's not always the case."

All interview transcripts can be found in Appendix J.2.2.

3.4.5.4.2 Stage 2 - Familiarisation with the Data

As the analyst also served as the interviewer, there was already a degree of familiarity with the data. However, to enhance familiarity, the videos of the transcripts were watched multiple times during the cleaning process, and the transcripts themselves were thoroughly reviewed.

3.4.5.4.3 Stage 3 - Generating Deductive Codebook

A deductive coding framework was employed based on a pre-defined codebook developed from the research and interview questions (see Table 3.7). This codebook included codes for familiarity with personas, awareness of data-driven personas, perceptions of the new tool, its impact on UX workflow, and willingness to adopt it.

The transcripts were then imported to NVivo, and each transcript went through a process of being coded based on the research questions.

Table 3.7: Initial Deductive Codebook for Thematic Analysis

Code	**Description**	**Purpose**
Background Knowledge	Captures information about participants' background, experience in UX and DCH, and familiarity with user profiling methods.	To understand the context each participant is coming from and their baseline knowledge.
Perceptions of Traditional User Groupings/Personas	Gathers participants' views on the effectiveness and limitations of traditional user grouping methods like personas.	To identify common criticisms or endorsements of current practices.
Thoughts of and Application of ISCP profiles and tool	Focuses on participants' perceptions of Information Seeking Clustering Profiles (ISCP), the tool for creating them, and their potential application in UX.	To assess initial reactions to ISCP (profiles and tool) and its perceived utility compared to traditional methods.
Integration Challenges	Identifies perceived challenges in integrating new user profiling methods into existing workflows.	To understand obstacles to adopting new methods like ISCP in practical settings.
Improvement Opportunities	Seeks suggestions for improving ISCP profiles, the tool and user profiling methods and integrating them into UX design processes.	To gather insights on potential enhancements and practical integration strategies.

3.4.5.4.3.1 Validating the Codebook Establishing a viable analysis framework requires determining the code's applicability to raw information (Boyatzis, 1998). Two interview transcripts were selected as test samples. After coding the transcripts using the predetermined codes, the research/supervisory team were asked to code them independently. Upon comparison, no adjustments to the predefined code template were deemed necessary.

3.4.5.4.4 Stage 4 - Coding

Rather than employing an iterative coding approach where codes are developed and refined as each interview is analysed, the thematic analysis was conducted by coding all ten interviews in their entirety after data collection was complete. This method was chosen to gain a holistic view of the dataset, allowing for simultaneous consideration of the entire body of data and more effective identification of thematic interrelationships (Richards and Morse, 2012). Additionally, it ensured consistent application of codes across all interviews, minimising potential bias from iterative coding practices. However, it is acknowledged that this method might have resulted in overlooking nuances that could have emerged through earlier coding iterations (Elo and Kyngäs, 2008). To mitigate this potential drawback, transcripts were systematically reviewed multiple times before applying codes, ensuring a thorough understanding of the data and allowing for adjustments to the coding scheme as needed following recommendations by (Braun and Clarke, 2006; Miles and Huberman, 1994).

3.4.5.4.4.2 Data Saturation One objective of data collection is to attain data saturation, which is reached when further data gathering no longer reveals new themes or insights, ensuring comprehensive and representative findings in qualitative research (Corbin and Strauss, 2008; Creswell, 2013). During coding, the researcher monitored theme emergence and redundancy closely. While thematic saturation appeared evident around participant 7, interviews continued to participant 10 for two main reasons.

Firstly, a larger dataset (10 participants) ensured sufficient information for robust analysis and comprehensive theme development. This allowed for deeper exploration of existing themes and potential identification of nuances or sub-themes. Secondly, continuing beyond initial saturation served to verify the overall consistency and representativeness of emerging themes across a wider participant pool, as suggested by Guest et al. (2006); Morse (2006). It is important to acknowledge that data saturation is not a fixed point but rather an iterative process. While no significant new themes arose after participant 7, the possibility exists that further interviews might have revealed additional variations or refinements. However, considering the

consistent repetition of key themes and the absence of novel insights in the final interviews, there is confidence that data saturation was achieved within the 10 participants interviewed. Should saturation not have happened by the 10th planned interview, more interviews would have been undertaken until saturation was evident.

The consistent emergence of key themes like Familiarity with personas, perceived value of generated profiles, perception of ISCP tool and change in UX process across interviews supports saturation achievement. Notably, the presence and characteristics of these themes stabilised after participant 7, with no major modifications arising in subsequent interviews. For example, in participant 10's interview, the theme of 'perceived value of the generated profiles' only reiterated the aesthetic appeal and visual value of the profiles and the data they offer demonstrated consistency with previous interviews, further solidifying its presence within the data.

3.5 Ethics

The University of Sheffield's Information School Ethics Review Process was applied to each of the phases of the study process. The researcher was required to submit a thorough research proposal, as well as copies of planned supporting documentation, during this phase. (Information Sheets and Consent Forms). The submission to the Ethics Review committee was approved (reference number 011190, approved on 14/11/2016 - see Appendix K K.1). A second ethics application for the follow up interviews study was also applied for and approved (reference number 055559, approved on 27/09/2023 - see Appendix K K.1).

3.6 Summary

The purpose of this chapter has been to present the methodology employed in this research. The methodology draws on a variety of disciplines to provide a more comprehensive understanding of the findings of the research studies. The methodology's rationale is to make it easier to capture user behaviour when interacting with a DCH website environment in response to a pop-up survey that asks questions about demographics, motivations, task, preferred content type, shared experience, technical and domain knowledge, engagement, usage, and self group classification, as well as

the unobtrusive collection of automatically recorded website interactions (log data) for the same time period.

Within this context, an adapted exploratory sequential multi-phased mixed methods approach becomes highly appropriate as a defining methodology. With the aim of determining the behaviour of users within a digital environment, this approach enables the identification of the behavioural patterns and characteristics that can be utilised to understand the use of pragmatism as the philosophical approach adopted in the study. To achieve a successful study, the data is captured through 4 phases with ethical consideration being made at each stage; the identification of those users as determined in literature (Phase 1); the completion of a survey by users who interact in real-time with a DCH environment (Phase 2); the examination of log data obtained within the same time period (Phase 3); and the application of a clustering algorithm on the survey data (Phase 4); application of heuristics to cluster results and presentation of results for different audiences (Phase 5).

Findings from each Phase were interpreted individually before progressing to the next phase, employing distinct analysis techniques for each. These findings are presented separately in the following chapters.

Chapter 4

PHASE 1: ESTABLISHING USER GROUPS THROUGH THE LITERATURE

4.1 Introduction

Virtual collections attract various website users, raising questions about their identities and how to ensure they have a meaningful experience. Previous studies simplify user profiles into generic categories based on factors like motivation, engagement, role during the visit, expertise, information needs, and chances of wandering in. Users are also categorised and modelled in fields like information-seeking and context-aware systems, reflecting diverse behaviours and contexts. However, modern web design aims to cater to the specific needs of end users, although developing an online resource for these audiences poses a significant challenge.

An initial review of the literature revealed a lack of comprehensive studies characterising DCH museum user groups. Specifically, there is no existing research identifying and defining user groups within the cultural heritage and wider Galleries, Libraries, Archives and Museums (GLAM) literature, along with their distinguishing characteristics.

The primary goal of this first phase of the thesis was to address this gap by identifying the main user groups utilised in previous Digital Cultural Heritage studies and their respective characteristics by answering the following research questions.

Phase 1 Research Questions:

RQ1: What are the main group categories present in the Digital Cultural Heritage literature?

RQ2: What are the similarities, shared characteristics, and differences of the identified groups?

Data collection for this phase employed an adapted PRISMA approach, facilitating the identification, screening, eligibility checks, data preparation, and data cleaning of potential publications (refer to Chapter 3, Section 3.4.1.3). This process produced a total of 95 user group labels, constituting the actual dataset (see Appendix C Table C.2).

The remainder of the Chapter is organised as follows: Section 4.2 provides an analysis of the dataset and describes the results and provides summaries of the user groups; Section 4.3 discusses and compares user groups and provides areas of discussion; finally Section 4.4 provides the foundation for the next quantitative study.

4.2 Analysis and Findings

The analysis of the collected data was performed in a two part structure. First examining the flat lists of user group names and performing a basic analysis, then secondly, exploring the definitions and details of each group to understand their uniqueness or duplication and to see if there are any relationships such as hierarchies that exist between the groups.

4.2.1 Temporal and Author Analysis of Data Set

In order to understand if the importance of using user groups names and definitions changed over time and also to ensure that the dataset was not biased to limited time frame, a temporal analysis was undertaken (see Chapter 3 Methodology). The inclusion of the year of publication revealed that the number and variety of user group labels had increased over time. The number of sources included, per year, that used a named user group label has steadily increased from the year 2003 and

peaked in the years of 2008 and 2012 which has shown that using named user groups has become more important over time (see Figure 4.1).

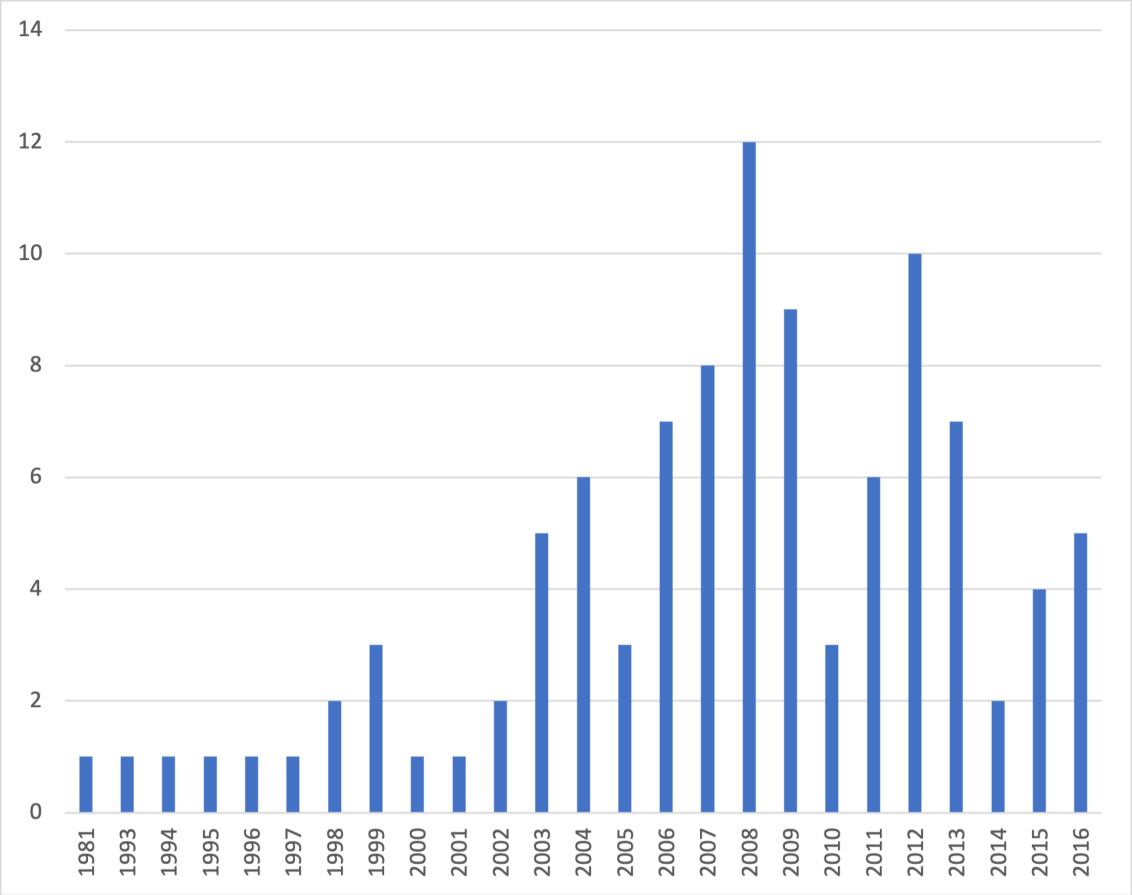


Figure 4.1: Number of sources per year

As well as proving that the dataset is not biased by publication date, and also understanding how the use of user group labels has grown in importance over time, it is important to also ensure that the data set is not biased by having too many publications from a single author. Of the forty-nine authors in the final set of publications, five authors had two publications each (3.3% each of all publications), one author (Marty) had three publications (5%) and one author (Skov) had four publications (6.7%). All other authors had a single publication included. This shows that the final data set is not skewed by any single author or by any single time frame.

4.2.1.1 Uses of User Groups Over Time

Analysis of publication counts by year in relation to the introduction of new group labels (Figure 4.2) reveals a slight upward trend in the usage of user group labels over time. This trend is supported by linear regression, which yields a statistically significant p-value of < 0.05 and a slope coefficient of 0.5251 ($t = 2.172, 95\%CI[0.017, 1.033]$), suggesting an increased adoption of these labels, thereby highlighting the relevance of this review. There is also an observed growth in the number of new user group labels being used. However, the trend line for new introductions of user group labels over time is not statistically significant (p-value of 0.4672), with a slope coefficient of 0.1326 ($t = 0.743, 95\%CI[-0.242, 0.508]$). This indicates that while new labels are being created over time, their introduction does not represent a statistically robust trend. Consequently, it is essential to discern whether these emergent labels reflect genuinely new groups or if they are reiterations of pre-existing categories with study-specific nomenclature.

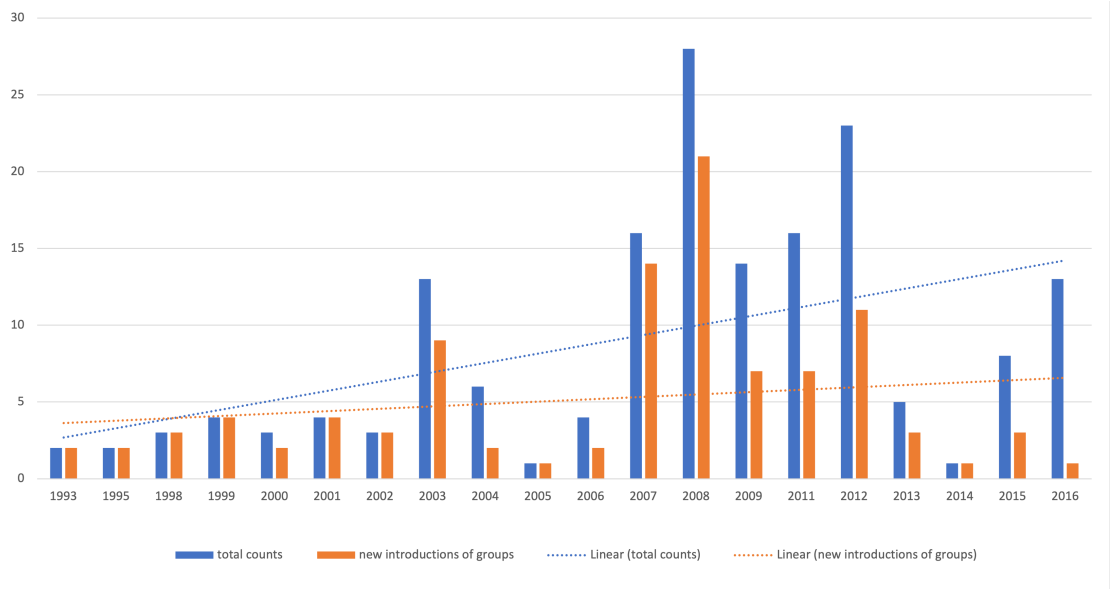


Figure 4.2: User groups used and introduced over time (year/user group mentions per year)

Upon reviewing the literature on user groups (see Figure 4.2 and Appendix C.3), it is observable that earlier studies tended to employ more generic or role-based labels such as 'educational visitor', 'general visitor', 'novice', 'expert', or 'curator'. These labels were typically used singularly, which may be attributable to the fewer

publications during the period from 1993 to 2008. Post-2009, literature reflects a diversification in the use of user group labels with multiple instances of labels like 'experience seeker', 'explorers', 'facilitators', 'hobbyist & professionals', and 'experts', the last of which has been consistently utilised throughout the studied timeframe. Notably, a surge in user group diversity was observed in 2008 with 28 distinct user group labels (of which only 7 were previously mentioned) and another in 2012 with 23 unique labels (13 previously identified). The years 2007 and 2008 mark a pivotal point, introducing 35 new user group labels (34% of the 102 user groups recognised after consolidation), indicative of an increased granularity in user group characterisation. This trend continues into the period between 2007 to 2012 with the introduction of 60 new labels (58% of the total identified). Pre-2007 data shows a lesser diversity with only 34 labels (33%). However, post-2012, there is a noticeable decline in both the use and the introduction of new user group labels, with merely 8 new groups (8%) coming to the fore.

4.2.1.2 Number and Frequency of User Groups Used

Analysing how many user groups were used and studied in each publication and how frequently each user group has been studied provides insights into the definitions. The largest number (n = 23, 39%) of papers referred to a single user group (see Table 4.1). This showed that these papers focused solely on a single user group label and, therefore, were not able to compare their study group against others, and when comparisons were made, they were against all other users that had been cast into a single group, often named *others* or sometimes *general public*. 10 publications (17%) referred to only 2 user groups, with a majority of these focusing on the expert and non-expert user group labels.

Table 4.1: Number of User groups in papers

No. User Groups	No. Publications
1	23
2	10
3	8
4	4
5	8
6	1
7	1
8	3
9	1
Total	59

There were 77 user groups (72.6%) that were only used once in all the publications. Table 4.2 shows that the majority (87.7%) of user groups were only used once or twice across all the publications, with only two user groups (Expert (10 times) and 'Hobbyist & Professionals' (7 times)) used more than 5 times (1.9%). (see Table E.1 in Appendix E for full breakdown by user group label.)

Table 4.2: Number of times a user group has been used in all publications

Times mentioned	User groups
1	77
2	16
3	4
4	4
5	3
7	1
10	1

The Expert was named in 10 sources and is the most used label. Followed by 8 instances of Hobbyist (including all variations of "hobbyist" and "hobbyist and professionals"). 6 instances of Novice (Novice, Novice user and Novice (end user searcher). There were 5 instances of Experience Seeker, Facilitator, Explorer, Recharger, and Student. This demonstrates that a number of authors are using Falk's (2009) work and applying it to the digital world. There were also 4 instances of Scholar. 3 instances of the groups of visitor, tourist, and virtual visitor. All the other groups had only one or two mentions (see Table E.1 in Appendix E).

4.2.2 User Group Label Comparative Analysis

The aim of this section of analysis is to compare the user group labels and see which are unique. This analysis saw the user group labels compared against each other in relation to the group name, group characteristics, the sample of participants used in each of the studies, as well as levels of domain knowledge, technical knowledge and motivations if indicated, as not all publications contained all the stated details.

4.2.2.1 User Group Label: Context and Participant

Establishing each publication's field of study and, where available, the background of the participants to whom the user group label was being applied, either reinforced or cast doubt on the viability of the user group label. This was achieved by identifying the outlet of the publication and focus of the research to identify the field of study, and then looking at the descriptions of the sample for the backgrounds. Interpreting the context that was behind these labels enabled the user group labels to either be merged or remain independent of each other (Section 3.4.1.4).

Examining the publications that were industry focused e.g., museums, the user group labels would often be associated with a work role e.g., curator. A curator would acquire, care for and develop a collection. They would also arrange displays of collection and loaned works and interpret the collection in order to inform, educate and inspire the public (Goodale et al., 2011; Skov, 2009). Whereas, a Museum information professional (MIP) would be expected to develop and manage information resources that help curators, directors, educators and registrars do their jobs and meet the expectations of other audiences (Marty, 2006). The majority of MIP usage, identified in the publications, tended to focus on the digital delivery of the museum experience (Marty, 2006; Gilliland-Swetland and White, 2004), while the Curator remained responsible for the overseeing and delivery of a collection (Goodale et al., 2011; Schouten, 1995; Li et al., 2007; Skov, 2009).

Other publications used user group labels associated with a learning role e.g. Teacher, or on behalf of a role subset e.g., school children (Sweetnam et al., 2012). Similarly, roles associated with higher education, despite having a different user group label, can be placed in a more appropriate group e.g., Uni Student can have 3 subsets reflecting the participants used in the study, namely Undergraduate, Post-

graduate Student (Masters), and PhD label (Nicholas et al., 2007). The user group labels of Field Independence /Dependence and Image Verbalizer can therefore be attributed to undergraduate students due to the role of the participants (Frias-Martinez et al., 2008). Whereas Expert communities can be aligned to Taught Postgraduate students (Masters Student) again based on the participant involved in the study (Srinivasan et al., 2009).

However, occasionally the details of the group label provided in the publications, or in some cases the participants used in the study, differed from the user group name (e.g. Ibrahim et al. was studying a virtual heritage website and names Casual user group, but participants of the study consisted of 9 postgraduate and 1 undergraduate student who had studied in this field and as such cast doubt on the use of a casual users label and were therefore re-classified as part of the Taught-Post-Graduate-Students group.). This process reduced the number of user group labels by 4. The dataset at this point comprised of 91 user group labels (see Appendix C Table C.2). Table 4.3 shows a snapshot of some of the merges that were made. All the columns to the right were merged with the label in the first column, for example, the General user and general visitor labels were merged into the General Public label.

Table 4.3: All merged groups based on similar or plurals

Removed duplicates and similar	**Similars**	
Academic	Lecturer/Academic	
Curator	Museum Curator	
Explorer	Explorers	
General Public	General User	General Visitor
Hobbyist	Hobbyist from Hobbyist & Professionals	
Lay Users	Laymen	
Novice	Novice User	
Postgraduate Student	Researcher Post Graduate Student	Expert Communities
Professional User	Professionals from Hobbyist & Professionals	
Researcher (Independent Researcher, Professional Researcher)	Researcher Group	
Scholar	Humanities Scholar	
Student	Student Group	
Virtual Museum Visitor	Virtual Visitor	Online Visitor

4.2.2.2 Classifying User Group Domain knowledge, Technical Knowledge and Motivations

All publications were annotated for domain knowledge, technical knowledge, and motivation. For every user group label that could be aligned to domain knowledge, technical knowledge and motivations, a low, medium, or high rating was assigned (Russell-Rose and Tate, 2012). The motivation was mapped based on how motivated the group would be in undertaking their task. In some cases, the dimensions and ratings were clearly identifiable from one or multiple of the associated papers. For example, ("Casual users are not domain or system experts" (Ibrahim et al., 2015) and "a novice user may become confused and give up using them [interfaces]" (Bohnert et al., 2012)). In some cases, the assessment had to be extrapolated from the text (for example, "Since novice end-user searchers typically do not possess much search experience, subject knowledge is their only asset" (Hsieh-Yee, 1993) where the subject knowledge was interpreted as medium). At the same time, any motivations for the users' engagement were also identified, as this was found to be another common attribute used to distinguish users (Fantoni et al., 2012).

4.2.2.3 User Group Analysis

The aim of this analysis is to understand the clarity of the user group labels from the literature that defines it, and then to look at how the groups relate to one another.

According to the sources from both the literature review (see Chapter 2) and the literature used in this systematic literature review phase (see Appendix A Table A.1), a user group label can be defined using many characteristics. However, much of this literature only focuses on a single characteristic that was determined by the author of a particular, usually single, publication, the availability and suitability of participants (demographics, locality, physical or digital presence) and the purpose of the study (academic, employment or commercial task based) all of which having contributed to the creation of multiple, yet differentiated versions of the same or standalone user group.

By combining all of these individually determined characteristics, it should be possible to create a much stronger description/characterisation of each user group label and therefore allow future researchers and development teams to have a greater

understanding of the groups they are dealing with.

The characteristics identified in this way are:

1. Demographic related (Pitkow and Recker, 1994; Hall et al., 2012b):
   (a) Age, gender, education, employment status, role in organisation.
   (b) Physical abilities.
2. Task related:
   (a) Goals and motivations (Spellerberg et al., 2016).
   (b) Tasks (europeana, 2014; Skov, 2009; Stiller, 2014; Goldman et al., 2004), (e.g. view artefacts or collections or research a civilisation).
   (c) Frequency of use and training of system or similar systems (from novice to expert) (Taylor, 2015).
   (d) Domain/Subject Expertise (Skov, 2009).
   (e) Equipment accessing system on to undertake task (Lapatovska et al., 2011; Pitkow and Recker, 1994).
3. Geographic Location. e.g (region, county, country).

This list also aligns with work in related areas of user centred design (UCD) (Noyes and Baber, 1999), task analysis (Redish and Wixon, 2003), user interface design (Shneiderman, 1998) and usability engineering (Wixon and Wilson, 1997); and also aligns closely with the findings of Kujala and Kauppinen (2004), who attempted to create a list of user group characteristics for UCD.

However, there are very few publications that detail most of these characteristics for the user groups. The characteristics that can most commonly identified are the technical and domain knowledge, which is why they were explicitly annotated in the earlier step.

4.2.2.3.1 Incomplete and Short Descriptions

None of the user group label definitions identified in any of the literature covered a full set of characteristics. The majority of the definitions identified were in fact

extremely brief and on their own not too descriptive, e.g. Warwick's (2012) Humanities Scholars, described as *"Higher Education use of educational portals in a specific domain"*; and Mason and McCarthy's (2008) Museum staff, described as *"Museum staff plan, implement and maintain museum websites. Often developing them in a haphazard fashion"*.

The shortest of descriptions is only 2 words long with 2 occurrences, e.g. Li et al.'s (2007) Spectator; described as *"passive user"*. The longest description is 179 words long, e.g. Marty's (2008) Museum visitor; described as *"the typical online museum visitor completing this online survey, therefore, visits museums approximately four times a year, visits museum websites approximately once a week, and considers it very important for museums to have museum websites. Online museum visitors are likely to use most types of online museum resources, especially online images and research materials. Most online museum visitors have very different expectations for museums and museum websites. Online museum visitors have a strong positive relationship with museum information resources in that they are frequent users of information resources in museums, and they expect museum information resources to be easily accessible online. Online museum visitors see museums and museum websites as complementary, where one is not likely to replace the other as users search for and access information. Online museum visitors are likely to make frequent visits to museums and museum websites, including visits to museum websites independent of planning or returning from museum visits. They are also likely to use a wide variety of digital museum resources in their daily lives, including online images of artefacts and online research materials"*.

Ten words is the most common length of the descriptions with 10 occurrences, but there are 63 descriptions that are 10 words or fewer. 166 definitions are shorter than half the word count of the longest (90 words) and only 3 are over half of the word count of the longest.

Some descriptions are fuzzy and ambiguous, e.g. Sweetnam et al.'s (2012) General Public description could realistically mean anyone as it covers all ages, all knowledge and technical levels as well as interests: (*"adults and children - a large, diverse set of users that bring a very wide range of interests, technical abilities, contextual and/or historical awareness to collections"*).

Fuzzy or ambiguous definitions like this are of little use in both practitioner and

research contexts, as they do not provide any guidance on how to design for such a group or what further research questions could or should be asked of the group. The lack of an explicit age range of the users impacts on the level of written content and the visual presentation; the unknown level of knowledge also affects the way the content or research question is created. The missing details on technical knowledge hinder the decisions on which features need to be built into the site (e.g. advanced search features, depth of site navigational structures), the unknown interests of the group (with relation to the content of the site) also would pose serious questions for the development team over the structure and interaction patterns to be implemented as well as the content creation and the missing information on tasks and motivations that the group want to undertake on the site limits the possibility of providing a meaningful feature rich system that the users want.

This inability to clearly define the vast majority of the groups, led to the initial approach of attempting to aggregate the groups further to improve the definitions, as discussed later in this chapter, and to the data-driven approaches discussed in the remaining chapters.

4.2.2.3.2 Definitions by Multiple Authors

There is a small set of user groups that are defined by multiple authors. Amongst others, these include:

1. **Novice User** - Rutledge et al.'s (2006): *"Can be hindered by lacking knowledge of rich and complex knowledge structures of museum collections"*; Srinivasan et al.'s (2009): *"Do not find systems easy to use because they are not familiar with the terms used for searches. Prefer content to be in a language they would use and not the language of experts"*; Warwick's (2012): *"May become confused and give up using them [applications] despite the rich possibilities they offer"*;

2. **Experts** - Paternò and Mancini's (2000): *"Experts want to have full access to all the information available. They need minimal support to formulate their requests and should be allowed to formulate such requests in a flexible way"*; and Hampson et al.'s (2012): *"Expert users already have a deep knowledge of the content of the digital collections, allowing their research to create new insights into these artefacts. Aware of the relevant search parameters to access the*

digital content and are motivated to interact with the system because of their research interest".

3. **General Visitor** - Booth's (1998): *"Requires information on opening hours, prices, the museum's facilities, what's on, notable exhibits and navigation aids"*; Skov and Ingwersen's (2008): *"Requires information on opening hours; prices and facilities etc".*

It is also apparent that the different definitions attempted to define the user groups from different angles/view points, relating to their own studies, for example, Novice user - according to Johnson's (2008) while focusing on digital archives: *"does not possess knowledge or experience to locate archival sources. Will often become frustrated early in the interaction process"*; and Warwick's (2012) who was looking at computer applications: *"May become confused and give up using them [applications] despite the rich possibilities they offer".* Both statements possess similarities that could if combined provide a richer and versatile definition of the Novice user group label that could be applied to a broader scope of areas of study, e.g., "Novice user group label shows a lack of knowledge in both the subject and search systems, causing the novice user to become confused and frustrated, resulting in them giving up easily."

However, even with multiple authors defining a single user group, there were no examples of an author definition covering all the characteristics. For example, the expert is the most prevalent of the user group labels indicated in the literature appearing 10 times (see Appendix E) and yet none of the available definitions cover motivation; an expert according to Isard et al. (2003) *"Might be interested in a list of references to published articles that discusses the selected exhibits"*; or Rutledge et al. (2006) *"Museum knowledge structure developed by experts such as art theoreticians and art historians"*; Pantano (2011) in Ibrahim et al. (2015) *"Experts are specialists in the field of cultural heritage"*; Paterno and Mancini (1999) *"Experts want to have full access to all the information available. They need minimal support to formulate their requests, and should be allowed to formulate such requests in a flexible way.".* Blending the definitions in a similar way to the Novice user group label illustrated above, provided a generic and non-operational definition of the Expert user group label, e.g., "Expert - knowledgeable in a specific subject area and is technologically

proficient".

4.2.2.3.3 Conflicting Definitions

Not all the definitions for one user group label provided by multiple authors align, instead, they occasionally appear to offer opposing or conflicting definitions. e.g. Expert - Pantano's (2011): *"Experts are specialists in the field of cultural heritage"*; compared to Amin et al.'s (2008): *"five expert roles - Researchers - Curator - Registrars - Teachers and students - IT personnel ..."* these definitions do not align, one suggesting that experts are specialists in a specific area while the other includes teachers and students who are not CH experts. A further example of conflicting definitions is: Student - Marchionini et al. (2003) describes students as having *"low domain knowledge, low library system skills"*. Whilst Srinivasan et al. (2009) describes students as being *"representative of the types of expert communities that are interested in the museum objects."*. Clearly they can not be both, therefore when the context of the participants of the study was looked at more closely it became clear that the authors were describing different age groups and educational experience, and yet they were defined under the same user group label.

The existence of diverging and or opposing characteristics for a user group label therefore presents the dilemma in terms of the accuracy of the characteristics that are to represent the user group label.

4.2.2.3.4 Uniquely Defined Labels

The dataset (see Appendix E) shows that 72% (n=77) of the user groups labels identified are only defined in a single publication. e.g. Hampson et al.'s (2012) Wider public, Górny and Mazurek's (2012) Prospective user, and Schweibenz's (2004) laymen.

This raises concerns about the user group label definitions, but more importantly it raises the issue of the creating of user group labels by authors for a particular study without understanding of how the user group label relates to previous work or literature e.g., Apprentice Investigator - Sweetnam et al. (2012); Casual User - Ibrahim et al. (2015); Incidental Users - Górny and Mazurek (2012). This ill-informed user group label creation expands the difficulties in achieving clarity of

description and appropriateness of user group labels, which is on of the core aims of this thesis. As each author defines their own user group without understanding or referencing previous uses of the user group, the potential for further distorting the user group definition expands and can cause confusion in researchers who later try to build systems for certain users groups and use the literature as a starting point for their user group descriptions. With 72% of the identified user group labels in thesis research only being used once (see Table 4.2), there is the likelihood of these groups not being significantly different from each other.

4.2.2.3.5 Related User Groups

In an attempt to develop coherent, recognisable and relevant user group definitions, concatenation of all determinable characteristics found in the publications was undertaken based on the user group labels and the definitions. The merging was a multistage process. Firstly, all the descriptions for each label were merged so that there was a single, albeit often repetitive description for each user group label. Then each definition was treated to a process of removing the repeating characteristics and forming the remaining text into a coherent description. The total number of publications used for each description can be seen in Table 4.4.

Table 4.4: The number of publications used to base the group definitions on.

User group	**Number of publications the definition is based on**
Academic	8
Novice	9
Expert	15
Professional researcher	5
Museum staff	3
Library staff	2
Curator	4
MIPS	3
Post doc researcher	2
Historian	1
Commercial publisher	1
General Public	15
Affinity seekers	1
Experience seeker	7
Virtual Museum visitor	10
Rechargers	5
Tourist	3
Student	5
School children	2
College Students	0
Uni student	1
Undergraduate student	4
Taught-Post-graduate students	6
PhD-student	1
Teacher	5
Hobbyist	25
Collectors	2
Facilitators	5
Respectful pilgrims	1

The concatenation of the user group label descriptions presented an opportunity for the creation of meaningful definitions that could readily be associated with a user group label to be developed through the synthesis of the descriptions provided from the publications, e.g., removal of duplicate words/phrases or given examples used to illustrate a label (see Appendix B) for example the Museum staff user group label was based upon the publications of Nicholas et al. (2007); Marty (2007b); Mason and McCarthy (2008); Srinivasan et al. (2009); Skov (2009). This produced an initial definition for the proposed label, for example,

"Museum professionals need to understand what museum visitors are looking for before and after a museum visit and make this information available to them. Mu-

seum staff plan, implement and maintain museum websites. Often developing them in a haphazard fashion. Museum documentation is traditionally written for and by museum professionals. Often containing professional jargon. Digitising cultural heritage collections involves opening up databases, previously the sole domain of the museum professionals. elite experts of the collections held."

The concatenated descriptions were then edited and developed into a singular, coherent and meaningful description for each user group label while capturing the relevant publication descriptors, e.g.,

"Museum staff need to understand what museum visitors are looking for before and after a museum visit and make this information available to them. Museum staff plan, implement and maintain museum websites, often developing them in a haphazard fashion. The digitisation of cultural heritage collections involves opening databases, containing documentation consisting of profession jargon; traditionally written for and by museum professionals"

These expanded definitions, in some instances, provide a more detailed description of the user group label as they are based on multiple sources, whereas some are still based on a single source. A complete list of the definitions can be found in Appendix F.1.

However, further examination of the concatenated list found that while some user group labels were clearly representative of what the label denoted, e.g., **Experience Seeker** - *"Experience seekers want to virtually explore the most renowned pieces and make memories. They enjoy art galleries and museums and aspire to be exposed to the things and ideas that exemplify what is best and intellectually most important within a culture or community. They see museums as important destinations and are collecting experiences."*;

Affinity Seeker - *"Affinity seekers are motivated to visit a museum because it speaks to their sense of heritage and bigger identity."*;

Recharger - *"Rechargers are motivated by the yearning to physically, emotionally and intellectually recharge in a peaceful refreshing environment. They seek to have an experience that allows them to reflect or rejuvenate."*;

Others presented with no descriptions e.g., PhD Student, Uni Student, Commercial Publisher, Post-Doc Researcher, or contained elements that were used in other

labels e.g., **Academic** - *"Academics are highly educated and motivated by work/research tasks to use DCH websites and collections. They can be highly focused on completing their tasks/research and can allocate significant time to the tasks. They have a high level of domain knowledge and can have a high level of library/collection system knowledge, but can require assistance when searching using some interfaces they have never used before. They often benefit from user friendly interfaces that enable easier access to the digital content, and can often turn to browsing for relevant content over search. Sometimes the Academic (historian, sociologist, anthropologist or author) acts as a mediator between the system and the general public as they are able to interpret the unnormalised collection."*;

Expert - *"Experts are specialists in their field and are motivated by improving their knowledge or achieving their task quickly. They have extensive training with the search systems (at least one year's experience) and could use these systems every day. Their high domain knowledge means that they need minimal support in formulating search requests, but they do expect access to all the information that is available. They can understand the complex knowledge structures of collections and would also be interested in a list of references to published materials. There are many roles that have been identified as experts including researchers, curators, registrars, art theoreticians, historians, teachers, IT personnel and some advanced students; all of which are linked to a formal career and as such enable the Expert to have clear goals for the task"*;

Professional Researcher - *"Researchers are people who do historical research. They have high motivation of furthering their knowledge in a specific historic field in order to advance their research and therefore have very high focus but medium levels of time allocations. Their domain knowledge ranges from medium to high, with the group label often being used to refer to academics and historical society personnel. Researchers are often highly educated with advanced degrees. They use search systems and DCH websites as filters to guide their research. They often use secondary sources to point them towards their most important sources: letters, journal's, diaries, historical newspapers, and government legal documents (e.g. deeds, wills and court record's). They often have a meaningful relationship to the collections."*.

This lack of clarity in the definitions, therefore, raises further questions regarding the relevance or validity of the user group label that was being used within the

publication, as well as its suitability to be representative of an entire grouping, for example Hsieh-Yee (1993), Novice (end user Searchers) - defined as non-professional searchers who have little or no search experience and have not taken courses on online searching or attended workshops provided by librarians or system vendors. However, the participants used in this study were 34 graduates recruited from Dept of Educational Administration, University of Wisconsin-Madison.

4.2.3 Comparison of User Groups After Concatenations

In order to establish any commonalities that may exist between user group labels, the literature points towards examining areas such as domain and technical knowledge (Russell-Rose and Tate, 2012) and motivation (Fantoni et al., 2012). By looking for these particular themes across the publications, the formation and identification of generic groupings or categories could be recognised.

Initially, each group was mapped against the two dimensions (domain and technical expertise) set out by Russell-Rose and Tate (2012). Figure 4.3 and Table 4.5 shows how there is significant overlap between these user group labels when evaluated on both dimensions, e.g., the user group labels of Academic, Expert, Library Staff, MIPS, Museum staff all map to high levels of domain and technical expertise.

The user group labels' motivations were also extracted, Table 4.5. However, it was not possible to extract a motivation for the majority of the user groups used in the publications and was deemed to only use this data as supplementary at this stage.

The inclusion of the number of publications that a user group label appeared also contributed to the notion that some user group labels are used more frequently than others, but the precise definition by which the label could be interpreted would still need to be established.

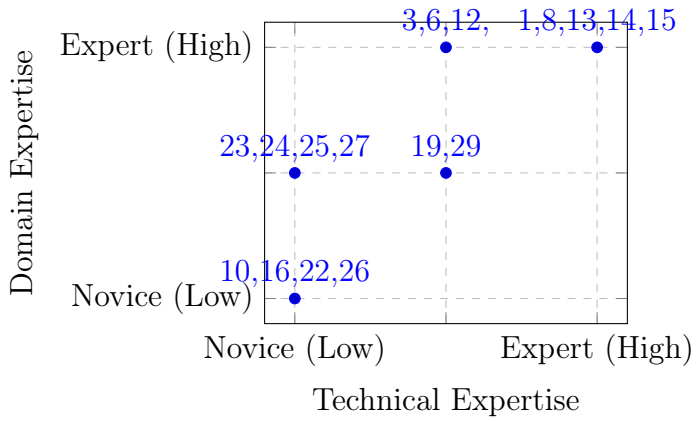


Figure 4.3: User groups (from Table 4.5) mapped onto Russell-Rose & Tate's dimensions of experience grid (Russell-Rose and Tate, 2012). The numbers in the grid indicate the numbers of the user groups identified in Table 4.5.

Table 4.5: Table showing the calculated domain, technical knowledge and motivations as well as the number of publications that contributed to that combined user group definition.

Representative No	User group	DK	TK	Motivation	No of publications
1	Academic	High	High	Work	8
2	Affinity seekers	Unknown/Unspecified	Unknown/Unspecified	Unknown/Unspecified	1
3	Collectors	High	Med	Unknown/Unspecified	2
4	College Students	Unknown/Unspecified	Unknown/Unspecified	Learn Captive - Study	2
5	Commercial publisher	Unknown/Unspecified	Unknown/Unspecified	Unknown/Unspecified	1
6	Curator	High	Med	Work	5
7	Experience seeker	Low	Unknown/Unspecified	Unknown/Unspecified	7
8	Expert	High	High	Unknown/Unspecified	15
9	Facilitators	Unknown/Unspecified	Unknown/Unspecified	Unknown/Unspecified	5
10	General Public	Low	Low	Unknown/Unspecified	12
11	Historian	Unknown/Unspecified	Unknown/Unspecified	Unknown/Unspecified	1
12	Hobbyist	High	Med	Unknown/Unspecified	20
13	Library staff	High	High	Work	2
14	MIPS	High	High	Work	3
15	Museum staff	High	High	Unknown/Unspecified	5
16	Novice	Low	Low	Unknown/Unspecified	9
17	PhD-student	Unknown/Unspecified	Unknown/Unspecified	Unknown/Unspecified	1
18	Post doc researcher	Unknown/Unspecified	Unknown/Unspecified	Unknown/Unspecified	2
19	Professional researcher	Med	Med	Work	5
20	Rechargers	Unknown/Unspecified	Unknown/Unspecified	Unknown/Unspecified	5
21	Respectful pilgrims	Unknown/Unspecified	Unknown/Unspecified	Unknown/Unspecified	1
22	School children	Low	Low	Learn Captive - Study	3
23	Student	Med	Low	Learn Captive - Study	6
24	Taught-Post-graduate students	Med	Low	Unknown/Unspecified	6
25	Teacher	Med	Low	Unknown/Unspecified	5
26	Tourist	Low	Low	Unknown/Unspecified	3
27	Undergraduate student	Med	Low	Learn Captive - Study	4
28	Uni student	Unknown/Unspecified	Unknown/Unspecified	Unknown/Unspecified	1
29	Virtual Museum visitor	Med	Med	Pleasure, Plan visit, or Curiosity	9

Mapping publications to the Russell-Rose and Tate model revealed that the majority of these user group labels fit into the high levels of domain expertise, while

also benefiting from a full spectrum spread of technical knowledge e.g., Academic label Domain Expertise = High Technical Expertise = High; Curator label Domain Expertise = High Technical Expertise = Med; Professional Researcher label Domain Expertise = Med, Technical Expertise = Med; Teacher label Domain Expertise = Med, Technical Expertise = Low; Novice label Domain Expertise = Low, Technical Expertise = Low.

In order to map the user groups to the Russel-rose and Tate model, initially each user group identified in the literature was examined and a rating of 1 (low), 2 (med) and 3 (high) was made for the technical knowledge and domain knowledge. Not all user group labels contained this information, and for those that did not declare or elude clearly to this, the field was noted as unknown/unspecified (see Table 4.5).

The ratings of the DK and TK for all the merged groups were evaluated, using the mode maths model (the most common rating for the DK and TK for those merged groups was assigned to represent the group). In those groups that had equal ones twos or threes (a tie situation with no mode), the highest value was adopted.

Therefore, the application of Russell-Rose and Tate model, mapping of the domain knowledge and technical knowledge, shows that all 18 of the 18 concatenated user groups identified share commonality either through design, coincidence or author subjectivity. For example, Table 4.5 presents a breakdown of the scores for each of the user group labels. There are 4 user group labels sharing the Low Low position (General Public, Novice, School Children and Tourist); 2 user group labels share the Medium Medium spot (Professional researcher and Virtual museum visitor); 3 user group labels share the High Medium position (Collectors, Curators, Hobbyist); 4 user group labels share the Medium Low spot (Student, Taught Postgraduate Student, Teacher, Undergraduate Student); 5 user group labels share the High High point (Academic, Expert, Library Staff, MIPs', Museum Staff).

Additional scrutiny of the mappings established the emergence of five main groupings of categories (see Table 4.6). These five groupings are based on similarities in domain knowledge and technical competence; for example, 'general public' and 'tourist' are similar in terms of domain and technical expertise (poor), while having distinct names.

Therefore, this means that either, there really are only nine possible groups of users (one for each position of the three by three grid); of which there have only

been five identified in the literature used in this review, a number of which have been given different labels; or that defining the user groups by the two dimensions of technical knowledge and domain knowledge does not allow enough scope for the true individuality of the groups to emerge and that this model is missing some additional dimensions like those outlined in Section 4.2.2.3.

Whilst there are a number of publications that have used the two factors of domain knowledge and technical knowledge as ways to separate and define groups, there are a third of the user groups identified in this research where these dimensions are not specified, which again raises the questions of is using this as a comprehensive method for understanding the groups based on literature feasible?

Expanding the model by adding another dimension of the users' motivations, does not help much as a clearly defined motivation is not available directly from most of the publications. Even with this data there could still be overlapping issues. For example 'academics' and 'Museum staff' may be considered similar with respect to domain and technical expertise (both being high) with a motivation of work being aligned to the Academic label and unknown to Museum Staff, although a reasonable assumption could be that of work also. Therefore, these two different role based groups would technically come out as the same using this three dimensional model.

It is not necessarily an issue that two groups have this overlap from an operationalisation point of view, as both groups may want the same levels of information and have similar technical abilities. But should their desired tasks be different, as one would assume they would be with the user groups of academics (searching for information) and museum staff (creating information) then using this model even with the motivation dimension added would mean that operationalising is still an issue.

A further issue with a model like this becomes apparent when looking at the 'Experience seekers' user group as it was identified in the literature, and described as having varying levels of the technical knowledge dimension from the multiple authors defining the group. The issue here is that because the user group label has multiple definitions that give it such a wide variation that they could sit anywhere on the scale. As a result, the experience seeker group has not been taken into the next level of grouping.

However, this may be evidence of this group being a sub-group that would sit

within one of the main five identified in the results of the Russel-Rose model, once the variation level has been set (e.g., if the Experience Seekers' technical knowledge was low then this would belong to Group 5; if the technical knowledge was high, then this would form a separate Group 6). Figure 4.4 and Table 4.6 shows these bigger merged groups applied to the (Russell-Rose and Tate, 2012) grid.

Whilst the findings have shown these five high-level groups, there is evidence to suggest that there are sub-groups within the groups. For example, Skov (2013) demonstrates that, even in rather narrow groupings, there are sub-groups: e.g., hobbyist may have sub-groups of 'collectors' and 'liberal arts enthusiasts'. The number of differently named and motivated roles in the fifth group of Table 4.6 (General Public, Tourist, School Children, Novice (Plus those merged with these groups earlier in the process - General user, General visitor, Experienced User, laymen, Incidental users, Potential visitor, spectators, surfers, Tentative observers, wider-public, captive user, Amateur, Digital archive user, Novice (end user searchers), Non-expert, Novice User) also highlights that there are potential sub-groups with only minor differences. Further work is needed to identify if these differences in motivation make these unique groups, or if they should be grouped together.

However, what is evident from this phase is either the user groups are not defined in the literature well enough, with the user group labels being unique, but the definitions of the labels are not as distinguishable from each other as initially thought, or the subjectivity of the publication's author could believe the formulation of a user group label to such a degree that what appears as a recognised definition for one label, e.g., novice who cannot use a system, may be labelled by another author who may use the same label but rather align it to a definition of a novice who can use a system. Equally, the user group labels of Expert and Academic also blur along the lines of definition. For example, if the definition of each was provided to a development team, they would struggle to distinguish between the labels and be able to provide the correct experience. It is this lack of clarity regarding definition appropriated to a user group label that leads to confusion and the inability to utilise these labels in a real world setting by demonstrating the potential fuzziness of using both the named labels of the groups to suit the research and also using only the two dimensions to distinguish between user groups. Or the alternative is that there are limitations of the model.

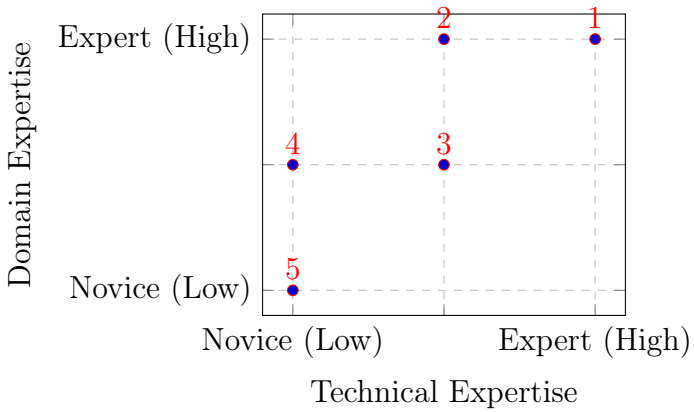


Figure 4.4: User Role groups (from Table 4.6) mapped onto Russell-Rose & Tate's dimensions of experience grid (Russell-Rose and Tate, 2012). The numbers in the grid indicate the numbers of the user groups identified in Table 4.6.

Table 4.6: User categories mapped to domain and technical expertise, and motivation

	User category	**Domain expertise**	**Technical expertise**	**Motivation**
1	Academic	High	High	Work
1	Expert	High	High	Not defined
1	Library Staff	High	High	Work
1	MIPs	High	High	Work
1	Museum Staff	High	High	Not defined
2	Hobbyists	High	Med	Not defined
2	Collectors	High	Med	Not defined
2	Curators	High	Med	Work
3	Professional researchers	Med	Med	Work
3	Virtual Museum Visitor	Med	Med	Pleasure, Plan visit, or Curiosity
4	Teacher	Med	Low	Not defined
4	Student	Med	Low	Learn Captive - Study
4	Undergraduate Student	Med	Low	Learn Captive - Study
4	Taught Post-Graduate Student	Med	Low	Not defined
5	General Public	Low	Low	Not defined
5	Tourist	Low	Low	Not defined
5	School Children	Low	Low	Learn Captive - Study
5	Novice	Low	Low	Not defined

Table 4.6: End of table from the previous page.

4.2.4 Establishing User Group Label Relationships and Hierarchy

Examination of the publications in terms of context and participants, extracting user group label characteristics and or definitions from the publications and literature (see Chapter 2) allowed for establishing relationships between the 27 user group labels created in the merging process described above (see Table 4.4). Based on the labels and definitions used by the publication author for each user group label, an overarching umbrella label could be identified that could encapsulate all of the labels that possessed either a similar name or a characteristic, e.g., General Public, Researcher, Educational Visitor.

Once the main themes (General Public, Educational Visitor, Researcher) were established the remaining labels were then allocated to a theme associated with the traits and characteristics that were found during the concatenation process (see Appendix F), and or the publication itself alluding to a particular theme or the possible relationships that could naturally be seen between the user group labels.

The Educational Visitor label consists of those user group labels that had an association with some form of educational experience. Labels that reflect the nature of those who take a significant part in either undertaking learning or a role in providing learning can be captured in the user group labels of Student, Teacher, Academic. The natural evolutionary sequence for undertaking education as a learner typically ranges from School Children to College Student to University Student, yet all can be considered to belong to a class of student. The sequence of undertaking higher education qualification's also lends itself to an ordering of supposed knowledge acquirement e.g. undergraduate, graduate, post-graduate, PhD; all being forms of a university student. The Teacher label captures all experiences aligned to the provision of a learning opportunity to others, while the Academic label reflects upon an individual task based focus and research activity. The sub-group of label Post-Doc Researcher aligns to the Academic label through the natural relationships that occur in the real world. Often, academic's will have assistants who are highly educated (PhD holders) who may undertake research activities at the behest of the academic.

According to the Cambridge Dictionary, a researcher is "someone whose job is

to study a subject carefully, especially in order to discover new information or understand the subject better". The use of a Researcher label therefore, allowed the sub-division of the category into a Professional Researcher whereby those labels that align specifically to museum work/places of employment e.g., Museum Staff, Curator, MIP's, Library Staff; while user group labels such as Hobbyist and Collector, may share an interest in a particular subject area, this is not as a form of employment and so can be categorised under a non-professional Researcher label.

The remaining high level user group label, General Public, consists of "ordinary people, especially all the people who are not members of a particular organisation or who do not have any special type of knowledge" (Cambridge Dictionary). The General Public label captures "those who usually visit the website more for some passing curiosity rather than to retrieve information to improve their knowledge of cultural heritage" (Ibrahim et al., 2015) and as such consist of varied and numerous characteristics that could not neatly occupy an educational or research associated label. However, some specific user group labels did emerge and have been aligned to the GP label on the basis that the participants indicated in the publications were from none of the other categories. For example, Osman (2016), used the labels of an Affinity Seeker who visit a museum because it speaks to their sense of heritage and a Respectful Pilgrim who goes to museums as a sense of duty, characteristics that are based in the emotions of a visitor, in order to explore the visitor roles to the museum that would aid the museum designers to create a better experience; Falk (2009), used the labels of an Experience Seeker who wants to see the most renowned pieces and make memories; a Recharger who wants to relax in a peaceful atmosphere seeking a restorative experience; a Facilitator who wishes to engage in a meaningful social experience with someone who they care about, from participants of a physical museum in order to establish the reasons they go, which was then mirrored by Fantoni et al. (2012), who conducted a survey to online museum visitors for the same purpose; Marty (2007b), used the label Online Museum Visitor who has specific and different needs and expectations of museum websites before and after a physical museum visit. Prior to the visit they seek information about the physical visit, online tours, but afterwards they additional information and images about collections and artefacts and future exhibitions and events, by surveying those people who had attended a physical museum in relation to their engagement activities with

the online version.

Combining the above into a hierarchy diagram produced Figure 4.5. This diagram shows a hierarchy within the three main categories, and whilst the top three categories are considered to be an equivalent level, it was not feasible to assign levelness across all the sub-groups.

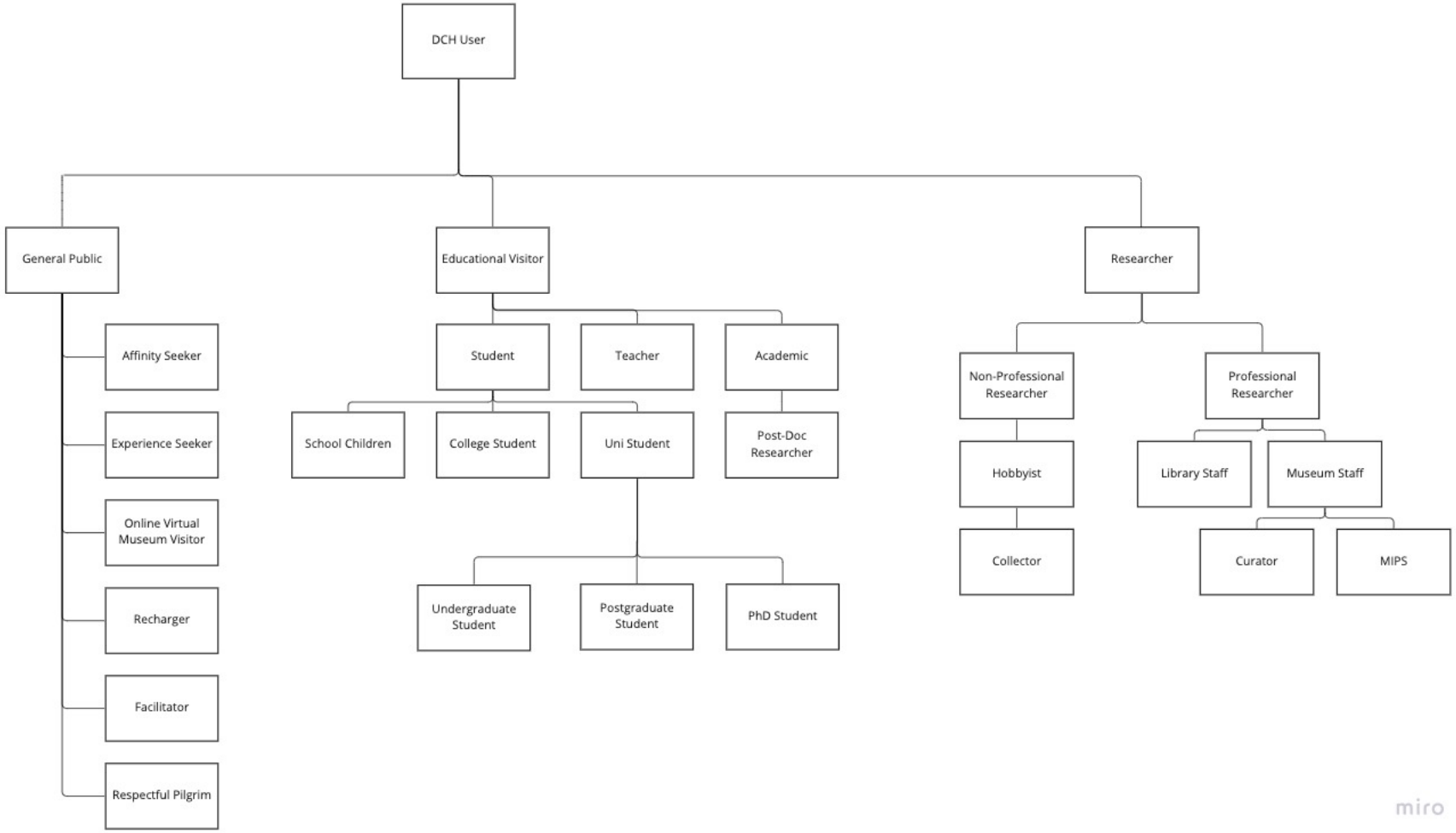


Figure 4.5: Devised Relationship Hierarchical chart of user group labels

4.2.5 The Novice and Expert Continuum

While Figure 4.5 depicts how the relationships between user group labels can be visualised, two user group labels have been omitted from the hierarchy, namely that of Novice and Expert. The user group labels of Novice and Expert, and the determining characteristics, were also established through the concatenation process. The Novice user label (Hsieh-Yee, 1993; Paterno and Mancini, 1999; Rutledge et al., 2006; Johnson, 2008; Srinivasan et al., 2009; Warwick, 2012) was the epitome of a user that has minimal knowledge, weak system experience and gives up easily, while the Expert user label (Lin and Gregor, 2006; Amin et al., 2008; Hampson et al., 2012; Ibrahim et al., 2015) roamed from associations with museum staff e.g., Curators, registrars and researchers, to detailed subject knowledge and extensive experience with search systems.

However, existing as a label in its own right or as a possible extension to any user group label of "novice" or "expert" was also seen when drilling down into the publications, e.g., Amin et al. (2007), uses the label Expert to capture the roles of those who worked in 4 different CH institutions (Researchers, Curators, Registrars) while Duff and Johnson (2002), uses the label Historian with the participants used in the publication consisting of assistant and associate professors who are clearly expert users of texts. Similarly, Srinivasan et al. (2009), uses the label Novice User in the publication, on the basis that they do not find systems easy to use because they are not familiar with the terms used for searches and would prefer content to be in language they would use and not the language of experts. Yet the publication participants consist of groups of masters students from the Dept of Information Studies at UCLA and a group of Inuit high-school students from Inukshuk High School, both groups being considered as types of experts in museum objects and their representation in catalogues and as such would be familiar with language used in their field.

This interchangeability of meaning behind these skill based labels evidenced through the literature further blurs what are already fuzzy user group definitions. Ordinarily, people develop skills through years of experience and by progressing from novice, advance beginner, proficient, competent, and fully expert. These stages being contingent to progressive problem-solving that differentiate experts from novices.

However, with the flexibility of scope applied by authors to these labels by either capturing other distinct labels e.g., Curator, Museum Information Professional, through use of Expert as a generic term or by the inclusion of participants who do not reflect the label they are representing in the publication e.g., Marchionini et al. (2003) uses the label Hobbyist while using participants from LC staff characterised as experts in both library system and content in addition to others; casts doubt on the suitability of these labels to become unique singular labels.

It is therefore clear that while these are not user groups in their own right, they can instead be considered as extensions to all the user groups identified in the hierarchy. As extensions, "expert" and "novice" can be indicative of the level of competence that may be applied to a label which can vary along a knowledge or technology based continuum that, was not specifically addressed in the publications, and hence can change the dynamics of the user group label.

However, the hierarchy is already fuzzy with the sub groups having most of the characteristics of the parent plus an addition, but then when adding in a novice to expert scale on each of these groups the lines between the groups and the group definitions become even more fuzzy.

4.3 Discussion

There are a growing number of studies that focus on individual virtual museum visitor groups in specific scenarios but most specify the division of the users into groups by single aspects. A large number of these studies casually reuse the user group labels and definitions from the physical museum setting with no regard to the entirely different setting and experience. The overall impact is that of vagueness in the ability to be able to define a DCH user, which despite the volume of publications, is evident through a lack of holistic studies of the DCH user groups.

This therefore posed the following questions: RQ1: What are the main user group categories present in the Digital Cultural Heritage literature? and RQ2: What are the similarities, shared characteristics, and differences of the identified groups?; identifying the user groups who engage with DCH is the premise of this research phase.

Establishing those user labels that have been featured in the literature (RQ1) to

date and understanding the characteristics and or traits of each as per the publication's author was the first stage in the development of a hierarchical model. The data results (Appendix C.3) showed that there was a growing rise in the use of authors/researchers studying users and documenting the user groups they studied. This rise in user groups being identified in the literature aligns with the rise of the user-centred design process, which later grew to be the larger industry-standard field of practice called user experience design. The growth in this field highlights the importance of developing for actual end users and, more importantly, the requirement to understand these users before design or development starts.

However, not all traits that could be recognised from the publication could be cleanly allotted to a user group label. Contradictions (RQ2) regarding a label name and its characteristics were apparent when the authors themselves discussed how the difference between Expert and Academic was not really that different or authors (Bohnert et al., 2012) using a Novice user group label stated that they could not use systems while others stated they do. Therefore, attempting to formulate the standardisation of user group labels that could be organically used by others in a similar way as used through personas (Cooper, 2004) (a standardised descriptor is provided) would be impossible, as they would struggle to distinguish between some of the group labels caused by this ambiguity.

What did become clear, while looking for similarities (RQ2) in terms of user group label name, or label characteristics, was the need to, where possible, create user group label definitions from the literature, which in turn allowed for an attempt at standardisation. Once irrelevant sentences and words were removed from the complied definitions a more meaningful definition was established (see Appendix F.1).

However, what did become apparent while performing this activity was the breadth of scope, that was attributed to the labels e.g., User group label Expert according to Amin et al. (2007), incorporates Researchers, Curators and Registrars who work within the museum environment. Therefore, the inclusion of the Russell-Rose and Tate (2012) model and the subsequent mapping of the labels against the criteria for domain and technical expertise that has been used as a defining grouping factor in previous literature, was performed in order to establish any commonalities that may exist between the labels and as such reduce the overall number. How-

ever, what was found was an added level of variability regarding the reliability and validity of said user group labels.

According to the Russell-Rose and Tate (2012) model, the dimensions of: *technical expertise* (e.g., proficiency in using technology, computers, the Internet, search engines etc) and *domain expertise* (e.g., familiarity with a given subject matter) suggests that some labels could be more heterogeneous than implied in the publication e.g., Table 4.6 revealed that labels such as Professional Researcher, Teacher and Taught Post-Graduate Student are considered to have a medium level of domain expertise but with also a medium to low technical expertise level, although when considering the work role of this type of label, reality would contradict this result and possibly place domain and technical expertise as high, which again would increase the likelihood of casting doubt on the reliability of the label definition.

Ultimately, while this phase has provided insight into the use and definitions of the labels that are currently being used in the field of CH, what cannot be ignored is the lack of uniformity into how the labels are created, used, interpreted and applied throughout the literature. It was found that essentially none of the definitions of the user group labels discovered in this phase adequately covered the characteristics of: Demographics (Age, gender, education, employment status, role in organisation, location and Physical ability); Goals and motivations; Tasks (e.g. view artefacts or collections or research a civilisation); Technical expertise; Domain/Subject Expertise; appropriateness to the user group label or the subject areas that were being investigated. It was found that while some labels could be aligned to the formation of a hierarchy, crossover between some labels was apparent, e.g., Professional researcher "They are highly motivated and focused on furthering personal knowledge in a specific historic field and/or to advance their research, although domain knowledge ranges from medium to high levels" and that of the hobbyist who are "motivated and driven by curiosity and interest in an intellectually challenging environment with a specific subject. They are highly focused and possess high levels of motivation and domain knowledge" (see Appendix F). Also, labels of the same name contradicted each other in terms of definition, e.g., the Student user group definitions from Marchionini et al. (2003) showed the group having "low domain knowledge and low library skills", whilst Srinivasan et al. (2009) described them as being "representative of experts". While those that had very similar definitions had

differing names, e.g., Collector and Hobbyist. Also, whilst a few have a very vague indication of levelness (low or high) to each characteristic, many had no indication at all and for those that did it is difficult to establish how these differ from label to label. The overall result impacted the lines separating the overall definitions, which in their current form are blurred and non-operational, especially by a team of developers.

In summary, the research contribution is understanding that the definitions are not usable for research or practice, but that the labels may be. The problems that arose in the analysis all point to a lack of formality in the group definitions: contradictions, holes, fuzziness, brevity, and genericness. The aim of extracting groups and definitions is clearly not achievable, but very necessary. It is possible to extract a set of more general user group labels and arrange them hierarchically, and the next step is to try to put some formality in the definitions of those labels. This will be crucial in order to make progress in the field of Digital Cultural Heritage and develop more effective and inclusive user-centred designs.

4.4 Summary

In this study, the initial aim was to investigate how to support multiple user groups with their access to digital cultural heritage (DCH) collections, based on the assumption that user groups were well-defined in the literature. However, a knowledge gap around the users of DCH websites was discovered, resulting in a pivot to identify the information-seeking characteristics of those user groups of museum websites. This involved extracting groups and definitions from literature, but it was found that the aim of doing so was not achievable due to the inconsistencies in the usage of group labels and definitions. As a result, it was recognised that the definitions are not usable for research or practice, but the labels may be.

The focus then shifted to identifying user group labels used in the literature and establishing the rationale for their use, as well as refining and categorising them based on their characteristics. This examination allowed the identification of overarching themes and relationships, and an attempt was made to create a user label hierarchy. However, the limitations of the current user group labels and the need for clarification to prevent random creation and usage were realised.

To address these issues, the thesis has pivoted to gather actual DCH user data to establish the characteristics of user groups that engage with a real-world museum and gallery website for the National Museums Liverpool. This was crucial for making progress in the field of DCH and developing more effective and inclusive user-centred designs. The research contribution lies in understanding the limitations of the existing group definitions, highlighting the importance of formalising user group labels and their definitions, and using this insight to inform future research and practice.

Chapter 5

PHASE 2: SURVEY OF USERS AND GROUPS FROM NML

5.1 Introduction

The literature review and Phase 1 revealed numerous studies on DCH website users, each with specific focuses such as motivation or experience, targeting single user groups such as professionals or hobbyists. However, existing definitions fail to encompass all user group characteristics, and aggregating them does not yield a cohesive set of definitions usable for research or practice.

Therefore, developing a system for user groups by merging characteristics from various literature sources or relying on incomplete definitions could pose significant challenges. As emphasised by the UK Government service manual, failing to understand users' needs can result in building inadequate services (UK, 2017).

The main objective of this second phase was to collect data on diverse characteristics from a broad spectrum of live DCH website users, aiming to enrich the existing characteristics and group classifications identified in Phase 1 and/or establish new groups by answering the following research questions:

Phase 2 Research Questions:

RQ3: How are user groups defined based on the survey responses?

RQ4: How do the aspects of the user groups defined in the survey compare to those in the literature?

Data collection for this phase involved designing and implementing a pop-up online survey to gather individualised opinions involving motivation, task, engagement, domain and technical expertise, shared experiences, and general demographic information. Users were also asked to self-classify from a provided list of user group labels or suggest their own. The survey ran for 28 days, during which 663 responses were received. Subsequently, the collected data was cleaned to ensure compliance with consent, age, completeness, personal data, or abuse criteria (see Chapter 3 Methodology, Section 3.4.2). This process resulted in a total of 514 survey responses constituting the actual dataset.

The remainder of the Chapter is organised as follows: Section 5.2 details the participant responses; Section 5.3 survey results and analysis; Section 5.4 an overall discussion and culminating with Section 5.5 an overview of the Phase 2 findings.

5.2 Responses

This section presents the participant responses for all survey questions. The results are reported as the raw number of answers and as percentages, rounded up to the nearest whole number.

5.2.1 Demographics: Location, Gender, Age and Education

Establishing the age of participants interacting with the NML website revealed that 184 (36%) of the respondents were in the 35-54 age bracket, while 141 (27%) were in the 55-64 group, 90 (18%) were in the 18-34 group, 81 (16%) were in the 65-74 group and 18 (3%) were 75 or over (see Table 5.1).

Table 5.1: Q16 - Ages.

Age Group	Number	Percentage
35 - 54	184	36%
55 - 64	141	27%
18 - 34	90	18%
65 - 74	81	16%
75+	18	3%

While the majority of the survey respondents were Female (311 or 60%), 193 (38%) classified themselves as male and the remaining 10 (2%) opted for "Rather

not say" (see Table 5.2).

Table 5.2: Q17 - Gender.

Gender	**Number**	**Percentage**
Female	311	60%
Male	193	38%
Rather not say	10	2%

The educational attainment level (see Table 5.3) of the participants revealed that 65 (13%) had successfully completed an undergraduate degree; with a further 10% having attained Masters level or college diploma. Other participants indicated the completion of a variety of educational qualifications, ranging from GCSE - or equivalent 10% (50 participants) to a combination of school leaver qualifications through to professional CPD.

Table 5.3: Q18 - Education programs or courses completed.

Education Level Completed	**Number**	**Percentage**
Undergraduate Degree	65	13%
Further education / college diploma	53	10%
Master Degree	52	10%
Secondary School (GCSE's, O-levels)	50	10%
Secondary School (GCSE's, O-levels), Further education / college diploma	45	9%
Secondary School (GCSE's, O-levels), Further education / college diploma, Undergraduate Degree, Professional Certifications	39	8%
Secondary School (GCSE's, O-levels), Further education / college diploma, Undergraduate Degree	29	6%
Undergraduate Degree, Professional Certifications	25	5%
Professional Certifications	21	4%
Secondary School (GCSE's, O-levels), Further education / college diploma, Professional Certifications	19	4%
Doctorate	18	4%
Secondary School (GCSE's, O-levels), Further education / college diploma, Undergraduate Degree, Master Degree	16	3%
Secondary School (GCSE's, O-levels), Further education / college diploma, Undergraduate Degree, Master Degree, Professional Certifications	13	3%
Master Degree, Professional Certifications	11	2%
Further education / college diploma, Professional Certifications	9	2%
None	8	2%
Secondary School (GCSE's, O-levels), Undergraduate Degree, Professional Certifications	7	1%
Secondary School (GCSE's, O-levels), Professional Certifications	4	1%
Secondary School (GCSE's, O-levels), Undergraduate Degree	4	1%

Secondary School (GCSE's, O-levels), Further education / college diploma, Undergraduate Degree, Master Degree, Doctorate	3	1%
Secondary School (GCSE's, O-levels), Undergraduate Degree, Master Degree	3	1%
Doctorate, Professional Certifications	2	0%
Undergraduate Degree, Master Degree, Professional Certifications	2	0%
Further education / college diploma, Undergraduate Degree	2	0%
Secondary School (GCSE's, O-levels), Undergraduate Degree, Master Degree, Professional Certifications	2	0%
Secondary School (GCSE's, O-levels), Further education / college diploma, Undergraduate Degree, Doctorate, Professional Certifications	1	0%
Secondary School (GCSE's, O-levels), Master Degree, Professional Certifications	1	0%
Secondary School (GCSE's, O-levels), Further education / college diploma, Undergraduate Degree, Master Degree, Doctorate, Professional Certifications	1	0%
Secondary School (GCSE's, O-levels), None	1	0%
Master Degree, Doctorate, Professional Certifications	1	0%
Secondary School (GCSE's, O-levels), Undergraduate Degree, Master Degree, Doctorate, Professional Certifications	1	0%
Secondary School (GCSE's, O-levels), Further education / college diploma, Master Degree	1	0%
Secondary School (GCSE's, O-levels), Master Degree	1	0%
Secondary School (GCSE's, O-levels), Undergraduate Degree, Master Degree, Doctorate	1	0%
Further education / college diploma, Undergraduate Degree, Professional Certifications	1	0%
Undergraduate Degree, Doctorate	1	0%
Undergraduate Degree, Master Degree, Doctorate	1	0%

According to the data, 58% of the respondents (290) were currently not studying, while 42% (224) were studying at least one qualification (see Table 5.4). Of those currently studying, 155 (31%) were studying a single qualification and 55 (11%) claimed to be studying multiple qualifications simultaneously. There were 14 (3%) participants that provided "other options" these mainly covered undertaking non-formal online courses/MOOC's, local night school practical courses or PostDoc research.

Table 5.4: Q19 - Education in Progress.

Education in Progress	**Number**	**Percentage**
None	290	56%
Professional Certifications	34	7%

Undergraduate Degree	34	7%
Further education / college diploma	31	6%
Master Degree	25	5%
Secondary School (GCSE's, O-levels)	19	4%
Secondary School (GCSE's, O-levels), Further education / college diploma	15	3%
Other	14	3%
Secondary School (GCSE's, O-levels), Further education / college diploma, Professional Certifications	9	2%
Doctorate	9	2%
Undergraduate Degree, Professional Certifications	7	1%
Secondary School (GCSE's, O-levels), Further education / college diploma, Undergraduate Degree	5	1%
Secondary School (GCSE's, O-levels), Further education / college diploma, Undergraduate Degree, Professional Certifications	4	1%
Further education / college diploma, Undergraduate Degree	2	0%
Secondary School (GCSE's, O-levels), Professional Certifications	2	0%
Secondary School (GCSE's, O-levels), Further education / college diploma, Undergraduate Degree, Master Degree, Professional Certifications	2	0%
Further education / college diploma, None	2	0%
Secondary School (GCSE's, O-levels), Undergraduate Degree	2	0%
Master Degree, Professional Certifications	1	0%
Doctorate, Professional Certifications	1	0%
Further education / college diploma, Professional Certifications	1	0%
Secondary School (GCSE's, O-levels), Further education / college diploma, Undergraduate Degree, Master Degree	1	0%
Secondary School (GCSE's, O-levels), Master Degree	1	0%
Secondary School (GCSE's, O-levels), Further education / college diploma, Undergraduate Degree, Master Degree, Doctorate, Professional Certifications	1	0%
Undergraduate Degree, Master Degree, Professional Certifications	1	0%
Master Degree, None	1	0%

It appears that this question may have been misunderstood by the participants. The question was seeking for the respondents to simply indicate the current qualifications they were studying, however, because the participants were allowed to select

multiple options, they selected all the qualifications they had previously completed and also included any that were currently being studied. Table 5.5 shows the same results as in Table 5.4 but with only the highest qualification listed in each option selected as the current one being studied. Where the highest is "Secondary School" these have been marked as "None" due to only over 18's being in the result and the participant has already agreed in Q1 to be 18 or over and also placed themselves in an age category over 18 in Question 16. The secondary Professional certifications were included with the professional certs.

Table 5.5: Q19 - Education in Progress - Compressed.

Education in Progress	**Number**	**Percentage**
None	309	60%
Professional Certifications	36	7%
Undergraduate Degree	43	8%
Further education / college diploma	48	9%
Master Degree	28	5%
Doctorate	9	2%
Undergraduate Degree Professional Certifications	11	2%
Further education / college diploma Professional Certifications	10	2%
Master Degree Professional Certifications	4	1%
Doctorate Professional Certifications	2	0%
Other	14	3%

Establishing the location of participants as they completed the survey (see Table 5.6) indicated that 59% of the participants were within a day trip's driving distance to the physical museums. Overall, 84% were from the UK (76% were from England, 6% were from Wales, Scotland and Ireland combined); while only 16% were from outside the UK.

Respondents visited from all over the world, but the majority were local to Liverpool. The percentages then decrease as the distance increases from Liverpool. That is apart from the overseas users, which accounted for 16% of respondents.

The complete locale figures are: 35% of respondents were from the Merseyside area (local to the museum). 24% from the North West of England. 18% from the rest of England, 7% from Wales, Scotland or Ireland and 16% from the rest of the world. The most popular locations for respondents to reside are: UK 84% (n = 498); USA 9% (n = 58) and 2% with an unknown location.

Table 5.6: Q20 - Location.

Location	Number	Percentage
Merseyside	182	35%
Northwest England	121	24%
Rest of England	93	18%
Outside of the UK	84	16%
Wales, Scotland or Ireland	34	7%

Table 5.7 shows that the majority of participants (297 - 57%) were in employment; 38% in full-time and the rest in part-time. Only 7% (36) identified as being in education. Just under a quarter (119 - 23%) identified themselves as retired. 9% (50) identified as being not employed and 3% (17) identified as being disabled and not able to work.

Table 5.7: Q21 - Employment status.

Employment Status	Number	Percentage
Employed, working full-time	193	38%
Retired	119	23%
Employed, working part-time	99	19%
Student (Full time)	31	6%
Not employed, NOT looking for work	33	6%
Not employed, looking for work	17	3%
Disabled, not able to work	17	3%
Student (Part time)	5	1%

5.2.2 Validating Survey Respondents using Google Analytics

To confirm that the survey response data was not influenced by those who hold NML memberships or consider themselves as friends of NML and would therefore complete all surveys in order to show support for NML and as such could skew the data, the response demographics data was compared to the demographics data captured by NML's Google Analytics (GA) data.

The GA data were captured from the same period as when the survey was active (1st February to 1st March 2017). The survey received a 50.4% response rate. This is comparable to the 52% bounce rate the website achieves. Whilst comparing the survey's response rate to the website's bounce rate (as identified by Google Analytics) is potentially a tenuous link, it does show that similar percentages are passing over and accepting.

Similarly, a comparison of the age groups illustrated that the survey received a good proportion of users from across the population (see Table 5.8):

Table 5.8: Survey response ages (Q16) compared to GA ages

Age Group	**Survey Percentage**	**GA Percentage**
18 - 34	18%	39%
35 - 54	36%	36%
55 - 64	27%	13%
65+	19%	12%
Total	100%	100%

as well as by Gender breakdown (see table 5.9):

Table 5.9: Survey response gender (Q17) compared to GA gender

Gender	**Survey Percentage**	**GA Percentage**
Female	60%	58%
Male	38%	42%
Rather not say	2%	N/A

GA reported a total number of website users as 174,457. 81% being new visitors and 19% being returning visitors.

Additionally, GA reports that the biggest majority of the audience accessing the website is from the UK (65%), with 17% from the US, 2% from Australia, 1% from Canada and 1% from Germany and all other locations being reported as being less than 1 %.

The Poll Daddy system (survey tool) automatically produced a world map of respondent locations (see Figure 5.1). *Note that the numbers in the world map image are not the same as in Table 5.10 due to it being from the Poll Daddy system, which captures all respondents and not just those in the cleaned dataset.*

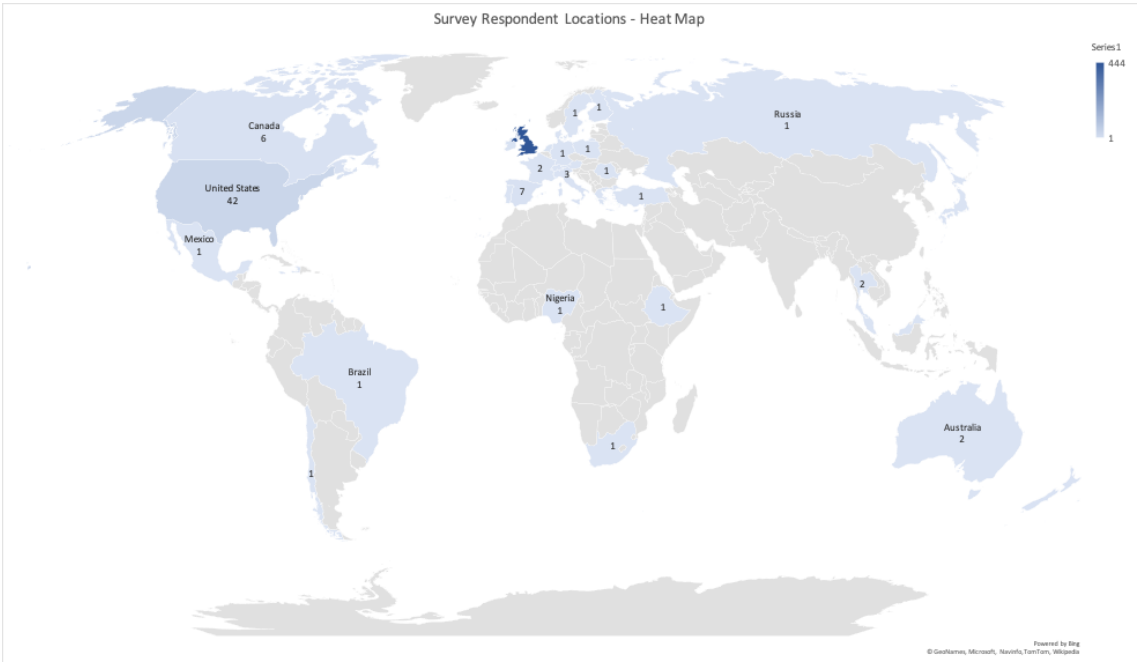


Figure 5.1: Q20 - World map of respondents (Poll Daddy Generated).

As well as the world map (see Figure 5.1), the Poll Daddy system also captures the IP address of the respondent's PC. Using the IP2Location free database, it is possible to identify which country the IP address is associated with. The results for all of the Poll Daddy captured data are shown in Table 5.10.

Table 5.10: Countries of respondents

Country	**Number**	**Percentage**
United Kingdom	430	84%
United States	39	8%
Spain	7	1%
Canada	6	1%
Netherlands	4	1%
New Zealand	2	0%
Thailand	2	0%
France	2	0%
Australia	2	0%
Italy	2	0%
Portugal	2	0%
Denmark	2	0%
Malaysia	1	0%
Greece	1	0%
Nigeria	1	0%
Japan	1	0%
Norway	1	0%
Germany	1	0%
Cyprus	1	0%
Ethiopia	1	0%
Austria	1	0%
Chile	1	0%
Sweden	1	0%
Trinidad and Tobago	1	0%
Poland	1	0%
Jamaica	1	0%

As is apparent from the above Poll Daddy data compared to the GA location data, there are some variation in results (for example: GA UK = 65% vs PD UK = 84%). There are two reasons for this variation, 1) the GA data captures the whole of the NML website (all sections) for the whole month period, but the survey only ran on part of the website for the first two weeks and another part for the following

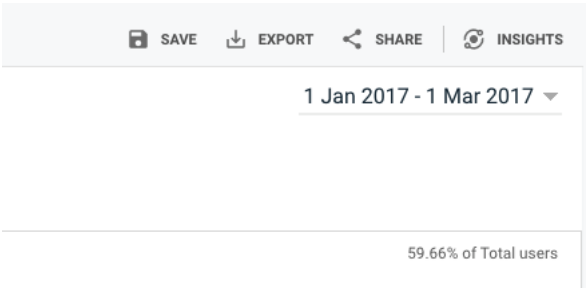


Figure 5.2: A close up of the GA estimate data figure

two weeks (as explained above). The second reason for the variation is due to GA not actually logging every user, but instead only capturing a percentage of the actual users that are trackable by being logged into a Google or partner account and also allowing the sharing of their location with these companies. Google Analytics does estimate where it can not calculate the accurate data, and it does show this percentage on the GA page of what the figure of users the estimate is based on. In this instance, locations were based on 59.66% of users (see Figure 5.2).

Whilst the results are not identical due to the reasons noted above, the results are close enough to show that the sample is a representative sample of users.

5.2.3 Motivation and Task

According to the literature (see Chapter 2 Literature Review) one of the significant determinants of user interaction was that of motivation. The survey captured both the driving force (motivation) behind the users' visit and also the task (reason) they wanted to do on the site during their visit.

5.2.3.1 Question 2 - Today I am visiting the NML website for?

The survey revealed that whilst all users want to find the information they want, they also want to do so as quickly as possible. It is thought that there are significant differences that users who seek for work or academic purposes expect in the range of tools (e.g., search techniques), layout of the site, language used on screen and speed of identifying the data wanted, compared to users who visit for personal reasons and possibly have a little more time, less knowledge or experience of the tools/site.

Table 5.11 lists the reasons provided by the participants. Of the 514 surveys, 70% cited personal reasons as the reason for interacting with the website. While

13% were there for no specific reason or merely passing the time; 9% were engaged for the purposes of study and 8% for work reasons. The results show that 70% of visitors come for personal reasons, but the majority of the literature studied shows people visiting for professional reasons.

Table 5.11: Q2 - Reason

Reason for visit	**Number**	**Percentage**
For personal reasons	362	70%
For no specific reason or to pass time	67	13%
For study (College, Further Education, University)	45	9%
For work	40	8%

5.2.3.2 Question 3 - What is the primary purpose of your visit to the NML website today?

Establishing the reason for which the interaction was being undertaken revealed that in preparation for a visit (physical) was the primary purpose for the engagement with the website; 48% followed by 19% for other purposes. Those who opted to choose the other purposes option and write their own reason e.g. 'family history', 'past exhibitions', 'to help recall previous visit', 'learn about a slave trade', 'preparing teaching materials' etc. clearly saw these as different even though they could actually be linked back to at least one of the names options. The remaining 33% was distributed between gaining knowledge of the collections that the museums exhibit (10%) gaining general knowledge of a collection (7%), specific museum artefact information (6%), keeping up to date with museum news (4%), gain knowledge of a specific type of object (4%), to those who did not know (1%) and to make an online purchase from the gift shop (1%) (see Table 5.12). These results again suggest that the main audience would not be professional based as the literature would have us believe. The people visiting for high-level knowledge, e.g. about collections or general knowledge of a collection, would also suggest that the majority of these knowledge seekers are non-professional and are not familiar with the museum and its holdings. This could be a potential issue for search feature usage.

Table 5.12: Q3 - Purpose

Purpose of visit	**Number**	**Percentage**
Prepare for a visit to the museum (Pre-visit)	249	48%
Other purposes	100	19%
Gain knowledge of the collections that the museums exhibit (Museum Overview)	50	10%
Gain general knowledge of a collection (our art collection, the Egyptian or horology collection) (collection overview)	34	7%
Gain knowledge of a specific museum object (a certain clock (The Barry astronomical clock) or statue (Huskisson Statue) for example) (known item)	30	6%
To keep up to date with the museums' news via the blogs	23	4%
Gain knowledge of a specific type of object (Egyptian burial objects) (known collection)	19	4%
I don't know	5	1%
To buy an item from the online gift shop	4	1%

The survey provided all the possible options as to why people were visiting NML website, with the biggest reason being to prepare for a visit to the physical museum (see Table 5.11) The second-biggest selection chosen by the respondents was the "other option" because the offered selections did not fit their perceived reason for their visit. Upon further investigation of the "other options" submitted, a number of them are specific versions of the options offered; e.g. to look at the Egyptian animal exhibition or to find out about the slave trade, but others were options not considered or identified from research during the devising of the question, e.g. recall previous visit, study family history, researching for a friend or relative. There was also one option of "just to admire" submitted, which could possibly be an equivalent of Falk's (2009) Recharger.

5.2.4 Shared Experience, Engagement and Usage

5.2.4.1 Question 5 - Are you sharing this website experience?

Establishing if the participants' engagement with the NML website was a shared experience which, for them, could make the interaction more meaningful revealed that while 89% or 459 of the participants viewed the website alone, 30 (6%) of respondents were looking with a child by their side and 9 (2%) with another person

remotely. A further 16 (3%) respondents chose the "other" option, albeit they were still performing the activity within close proximity of another individual (see Table 5.14) or basically to searching with an adult by their side. Which is one option to this question that should be included in future surveys using this question. Examples of these responses include: "*Looking with my neighbour who is 97 yrs old*" and "*...looking with my wifr*" and there is one response where they are searching with their cat "*With my cat, but she doesn't seem particularly interested.*" (see Table 5.13).

Table 5.13: Q5 - Sharing the Online Experience

Sharing options	**Number**	**Percentage**
No, I am on my own.	459	89%
I am looking with a child/children by my side	30	6%
I am looking with another person remotely	9	2%
Other	16	3%

Table 5.14: Sharing other options

Sharing Other Options	**Number**
Looking with my neighbour who is 97 yrs old	1
I am looking with my wifr, for another future visit to the museum	1
Looking with an adult by my side	1
no no no	1
Looking for us to vist exhibition	1
With my cat, but she doesn't seem particularly interested.	1
colleague	1
Looking good with friend by my side	1
Wife	1
Yes, I am with a class of FE students	1
depends on whether i find the information I'm looking for	1
My father and I are looking for information	1
I'll refer to it in my lecture	1
will share with friends on Facebook	1
Discussing with partner	1
i not sharing yet	1

These results are rather interesting as web use is usually considered to be a solitary task. For the majority of the sample, this does ring true. However, the survey indicates the contrary (11% of the sample), for example there is evidence of parents and children searching together, which means that visuals and clear text are important. There are potentially people here that are similar to Falk's (2009) 'Facilitator' who are helping others, e.g. helping elderly neighbours. But also lots of couples exploring the site together, potentially preparing for a physical visit. Another fascinating, albeit very small part of the sample are those that are exploring with remote people. Remote web experiences are not that greatly explored, but maybe there is space for this to be explored further. Also, there is clear evidence in the 'Other notes' of teachers/lecturers using the site live with a whole class, suggesting that Spellerberg et al.'s (2016) splitting of "Falk's 'Facilitator' motivation into two. The adult, such as a parent or grandparent, who is facilitating an experience for children, became 'Parent'; and the adult who is facilitating an experience for other adults became 'Socializer'." was partially correct but maybe the 'Parent' does not cover the whole scenario and is not entirely correct as there are other instances of facilitation happening with children but not by parents and not by teachers. Maybe something like 'educational guide' would be a better term. Or maybe this is a third category.

5.2.4.2 Question 6 - How frequently do you visit the NML website?

Most survey respondents were using the NML website for the first time 268 (52%) while 163 (32%) claim to use the website a couple of times a year. A further 60 (11%) said they use the website monthly, with 19 (4%) respondents using the site weekly and 4 (1%) stating that they use the site daily. (see Table 5.15).

Table 5.15: Q6 - Frequency of visits to NML website

Frequency	**Number**	**Percentage**
This is the first time	268	52%
A couple of times a year	163	32%
Monthly	60	11%
Weekly	19	4%
Daily	4	1%

Repeat visits followed a similar pattern to those identified in the Ciber (2013) report and in Hall et al. (2012b). The number of respondents who visited the NML website for the first time equal 52% (n = 298) with 31% (n = 179) revisiting a couple of times a year, 12% (n = 69) revisited monthly, 4% (n = 21) visited weekly and only 1% (n = 6) visited daily.

Comparing the frequency of visits to the data from Google Analytics (GA), revealed that GA figures indicated more first time visitors than the survey data suggested. GA claims 81% of visitors over the 4-week period the survey was active, were new visitors to the site, and that only 19% of visitors were returning visitors (see Figure 5.3). This presented a 29% disparity between the survey results and the GA data for new visitors, and returning visitors. While this seems like the survey could be capturing a slightly biased pool of respondents, it has to be noted that GA is far from 100% accurate. Rogina (2022) outlines a number of reasons for this:

1. A user access the site from a different device.
2. A user switches browser.
3. The user clears their cookies.
4. The user is privately browsing (using incognito mode or private browsing sets no cookie)

As the survey results are actual responses from people, it is fair to assume that these are more accurate in this case than Google Analytics data.

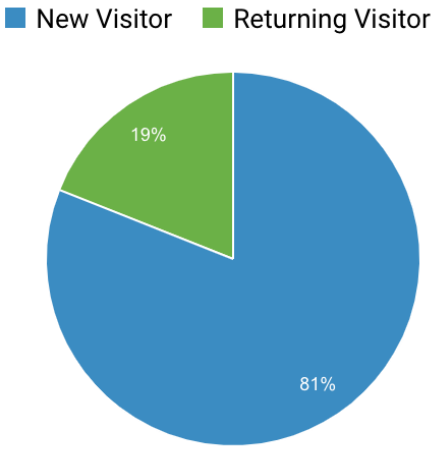


Figure 5.3: Q6 - Frequency of visits to NML website according to Google Analytics

5.2.4.3 Question 7 - Roughly how many web pages have you viewed on the NML website during this visit?

Establishing the longevity of interaction, behaviour of respondents indicated that a significant majority accessed between 1 and 10 pages during the visit, 245 (48%) visiting only one page and 243 (47%) visited between 2 and 10 pages. A further 20 (4%) visited between 11 and 20 pages, and 6 participants (1%) used 21 pages or more. (see Table 5.16).

Table 5.16: Q7 - Number of web pages accessed during this visit.

Number of pages accessed	Number	Percentage
2 - 10	243	47%
Just the one	245	48%
11 - 20	20	4%
21+	6	1%

The high percentages of low numbers of page views may be a result of the pop-up style delivery of the survey. The survey appeared to a user 10 seconds after being on a page. Therefore, if a user was particularly quick browsing through the web pages, it would only appear if they lingered. 10 seconds may appear to be too quick, but the survey needed to capture all users, including those that flit in and out of the site, without appearing demanding and asking immediately.

Alternatively, this could also be representative of the reported 66% of users that leave after one page within a short period of time. 6% identified that they had visited over 11 pages, and from those 2% (n = 9) identified that they had looked at over 21 pages during this single visit.

5.2.4.4 Question 8 - When on a web page about an object or collection, do you typically?

How the survey respondents reacted to a web page provided a number of typical techniques users perform; scanning, looking for relevant pictures, reading the entire page for options for selection. Scanning the page for relevant words was the primary technique used by the participants 240 (47%). While looking for pictures is the second 160 (31%) and finally reading the whole page 114 (22%) (see Table 5.17).

Table 5.17: Q8 - How people use content

Method of using web pages	**Number**	**Percentage**
Scan the page to find the relevant words you are looking for?	240	47%
Look for pictures first and then read the text	160	31%
Read the whole page from top to bottom in depth	114	22%

The results suggest that users tend to use different methods to consume content on web pages. It is interesting that 78% of participants scan the page for text or images, compared to the 22% who read sequentially and in-depth. This result suggests that users are coming with some ideal or maybe a goal in mind for what they are seeking. It also provides some limited insight into the types of users for example, users who prefer to scan web pages to find the relevant information they are looking for may be busy and looking for specific information quickly. They may not have a lot of time to spend reading entire web pages and may prefer to quickly scan for keywords or phrases. Given that the majority of the sample claimed to visit for personal reasons and not work or study; and the main reasons were for preparing for a visit or gaining knowledge of the collections on offer, it suggests that these leisure/general public users still want results quickly and whilst they may not know exactly what they are looking for they are aware of the sort of things they

are seeking or what they are not seeking. Users who prefer to look for pictures first and then read the text may be more visually-oriented and may be looking for a more engaging and visually appealing experience. Users who prefer to read the whole page from top to bottom in depth may be more thorough and analytical in their approach to consuming information, suggesting these could be more academic, professional or research users. This information may be useful for website designers and developers to optimize the layout and design of web pages and the content to meet the needs of different types of users.

5.2.4.5 Question 14 - What time of day do you usually undertake the type of activity you are primarily using the NML website for today?

The most popular NML website access time indicated by the survey participants was during the evening, with 196 (38%) participants indicating this as their only access time. There were 75 (15%) of participants indicating that they accessed the website during the morning, with a further 65 (13%) who accessed the website in the afternoon. A further 21 (4%) of respondents indicated a lunchtime access.

However, some participants indicated multiple access times and as such undertook these at various points during the day. Some 157 (30%) selected a combination of times, the most popular combination being all four times 52 or 10%) (see Table 5.18).

Table 5.18: Q14 - Preferred access times

Preferred access times	Number	Percentage
Evening	196	38%
Morning	75	15%
Afternoon	65	13%
Morning, Lunchtime, Afternoon, Evening	52	10%
Afternoon, Evening	29	6%
Morning, Evening	27	5%
Lunchtime	21	4%
Morning, Afternoon, Evening	16	3%
Morning, Afternoon	11	2%
Morning, Lunchtime, Afternoon	7	1%
Morning, Lunchtime	6	1%
Lunchtime, Evening	4	1%
Lunchtime, Afternoon, Evening	3	1%
Lunchtime, Afternoon	1	0%
Morning, Lunchtime, Evening	1	0%

The times of respondent interactivity provide some insight into the type of user that engages with the website. For example, it would not be expected to see much work activity in the evening, and transversely it may not be expected that much non-work activity occurs during the day-time. What is seen is that the majority of participants preferred to conduct the types of activities they were doing in the evening, with the morning and afternoon being joint second. However, the combined lunchtime interaction of only 12%, which may have been expected to have been a little higher with being a pleasure activity (see Table 5.18) This result links well with the responses to the visit reason and motivations, with users claiming to be preparing for a visit the night before or looking for somewhere to go that day because it may be rainy.

5.2.4.6 Question 15 - What type of device are you using for this visit to NML website?

The types of technological device the survey participants used to access the NML website revealed that most of the respondents, 290 (56%) were using a PC (Desktop or Laptop), 122 (24%) were using a tablet and 102 (20%) were using a mobile phone (see Table 5.19).

Table 5.19: Q15 - Type of device.

Device Type	Number	Percentage
Desktop or Laptop PC	290	56%
Tablet PC (iPad, Amazon fire, Android tablet etc)	122	24%
Mobile Phone	102	20%

This provides a further variable that may assist in identifying the type of user accessing the website from the device being used. It is not expected that a mobile will be used in a collection area of a website for a work task, due to the amount of information need on screen to complete the task.

5.2.5 Domain Knowledge

5.2.5.1 Question 9 - In the context of cultural heritage and your current visit to the NML website, please select the appropriate statement:

The purpose of this question was to establish if the participants understood the concept of cultural heritage and whether this played a part in their website interaction. As can be seen from the responses in Table 5.20 the majority of participants, 307 (60%) only possessed some knowledge and or experience with the cultural heritage, while 134 (26%) of participants classed themselves as novices. A further 51 (10%) of participants claimed to be highly experienced, although 22 participants (4%) indicated that they did not know what their current levels of knowledge and experience were. This result again suggests that the majority of users are not the professional users the literature describes.

Table 5.20: Q9 - Participants Domain Knowledge

Domain Knowledge	Number	Percentage
I have some experience and background knowledge	307	60%
I'm a novice with little knowledge within this area	134	26%
I'm highly experienced and have extensive background knowledge	51	10%
I don't know	22	4%

5.2.5.2 Question 10 - Rate your Cultural heritage knowledge (levels 1-5):

This question further explored the participant's self-perceived level of cultural heritage knowledge using a 5-point Likert scale, with Level 1 being the lowest (e.g. no CH knowledge) and Level 5 (extensive CH knowledge) the highest (see Chapter 3 Methodology).

The majority of participants, 209 (41%) chose level 3, which is the mid-way point in the given scale and is considered a neither agree nor disagree option in relation to what they know in relation to CH. Level 4 revealed 155 (30%) have determined that they are knowledgeable regarding CH, and Level 2 recorded 74 (14%) may have some CH knowledge and Level 5 recorded 63 (12%) participants with extensive CH knowledge. However, those who have self-determined as having no CH knowledge (Level 1) recorded 13 (3%) participants (see Table 5.21). The responses here suggest that most users consider themselves as having some level of general CH knowledge, indicating that there is a general interest in CH for nearly all users.

Table 5.21: Q10 - General CH Knowledge

CH Knowledge Rating	Number	Percentage
3	209	41%
4	155	30%
2	74	14%
5	63	12%
1	13	3%

5.2.6 Groups

5.2.6.1 Question 22 - Which of the following groups would you place yourself in for this visit to the NML website?

This question asked the participants to self determine their grouping category for the purpose of the immediate NML website interaction. Table 5.22 shows the self selected groups participants placed themselves into. 95.4% of the participants placed themselves into one or a combination of the provided options, and 4.6% chose to create their own using the 'other' option. There were 41 combinations of the provided groups created by the participants (see Table 5.22).

Table 5.22: Q22 - Self Selected Grouping

Self Selected Group	**Number**	**Percentage %**
GP : General public/user	253	44.9
NP : Non-professional researcher (hobbyist, amateur historian)	89	15.8
NP/GP: Non-professional researcher (hobbyist, amateur historian), General public/user	48	8.5
S: Student (college, university, further education)	33	5.9
T: Teacher	18	3.2
A: Academic (scholar, post doc researcher, academic support)	16	2.8
M: Museum staff (curator, archivist)	10	1.8
A/T: Academic (scholar, post doc researcher, academic support), Teacher	9	1.6
NP/T/GP: Non-professional researcher (hobbyist, amateur historian), Teacher, General public/user	7	1.2
P: Professional researcher (historian, genealogist)	5	0.9
S/GP: Student (college, university, further education), General public/user	5	0.9
NP/S/GP: Non-professional researcher (hobbyist, amateur historian), General public/user, Student (college, university, further education), General public/user	3	0.5
A/S/GP: Academic (scholar, post doc researcher, academic support), Student (college, university, further education), General public/user	3	0.5
NP/A/GP: Non-professional researcher (hobbyist, amateur historian), Academic (scholar, post doc researcher, academic support), General public/user	2	0.4
NP/A/S/GP: Non-professional researcher (hobbyist, amateur historian), Academic (scholar, post doc researcher, academic support), Student (college, university, further education), General public/user	2	0.4
T/GP: Teacher, General public/user	2	0.4
A/S: Academic (scholar, post doc researcher, academic support), Student (college, university, further education)	2	0.4
A/T/S: Academic (scholar, post doc researcher, academic support), Teacher, Student (college, university, further education)	2	0.4
NP/A/S: Non-professional researcher (hobbyist, amateur historian), Academic (scholar, post doc researcher, academic support), Student (college, university, further education)	2	0.4
NP/S: Non-professional researcher (hobbyist, amateur historian), Student (college, university, further education)	2	0.4
P/M/GP: Professional researcher (historian, genealogist), Museum staff (curator, archivist), General public/user	2	0.4
P/A: Professional researcher (historian, genealogist), Academic (scholar, post doc researcher, academic support)	2	0.4
P/A/T: Professional researcher (historian, genealogist), Academic (scholar, post doc researcher, academic support), Teacher	2	0.4

Table 5.22 continued from previous page

Self Selected Group	Number	Percentage %
NP/P: Non-professional researcher (hobbyist, amateur historian), Professional researcher (historian, genealogist)	2	0.4
P/M: Professional researcher (historian, genealogist), Museum staff (curator, archivist)	2	0.4
P/T/S: Professional researcher (historian, genealogist), Teacher, Student (college, university, further education)	1	0.2
T/M: Teacher, Museum staff (curator, archivist)	1	0.2
P/A/T/S/M: Professional researcher (historian, genealogist), Academic (scholar, post doc researcher, academic support), Teacher, Student (college, university, further education), Museum staff (curator, archivist)	1	0.2
A/T/S/M: Academic (scholar, post doc researcher, academic support), Teacher, Student (college, university, further education), Museum staff (curator, archivist)	1	0.2
S/M: Student (college, university, further education), Museum staff (curator, archivist)	1	0.2
P/A/S/M: Professional researcher (historian, genealogist), Academic (scholar, post doc researcher, academic support), Student (college, university, further education), Museum staff (curator, archivist)	1	0.2
T/S/GP: Teacher, Student (college, university, further education), General public/user	1	0.2
A/GP: Academic (scholar, post doc researcher, academic support), General public/user	1	0.2
NP/P/A/S/M/GP: Non-professional researcher (hobbyist, amateur historian), Professional researcher (historian, genealogist), Academic (scholar, post doc researcher, academic support), Student (college, university, further education), Museum staff (curator, archivist), General public/user	1	0.2
P/T/M/GP: Professional researcher (historian, genealogist), Teacher, Museum staff (curator, archivist), General public/user	1	0.2
A/M: Academic (scholar, post doc researcher, academic support), Museum staff (curator, archivist)	1	0.2
M/GP: Museum staff (curator, archivist), General public/user	1	0.2
A/S/M: Academic (scholar, post doc researcher, academic support), Student (college, university, further education), Museum staff (curator, archivist)	1	0.2
T/S: Teacher, Student (college, university, further education)	1	0.2
NP/A: Non-professional researcher (hobbyist, amateur historian), Academic (scholar, post doc researcher, academic support)	1	0.2
O: Other	26	4.6
Total	564	100

5.2.6.2 Q22 - Other groups

There were 26 (4.6%) of participants that chose to create their own group/category of user, as they did not consider themselves to be part of one of those provided. These other groups can be seen in Table 5.23.

Table 5.23: Q22 - Other groups.

Other Group	**Number**	**Percentage of the 26 other options**
Parent	3	12%
artist	2	8%
Writer	2	8%
gallery/museum trustee	1	4%
Leader of a Guide group planning a visit	1	4%
Newsletter editor	1	4%
Previously lived in Liverpool	1	4%
I am a researcher but this visit relates to my personal family research	1	4%
journalist checking a story for factual accuracy	1	4%
Tourist	1	4%
Leisure visitor	1	4%
researcher / economist	1	4%
Educator in museums	1	4%
finding exhibition details for U3A group	1	4%
Good Lord!	1	4%
ex museum employee both in UK and internationally	1	4%
As a mother keen to teach her daughter more about the history of slavery	1	4%
Youth leader	1	4%
Librarian	1	4%
freelance music producer and curator of site specific installations	1	4%
Artist	1	4%
Retired	1	4%

However, while a tenuous link could be made to one or two groups identified in Phase 1 (see Figure 4.5) e.g., Tourist, researcher/economist, gallery/museum trustee, it was impossible to establish any meaningful determinable characteristics behind the label the participant created (see Phase 1). Therefore, as these numbers are too small and cannot be cleanly associated with the user group labels found in Phase 1, they were removed and not used in the analysis from this point onward.

5.2.7 Preferred Content type and Technical Expertise

5.2.7.1 Question 4 - Please rank the following content types in order of your preferred preferences

The visual content published on the NML website is intended to intrigue and attract visitors into wanting to know more and therefore engage with more pages/content. Content, in terms of the website, means any creative element, for example, text, applications, images, data, e-services, audio and video files, and so on. The survey listed A-typical content choices which are already present on the NML website of; Text; Pictures; Videos; Audio (speech or music); Interactive / hands on content (quizzes, virtual tours) and asked the participants to put them into their own personal preferred order which could culminate in some 720 possible permutations.

Upon examining the data, the participants generated 123 different combinations of the six content types (see Appendix H Table H.1). This was ranked using the number of participants per combination from high to low. It was decided that only those combinations that had 10 or more participants aligned against it would be explored further, in order to ensure the production of meaningful findings; resulting in 12 viable permutations (see Table 5.24). The top combination with 46 respondents (8%) was Text, Pictures, Videos, Audio (speech or music), Interactive / hands on content (quizzes, virtual tours).

Table 5.24: Q4 - Content (Top 12 Permutations)

Content Preference Permutation	**Number**	**Percentage**
Text, Pictures, Videos, Audio (speech or music), Interactive / hands on content (Quizzes, virtual tours)	46	8%
Pictures, Text, Videos, Audio (speech or music), Interactive / hands on content (Quizzes, virtual tours)	45	8%
Pictures, Text, Videos, Interactive / hands on content (Quizzes, virtual tours), Audio (speech or music)	35	6%
Pictures, Text, Interactive / hands on content (Quizzes, virtual tours), Videos, Audio (speech or music)	29	5%
Text, Pictures, Videos, Interactive / hands on content (Quizzes, virtual tours), Audio (speech or music)	26	5%
Text, Pictures, Audio (speech or music), Videos, Interactive / hands on content (Quizzes, virtual tours)	20	4%
Pictures, Text, Audio (speech or music), Videos, Interactive / hands on content (Quizzes, virtual tours)	18	3%

Text, Pictures, Interactive / hands on content (Quizzes, virtual tours), Videos, Audio (speech or music)	17	3%
Pictures, Text, Interactive / hands on content (Quizzes, virtual tours), Audio (speech or music), Videos	15	3%
Pictures, Text, Audio (speech or music), Interactive / hands on content (Quizzes, virtual tours), Videos	14	2%
Pictures, Videos, Text, Audio (speech or music), Interactive / hands on content (Quizzes, virtual tours)	13	2%
Pictures, Videos, Interactive / hands on content (Quizzes, virtual tours), Text, Audio (speech or music)	10	2%
hline		

Table 5.25 shows how each of the content types was ranked. What is evident is that pictures have been identified as the most important content type format followed closely by text, whilst interactive and audio were placed last the most.

Table 5.25: Q4 - Ranked Content Preferences

	First	**Second**	**Third**	**Forth**	**Fifth**
Audio	32 (6%)	34 (7%)	112 (22%)	168 (33%)	163 (32%)
Interactive	55 (11%)	51 (10%)	108 (21%)	118 (23%)	177 (35%)
Pictures	227 (45%)	162 (32%)	59 (12%)	37 (7%)	25 (5%)
Text	163 (32%)	177 (35%)	54 (11%)	61 (12%)	56 (11%)
Video	37 (7%)	90 (18%)	181 (36%)	130 (26%)	68 (13%)

Knowing that the visuals are highly important to the audience shows that they may be more visually oriented experiences. By using the visuals to capture their attention and interest, at which point they then use the text to learn.

5.2.7.2 Question 11 - When seeking information on a website, which method do you prefer?

The publications found through the literature review (Chapter 2) and Phase 1 (Chapter 4) established that the majority of publications focused upon the use of Search systems. This question focused on how the participant would seek out information on the NML website and presented the options that are available, e.g. a Search box or content links and navigation. When asked if they preferred to use the search box or browse by the navigation buttons and links in content, 322 participants (63%) chose to browse while 192 (37%) chose to use the search box (see Table 5.26).

Table 5.26: Q11 - Search or Browse Preference

Search or Browse preference	Number	Percentage
Using content links or navigational buttons	322	63%
The search box	192	37%

knowing that the majority of users prefer browsing to searching links with the previous questions result in that they prefer the images before text. This indicates that the users are potentially seeking quickly and are also looking for more engaging, user-friendly and intuitive browsing experiences that allow them to discover and engage with content in a more natural way. For example, if users prefer browsing and navigating, they may be more likely to engage with pictures, videos, and interactive content that catches their attention and provides an engaging browsing experience. Understanding the preferences of users for different types of content can help website designers and developers to create a website that is more user-friendly and meets the needs and preferences of a wide range of users.

The data also suggests that a decent browse system would outperform a good search system in terms of satisfying users. Whilst there has been a stream of work looking into this field of work (Rich Prospect browsing (Ruecker et al., 2011; Giacometti, 2009; Morse et al., 2021) and generous interfaces (Whitelaw, 2015)), this area is still hugely understudied, and none of the research systems proposed has yet been adopted by museums suggesting that either museums are not understanding their users and/or that there still needs to be significant work undertaken in this area (Hall and Walsh, 2021).

Interacting with the site/collection also needs to be understood in order to deliver an environment that is suitable for the audience. Navigational browsing is not mentioned that often in the literature, but the findings in this survey show that they are the preferred method of interacting within a site, over search. The survey respondents identified that they preferred to use navigational links or buttons to seek information (63%) as opposed to using search boxes. This could be linked to their lack of knowledge about what is held in the museum and so do not know what to search for. Again, this backs the idea that the main users are not experts.

5.2.7.3 Question 12 - When using the search box to search for content that you do not find in the first set of results, do you?

Expanding on the previous question, this question focused on the reaction of the participant if they did not find what they required at the first attempt of undertaking a search. Most of the participants, 288 (56%) would search a further 2 or 3 times if the initial search results showed nothing of interest/relevance. 182 (35%) participants would continue searching until they found the material/answer they were looking for, and 44 (9%) claimed they would give up after the first search attempt (see Table 5.27).

Table 5.27: Q12 - Search Persistence

Search Persistence	**Number**	**Percentage**
Search a further 2 or 3 times	288	56%
Persist until you find the material/answer you were looking for	182	35%
Give up	44	9%

Given that the majority of users claim to prefer browsing to searching, it is interesting that when these users do use the search features; and do not find the results they perceive as relevant the first time, they do continue searching as opposed to giving up. This suggests that museums can't simply focus on one mechanism but need to improve both simultaneously.

5.2.7.4 Question 13 - Please rate your proficiency with the internet tools and services (1(low) - 5(high))

This question focused on the participant's self-perceived proficiency when interacting with technology, particularly with sites such as Facebook, Twitter, Instagram, LinkedIn, Internet banking, Online Shopping, streaming music services such as Spotify, Streaming Video (YouTube, BBC iPlayer), document creation and reading (Google Docs). These questions were presented as Likert scales, with 1 being low and 5 being high.

Data revealed that the majority of participants, 185 (36%) were experienced using Facebook, but significantly 145 (28%) claimed they were very weak at using Facebook (see Table 5.28). Of the participants, 294 (57%) indicated a lack of proficiency with Twitter, with only 58 (11%) claiming to have a high level of expertise.

Instagram proved less popular among the participants, with 363 (71%) choosing Level 1. Only 57 (11%) claimed to have high proficiency in using Instagram.

Of the participants, 298 (58%), mostly did not feel confident with LinkedIn choosing Level 1, while only 34 (7%) claimed to have high proficiency (choosing Level 5). Online banking was something that 225 (44%) of the participants felt proficient at by choosing Level 5. 111 (22%) chose Level 4, 60 (12%) chose Level 3; 26 (5%) chose Level 2 but 92 (18%) chose Level 1.

Online shopping was also something that 232 (45%) of the participants felt proficient at by choosing Level 5. 121 (24%) chose Level 4. 94 (18%) chose Level 3. 34 (7%) chose Level 2 and 33 (6%) chose Level 1.

Online music was something that 221 (43%) of the participants did not feel comfortable with in choosing Level 1. 97 (19%) were at the other end of the scale and were very confident using music online. 78 (15%) chose Level 3, 69 (13%) choosing Level 2 and 49 (10%) choose Level 4.

Streaming video was something that over half of the participants felt comfortable using 319 (62%) by choosing Level 5 or Level 4. A further 155 (23%) chose level 1 or 2, while 77(15%) participants chose Level 3. Online Document writing using tools such as Google Docs or Google Sheets etc showed that there was a quite consistent spread right across the scope, but with slightly more participants feeling proficient (247 (48%) choosing level 4 or 5) at using these types of tools.

This question was an attempt to gain a singular level of technical web competency. By ascertaining the users' perceived proficiency levels from all of these everyday online tools, it is hoped to gain a sense of how their web skills stood.

Table 5.28 shows all results for Q13.

Table 5.28: Q13 - General Internet Proficiency - All Results

Proficiency	**Facebook**	**Twitter**	**Instagram**	**LinkedIn**	**Banking**	**Shopping**	**Music**	**Video**	**Docs**
1	145 (28%)	**294 (57%)**	**363 (71%)**	**298 (58%)**	92 (18%)	33 (6%)	**221 (43%)**	78 (15%)	115 (22%)
2	29 (6%)	53 (10%)	29 (6%)	52 (10%)	26 (5%)	34 (7%)	69 (13%)	40 (2%)	40 (8%)
3	53 (10%)	62 (12%)	39 (8%)	75 (15%)	60 (12%)	94 (18%)	78 (15%)	77 (35%)	112 (22%)
4	102 (20%)	47 (9%)	26 (5%)	55 (11%)	111 (22%)	121 (24%)	49 (10%)	128 (45%)	89 (17%)
5	**185 (36%)**	58 (11%)	57 (11%)	34 (7%)	**225 (44%)**	**232 (45%)**	97 (19%)	**191 (57%)**	**158 (31%)**

From the results in Table 5.28 it is possible to calculate a weighted average

proficiency level for each type of tool.

By assigning weights to each proficiency level (e.g., 1 for level 1, 2 for level 2, etc.) and then calculating the weighted average proficiency level across the tool. For example, the weighted average proficiency level for facebook would be: (145 x 1 + 29 x 2 + 53 x 3 + 102 x 4 + 185 x 5) / (145 + 29 + 53 + 102 + 185) = 3.30

Table 5.29 shows the full set of Weighted average proficiency calculations.

Table 5.29: Q13 - Weighted average proficiency level for web tools

Web System	Weighted average proficiency level
Facebook	3.30
Twitter	2.07
Instagram	1.80
LinkedIn	2.00
Banking	3.68
Shopping	3.94
Music	2.48
Video	3.61
Docs	3.26

The results suggest that, on average, respondents have the highest proficiency levels for shopping and banking, with average scores of 3.94 and 3.68, respectively. Music and Twitter appear to be the tools for which respondents have the lowest proficiency levels, with average scores of 2.48 and 2.07, respectively.

Overall, the results suggest that respondents have a range of proficiency levels across different web tools, and that there is variation in the levels of proficiency depending on the specific tool. This information could be useful for designing web interfaces and content that is more accessible and usable for users with different levels of proficiency.

Trying to make sense of this data with respect to Russell-Rose and Tate's (2012) model is far more difficult than expected when this question was devised. Calculating an overall average score from all the weighted average scores is 2.91. Then applying the same method for question 9, which was the domain knowledge question, produces an average score of 1.59. Plotting all NML users onto the Russell-Rose and Tate (2012) rose plot sees then sit within the Expert Technical Expertise and Novice Domain knowledge quadrant (see Figure 5.4).

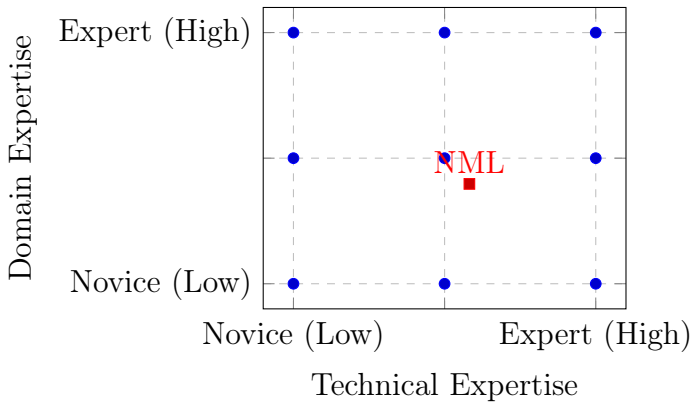


Figure 5.4: NML users mapped to Russell-Rose & Tate's dimensions of experience grid Russell-Rose and Tate (2012)

5.2.8 Other Data Gathered by the Poll Daddy System

As well as capturing the survey responses, the Poll Daddy system also captured data automatically from the request being made. This data is analysed separately from the survey responses as it is only additional or validating data and not part of the original question set.

5.2.8.1 Device Type Participant Using

The survey participants were asked in Q15 to reveal the type of device they were using at that point in time. The Poll Daddy System automatically logged the device that was being used to access the survey (see Table 5.30)

Table 5.30: Participant Device Type

Device Type	Number	Percentage
Browser, Slider	366	71%
Mobile Device, Direct	148	29%

Comparing this data to Q15 asked directly to the respondent shows some weaknesses in automatically collecting this type of data as the Poll Daddy system was only able to collect two different types of devices, Browser, Slider (or Desktop or Laptop PC) and mobile Device, direct. The survey question allowed for a deeper exploration, looking into if the mobile device was a Tablet or mobile phone. This difference could be the reason for the disparity between the two sets of results, with

Desktop or Laptop PC use being 71% from the Poll Daddy system's automated results compared to the 56% of users responses in the survey. Even if the Desktop and Laptop PC results are joined (totalling 80%) with the Tablet PC, there is still a small difference of 9%. Whilst the mobile results are closer with 20% of respondents choosing mobile phone and the Poll Daddy system claiming 29% there is still a 9% difference. However, because of limitations in some device fingerprinting code, such as: Dynamic IP addresses which can be shared among multiple devices; Browser settings: users can change settings in order to hide themselves, or they may use different browsers or devices at different times; Network configurations: using proxies, VPNs, and firewalls and Code updates: Device fingerprinting code is often updated over time to improve its accuracy and to account for new devices and technologies; All of these can lead to different fingerprinting results and so 9% is explainable.

5.2.8.2 Referral Link Data

The following tables show a summary of the museum/gallery and the type of page in the NML website the participant was on when they decided to complete the survey. This data has been captured and summarised from the refer link captured (see Appendix H Table H.3). As the survey was only hosted on the NML website, all the links are from the *http://www.liverpoolmuseums.org.uk/* domain. To keep the table readable, this high-level domain part has been omitted from each of the Uri's.

5.2.8.2.1 Page Type Entered the Survey From

The data in Table 5.31 was calculated from the referral Uri. An example URL is "*http://www.liverpoolmuseums.org.uk/walker/exhibitions/victorian-treasures/index.aspx*". Specific elements of the URL reveal where the participant is looking on the website. For example the "*http://www.liverpoolmuseums.org.uk/walker*" indicates that the participant is looking at content relating to the Walker Gallery. The next word indicates the type of page the user was on, and in this example "*exhibitions*" indicates an exhibition page. The next part of the URL reveals the type of exhibition, which turns out to be the "*victorian-treasures*" exhibition. The last part of the URL tells us which exact page in this exhibition, in this example, is the default home page

("*index.aspx*").

Within the NML website, there are certain page types: The landing page, which is a high level page in the site, such as the main home page. Exhibition pages which have just been discussed. Pages that show what events the museum is going to host. Visit type pages which inform museum website visitors of the facilities and opening times and other information to prepare them for a physical visit. Collection pages where the visitor can freely explore the collections artefacts online. There are also a few curated collections which do not allow for free exploration but instead provide a story and insight into the selected artefacts, with all information being created by the curators. Other page types include competition pages, talks, education, and a contact page. There is also a *kids* section, but because of the age requirement for this survey, no referral links from this part of the site are present.

Table 5.31: Entry Page Type

Page Type	**Number**	**Percentage**
Landing	125	24%
Exhibition	109	21%
Events	93	18%
Visit	88	17%
Collection	51	10%
Curated	23	4%
Archive	9	2%
Competition	6	1%
Talks	5	1%
Contact	3	1%
Education 2		0%

This result shows that the majority of users are entering on high-level, heavily crafted pages as opposed to deeper-level collection or item pages that are not crafted but instead where the content is drawn from the collection database. This also links to previous findings of users not being aware of what is held within the collections or is on show in the exhibitions, and also backs the high result of first-time visitors.

5.2.8.2.2 Museum Website Entered the Survey From

The data in Table 5.32 was calculated from the referral URL. An example URL is "*http://www.liverpoolmuseums.org.uk/walker/exhibitions/victorian-treasures/index.aspx*" the first segment after the top level domain of "*http://www.liverpoolmuseums.org.uk/*"

indicates which museum or gallery the user was on. In this example, it is "*walker*" for the Walker art gallery.

Table 5.32: Entry Museum Website

Museum/Gallery	**Number**	**Percentage**
Walker Art Gallery (walker)	179	35%
World Museum Liverpool (wml)	107	21%
Maritime Museum (maritime)	101	20%
International Slavery Museum (ism)	60	12%
Lady Lever Art gallery (ladylever)	44	9%
Sudley House (sudley)	22	4%

The Walker Art Gallery and the World Museum are the most well-known, larger and most physically visited of the NML museums/galleries and so seeing that these attracted more entries is not a shock considering the previous results have shown that a lot of users were visiting the website to plan for a physical visit.

5.3 Phase 2 : Survey Results Analysis

The aim of this work was to develop detailed characteristics for the user groups identified from the literature. To achieve that, the results were grouped using the self-categorisation data (Q22) and then tested to see which aspects and responses described and distinguished the groups.

5.3.1 Determining the user Group labels

Participants in the poll were asked to self-classify into one or more of seven groups or to submit a free-text "other" response (see Table 5.22 and Table 5.23). In total, there were 39 different combinations of user group selections.

Table 5.33: Most frequently selected user groups, before applying the rules merging the multi-selection responses (**pre-merging**) and after (**merged**). The "Other" group has been omitted and has not been subjected to further analysis.

Group	**Respondents (n)**	**Merged (n)**
GP: General Public	253 (49.2%)	253 (49.2%)
NP: Non- Professional (hobbyist, amateur historian)	89 (17.3%)	137 (26.7%)
NP/GP: Non-Professional and General Public	48 (9.3%)	-
S: Student	33 (6.4%)	33 (6.4%)
T: Teacher	18 (3.5%)	25 (4.9%)
A: Academic	16 (2.8%)	25 (4.9%)
MS: Museum Staff (curator, archivist)	10 (1.9%)	10 (1.9%)
A/T: Academic/Teacher	9 (1.8%)	-
NP/T/GP: Non-professional/Teacher/General Public	7 (1.4%)	-
P: Professional researcher	5 (1.0%)	5 (1.0%)
O: Other	26 (5.1%)	26 (5.1%)
Total Respondents	514 (100%)	514 (100%)

Before investigating the groups in more detail, it's crucial to address the approach taken to ensure the statistical integrity of the findings due to multiple comparisons. Given the multiple comparisons conducted in this analysis, the Bonferroni correction has been applied to adjust the significance levels of our p-values to mitigate the risk of Type I errors (i.e., falsely declaring a result significant due to chance). This conservative statistical method addresses the problem of multiple comparisons by dividing the desired significance level (typically $\alpha = 0.05$) by the number of tests performed, thereby maintaining the integrity of our findings by ensuring that the

overall chance of observing at least one false positive is kept under the desired significance level (Jafari and Ansari-Pour, 2019).

In practical terms, for each set of comparisons made within this analysis, the p-value required for significance has been adjusted according to the number of tests conducted. For example, if five comparisons are made, a single comparison would need to achieve a p-value less than 0.01 (0.05/5) to be considered significant after the Bonferroni correction. This method has been consistently applied throughout the analysis whenever multiple χ^2 tests or other statistical tests are performed on grouped data to ensure that the reported significance levels accurately reflect the heightened criteria for statistical evidence in the presence of multiple comparisons.

Following this approach, the multi-selection groups' responses were compared to their constituent groups using χ^2 tests. For example, the responses from the "non-professional/general public" group were individually compared to those from both the "non-professional" and "general public" groups. The rationale behind this comparison was to assess whether the multi-selection groups' responses could be attributed to ambiguity in group boundaries, hypothesising that their responses would align closely with one of the constituent groups while differing from the other(s).

For the 'non-professional/general public' group, significant differences were observed compared to the 'general public' group, even after applying the Bonferroni correction to adjust for multiple comparisons ($p < 0.05$ for all non-demographic questions), whereas no significant differences were found compared to the 'non-professional' group. Similarly, the 'academic/teacher' group did not differ significantly from the 'academic' group but showed significant differences from the 'teacher' group after adjustment ($p < 0.05$). The 'non-professional/teacher/general public' group exhibited no significant differences compared to the 'teacher' group but was significantly different from both the 'non-professional' and 'general public' groups after applying the Bonferroni correction ($p < 0.05$).

Consequently, the three multi-select groups were merged with the constituent group to which they showed no significant differences, streamlining the analysis to focus on these consolidated groups as illustrated in Table 5.33.

The multi-selection groups with less than 6 participants have not been processed in this manner, due to the group size being too small for statistical validity. Similarly, the "other" group has not been analysed in more detail and is not taken into

account for the further analysis. The remaining analysis is thus conducted on the groups "General Public" (GP), "Non-professional" (NP), "Student" (S), "Teacher" (T), "Academic" (A), "Museum Staff" (S), and "Professional" (P) with a participant sample size of 488 (see Table 5.34).

Table 5.34: Merged groups

Group	Respondents (n)	Percentages (%) based on 488 merged groups
General Public	253	51.8%
Non-Professionals	137	28.1%
Students	33	6.8%
Teachers	25	5.1%
Academics	25	5.1%
Museum Staff	10	2.0%
Professionals	5	1.0%
Total Respondents	488	100%

This covers 88% of respondents, with all other groups being multi-selected groups each with less than 1% and so these have been omitted (see the complete list of participant group selections in Appendix H, Table H.2).

In contrast to the findings in current literature (see Chapters 2 and 4) in that the Professional, and Museum Staff are the dominant user group label that are the focus of many studies and subsequently could be presumed to be the main audiences of the museum sites; Table 5.22 indicates that these are in-fact the smallest audiences and clearly demonstrates the issue of the majority of users (general public, non-professional researcher) being an under-represented group in the literature.

The Professional and Museum staff groups are possibly highly studied due to the ease of access to the participants for researchers. The General public, non-professional groups may not be as easy or willing to partake in lab studies and so are under-represented in the literature. The results from this survey shows that there needs to be more studies of these highly important sets of users.

5.3.2 Demographics: Location, Gender, Age and Education by Group

Whilst the demographics earlier allowed for a picture to be drawn of the survey respondents, here they are being discussed as potential defining characteristics of the groups as opposed to just demographic data. Starting with a cross tabulation analysis of the self selected user groups against the demographic questions allows for the discovery of any patterns that emerge.

5.3.2.1 Question 16 - How old are you?

The distribution of participants among self-assigned groups across different age bands highlights varying demographic trends within the study (see Table 5.35). In total, 488 participants self categorised into seven distinct groups, each exhibiting unique age-related characteristics.

The General Public was the largest group, comprising 253 individuals. It was fairly evenly distributed across age bands, with a noticeable emphasis on the 35-54 (n = 94, 37%) and 55-64 (n = 84, 33%) age brackets, indicating a middle-aged skew. Notably, there were also 40 participants (16%) aged 65-74 and 9 participants (4%) over 75 years old within this group, pointing to a significant representation of older generations.

The Non-Professionals, consisting of 137 participants, also displayed a broad age range, with the highest numbers in the 35-54 (n = 42, 31%) and 55-64 (n = 37, 27%) brackets. This group, too, had a notable presence of older individuals, with 30 participants (22%) in the 65-74 age band and 4 (3%) over 75.

Students and Museum Staff showed a pronounced youthful demographic, with 70% of both groups falling within the 18-34 age band (n = 23 for Students and n = 7 for Museum Staff). This starkly contrasts with the Professionals and Teachers, where the majority were in the working age of 35-54, with 60% (n = 3) of Professionals and 68% (n = 17) of Teachers occupying this bracket.

Academics (n = 25) presented a more even distribution across the age bands, with a modest presence in the older age categories, including 20% (n = 5) in the 65-74 range and 8% (n = 2) over 75, which is noteworthy for its inclusion of senior individuals.

In contrast, the Professional group, albeit small (n = 5), had three-quarters of its members in the 35-54 age band and one-fifth in the 65-74 age band, suggesting a lesser representation of the youngest and oldest age groups.

A Chi-squared test for independence applied to these data indicated that there is something significant ($p < 0.001, \chi^2 = 113.75, df = 24$). This confirms a significant association between participants' ages and the groups they self-assigned to, indicating that the distribution is non-random. Such diversity in age distribution across groups, validated by the Chi-squared test, could have implications for group dynamics, preferences, and responses within the study context. These findings underscore the importance of considering demographic factors when analysing group-related behaviours and preferences.

Table 5.35: Ages by user groups

Age	**18 - 34**		**35 - 54**		**55 - 64**		**65 - 74**		**75+**		**Group Total**	
Groups	n	%	n	%	n	%	n	%	n	%	**n**	**%**
Academics	4	16	7	28	7	28	5	20	2	8	**25**	**100**
General Public	26	10	94	37	84	33	40	16	9	4	**253**	**100**
Museum Staff	7	70	2	20	1	10	0	0	0	0	**10**	**100**
Non-Professionals	24	18	42	31	37	27	30	22	4	3	**137**	**100**
Professionals	1	20	3	60	0	0	1	20	0	0	**5**	**100**
Students	23	70	7	21	1	3	2	6	0	0	**33**	**100**
Teachers	3	12	17	68	5	20	0	0	0	0	**25**	**100**
Age Category Total (n)	**88**		**172**		**135**		**78**		**15**		**488**	

5.3.2.2 Question 17 - What gender best describes you?

The data showed (see Table 5.36) that of the 488 surveys, 294 (60 %) of the participants indicated that they were female; 185 (38 %) participants indicated that they were male; and the remaining 9(2 %) preferred not to answer the question. The gender by group (see Table 5.37) predominantly indicates that whilst there is a slight skew in the dominance of the female participants in the sample as a whole, the split of gender is pretty equal across the groups, as confirmed by a chi^2 test (p-value of $p = 0.344$ $\chi^2 = 13.34, df = 12$).

Table 5.36: Q17 - Gender by user groups

Gender	**Female**		**Male**		**Rather not say**		**Group Total**	
Group	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
Academic	16	64	8	32	1	4	**25**	**100**
General Public	154	61	95	38	4	2	**253**	**100**
Museum Staff	8	80	2	20	0	0	**10**	**100**
Non-Professional	71	52	63	46	3	2	**137**	**100**
Professional	3	60	2	40	0	0	**5**	**100**
Student	21	64	11	33	1	3	**33**	**100**
Teacher	21	84	4	16	0	0	**25**	**100**
Gender Total (n)	**294**		**185**		**9**		**488**	

The data revealed a gender disparity among the 488 respondents, with 294 (60%) identifying as female, 185 (38%) as male, and the remaining 9 (2%) preferring not to specify their gender (see Table 5.36). Despite the overall dominance of female participants in the sample, the gender split within most groups was relatively proportional.

Table 5.37: Age and gender by user group

Gender	**Female**							**Male**							**Rather not say**				**Age Group Total**
Group	**A**	**GP**	**MS**	**NP**	**P**	**S**	T	**A**	**GP**	**MS**	**NP**	**P**	**S**	**T**	**A**	**GP**	**NP**	**S**	
Age																			
18 - 34	4	21	6	17	1	15	3	0	5	1	6	0	7	0	0	0	1	1	88
35 - 54	3	62	2	27	2	4	13	4	31	0	14	1	3	4	0	1	1	0	172
55 - 64	7	43	0	14	0	0	5	0	39	1	22	0	1	0	0	2	1	0	135
65 - 74	2	25	0	13	0	2	0	3	15	0	17	1	0	0	0	0	0	0	78
75+	0	3	0	0	0	0	0	1	5	0	4	0	0	0	1	1	0	0	15
Gender Group Total	16	154	8	71	3	21	21	8	95	2	63	2	11	4	1	4	3	1	488

In a detailed analysis of gender distribution by age and user group, no statistically significant differences were observed after applying a Bonferroni correction to account for the multiple comparisons made across the seven group categories. The Bonferroni correction, a conservative method to control the familywise error rate, adjusted our significance level to 0.00714, given the seven tests conducted.

Further to this, an additional layer of analysis was conducted to understand the interaction between age, gender, and group affiliation among survey participants. While the Chi-squared test of independence revealed a statistically significant association between these variables for the dataset as a whole ($p = 5.26 \times 10^{-10}$, $\chi^2 = 197.32$, $df = 90$), the granularity provided by examining individual group categories with the Bonferroni adjusted α level did not yield statistically signifi-

cant results. This indicates that the patterns observed in the aggregated data do not uniformly persist across all groups when considering the stricter criterion for significance.

In a more granular examination of the gender distribution within individual groups, the following outcomes were noted, none of which reached statistical significance under the Bonferroni corrected α level.

These findings suggest that, despite initial indications of significant associations in the overall sample, the distribution of age and gender within specific user groups does not significantly deviate from what would be expected by chance when a rigorous correction for multiple comparisons is applied. This differentiation underscores the importance of considering the potential for Type I errors in statistical analyses involving multiple tests and highlights the need for careful interpretation of results in studies with multiple group comparisons.

5.3.2.3 Question 18 - Please select completed educational programs/-courses?

A chi^2 test was performed against the self-selected groupings that suggest there is something significant within this data ($p < 0.001$, $\chi^2 = 386.54, df = 216$). Table 5.38 shows that the most educated participants self categorised as general public consisting of the largest number of participants, 29%, who have a PhD, Masters and Undergraduate degrees. The non-professional group is ranked second and consists equally of relatively high numbers, the exception being the Academic category which, in a single grouping, holds 50% more in PhD's, but this raises the question of were some academics self classifying as general public because of the task they were doing, with the survey specifically asking about the specific visit when they completed the survey? Overall, the general public group appears to be the most qualified group of all. Table 5.38 also shows that the museum staff and professionals groups are in the bottom position, ranking 6 and 7 respectively, of the seven groups although 86% of these combined groups are educated to degree level or above which favours positively in comparison to the 54% of the combined general public and non-professionals groupings.

What is interesting to note from the data is that 88% of all the participants have

more than a basic (GCSE level) education and 68% have higher (undergraduate/-masters/doctorate level) education qualifications.

Table 5.38: Q18 - Completed education by group

Groups	**General Public**		**Academics**		**Non-Professionals**		**Teachers**		**Museum Staff**		**Professionals**		**Students**	
Completed Education	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
PhD	11	4	10	40	4	3	1	4	1	10	0	0	0	0
MSc	47	19	8	32	22	16	6	24	5	50	3	60	4	12
BSc	83	33	5	20	46	34	15	60	2	20	2	40	10	30
Prof	27	11	1	4	20	15	1	4	0	0	0	0	1	3
FE	52	21	0	0	29	21	2	8	1	10	0	0	11	33
Secondary	27	11	1	4	14	10	0	0	1	10	0	0	7	21
None	6	2	0	0	2	1	0	0	0	0	0	0	0	0
Column Total	**253**	**100**	**25**	**100**	**137**	**100**	**25**	**100**	**10**	**100**	**5**	**100**	**33**	**100**

5.3.2.4 Question 19 - Please select in progress educational program/-courses?

In addition to establishing the educational level of the participants, the survey also gathered data regarding those participants who were still studying (see Table 5.39) The chi^2 for groups against education in progress produces a result that appears significant $p < 0.001$ ($\chi^2 = 336.32, df = 144$), however this is mostly down to the large numbers of non-professionals (n = 75) and general public (n = 175) who are not studying anything currently.

The survey assessed the educational qualifications of participants who are currently pursuing further education, revealing a diverse array of academic pursuits across the different groups. Of the 488 respondents, a substantial portion of the General Public (n = 78, 31%) and Non-Professionals (n = 62, 45%) are engaged in further education, with Further Education (FE) and Bachelor's degrees (BSc) being the most common qualifications being pursued.

Academics, although a smaller group overall (n = 25), show a significant inclination towards postgraduate education with 12% (n = 3) pursuing a PhD and another 12% (n = 3) working towards a Master's degree (MSc). Similarly, in the Non-Professional group, 12% (n = 17) are studying for professional qualifications.

Teachers and Students indicate a strong engagement with further education: 48% (n = 12) of Teachers are pursuing additional qualifications, with Professional development and FE courses being the most prevalent. Unsurprisingly, 100% (n = 33)

of Students are currently in education, with a majority (55%, n = 18) undertaking undergraduate studies.

In contrast, Museum Staff and Professionals demonstrate lower levels of current educational engagement, with 70% (n = 7) of Museum Staff and 60% (n = 3) of Professionals not pursuing further education. This indicates that members of these groups are less likely to be enrolled in educational programmes at the time of the survey.

Although the data shows that 59% (n = 287) of the participants are not currently engaged in any form of study, 41% (n = 201) were, with at least 19% (n = 91) studying at undergraduate level or above. This would indicate that the participants of the survey are generally individuals who are able to use their own ideas and research in response to complex problems and situations.

Overall, the data suggests that while a significant proportion of the survey population is engaged in further education, there is variability across groups, with academics and students being the most educationally active.

Table 5.39: Q19 - Education in progress by group

Groups	**General Public**		**Academics**		**Non-Professionals**		**Teachers**		**Museum Staff**		**Professionals**		**Students**	
In Progress Education	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
PhD	1	0	3	12	4	3	1	4	0	0	0	0	1	3
MSc	12	5	3	12	4	3	2	8	1	10	1	20	6	18
BSc	18	7	3	12	8	6	3	12	1	10	1	20	18	55
Professional	17	7	2	8	17	12	4	16	1	10	0	0	2	6
FE	20	8	0	0	20	15	2	8	0	0	0	0	6	18
Secondary	10	4	0	0	9	7	0	0	0	0	0	0	0	0
Sub-total	**78**	**31**	**11**	**44**	**62**	**45**	**12**	**48**	**3**	**30**	**2**	**40**	**33**	**100**
None	175	69	14	56	75	55	13	52	7	70	3	60	0	0
Column Total	**253**	**100**	**25**	**100**	**137**	**100**	**25**	**100**	**10**	**100**	**5**	**100**	**33**	**100**

5.3.2.5 Question 20 - Where in the world are you at this moment?

The location data indicated that 61% of the participants are local to the National Museums Liverpool (NML), 18% reside elsewhere in England, 6% in other parts of the UK, and 15% are from outside the UK. A χ^2 analysis of the contingency table for participant location and group membership revealed significant differences within the dataset ($p < 0.001$, $\chi^2 = 6.21, df = 24$), suggesting variations in the geographical distribution of different participant groups (see Table 5.40). Specifically, the "general public" group exhibited significant differences from both the "academic" and "professional" groups after adjusting for multiple comparisons with the Bonfer-

roni correction ($p < 0.05$, $\chi^2 = 6.21, df = 4$). The initial significance level was set at $\alpha = 0.05$, and the Bonferroni correction was applied due to 21 total pairwise comparisons, resulting in a corrected significance threshold of $\alpha = 0.00238$. In contrast, the "non-professional" group showed significant differences from the "general public", "academic", and "professional" groups before the application of the Bonferroni correction. However, these differences were not statistically significant after the correction was applied. This result highlights that the "general public" group tends to be more locally based compared to the "academic" and "professional" groups.

Table 5.40 also shows that 48% of Academics are from the UK and the 52% are from the rest of the world. As would be expected, 80% of the museum staff are in the immediate Merseyside area, 10% being in the same geographical location and none outside of England. Similarly, teachers are mostly really local, which is to be expected when planning a physical visit to the museum. Students are predominantly from the UK (73%), although 27% are situated in the rest of the world.

Table 5.40: Responses to Question #20 "Where in the world are you at the moment?"

Location	**Merseyside**		**Northwest**		**England**		**UK**		**World**		**Row Total**	
Groups	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
Academics	4	16	3	12	4	16	1	4	13	52	**25**	**100**
General Public	95	38	73	29	44	17	18	7	23	9	**253**	**100**
Museum Staff	8	80	1	10	1	10	0	0	0	0	**10**	**100**
Non-Professionals	47	34	25	18	25	18	10	7	30	22	**137**	**100**
Professionals	0	0	0	0	0	0	2	40	3	60	**5**	**100**
Students	8	24	8	24	7	21	1	3	9	27	**33**	**100**
Teachers	13	52	3	12	4	16	1	4	4	16	**25**	**100**
Column Total (n)	**175**		**113**		**85**		**33**		**82**		**488**	

5.3.2.6 Question 21 - Which of the following categories best describes your employment status?

Employment by group (see Table 5.41) has a $p-value < 0.001$, $\chi^2 = 263.48, df = 42$ and shows a significant difference in employment status. This could be due to the majority of the academics, teachers and all the professionals and Museum staff are working, which is what would be expected. The interesting split here is the general public and non-professional groups, which have the majority of the group classed as working, but both have rather large retired samples and similar-sized non-

working groups. After conducting pairwise Chi-Square tests to explore differences between groups across employment. A total of 21 comparisons were performed, with Bonferroni correction applied to adjust for multiple testing. The corrected alpha level was set at $\alpha = 0.00238$.

Significant differences were observed in the following pairs:

- Non-Professionals vs. Students: The Chi-Square statistic was 84.9047 with 7 degrees of freedom (df), and an original p-value of <0.00001, which remained significant after Bonferroni correction (corrected p-value < 0.00001).

- General Public vs. Academics: The Chi-Square statistic was 29.6251 with 7 df, and an original p-value of 0.00011, which remained significant after correction (corrected p-value = 0.00234).

- General Public vs. Students: The Chi-Square statistic was 152.9647 with 7 df, and an original p-value < 0.00001, which remained significant after correction (corrected p-value < 0.00001).

- Teachers vs. Students: The Chi-Square statistic was 31.6635 with 7 df, and an original p-value of 0.000019, which remained significant after correction (corrected p-value = 0.00040).

Other comparisons, although initially showing significance, did not maintain their significance after adjusting for multiple comparisons. These include non-professionals vs. academics, non-professionals vs. teachers, and academics vs. students, among others, which demonstrated initial significance but did not remain significant after Bonferroni correction. Whilst there are clear differences between the groups, the nature of the groupings and what they do for employment means there are no real surprises in these results.

Table 5.41: Q21 - Employment status by group

	Full-time Work		Part-time Work		Retired		Student Full-time		Student Part-time		Unemployed		Looking for work		Disabled		Row Total	
Groups	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
Academics	13	52	4	16	2	8	4	16	0	0	0	0	0	0	2	8	**25**	**100**
General Public	95	38	56	22	70	28	3	1	1	0	13	5	9	4	6	2	**253**	**100**
Museum Staff	8	80	2	20	0	0	0	0	0	0	0	0	0	0	0	0	**10**	**100**
Non-Professionals	42	31	22	16	41	30	3	2	3	2	13	9	4	3	9	7	**137**	**100**
Professionals	4	80	0	0	0	0	0	0	0	0	0	0	1	20	0	0	**5**	**100**
Students	5	15	2	6	2	6	21	64	1	3	1	3	1	3	0	0	**33**	**100**
Teachers	17	68	5	20	0	0	0	0	0	0	2	8	1	4	0	0	**25**	**100**
Column Total (n)	**184**		**91**		**115**		**31**		**5**		**29**		**16**		**17**		**488**	

5.3.3 Motivation and Reason for visit

5.3.3.1 Question 2 - Today I am visiting the NML website for?

Establishing the reason why the participants visited NML suggests that there is something significant in the data ($p < 0.001$ $\chi^2 = 317.15, df = 18$). To investigate this significance for visiting museums among different groups, pairwise Chi-square tests were conducted to examine the association between group membership and visitation reasons. Given the multiple comparisons involved, the Bonferroni correction was applied to adjust for the increased risk of Type I error. A total of 21 comparisons were made, setting the Bonferroni adjusted alpha level at $\alpha = 0.00238$.

The analysis revealed significant differences in the reasons for museum visits among several groups. Notably, the comparison between groups non-professional and academics groups yielded a $\chi^2 = 50.625, df = 3$ (original $p < 0.001$, corrected $p < 0.001$), indicating a significant difference even after adjusting for multiple comparisons. Similarly, significant differences were observed between groups non-professionals and students (corrected $p < 0.001$ $\chi^2 = 85.023, df = 3$), general public and academics (corrected $p < 0.001$ $\chi^2 = 96.978, df = 3$), and between groups general public and students (corrected $p < 0.001$ $\chi^2 = 146.558, df = 3$), among others. While several comparisons revealed significant differences, others, such as the comparison between groups academic and professional (original $p = 0.339$, corrected $p = 1.000 \chi^2 = 3.360, df = 3$), did not yield significant associations after adjusting

for multiple comparisons. The data in table 5.42 shows that 70% of the total number of participants reason for the visit are personally motivated visits to the website, of which 58% were from the general public group and 33% from the non-professional group. Some 70% of students predominantly used the site for study, which is pretty much expected. However, the academics show almost equal amounts coming to the NML site for work, study, and personal reasons (which raises the question of whether an academic self-classifies as an academic or not when doing a non-work task). Museum staff (60%) were using the site mostly for work, but there are instances where some were using it for personal reasons and to simply pass some time. The professionals are also all using it for work (60%) or for personal reasons (40%). While Teachers are mostly saying they were using it for personal reasons (44%) or for work (32%).

However, what is interesting is the category to "pass time". Looking at all groups, in Table 5.42, shows that 14% of the participants visit the site to pass time. The general public had 43 participants and the non-professionals had 15, which is something that can be readily understood. Surprisingly, the most significant group that indicated "pass time" was that of museum staff, with some 20% (2 of 10) of those participants self-identified as museum staff, followed by teachers with 16% (4/25). The findings regarding the general public users correspond with the observations made by Walsh and Hall (2015); Mayr et al. (2016). Specifically, a significant number of these general public users visit the website to pass time, engaging in what is commonly referred to as 'casual browsing.' This behaviour, prevalent among visitors to physical museums, is similarly reflected in the responses given by survey participants about their online use.

Table 5.42: Responses to question #2 "Today I am visiting the NML website:"

Reason for Visit	**Personal**		**Study**		**Pass Time**		**Work**		**Row Total**	
Groups	n	%	n	%	n	%	n	%	**n**	%
Academics	8	32	9	36	1	4	7	28	**25**	**100**
General Public	200	79	6	2	43	17	4	2	**253**	**100**
Museum Staff	2	20	0	0	2	20	6	60	**10**	**100**
Non-Professionals	112	82	5	4	15	11	5	4	**137**	**100**
Professionals	2	40	0	0	0	0	3	60	**5**	**100**
Students	7	21	23	70	2	6	1	3	**33**	**100**
Teachers	11	44	2	8	4	16	8	32	**25**	**100**
Column Total (n)	**342**		**45**		**67**		**34**		**488**	

5.3.3.2 Question 3 - What is the primary purpose of your visit to the NML website today?

A χ^2 on the responses to Q3 the purpose for their visit to the website today, against the established groups, reports that there is something interesting in this data ($p < 0.001$ $\chi^2 = 108.97, df = 48$).

To explore this significance in the different purposes groups visit museums, pairwise Chi-square tests were employed. Considering the multiple comparisons made (totalling 21), the Bonferroni correction was applied to mitigate the risk of Type I error, setting an adjusted significance level at $\alpha = 0.00238$.

A total of 21 pairwise comparisons were conducted with the Bonferroni correction, applied to adjust for the multiplicity of tests. Significant associations were observed in the following comparisons even after Bonferroni adjustment:

- Non-professional and general public groups demonstrated a significant difference with a corrected $p < 0.01$, $\chi^2 = 28.214, df = 8$.
- General public and academic groups showed a profound association (corrected $p < 0.001$ $\chi^2 = 41.344, df = 8$), indicating a significant variance in their museum visitation purposes.
- The comparison between the general public and professional groups also yielded a significant result (corrected $p < 0.01$ $\chi^2 = 31.870, df = 8$).
- Another notable difference was observed between the general public and student groups (corrected $p < 0.05$ $\chi^2 = 25.020, df = 8$).

Several comparisons initially indicated significance before the Bonferroni correction but did not maintain this status upon adjustment. For example, the comparison between groups non-professional and academic groups (original $p = 0.034$) and between the general public and museum staff (original $p = 0.036$) showed no significant differences after correction.

The data in table showed (see Table 5.43) that of the 488 participants, 49% were primarily looking at the NML website in preparation of a physical visit; 10% were viewing the website for an overview of the NML museum; 7% were looking at a Collection overview; 4% were there looking for a Known Collection; 6% were

looking for a Known Item; 0.4% were looking for the Shop; 5% for News; 19% indicated Other; and finally 1% indicated Unknown.

A closer examination of the data focusing on the particular groupings, found that the purpose of the visit by group, highlights that the main focus of the general public group was to prepare for a physical visit (61%, n = 154 of 253) as was the teacher group (60%, n = 15 of 25). To a much smaller scale, the non-professional participants also have this selected the most (36%, n = 49 of 137).

The general public group also seems to want to gain an overview of what collections are on exhibit. The professionals, however, were not preparing for a visit or trying to gain an overview of the museum or the collections at all; similarly, they were not interested in the peripherals of the website, such as museum news or shop. Instead, they sought known items from the collections or known collections. The museum staff were interested in known collections (20%) but were also preparing for visits (40%). This may suggest that the participants who self-identified as museum staff were not staff at the NML museums but may either work at other museums or another part of the NML museums. Generally, it is difficult to think that someone who ordinarily works for a museum would want to visit the museum during their off time, but it is possible.

Table 5.43: Responses to question #3 “What is the primary purpose of your visit to the NML website today?”

Visit Purpose	**Pre-Visit**		**Museum Overview**		**Collection Overview**		**Known Collection**		**Known Item**		**Shop**		**News**		**Other**		**Unknown**		**Row Total**	
Groups	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
Academics	5	20	0	0	2	8	3	12	5	20	1	4	1	4	8	32	0	0	**25**	**100**
General Public	154	61	23	9	13	5	4	2	8	3	1	0	12	5	37	15	1	0	**253**	**100**
Museum Staff	4	40	1	10	1	10	2	20	0	0	0	0	0	0	2	20	0	0	**10**	**100**
Non-Professionals	49	36	17	12	9	7	5	4	12	9	0	0	7	5	35	26	3	2	**137**	**100**
Professionals	0	0	0	0	0	0	1	20	2	40	0	0	0	0	2	40	0	0	**5**	**100**
Students	11	33	6	18	6	18	3	9	1	3	0	0	0	0	5	15	1	3	**33**	**100**
Teachers	15	60	3	12	1	4	0	0	1	4	0	0	2	8	3	12	0	0	**25**	**100**
Column Count (n)	**238**		**50**		**32**		**18**		**29**		**2**		**22**		**92**		**5**		**488**	

This highlights that preparation for a visit is a primary characteristic for most groups, including both the general public and the non-professional groups, aligning with previous studies that identified this as the most frequently mentioned motivation for visiting a museum website Booth (1998); Goldman et al. (2004); Marty (2007b).

The findings illustrate the variability in reasons for museum visits across different demographic groups, with significant differences detected in four out of 21 comparisons after adjusting for multiple tests. These results highlight museum-goers' nuanced preferences and motivations, offering valuable insights for museum web developers and marketers to better understand the website visitors.

5.3.4 Shared Experience, Engagement and Behaviour

5.3.4.1 Question 5 - Are you sharing this website experience?

When it came to viewing the NML website (see Table 5.44) the data showed that of the 488 participants, 89% were viewing the website alone; 8% viewed with others, whether children or remotely with someone else. This left 3% who indicated "Other".

While the χ^2 shows the is nothing significant ($p-value = 0.11, \chi^2 = 25.531, df = 18$) sharing by group data, there are a number of results that are still interesting. In the groupings general public (n = 17), non-professionals (n = 9) and teachers (n = 3), group participants claimed to be visiting the website with a child at their sides. None of the professionals, museum staff, academics or students claimed this, which is what would be expected if they were there for work or study reasons or during work time. What is interesting is that 2 academics and 1 professional were viewing with another adult by their side. Only the general public and non-professionals group had participants claim to be viewing with another adult remotely from them. But both also had a number claim to be viewing with another adult during this website visit. All the students and museum staff were on their own.

Table 5.44: Q5 - Sharing the Online Experience by group

Sharing the online visit	**I am looking with a child/children by my side**		**Other**		**I am looking with someone else who is remote**		**No, I am on my own**		**Row Total**	
Groups	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
Academics	0	0	2	8	0	0	23	92	**25**	**100**
General Public	17	7	5	2	3	1	228	90	**253**	**100**
Museum Staff	0	0	0	0	0	0	10	100	**10**	**100**
Non-Professionals	9	7	6	4	6	4	116	85	**137**	**100**
Professionals	0	0	1	20	0	0	4	80	**5**	**100**
Students	0	0	0	0	0	0	33	100	**33**	**100**
Teachers	3	12	0	0	0	0	22	88	**25**	**100**
Column Total (n)	**29**		**14**		**9**		**436**		**488**	

5.3.4.2 Question 6 - How frequently do you visit the NML website?

A χ^2 of frequency of visit against group returns a $p < 0.001$ ($\chi^2 = 100.859, df = 24$) showing that there is something significant happening within this data. The frequency of visit (Table 5.45) shows a similar picture to table 5.44, although with significant differences after the Bonferroni correction was applied to adjust the significance level for multiple comparisons (21 comparisons) to $\alpha = 0.00238$.

A significant association between groups of non-professional and museum staff (corrected $p < 0.001$ $\chi^2 = 46.478, df = 4$), suggesting a substantial difference in website visitation frequency between these groups. comparison between the general public and academics also showed a significant difference (corrected $p < 0.001$ $\chi^2 = 21.976, df = 4$). The comparison between groups general public and museum staff yielded the most significant difference (corrected $p < 0.001$ $\chi^2 = 67.964, df = 4$). Another significant result was observed between the student and museum staff groups (corrected $p < 0.001$ $\chi^2 = 22.544, df = 4$).

Several group comparisons initially indicated potential significance but did not show significance after the Bonferroni correction adjustment. For instance, the comparison between groups non-professional and academic groups (original $p = 0.012$) and academic and museum staff (original $p = 0.014$) were not considered significant

after correction.

The results of this question clearly show that the site attracts a lot of users for the first time. This could be a potential issue of failing to gain repeat visits from visitors. Still, the content is not necessarily the type of content that would see a lot of daily or weekly repeat visits from the general public and non-professional groups, who were mostly looking for pre-visit planning. All the professionals, a majority of the students and academics were first-time visitors. The museum staff were all repeat visitors, although the frequency ranged from a couple of times a year to daily. The academics are the only other group to have a couple of participants claim to visit daily.

Table 5.45: Responses to question #6 "How frequently do you visit the NML website?"

Frequency	**First Visit**		**Yearly**		**Monthly**		**Weekly**		**Daily**		**Row Total**	
Group	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
General Public	133	53	82	32	32	13	6	2	0	0	**253**	**100**
Non-professional	78	57	40	29	13	9	6	5	0	0	**137**	**100**
Student	22	67	7	21	2	6	2	6	0	0	**33**	**100**
Academic	10	40	10	40	2	8	1	4	2	8	**25**	**100**
Teacher	11	44	8	32	3	12	3	12	0	0	**25**	**100**
Professional	5	100	0	0	0	0	0	0	0	0	**5**	**100**
Museum Staff	0	0	2	20	5	50	1	10	2	20	**10**	**100**
Column Total	**254**	**52**	**149**	**31**	**57**	**12**	**19**	**4**	**4**	**1**	**483**	**100**

5.3.4.3 Question 7 - Roughly how many web pages have you viewed on the NML website during this visit?

The number of pages viewed during the visit (see Table 5.46) shows that all the groups have a lot of single-page visitors. The χ^2 result for the contingency table of the number of pages accessed by the group returns a $p - value < 0.05$ ($\chi^2 = 38.103, df = 18$), suggesting that there is something significant in the data. After the Bonferroni correction was used to adjust the significance levels for the 21 comparisons, there is only a significant difference between the general public and academic groups (corrected $p < 0.05, \chi^2 = 22.748, df = 3$). While the comparison between groups non-professional and academic showed some initial indication of a difference (original $p = 0.035$), this did not hold after applying the Bonferroni correction (corrected $p = 0.729$). Similarly, other comparisons, such as general public

and museum staff (original $p = 0.016$), did not maintain their significance after correction. The data in table 5.46 shows that a single website page is viewed by 48% of the participants across all groups; a further 2 - 10 pages are then viewed by 47%. A dramatic reduction can then be seen when looking at the remaining number of participants, some 5% who view more than 11+ pages. The interesting results in this table are that a few general public (n = 8) and non-professionals (n = 8) view a lot of pages (11+). This behaviour is more than what is expected to be seen by professionals who have not gone beyond viewing 10 pages. Museum staff tended to only visit between 2 and 10 pages, and academics who, similarly to the museum staff, tended to remain at 10 pages, although 4 of the 25 participants have indicated that they view 11+ pages. To investigate the notable differences between the general public and academic groups, it was observed that academics have a relatively higher frequency of participants accessing '21+' pages compared to their group size, in contrast to the general public. This indicates a higher likelihood of academics accessing a more extensive number of pages, which could suggest a deeper engagement with the site, potentially for scholarly or educational purposes. On the other hand, the general public's page access is more evenly spread, primarily within the 'Just the one' and '2 to 10' page categories, which could suggest a trend towards more casual browsing behaviour.

Table 5.46: Q7 - Number of web pages accessed during this visit by group.

Number of pages accessed	**Just the one**		**2 to 10**		**11 to 20**		**21+**		**Row Total**	
Groups	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
Academics	13	52	8	32	1	4	3	12	**25**	**100**
General Public	123	49	122	48	7	3	1	0	**253**	**100**
Museum Staff	2	20	6	60	2	20	0	0	**10**	**100**
Non-Professionals	66	48	63	46	6	4	2	1	**137**	**100**
Professionals	3	60	2	40	0	0	0	0	**5**	**100**
Students	17	52	15	45	1	3	0	0	**33**	**100**
Teachers	11	44	13	52	1	4	0	0	**25**	**100**
Column Total (n)	**235**		**229**		**18**		**6**		**488**	

5.3.4.4 Question 8 - When on a web page about an object or collection, do you typically?

The analysis of how participants interact with website content (see Table 5.47) reveals that, while there are minor variations in the methods of engagement across different groups, none of these differences reach statistical significance ($p = 0.2, \chi^2 = 14.988, df = 12$). Approximately 47% of participants favour scanning the page for keywords, the most preferred method across all groups. The second most common approach is to look at pictures before reading the text, accounting for 31% of participants. The remaining 23% prefer to read the entire page from top to bottom. There is a marginal trend where groups not engaged in work-related tasks are slightly more inclined to look for images first.

Specifically, academics and teachers are more prone to scan for relevant words, at 44% and 56%, respectively. This approach is also dominant among the general public, non-professionals, and students, with around 45-47% favouring this method. Reading the page from top to bottom in depth is a method chosen by a significant number of non-professionals (30%) and museum staff (30%) but less so by the general public (18%).

Professionals display an even split between scanning for words and reading in-depth, both at 40%, suggesting a balanced approach to engaging with the web pages. Overall, the preference for scanning for keywords may indicate a targeted and efficient approach to information retrieval on the web. These patterns are reflective of the pragmatic way in which users navigate digital content.

Table 5.47: Q8 - How people use content by group

Method of using web pages	**Look for pictures first and then read the text**		**Read the whole page from top to bottom in depth**		**Scan the page to find the relevant words you are looking for?**		**Row Total**	
Groups	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
Academics	7	28	7	28	11	44	**25**	**100**
General Public	92	36	46	18	115	45	**253**	**100**
Museum Staff	1	10	3	30	6	60	**10**	**100**
Non-Professionals	32	23	41	30	64	47	**137**	**100**
Professionals	1	20	2	40	2	40	**5**	**100**
Students	11	33	7	21	15	45	**33**	**100**
Teachers	7	28	4	16	14	56	**25**	**100**
Column Total (n)	**151**		**110**		**227**		**488**	

5.3.4.5 Question 14 - What time of day do you usually undertake the type of activity you are primarily using the NML website for today?

The χ^2 returned $p < 0.01$ ($\chi^2 = 127.739, df = 84$), suggests that there is some interesting data (see Table 5.48). All groups have some participants claiming to use the site at all times of the day. However, when considering the combinations of options made available to the participants, some 308 (63%) participants, that spanned all groups, chose options that offered/contained an evening interaction. Examining this further identified 38% (n = 187) of the 488 participants across the teachers, students, non-professionals and general public groups stand out as being predominantly evening users. This includes quite significant amounts of some of the groups: 52% of teachers, 41% of general public, 40% of non-professional, and 33% of student users. The academics, professionals and museum staff are, as expected, primarily daytime users (60% of those participants in these groups indicated a preference for a daytime interaction and excluded evening). However, upon closer examination of the timings indicated by the Academics, there is a fairly even split between daytime and evening (48% vs 52%), which may be linked to the fluidity of their job.

After the Bonferroni correction was applied, adjusting the α level to 0.00238

for the 21 comparisons, significant differences were observed in the preferred access times between the non-professional and museum staff groups (corrected $p < 0.01, \chi^2 = 35.556, df = 12$), as well as between the general public and museum staff groups (corrected $p < 0.01, \chi^2 = 35.092, df = 12$). The key to these differences appears to be the distinct temporal patterns of web access: museum staff have a notable preference for afternoon access (30%), which contrasts with the evening access favoured by the majority of the general public (41%) and non-professionals (40%). This suggests that museum staff's website usage is more aligned with their professional hours, while the general public and non-professionals tend to access the website outside of typical work hours, reflecting their personal or leisure-related usage patterns.

The comparisons between the general public and academic groups (original $p = 0.016$) and the general public and student groups (original $p = 0.028$) initially suggested differences. However, these did not remain significant after applying the Bonferroni correction (corrected $p = 0.326$ and 0.582, respectively).

The data, therefore, suggests that the primary interaction with the NML website takes place during the evening, which could essentially be classed as the participants' leisure time. This could also explain the significant difference with the museum staff who would access during the working day.

Table 5.48: Q14 - Preferred access times by group

Preferred Access Times	Groups	Academics	General Public	Museum Staff	Non-Professionals	Professionals	Students	Teachers	Row Total (n)	Row Total (%)
Afternoon	**n**	4	28	3	18	1	5	3	**62**	
	%	16	11	30	13	20	15	12		**12.7**
Afternoon and Evening	**n**	3	7	0	10	0	5	3	**28**	
	%	12	3	0	7	0	15	12		**5.7**
Evening	**n**	3	104	1	55	0	11	13	**187**	
	%	12	41	10	40	0	33	52		**38.3**
Lunchtime	**n**	1	12	1	2	1	2	2	**21**	
	%	4	5	10	1	20	6	8		**4.3**
Lunchtime and Afternoon	**n**	0	0	0	0	0	1	0	**1**	
	%	0	0	0	0	0	3	0		**0.2**
Lunchtime and Afternoon and Evening	**n**	0	1	0	2	0	0	0	**3**	
	%	0	0	0	1	0	0	0		**0.6**
Lunchtime and Evening	**n**	0	4	0	0	0	0	0	**4**	
	%	0	2	0	0	0	0	0		**0.8**
Morning	**n**	5	38	0	23	2	5	1	**74**	
	%	20	15	0	17	40	15	4		**15.2**
Morning and Afternoon	**n**	2	4	1	3	0	0	0	**10**	
	%	8	2	10	2	0	0	0		**2.0**
Morning and Afternoon and Evening	**n**	2	4	0	3	0	1	1	**11**	
	%	8	2	0	2	0	3	4		**2.3**
Morning and Evening	**n**	0	19	0	5	0	1	0	**25**	
	%	0	8	0	4	0	3	0		**5.1**
Morning and Lunchtime	**n**	0	3	1	1	0	1	0	**6**	
	%	0	1	10	1	0	3	0		**1.2**
Morning and Lunchtime and Afternoon	**n**	0	3	2	1	0	0	0	**6**	
	%	0	1	20	1	0	0	0		**1.2**
Morning and Lunchtime and Afternoon and Evening	**n**	5	26	1	13	1	1	2	**49**	
	%	20	10	10	9	20	3	8		**10.0**
Morning, Lunchtime, Evening	**n**	0	0	0	1	0	0	0	**1**	
	%	0	0	0	1	0	0	0		**0.2**
Columm Total	**n**	**25**	**253**	**10**	**137**	**5**	**33**	**25**	**488**	
	%	**100**	**100**	**100**	**100**	**100**	**100**	**100**		**100.0**

5.3.4.6 Question 15 - What type of device are you using for this visit to NML website?

The survey provided an opportunity to establish how the NML website was being accessed. Performing a χ^2 $p = 0.001$ ($\chi^2 = 32.376, df = 12$) suggested that there is

something of significance in the device type by group data that was gathered.

The data (see Table 5.49) showed that 56% (273) of the participants from all groups accessed the website via a desktop/laptop PC. Accessing via a tablet PC was undertaken by 24% (117), excluding museum staff and professionals, which recorded zero participants for this option, while mobile phone access was 20%, excluding professionals, which recorded zero participants also for this option.

A closer examination of the data for each individual group revealed that the majority of academics from this group, (72%) used a PC, 16% tablet and 12% mobile. Similarly, the general public (50% PC, 28% a tablet and 21% a mobile); non-professionals (56% PC, 24% tablet and 20% mobile), and students (76% PC, 21% tablet and 3% mobile) also indicated this order of device preference. The professional groups preference was 100% for using desktop computers.

The data reveals interesting patterns in the use of mobile and tablet devices. Notably, 20% of participants across all groups, with the exception of professionals, preferred to access the website using mobile technology. The museum staff group demonstrated a significant inclination towards PC use (80%), with mobile use at 20% and no tablet use reported. Similarly, the teacher group showed a preference for PCs (52%), followed by mobiles (44%) and minimal use of tablets (4%). Despite being the third choice among the general public, mobile usage was still notable at 21%. The higher engagement with technology among students was expected, given their regular interaction with digital devices. Likewise, professionals were anticipated to exhibit higher usage, primarily for work-related tasks.

It's interesting that the non-professional and general public users seem indifferent to the type of device they use, with a slight preference for desktops. Perhaps this apparent flexibility in device choice might be attributed to their circumstances at the time of accessing the information. For example, if they are out of the house for the day and decide to see if the museum is open, then they would likely use their phone. but if they were home then they would use a PC. However, it's important to note this is purely speculative and lacks concrete data and further research about specific device use, task and location. It could also be that this stands out for no more reason than that these groups have significantly more members, so there is more chance of them choosing the full range of devices.

The only statistically significant differences between groups after the Bonfer-

roni correction has been applied to account for the 21 comparisons are between the teacher and the student groups (corrected $p < 0.05, \chi^2 = 15.820, df = 2$). This could suggest a generational/age technology adoption difference between the students (70% under the age of 35) and teachers (88% over the age of 34, 20% being over 54).

Table 5.49: Q15 - Type of device by group

Groups	Academic		General Public		Museum Staff		Non-Professional		Professional		Student		Teacher		Row Total	(All groups)
Device	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
Desktop or Laptop PC	18	72	127	50	8	80	77	56	5	100	25	76	13	52	**273**	**56**
Mobile Phone	3	12	54	21	2	20	27	20	0	0	1	3	11	44	**98**	**20**
Tablet PC (iPad, Android tablet etc)	4	16	72	28	0	0	33	24	0	0	7	21	1	4	**117**	**24**
Total Column	**25**	**100**	**253**	**100**	**10**	**100**	**137**	**100**	**5**	**100**	**33**	**100**	**25**	**100**	**488**	**100**

5.3.5 Domain Knowledge

5.3.5.1 Question 9 - In the context of cultural heritage and your current visit to the NML website, please select the appropriate statement:

Performing a χ^2 of domain knowledge by group returned a p-value of $p < 0.05$ ($\chi^2 = 148.487, df = 18$), suggesting that there is something significant within this data.

The participants were asked this question to establish how knowledgeable they felt they were at the time of completing the survey and were provided with 4 options (see Table 5.50). The data indicated that 96% (468) of the participants possessed some domain knowledge, albeit to differing levels, while 4% had stated that they did not know.

The groups that showed that they had higher domain experience levels for the reason they were visiting were, as expected, the Academics and Museum staff groups. However, the Non-Professionals, General Public and Teachers groups indicate a higher percentage than the Museum Staff group for this particular option, albeit this could be skewed by the volume of participants for these groups in comparison to the Museum Staff group.

The majority of participants (59.8%) indicated that they have some experience and knowledge. The higher concentration of participants can be seen as the General Public (31.4%) and the Non-Professionals (20.1%) groups, while Students (3.1%), Academics (2%) and Teachers (2%) are all fairly similar.

Those who selected the novice option consisted of 26.2% of the total number of participants. This consisted mainly of the general public (16%), non-professionals (5.9%) groups that comprised of the majority of survey respondents and as such is expected. The students' and Teachers' roles are more generally about learning, so categorising themselves as novices is not surprising. However, what is surprising is the selection of this option by the Professionals group, and then to also select the "don't know" option, which when considering the definition provided under Phase 1 (see Appendix F) requires further investigation.

Table 5.50: Responses to question Q9 "In the context of cultural heritage and your current visit to the NML website, please select the appropriate statement"

Domain Knowledge	**Groups**	**Academics**	**General Public**	**Museum Staff**	**Non-Professionals**	**Professionals**	**Students**	**Teachers**	**Row Total (n)**	**Row Total (%)**
I'm highly experienced and have extensive background knowledge	n	15	6	5	10	1	3	8	**48**	
	%	60	2	50	7	20	9	32		**10**
I have some experience and background knowledge	n	10	153	5	98	1	15	10	**292**	
	%	40	60	50	72	20	45	40		**60**
I'm a novice with little knowledge within this area	n	0	78	0	29	2	14	5	**128**	
	%	0	31	0	21	40	42	20		**26**
I don't know	n	0	16	0	0	1	1	2	**20**	
	%	0	6	0	0	20	3	8		**4**
Column Total	**n**	**25**	**253**	**10**	**137**	**5**	**33**	**25**	**488**	
	%	**100**	**100**	**100**	**100**	**100**	**100**	**100**		**100**

When considering the results from the data for all questions up to this point, the outcomes for the general public and non-professional groups have been similar, however, as for domain knowledge there are some differences between the two (see Table 5.50). For the domain knowledge when using the NML website (see Table

5.50), the general public and non-professionals are significantly different after adjusting for multiple (21) comparisons corrected $p < 0.001, \chi^2 = 18.637, df = 3$. The general public group also showed significant differences when compared with academics, teachers, and museum staff, among others, with corrected $p < 0.001, \chi^2 = 110.942, df = 3$ for general public vs. academics, showcasing the substantial variation in domain knowledge. The non-professionals also showed significant differences to the academics, professionals, teachers and museum staff groups with corrected P-values indicating robust significance in these comparisons (e.g., non-professionals vs. academics with a corrected $p < 0.001, \chi^2 = 46.495, df = 3$). Academics as well as showing the above difference to the general public and non-professional groups also displayed a significant difference to the professional group (corrected $p < 0.05, \chi^2 = 16.705, df = 3$).

Comparisons such as non-professional vs. student, general public vs. professional, and academic vs. teachers, while initially showing significance, did not maintain this after the Bonferroni correction.

5.3.5.2 Question 10 - Rate your Cultural heritage knowledge

The participants were asked to rate their general CH knowledge on a scale from 1 to 5, with 5 indicating the highest level of knowledge and 1 the lowest (see Table 5.51). Statistical analysis, applying a χ^2 test, revealed significant variations in the self-assessment of CH knowledge among different groups ($p < 0.001, \chi^2 = 77.546, df = 24$). After implementing the Bonferroni correction for multiple comparisons, the pattern observed for the general public in relation to specific CH knowledge remained consistent. However, the non-professional group's difference was only significantly discernible from the academic group (corrected $p < 0.05, \chi^2 = 26.735, df = 4$), with the variance to the general public being borderline and not statistically significant (corrected $p = 1.00, \chi^2 = 8.807, df = 4$). Significant differences were noted between the general public and the academic (corrected $p < 0.05, \chi^2 = 52.429, df = 4$), teacher (corrected $p < 0.05, \chi^2 = 16.624, df = 4$), and museum staff groups (corrected $p < 0.05, \chi^2 = 20.831, df = 4$), as well as between the academic and student groups (corrected $p < 0.05, \chi^2 = 22.189, df = 4$).

A significant portion of participants, 40.8% (199 out of 488), chose the mid-

point (3) on the Likert scale, typically considered a neutral position regarding CH knowledge. This outcome was anticipated given the scale's design.

At the scale's lower end, combining ratings 1 and 2, 17% (83 out of 488) of participants across all groups, with the exception of Museum Staff, expressed a lower level of CH knowledge. Notably, the General Public and Non-Professionals were expected to form the majority of this segment, and indeed, 22% of the General Public (55 out of 253) and 12% of Non-Professionals (17 out of 137) reported lower CH knowledge. Surprisingly, individuals from Academics, Professionals, and Teachers (1 each), alongside Students (8), also demonstrated lower levels of CH knowledge, prompting further examination into their learning roles and professional definitions provided in Phase 1 (see table 5.51 for detailed data).

Conversely, at the scale's upper end (ratings 4 and 5), 11.7% (57 out of 488) of participants from all groups indicated higher levels of CH knowledge. The Academics (88%, with 22 out of 25) and Museum Staff (90%, with 9 out of 10) confirmed the expected significant CH knowledge. The Non-Professionals (47%, with 64 out of 137) and Teachers (52%, with 13 out of 25) exhibited mid-range scores, suggesting comparable knowledge levels between Non-Professionals and paid professionals and that Teachers may have developed their expertise through curriculum-related materials. The General Public (34%, with 86 out of 253), Professionals (40%, with 2 out of 5), and Students (30%, with 10 out of 33) were identified at the lower spectrum of higher ratings, indicating considerable CH knowledge possibly tempered by a lack of confidence to rate higher.

The professionals again did not claim the highest levels but were threes and fours. The teachers and students are again spread across all options, with the majority claiming mid-level knowledge, but a few selected five in both categories.

Table 5.51: Responses to question Q10 "Rate your general Cultural Heritage knowledge" (Likert-like scale, 1 - low, 5 - high)

CH knowledge	**1 (Low)**		**2**		**3**		**4**		**5 (High)**		**Row Total**	
Groups	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
Academics	1	4	0	0	2	8	10	40	12	48	**25**	**100**
General Public	8	3	47	19	112	44	70	28	16	6	**253**	**100**
Museum Staff	0	0	0	0	1	10	5	50	4	40	**10**	**100**
Non-Professionals	3	2	14	10	56	41	49	36	15	11	**137**	**100**
Professionals	0	0	1	20	2	40	2	40	0	0	**5**	**100**
Students	1	3	7	21	15	45	7	21	3	9	**33**	**100**
Teachers	0	0	1	4	11	44	6	24	7	28	**25**	**100**
Column Total (n)	**13**		**70**		**199**		**149**		**57**		**488**	

5.3.6 Preferred Content Type and Technical Expertise

5.3.6.1 Question 4 - Please rank the following content types in order of your preferred preferences

Respondents were asked to rank their preference of types (text, pictures, video. audio, interactive) of content when accessing a museum website. The χ^2 tests show that there was nothing significant for any group against any content type:

- Content types videos across groups not significant ($p = 0.6$).
- Content types text across groups not significant ($p = 0.3$).
- Content types interactive across groups, not significant ($p = 0.14$).
- Content types audio across groups not significant ($p = 0.6$).
- Content types pictures across groups not significant ($p = 0.6$).

A χ^2 test of all the groups together also indicates that there is nothing significant in the data ($p = 0.99$). However, when analysing the data, it is clear to see that some interesting patterns emerge. Table 5.52 demonstrates with the colour coding that the preference of nearly all groups (with the exception of the museum staff group that preferred text), preferred pictures the most, followed by text then videos. Nearly all groups then preferred interactive elements and ranked audio as last. The

teacher group, however, was slightly different from the other groups as they preferred interactive text and videos on a pretty even level after text and put audio last. A possibility for this different preference is due to their role and their reason for being at the site, e.g. are they looking for material to show to a class of students or direct students to learn from independently in the future?

Table 5.52: Average scores of rankings of content types (the lower the number, the higher it was ranked by participants)

Group/Content type	Text	Pictures	Videos	Audio	Interactive
Academics	1.00	0.67	2.12	**2.92**	3.13
General Public	1.40	0.97	2.23	2.78	**2.48**
Museum Staff	0.60	0.90	2.60	**2.90**	3.00
Non-Professionals	1.35	0.93	2.28	**2.66**	2.76
Professionals	1.00	0.60	1.40	3.80	**3.20**
Students	1.45	1.27	1.88	2.72	**2.58**
Teachers	2.00	1.12	2.08	2.75	**1.92**
All	1.26	0.92	2.08	2.93	**2.72**

5.3.6.2 Question 11 - When seeking information on a website, which method do you prefer?

When interacting with web pages, people will generally adopt an approach that they are comfortable with and is usually based on social, learning or work experiences. The outcome of performing a χ^2 shows that there is nothing of significance in this data ($p = 0.14$) even though it is interesting that most prefer browsing to searching.

The participants were offered two options to select: browse or search. 63% of the participants across all groups selected browse as the preferred method of finding information on a website, while the remaining 37% opted for search. While it is clear that the majority of participants prefer to browse over search, this could be because the majority of knowledge levels average at about mid-level, and so they may not be aware of what is within the website.

When looking at the data per group, the Professionals group is the only group to prefer Search to Browse, while Museum Staff and Students are fairly evenly balanced across both options.

Table 5.53: Q11 - Search or Browse Preference by group

Preferred Info Seeking Method	**Browse**		**Search**		**Row Total**	
Groups	**n**	**%**	**n**	**%**	**n**	**%**
Academic	16	64	9	36	**25**	**100**
General Public	170	67	83	33	**253**	**100**
Museum Staff	5	50	5	50	**10**	**100**
Non-Professional	81	59	56	41	**137**	**100**
Professional	1	20	4	80	**5**	**100**
Student	17	52	16	48	**33**	**100**
Teacher	17	68	8	32	**25**	**100**
Column Total (n)	**307**		**181**		**488**	

Looking at the users' preferred information seeking techniques, Table 5.53 reveals that accessing navigational links is the preferred interaction method over search for all groups. One possible explanation is that these groups' lower CH and Domain knowledge makes it more difficult for them to phrase successful searches, therefore they prefer the guided nature of navigational links. Although browsing is preferred above searching by almost all groups.

5.3.6.3 Question 12 - When using the search box to search for content that you do not find in the first set of results, do you?

Expanding on Q10, this question sought to find out the habits of how participants engage with search features. While the data shows that the majority of participants will search a couple of times if they do not get the results they seek in the first set of search engine results pages (SERPS). However, the results of the chi^2 show nothing significant in this data ($p = 0.23$).

Table 5.54 illustrates search behaviours across different groups. A small percentage (9%) might give up quickly, but notably, over half (56%) of participants would attempt searching 2 or 3 additional times. A significant portion (35%) across all groups would persist until finding the needed material, except for the Professionals.

Table 5.54: Q12 - Search Persistence by group

Group	Academic		General Public		Museum Staff		Non-Professional		Professional		Student		Teacher		Row Total	
Search Persistence	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
Persist until you find the material/answer you were looking for	8	32	78	31	1	10	61	45	0	0	15	45	9	36	**172**	**35**
Give up	2	8	24	9	1	10	9	7	1	20	3	9	2	8	**42**	**9**
Search a further 2 or 3 times	15	60	151	60	8	80	67	49	4	80	15	45	14	56	**274**	**56**
Column Total (n)	**25**		**253**		**10**		**137**		**5**		**33**		**25**		**488**	
Columns Total (%)		**100**		**100**		**100**		**100**		**100**		**100**		**100**		**100**

The influence of CH knowledge on interactions with the search system is also visible in Table 5.54, which shows the groups' search use patterns. "Non-professional" users are significantly more likely ($p = 0.02, \chi^2 = 7.418, df = 2$) to persist with query reformulations until they successfully find what they are looking for compared to the general public users. It is likely that their higher CH knowledge allows them to modify their searches to successfully retrieve the information they are looking for.

5.3.6.4 Please rate your proficiency with the internet tools and services (1(low) - 5(high))

Engaging with technology and integrating it into every day activities, either through work, education or social experiences, is part of today's society. Understanding how proficient people are in terms of using internet tools and services across common social platforms and apps allows for gaining an understanding of the technical abilities of the participants when aligned to the various groups.

The Likert scale that was used for this question provided a range of possible choices from 1 to 5 with 1 representing either a "do not use" to a low proficiency level or 5 representing what could be coined as expert level at using the full range of features offered by the various platforms.

5.3.6.4.1 Facebook

Facebook, founded in 2004, is defined as an online social networking website where people can create profiles, share information about themselves, and respond to the information posted by others. It is one of the oldest social media platforms.

While Facebook proficiency across groups (see Table 5.55) is not significant ($p = 0.35$), what is interesting from looking at the data is the Teachers, Students, Professionals and Museum Staff groups rate themselves as highly proficient (rating 5), although Museum Staff does not rate below 4, while Academics, General Public, Non-Professionals group data shows comparable figures at both ends of the scale (1 and 5). Overall, 36% of all participants across all groups claim to be highly proficient with the platform.

Table 5.55: Q13 - General Internet Proficiency - Facebook by group

Facebook Proficiency	**1**		**2**		**3**		**4**		**5**		**Row Total**	
Groups	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
Academics	7	28	1	4	2	8	8	32	7	28	**25**	**100**
General Public	80	32	13	5	25	10	50	20	85	34	**253**	**100**
Museum Staff	0	0	0	0	0	0	4	40	6	60	**10**	**100**
Non-Professionals	40	29	8	6	16	12	24	18	49	36	**137**	**100**
Professionals	0	0	1	20	1	20	0	0	3	60	**5**	**100**
Students	4	12	2	6	5	15	7	21	15	45	**33**	**100**
Teachers	5	20	1	4	1	4	6	24	12	48	**25**	**100**
Column Total (n)	**136**		**26**		**50**		**99**		**177**		**488**	
Column Total (%)		**28**		**5**		**10**		**20**		**36**		**100**

5.3.6.4.2 Twitter

Twitter was founded in 2006 and is a social networking and micro-blogging online service that allows users to send and receive text-based messages or posts consisting of a restricted number of characters called "tweets."

Completing a χ^2 test of the Twitter proficiency across groups (see Table 5.56) suggests there is something significant ($p = 0.05$). The data reveals that the majority of participants (58%) rate themselves at the lowest end of the scale for all groups, while the remaining 42% are spread fairly evenly across the other scales.

The data shows that scale 3 was selected by 12% (60) of the participants across the majority of groups, with the exclusion of the Professionals group (0%). By considering the percentages of each group and ranking them in order, General Public and Non-Professionals are again first and second of the seven groups, while Students, Teachers and Academics remain in the middle but change places when compared to Facebook data. Museum Staff are ranked at the bottom of the groups.

The data also shows that at the higher end of the scale (5), the number of participants belonging to all groups has reduced to 11% (54). General Public and Museum Staff are in first and second place of the seven groups, although this may be due to these groups having more members. However, when this data is compared to the same rate for Facebook, a considerable difference (25%) can be seen and may indicate that the participants are a lot less comfortable with this platform.

However, nothing is showing as significant after conducting a series of pairwise χ^2 tests and correcting for multiple comparisons using the Bonferroni correction.

Table 5.56: Q13 - General Internet Proficiency - Twitter by group

Twitter Proficiency	**1**		**2**		**3**		**4**		**5**		**Row Total**	
Groups	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
Academics	15	60	1	4	2	8	1	4	6	24	**25**	**100**
General Public	152	60	26	10	28	11	21	8	26	10	**253**	**100**
Museum Staff	3	30	0	0	1	10	5	50	1	10	**10**	**100**
Non-Professionals	78	57	14	10	21	15	13	9	11	8	**137**	**100**
Professionals	3	60	0	0	0	0	0	0	2	40	**5**	**100**
Students	16	48	4	12	5	15	3	9	5	15	**33**	**100**
Teachers	14	56	3	12	3	12	2	8	3	12	**25**	**100**
Column Total (n)	**281**		**48**		**60**		**45**		**54**		**488**	
Column Total (%)		**58**		**10**		**12**		**9**		**11**		**100**

5.3.6.4.3 Instagram

Instagram was founded in 2010 and was acquired by Facebook in 2012. Instagram is a free, online photo-sharing application and social network platform.

Instagram proficiency across groups (see Table 5.57) suggested that there is something significant ($p < 0.001, \chi^2 = 57.189, df = 24$). Initial findings suggested differences in Instagram usage between the non-professional and general public groups (original $p = 0.021$), non-professional and student groups (original $p = 0.044$), and the teacher and museum staff groups (original $p = 0.023$). However, these differences were not considered significant after applying the Bonferroni correction. There was, however, a pronounced difference observed between groups general public and student groups (corrected $p < 0.05, \chi^2 = 21.087$, $df = 4$) and between the general public and museum staff groups (corrected $p = 0.05, \chi^2 = 21.574$, $df = 4$), indicating significant disparities in usage between these groups even after adjusting for

multiple comparisons.

The data shows that the majority of participants (71% or 346) across all groups rated themselves at the lowest end of the scale (1), while the distribution of the remaining participants fluctuates across scales 2 - 5, with scale 2 at 6%, scale 3 at 8% scale 4 at 5% and finally scale 5 at 11%.

A closer examination of the groups shows that the majority of General Public members, 79% (n = 200) gave themselves a scale of 1, with the remaining 53 participants being evenly spread across the other scale options with the exception of scale 5. Scale 5 is considered an expert level, of which 8% (n = 20) of General Public claimed to be. Similarly, Non-Professionals also had a substantial number of participants 67% (n = 92) rated at scale 1, the difference when comparing Non-Professionals to General Public is the spread of the remaining 45 participants in which the fluctuation occurs at every other scale e.g., scale 2 = 6%, scale 3 = 12%, scale 4 = 14% and scale 5 = 10%. The Academics, Students and Teachers groups are fairly equally placed in the middle of this lower scale, although it is surprising that the students have such a high number in this scale considering that most would have grown up with this platform. Museum Staff and Professionals are at the bottom of the seven groups, which is reflective of the general picture relating to engaging with this platform.

The proficiency level of the groups when interacting with this platform even at scale 5, expert, generally follows the same sequence as Twitter: General Public = 8%, Non-Professionals = 10%, Students = 27%, Academics = 20%, Museum Staff = 50% and Teachers = 4%, Professionals = 0%.

Table 5.57: Q13 - General Internet Proficiency - Instagram by group

Instagram Proficiency	**1**		**2**		**3**		**4**		**5**		**Row Total**	
Groups	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
Academics	13	52	2	8	3	12	2	8	5	20	**25**	**100**
General Public	200	79	13	5	10	4	10	4	20	8	**253**	**100**
Museum Staff	4	40	0	0	1	10	0	0	5	50	**10**	**100**
Non-Professionals	92	67	8	6	17	12	6	4	14	10	**137**	**100**
Professionals	3	60	1	20	0	0	1	20	0	0	**5**	**100**
Students	17	52	0	0	4	12	3	9	9	27	**33**	**100**
Teachers	17	68	3	12	3	12	1	4	1	4	**25**	**100**
Column Total (n)	**346**		**27**		**38**		**23**		**54**		**488**	
Column Total (%)		**71**		**6**		**8**		**5**		**11**		**100**

5.3.6.4.4 LinkedIn

LinkedIn was launched in 2003 as a social platform designed specifically for the business community.

Performing a chi proficiency test across groups (see Table 5.58) suggested that there is something significant ($p < 0.001, \chi^2 = 49.857, df = 24$). The data shows that the majority of the participants across all groups (58.4%), excluding Professionals, have rated themselves at the lowest end of proficiency.

The General Public predominantly rates at the lowest proficiency level (64%, n = 162). In contrast, Academics and Teachers tend to spread across the scale, with a notable 20% of Academics (n = 5) and 8% of Teachers (n = 2) at the highest proficiency. Museum Staff and Non-Professionals exhibit a bimodal distribution, with peaks at levels 1 and 3 for Museum Staff (40% each, n = 4) and levels 1 and 3 for Non-Professionals (56%, n = 77 and 19%, n = 26, respectively). Professionals, though a small sample (n = 5), show an even distribution across levels 2, 4, and 5. Students are primarily at the lower end of proficiency (52%, n = 17 at level 1). Overall, the majority of responses across all groups fall into the lowest proficiency category (58%, n = 285).

Again when comparing this platform with the previous platforms, the majority of participants are clearly not as proficient with LinkedIn as the other platforms.

Table 5.58: Q13 - General Internet Proficiency - LinkedIn by group

LinkedIn Proficiency	**1**		**2**		**3**		**4**		**5**		**Row Total**	
Groups	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
Academics	7	28	6	24	4	16	3	12	5	20	**25**	**100**
General Public	162	64	19	8	30	12	24	9	18	7	**253**	**100**
Museum Staff	4	40	1	10	4	40	1	10	0	0	**10**	**100**
Non-Professionals	77	56	14	10	26	19	16	12	4	3	**137**	**100**
Professionals	0	0	2	40	0	0	2	40	1	20	**5**	**100**
Students	17	52	6	18	5	15	3	9	2	6	**33**	**100**
Teachers	18	72	1	4	2	8	2	8	2	8	**25**	**100**
Column Total (n)	**285**		**49**		**71**		**51**		**32**		**488**	
Column Total (%)		**58**		**10**		**15**		**10**		**7**		**100**

After conducting a series of pairwise χ^2 tests and applying the Bonferroni correction for 21 comparisons, significant differences were observed between the academic group and the general public ($p < 0.05, \chi^2 = 16.771, df = 4$). Also, a significant difference was found between the non-professional and academic groups ($p < 0.05, \chi^2 = 17.701, df = 4$).

Given the work role nature of the academic group, it is understandable why this would differ from the non-work related groups on LinkedIn.

5.3.6.4.5 Online Banking

Online banking enables customers of a bank to conduct a range of transactions through their bank's website and which, according to the Office for National Statistic (Gov UK, 2019), is utilised by the majority of the population, although there is some decline in the 80+ age bracket.

Performing the chi proficiency test across groups suggests there is nothing significant ($p = 0.9$). The data (see Table 5.59) shows that The General Public predominantly reports the highest proficiency (46%, n = 117), followed by Non-Professionals with 41% (n = 56) at level 5. Academics and Teachers show the greatest proficiency at level 5 (32%, n = 8 and 44%, n = 11, respectively), while Museum Staff demonstrates a concentration at level 4 (40%, n = 4). Students and Professionals exhibit a more balanced distribution across proficiency levels, with the highest percentages at level 5 for both groups (33%, n = 11 and 4%, n = 2, respectively). The majority of responses across all groups are at the highest proficiency level (43%, n = 208),

while the lowest proficiency level has the fewest (18%, n = 89).

Table 5.59: Q13 - General Internet Proficiency - Online Banking by group

Banking Proficiency	**1**		**2**		**3**		**4**		**5**		**Row Total**	
Groups	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
Academics	4	16	2	8	4	16	7	28	8	32	**25**	**100**
General Public	44	17	12	5	25	10	55	22	117	46	**253**	**100**
Museum Staff	1	10	0	0	2	20	4	40	3	30	**10**	**100**
Non-Professionals	29	21	9	7	15	11	28	20	56	41	**137**	**100**
Professionals	1	20	0	0	1	20	1	20	2	40	**5**	**100**
Students	7	21	1	3	7	21	7	21	11	33	**33**	**100**
Teachers	3	12	1	4	3	12	7	28	11	44	**25**	**100**
Column Total (n)	**89**		**25**		**57**		**109**		**208**		**488**	
Column Total (%)		**18**		**5**		**12**		**22**		**43**		**100**

5.3.6.4.6 Online Shopping

Online shopping allows for the purchasing of goods and services over the internet through the use of a web browser. According to Coppola (2021) the older the age group, the fewer online purchases are likely to be made, which is fairly reflective of the data found regarding online banking (see 5.3.6.4.5) and is likely to be in response to cognitive decline.

The chi test performed indicated that the Online Shopping proficiency across groups suggests there is nothing significant $p = 0.4$). The data (see Table 5.60) shows that the General Public has the highest number of proficient users, with 48% (n = 121) rating themselves at level 5. Museum Staff, while a smaller group, has 60% (n = 6) reporting the highest proficiency. Non-Professionals also have a significant portion at level 5 (42%, n = 58). For Academics and Teachers, the largest group rates their proficiency at level 4 (44%, n = 11 and 20%, n = 5, respectively) and level 5 (24%, n = 6 and 56%, n = 14). Professionals are evenly split between levels 3 and 5 (40%, n = 2 each). Students show a distributed proficiency, with the highest concentration at level 4 (30%, n = 10). The overall majority (44%, n = 216) consider themselves highly proficient in online shopping.

Table 5.60: Q13 - General Internet Proficiency - Online Shopping by group

Online Shopping Proficiency	**1**		**2**		**3**		**4**		**5**		**Row Total**	
Groups	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
Academics	2	8	1	4	5	20	11	44	6	24	**25**	**100**
General Public	16	6	16	6	42	17	58	23	121	48	**253**	**100**
Museum Staff	0	0	0	0	1	10	3	30	6	60	**10**	**100**
Non-Professionals	8	6	12	9	31	23	28	20	58	42	**137**	**100**
Professionals	1	20	0	0	2	40	0	0	2	40	**5**	**100**
Students	4	12	3	9	7	21	10	30	9	27	**33**	**100**
Teachers	1	4	2	8	3	12	5	20	14	56	**25**	**100**
Column Total (n)	**32**		**34**		**91**		**115**		**216**		**488**	
Column Total (%)		**7**		**7**		**19**		**24**		**44**		**100**

5.3.6.4.7 Online Music

Streaming music is a method of feeding audio content to an electronic device, e.g., a mobile phone.

The χ^2 test performed against Streaming Music proficiency across groups suggests there is something significant ($p = 0.03, \chi^2 = 37.505, df = 24$). However, there is nothing of significance after conducting a series of pairwise tests and applying the Bonferroni correction to adjust for multiple comparisons.

The data (see Table 5.61) shows that the General Public has the largest proportion rating themselves at the lowest proficiency level (48%, n = 121). Academics and Teachers show a relatively even distribution across all proficiency levels, with a slight leaning towards higher proficiency (28% of Academics and 28% of Teachers at level 5). Non-Professionals also reported lower proficiency levels, with 45% (n = 61) at level 1. Both Museum Staff and Professionals have small sample sizes but show a spread across the proficiency levels, with Museum Staff indicating a higher proficiency (30% at level 5). Students indicated a greater proficiency, with 39% (n = 13) rating themselves at the highest level. Overall, the majority of participants across all groups rate themselves at the lowest level of streaming music proficiency (42%, n = 207), while fewer consider themselves highly proficient (19%, n = 93).

Table 5.61: Q13 - General Internet Proficiency - Streaming Online Music by group

Streaming Music Proficiency	**1**		**2**		**3**		**4**		**5**		**Row Total**	
Groups	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
Academics	10	40	1	4	6	24	1	4	7	28	**25**	**100**
General Public	121	48	28	11	40	16	25	10	39	15	**253**	**100**
Museum Staff	1	10	2	20	2	20	2	20	3	30	**10**	**100**
Non-Professionals	61	45	24	18	18	13	11	8	23	17	**137**	**100**
Professionals	1	20	2	40	1	20	0	0	1	20	**5**	**100**
Students	7	21	6	18	3	9	4	12	13	39	**33**	**100**
Teachers	6	24	3	12	6	24	3	12	7	28	**25**	**100**
Column Total (n)	**207**		**66**		**76**		**46**		**93**		**488**	
Column Total (%)		**42**		**14**		**16**		**9**		**19**		**100**

5.3.6.4.8 Online Video

Streaming is the continuous transmission of video files from a server to a client. It enables users to view videos on their personal devices, such as mobile phones and laptops, without having to download them.

Performing a chi test on the streaming video proficiency across groups suggests there is nothing significant ($p = 0.53$).

The data (see Table 5.62) shows that the majority of the General Public (38%, n = 95) report the highest proficiency, and a similar trend is observed among Non-Professionals (31%, n = 42) and Teachers (40%, n = 10). Museum Staff tend to rate themselves towards the higher proficiency levels (40% at both levels 4 and 5). Academics' responses are distributed, with the highest concentration at level 5 (36%, n = 9). Students predominantly consider themselves highly proficient (64%, n = 21 at level 5). Professionals have the least variation, with equal distribution at levels 4 and 5 (40% for each). Overall, 38% of the total responses indicate high proficiency in streaming video.

Table 5.62: Q13 - General Internet Proficiency - Streaming Online Video by group

Streaming Video	**1**		**2**		**3**		**4**		**5**		**Row Total**	
Groups	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
Academics	4	16	2	8	4	16	6	24	9	36	**25**	**100**
General Public	42	17	22	9	41	16	53	21	95	38	**253**	**100**
Museum Staff	1	10	0	0	1	10	4	40	4	40	**10**	**100**
Non-Professionals	24	18	9	7	21	15	41	30	42	31	**137**	**100**
Professionals	0	0	1	20	0	0	2	40	2	40	**5**	**100**
Students	1	3	1	3	4	12	6	18	21	64	**33**	**100**
Teachers	3	12	2	8	3	12	7	28	10	40	**25**	**100**
Column Total (n)	**75**		**37**		**74**		**119**		**183**		**488**	
Column Total (%)		**15**		**8**		**15**		**24**		**38**		**100**

5.3.6.4.9 Online Documents

An online document is generally a digital version of a print-based document used to facilitate or share information between parties. E-documents take the place of a physical document and generally serve the same purpose but in a digital format that is accessible from technological devices, e.g., Google Docs, Microsoft 365.

Performing the chi test against the Online documents proficiency across groups (see Table 5.63) suggests there is nothing significant ($p = 0.19$).

Academics show the highest proficiency with 56% (n = 14) rating themselves at level 5. The General Public presents a broad distribution but is most proficient at level 5 (29%, n = 73). Museum Staff, while a smaller group, are evenly distributed across levels 2 to 5. Non-Professionals also have a relatively even spread, yet the highest proportion is at level 5 (28%, n = 39). Both Professionals and Students show a tendency towards higher proficiency, with 40% of Professionals (n = 2) and 30% of Students (n = 10) at level 5. Teachers, similar to Academics, have a substantial number rating high proficiency (40%, n = 10 at level 5). The overall trend suggests a decent level of confidence in using online documents, with the majority across all groups rated at the higher proficiency levels (levels 4 and 5 combined constitute 48% of responses).

Table 5.63: Q13 - General Internet Proficiency - Using Online Documents by group

Using Online Documents	**1**		**2**		**3**		**4**		**5**		**Row Total**	
Groups	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
Academics	3	12	1	4	3	12	4	16	14	56	**25**	**100**
General Public	66	26	23	9	55	22	36	14	73	29	**253**	**100**
Museum Staff	0	0	3	30	2	20	3	30	2	20	**10**	**100**
Non-Professionals	29	21	9	7	33	24	27	20	39	28	**137**	**100**
Professionals	1	20	0	0	0	0	2	40	2	40	**5**	**100**
Students	6	18	2	6	7	21	8	24	10	30	**33**	**100**
Teachers	3	12	1	4	6	24	5	20	10	40	**25**	**100**
Column Total (n)	**108**		**39**		**106**		**85**		**150**		**488**	
Column Total (%)		**22**		**8**		**22**		**17**		**31**		**100**

5.3.7 Other Data gathered by the Poll Daddy System

As stated earlier (see Section 5.2.8) the Poll Daddy System captured additional data when a participant completed the online survey. This section analyses this data and where possible compares it to relevant survey questions. Some of the data appears incomplete due to the ability of some participants devices/browsers or operating systems blocking data collection of certain data, e.g. the device's operating system. Therefore, some of the data referred to in this section is based on the available Poll Daddy collected data of 292 participants.

5.3.7.1 Museum Website Entered the Survey From

NML museum consists of a number of separate places of interest around the Liverpool area. The data collected (see Table 5.64 and Appendix H.3) by the Poll Daddy System showed the starting point of each of the participant's online journey with the website. It captured this as a referrer URL to the survey system. The referrer URL is the http URL of the page the participant was on when they clicked the take survey pop-up button.

Performing the χ^2 result is $p < 0.01, \chi^2 = 1.925, df = 30$ revealed that there is something significant in this data.

Table 5.64: Entry Museum Website by group

Entry Museum/Gallery	Groups	Academics	General Public	Museum Staff	Non-Professionals	Professionals	Students	Teachers	Row Total (n)	Row Total (%)
International Slavery Museum (ism)	**n**	8	17	2	16	1	11	2	**57**	
	%	32	7	20	12	20	33	8		**12**
Lady Lever Art Gallery (ladylever)	**n**	3	24	0	13	1	0	1	**42**	
	%	12	9	0	9	20	0	4		**9**
Maritime Museum (maritime)	**n**	2	47	1	34	2	3	4	**93**	
	%	8	19	10	25	40	9	16		**19**
Sudley House (sudley)	**n**	0	13	0	7	0	2	0	**22**	
	%	0	5	0	5	0	6	0		**5**
Walker Art Gallery (walker)	**n**	7	93	4	43	1	9	10	**167**	
	%	28	37	40	31	20	27	40		**34**
World Museum Liverpool (wml)	**n**	5	59	3	24	0	8	8	**107**	
	%	20	23	30	18	0	24	32		**22**
Column Total	**n**	**25**	**253**	**10**	**137**	**5**	**33**	**25**	**488**	
	%	**100**	**100**	**100**	**100**	**100**	**100**	**100**		**100**

The data revealed that of the 6 places of interest, the Walker Art Gallery has the highest overall visitation (34%, n = 167), with the General Public constituting the largest visiting group (37%, n = 93). The Maritime Museum (maritime) also shows a notable visitation rate at 19% (n = 93), with high attendance from the General Public (19%, n = 47). The International Slavery Museum (ism) and the World Museum Liverpool (wml) have similar overall visitation rates (12% and 22%, respectively), but the Academic group shows a higher preference for the International Slavery Museum (32%, n = 8).

The Lady Lever Art Gallery (ladylever) and Sudley House (sudley) have the lowest visitation rates at 9% (n = 42) and 5% (n = 22), respectively, with Sudley House not being visited by Museum Staff, Professionals, or Teachers. Notably, Museum Staff show the highest visitation percentage to the Walker Art Gallery (40%, n = 4).

Overall, each museum/gallery's visitation reflects a diverse range of interests from the different groups, with General Public consistently representing a significant proportion of visitors across all venues. The column totals confirm that every group's responses sum up to 100%, providing a complete picture of the preferences within the sample size of 488 respondents.

Removing the general public and non-professionals groups and excluding Sudley House and the Lady Lever Gallery data presents a more evenly distributed data spread across all locations, with Academics, Students and Teachers aligning fairly closely across the remaining 4 locations. Museum staff and professionals are also fairly evenly distributed across the remaining locations. Still, professionals again have no presence at one of the sites (WML), which causes the overage average to decrease significantly compared to the museum staff.

After conducting a series of pairwise χ^2 tests and applying the Bonferroni correction to adjust for 21 comparisons, thereby reducing the risk of type I errors, only two pairs showed significant differences: the general public and the students (corrected $p < 0.01, \chi^2 = 27.205, df = 5$) and the general public and the academic groups (corrected $p < 0.05, \chi^2 = 19.640, df = 5$).

5.3.7.2 Page Type Entered the Survey From

Performing the χ^2 test yielded a result of $p < 0.001, \chi^2 = 142.647, df = 60$, indicating significant differences within the data. Subsequent pairwise χ^2 tests, adjusted for multiple comparisons using the Bonferroni correction, revealed significant differences between the general public and the professional groups, academic, non-professional, and student groups, all with $p - values < 0.05, df = 9$.

The data (see Table 5.65) shows that the majority of participants across all groups arrived at the survey more from the Exhibition and Landing (general information) pages than any of the collection related pages (Curated, Collection and Archive). However, academics buck this trend by coming more from collection related pages rather than any other type of page. The non-professionals tend to have an even spread across the collection and general information type pages.

Table 5.65: Entry Page Type by Group

Entry Page Type	**Archive**		**Collection**		**Competition**		**Contact**		**Curated**		**Education**		**Events**		**Exhibition**		**Landing**		**Talks**		**Visit**		**Row Total**	
Groups	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
Academics	0	0	6	24	0	0	0	0	6	24	1	4	4	16	3	12	3	12	1	4	1	4	**25**	**5**
General Public	1	0	14	6	3	1	3	1	3	1	0	0	51	20	56	22	68	27	0	0	54	21	**253**	**52**
Museum Staff	0	0	2	20	0	0	0	0	0	0	0	0	4	40	2	20	1	10	0	0	1	10	**10**	**2**
Non-Professionals	7	5	21	15	1	1	0	0	8	6	0	0	21	15	34	25	24	18	2	1	19	14	**137**	**28**
Professionals	0	0	1	20	0	0	0	0	0	0	0	0	0	0	1	20	1	20	1	20	1	20	**5**	**1**
Students	0	0	1	3	1	3	0	0	4	12	0	0	4	12	2	6	16	48	1	3	4	12	**33**	**7**
Teachers	0	0	2	8	0	0	0	0	2	8	1	4	6	24	4	16	5	20	0	0	5	20	**25**	**5**
Column Total (n)	**8**		**47**		**5**		**3**		**23**		**2**		**90**		**102**		**118**		**5**		**85**		**488**	
Column Total (%)		**2**		**10**		**1**		**1**		**5**		**0**		**18**		**21**		**24**		**1**		**17**		**100**

5.3.7.3 Detailed Locations

Performing the χ^2 test produced a result of $p < 0.001, \chi^2 = 257.757, df = 174$, indicating significant differences within the data. Further analysis through a series of pairwise χ^2 tests, with adjustments for multiple comparisons using the Bonferroni correction, revealed significant differences between the professional and both the non-professional (corrected $p < 0.001, \chi^2 = 48.582, df = 16$) and general public (corrected $p < 0.001, \chi^2 = 68.807, df16$) groups, as well as between the general public and academic groups (corrected $p < 0.001, \chi^2 = 50.983, df = 16$).

The data (see Table 5.66) showed that the United Kingdom has the highest number of respondents (n = 389), with a significant proportion from the General Public (87%, n = 220).

A closer examination of the data shows that 87% of the general public users and 74% of non-professional users are from the UK. The majority of academics are from outside of the UK (48% = UK, 24% = USA, 16% Europe and 8% are unknown) as are the professionals (20% = UK, 16% = Europe, 4% = Rest of the world (ROW)). Museum staff are entirely from the UK (100%). The majority of students and teachers are based in the UK (76% and 80% respectively); however, students 15% of participants are from the Rest of the World and none from Europe, while teachers have 8% from Europe and the Rest of the World while both groups also claiming 3% and 4% respectively of Unknowns.

Table 5.66: Countries of respondents by group

Country	Groups	Academics	General Public	Museum Staff	Non-Professionals	Professionals	Students	Teachers	Row Total
Australia	**n**	0	1	0	1	0	0	0	**2**
	%	0	0	0	1	0	0	0	**0**
Austria	**n**	0	0	0	1	0	0	0	**1**
	%	0	0	0	1	0	0	0	**0**
Canada	**n**	1	1	0	3	1	0	0	**6**
	%	4	0	0	2	20	0	0	**1**
Chile	**n**	0	0	0	1	0	0	0	**1**
	%	0	0	0	1	0	0	0	**0**
Cyprus	**n**	0	1	0	0	0	0	0	**1**
	%	0	0	0	0	0	0	0	**0**
Denmark	**n**	1	0	0	1	0	0	0	**2**
	%	4	0	0	1	0	0	0	**0**
Ethiopia	**n**	0	0	0	0	0	1	0	**1**
	%	0	0	0	0	0	3	0	**0**
Finland	**n**	0	1	0	0	0	0	0	**1**
	%	0	0	0	0	0	0	0	**0**
France	**n**	1	1	0	0	0	0	0	**2**
	%	4	0	0	0	0	0	0	**0**
Germany	**n**	0	1	0	0	0	0	0	**1**
	%	0	0	0	0	0	0	0	**0**
Greece	**n**	0	0	0	1	0	0	0	**1**
	%	0	0	0	1	0	0	0	**0**
Ireland	**n**	0	2	0	1	1	0	0	**4**
	%	0	1	0	1	20	0	0	**1**
Italy	**n**	0	0	0	1	0	0	1	**2**
	%	0	0	0	1	0	0	4	**0**
Jamaica	**n**	0	0	0	0	0	1	0	**1**
	%	0	0	0	0	0	3	0	**0**
Japan	**n**	0	0	0	1	0	0	0	**1**
	%	0	0	0	1	0	0	0	**0**
Malaysia	**n**	0	0	0	0	0	1	0	**1**
	%	0	0	0	0	0	3	0	**0**
Netherlands	**n**	0	0	0	3	0	0	0	**3**
	%	0	0	0	2	0	0	0	**1**
New Zealand	**n**	0	2	0	0	0	0	0	**2**
	%	0	1	0	0	0	0	0	**0**
Nigeria	**n**	0	0	0	0	0	1	0	**1**
	%	0	0	0	0	0	3	0	**0**
Norway	**n**	0	0	0	1	0	0	0	**1**
	%	0	0	0	1	0	0	0	**0**
Poland	**n**	0	1	0	0	0	0	0	**1**

Table 5.66 continued from previous page

Country	Groups	Academics	General Public	Museum Staff	Non-Professionals	Professionals	Students	Teachers	Row Total
	%	0	0	0	0	0	0	0	**0**
Portugal	**n**	0	1	0	0	1	0	0	**2**
	%	0	0	0	0	20	0	0	**0**
Spain	**n**	2	2	0	2	0	0	1	**7**
	%	8	1	0	1	0	0	4	**1**
Sweden	**n**	0	0	0	1	0	0	0	**1**
	%	0	0	0	1	0	0	0	**0**
Switzerland	**n**	0	1	0	0	0	0	0	**1**
	%	0	0	0	0	0	0	0	**0**
Thailand	**n**	0	2	0	0	0	0	0	**2**
	%	0	1	0	0	0	0	0	**0**
Trinidad and Tobago	**n**	0	0	0	0	0	1	0	**1**
	%	0	0	0	0	0	3	0	**0**
United Kingdom	**n**	12	220	10	101	1	25	20	**389**
	%	48	87	100	74	20	76	80	**80**
United States	**n**	6	14	0	14	1	2	2	**39**
	%	24	6	0	10	20	6	8	**8**
Unknown	**n**	2	2	0	4	0	1	1	**10**
	%	8	1	0	3	0	3	4	**2**
Column Total	n	**25**	**253**	**10**	**137**	**5**	**33**	**25**	**488**
	%	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**

5.4 Discussion

The formulation of the definitions/characteristics for each label was achieved through the interpretation of the participants' answers to the survey questions (see Appendix H Section H.4). This produced definitions that covered areas of motivation and task, engagement, technical expertise, domain and cultural heritage knowledge, shared experience, usage and demographics. However, when these Phase 2 definitions were compared to the definitions found or established in the literature or formulated in Phase 1, a lack of clarity and confusion was evident, particularly across the professionally focused labels.

Examining the traits that can be extrapolated from this characterisation of an Academic label and compared to the Phase 1 definition, revealed similarities in the areas of education, reason and purpose of the visit, domain and CH knowledge and

some technology proficiency (browsing over searching and general organisational website usage e.g., online banking, shopping). However, significant gaps were also apparent in areas concerned with technical expertise, usage and engagement. Additionally, the literature review and Phase 1 provided little by way of demographic detail relating to those who took part in those original studies, preferring to use generalised terms such as "student". An examination of the paper by Srinivasan et al. (2009) revealed that "student" was composed of "a group of masters students in the Dept of Information Studies at UCLA and a group of Inuit high-school students at Inukshuk High School" although the reality of the situation would dictate that masters students would be more mature and knowledgeable than high school students.

Similarly, examining the definitions from the literature and both the Phase 1 and Phase 2 (see Appendix F) for the General Public, revealed similarities in the areas of age and gender, reason for visit, purpose of visit, and domain knowledge. Comparisons were also noted in the areas of web page proficiency (general organisational website usage e.g., online banking, shopping) and the number of pages that may be visited on a specific website. However, significant gaps were also apparent in areas concerned with demographic details e.g., Education levels, Location, employment status; engagement, technical expertise, usage and CH knowledge. The diversity of this group is reflective of the variety of labels that appeared in the literature e.g. General Public (Schouten, 1995), General User (Sarraf, 1999), General Visitor (Booth, 1998), Experienced User (Cifter and Dong, 2008), laymen (Schweibenz, 2004), Incidental users (Górny and Mazurek, 2012), Potential visitor (Cunliffe et al., 2001), spectators (Li et al., 2007), surfers (Marchionini et al., 2003), Tentative observers (Templeton, 2011), wider-public (Hampson et al., 2012). Nevertheless, it is this breadth of variety in relation to motivation and task, shared experience, engagement, Technical expertise, domain and CH knowledge that prevents the recognition of an operational definition of what constitutes a General Public user label.

Comparing the Museum Staff label definition from this phase to the literature and Phase 1 also indicated a variety of different labels that were being coined by the authors to meet the specific needs of their publication e.g., Museum Professional (Marty, 2007a), senior staff (Nicholas et al., 2007), curator (Schouten, 1995), Museum curator (Gilliland-Swetland and White, 2004), Museum Information Profes-

sional (Marty, 2008). The common characteristic of working for a museum binding them together. Similarities can be found between the Phase 1 and Phase 2 definitions in relation to elements of motivation and task and engagement, with a minor reference to demographics (employment status) but gaps occur when looking at areas such as technical expertise, which can vary from possessing a basic rudimentary understanding to being highly proficient in a specific system; to being knowledgeable in a narrow area of focus e.g., a specific field domain and CH knowledge and as such preferring text as the medium to convey knowledge. This sometimes contradictory expansion of the definition, therefore, again prevents the general operationalisation of the label.

The comparison of the Non-Professionals group label to the literature and Phase 1 also revealed similarities with the Novice group label (Hsieh-Yee, 1993) in terms of motivation and task, although the survey indicates a willingness to perform a limited search before giving up; technical expertise in terms of using search techniques but still preferring pictures to text and scanning rather than reading; and possessing some level of domain and CH knowledge although not extensive. While these areas and others, e.g., demographic and usage, have been expanded, there remains a lack of clarity as to what a Non-Professionals label looks like in order to be operationalised when there is a strong resemblance to the traits of other labels, e.g., General Public and Novice.

The Professional label offered a number of possible similar derivatives that have been found in the literature for example, the Hobbyist and Professional label used by Falk (2009); Skov (2009); Stebbins (2007); Spellerberg et al. (2016), all of which present a definition of a label that has good domain knowledge and possess a technical proficiency at using appropriate systems to find what they want, which aligns to the findings of the Phase 2 survey. However, some elements of the Phase 2 findings offer expansion or contradiction on the composition of the label. For example, the survey indicates a varying web page usage proficiency, ranging from mixed ability for social platforms to high levels of proficiency when dealing with other online interactions such as banking or shopping; Technical expertise also varies with a preference of using a PC over more mobile technology and using keyword searching but abandoning the process of finding what they want if this is not achieved relatively quickly. There is also evidence that, similar to the other group labels, scanning

rather than reading and pictures over text is the preferred option. The overall result therefore being presented for a definition of this label is one of mixed messages and confusion.

The Student label has frequently appeared in the literature e.g., Paterno and Mancini (1999); Paternò and Mancini (2000), art student, followed by Marchionini et al. (2003) student(K-16), Nicholas et al. (2007) using PhD-student and Postgraduate student and finally Duff et al. (2012) who used Uni Student. In three of these instances, no definition regarding the characteristics of these labels was provided. However, the label of student is generally assigned to someone who is studying at university, and depending on what level of study they are at, an expectation of acquired knowledge and understanding can be made e.g., a PhD student would possess more subject related knowledge than a first year undergraduate student. Yet, when comparing the findings from literature and Phase 1 with Phase 2 areas of uncertainty can be seen. For example, similarities exist in areas of web page proficiency, elements of motivation and task and a relatively novice to mid-range level of domain knowledge and CH, but the gaps are significant. Phase 2 expands the original definition to include areas of demographics where it was once presumed, levels of technical expertise, both in terms of the how and what technology is being used, and a clearer understanding of the level of domain and CH knowledge that can be attributed to this group.

Similarly, the literature and Phase 1 found a few variations of the Teacher label, e.g., educational visitor (Skov and Ingwersen, 2008), School Enquirer (Cunliffe et al., 2001), but essentially the traits behind this label are focused towards a learning experience. It was found in this phase, that there is evidence of some alignment to areas such as motivation and task, elements of engagement and mid to high levels of domain knowledge, and an indication of some level of confidence when dealing with specific technology systems. However, it was also found that gaps exist in these same areas and in terms of demographics, technical expertise, CH knowledge.

It's important to acknowledge that while definitions for each group were extracted at this stage, the manual analysis required significant time and learning of analysis techniques, which may not be feasible for untrained museum practitioners needing quick actionable results. The definitions produced here represent an improvement over those found in the literature. However, for them to be action-

able and usable by other DCH researchers, the groups must be clearly defined and distinguishable from one another. Despite efforts the definitions from this phase exhibit some overlap between groups; for instance, the Museum Staff and Professionals share similarities in employment, education, and web proficiency. Similarly, Non-Professionals and Students exhibit overlapping demographics and motivations for website visits. While perfect uniqueness among groups isn't expected, clearer distinctions would be advantageous.

Therefore, while the survey provided valuable additional information for potential group labels, it also revealed uncertainty about the characteristics of each label. If even professional-based groups like Academics, Teachers, Museum Staff, and Professionals cannot agree on their self-categories, it's likely that UX or web development professionals would also have differing interpretations due to the lack of clear and consistent definitions. This ambiguity makes it challenging to operationalise the user groups effectively, potentially leading to confusion and disuse. To address this, a transaction log analysis will be conducted to identify these groups within server logs and potentially uncover additional behaviours for refining and differentiating these definitions further.

5.5 Summary

The literature review and Phase 1 (see Chapters 2 and 4) highlighted the differing interpretations of a website user group label that was used in a publication. Often the same label was used with a differing definition behind it, or the definitions were similar, but unique labels used, and has meant that any intention of creating a standardised usage of these labels would not be operationable.

The majority of previous research (see Chapter 2) and Phase 1 (Chapter 4) into the users of DCH websites has focused on those user groups that are easier to access (academics, museum staff, students and professionals), and potentially the easiest by researchers to comprehend the characteristics of the user groups. However, as the results of this survey show (see Table 5.34), they form only a small fraction of the total number of website visitors. The main user groups identified from this survey are the "general public" and "non-professional" visitors, who make up nearly 80% of all visitors. The survey data also allowed the definition of the criteria (motivation, task,

engagement, domain knowledge, and location) that distinguish these two groups from the other groups and also the criteria (domain knowledge and location) that distinguish the two groups from each other.

Also, contrary to the literature which is almost predominantly focused on searching and search systems the survey identified a preference in almost all the groups for browsing, suggesting that Domain knowledge is not a driver for search.

The differences between these two groups (General Public and Non-professional) , and those groups that have been studied more frequently (Academics, Students, MIP's) fall into two main categories: why they are visiting NML's website and how much CH and domain knowledge/experience they have. Additionally, location distinguishes users, with both groups more likely to be local to the museum's location (see Table 5.40). Both the General Public and the Non-Professional have a primarily personal reason for visiting NML. What is particularly interesting is that there is a significant fraction of the General Public, and a smaller subgroup, similar to the "loungers" (Hein, 1998), within the Non-Professionals who are visiting the website purely to pass the time.

Other notable characteristics of the General Public and the Non-Professional groups include lower levels of CH knowledge and competence, which may need the use of additional domain knowledge assistance in their interactions. While there is no difference in overall CH knowledge levels between both groups, when questioned about their specific visit to the NML website, the "General Public" reported a lower degree of knowledge than the "Non-Professionals". This corresponds nicely with the "Non-Professional" group's increased concentration on museum material, as users who come for specific content are more likely to be educated about what they are looking for. It is also supported by the responses to their search behaviour, where significantly more "Non-Professional" users reported that they would persist in searching until they found the information they were seeking. This indicates that they may be more familiar with CH search systems and understand that the desired information is often hidden deep in the search results, or that their increased CH knowledge means that they have the ability to develop more complex and numerous search terms to find what they are looking for. While in contrast this may also explain why DCH websites suffer such high bounce rates, as based on the survey results, those users who leave immediately are more likely to belong to the "general

public" and "non-professional" user groups, as the initial analysis of their responses indicates that they are more likely to give up quickly.

The findings of the survey also suggests that people self-classifying as non-professionals in this case, are closer to general public as opposed to previous research which showed them close to experts. This could be due to users having a different interpretation of the non-professional label; however this is unlikely due to the late sample sizes. This therefore suggests that when designing for domain knowledge, non-professionals should be treated like the general public.

This phase of work has demonstrated that users of a system can can be effectively studied across all information-seeking characteristics, which are typically studied in isolation. The survey design can be reused by others to gather comparable data. The derived definitions represent significant improvement over those found in the literature and are also comprehensive without gaps. However, the definitions based on the labels do show some significant overlap, indicating that further work is required to provide clarity.

While the survey analysis provides an initial view of NML website user characteristics, understanding how different user groups interact with the website, identifying patterns, and improving user support across various tasks and goals require further investigation. Matching the behaviours identified in the survey (Phase 5) with web server logs (Phase 6); which captures all NML website use, will offer insights into the broader audience and potentially identify additional behaviours linked to the group definitions.

Chapter 6

PHASE 3: DESCRIPTIVE WEBLOG ANALYSIS OF USERS FROM NML

6.1 Introduction

During Phase 2, the online NML user categories were determined through the responses from a comprehensive online survey. User groups were differentiated by factors including motivation, task, engagement, search strategy, domain knowledge and location. Findings revealed that the frequently understudied "general public" and "non-professional" groups made up the majority of users.

Phase 3 aims to deepen the understanding of NML users, exploring Phase 2 questions from a different perspective and ensuring the comprehensiveness of the survey used previously. Specifically, it focuses on studying NML user engagement and behaviour through the analysis of user-system interactions (server logs) (Q6, Q7, Q8, Q14, Q15 from the survey). While direct log traces of survey participants cannot be extracted, transaction logs covering the same time period (1/2/2017 to 1/3/2017) plus and minus one month were utilised. This ensures the inclusion of all survey respondents in the log data, validating survey findings and allowing for identification of characteristics within the log data. However, due to the larger number of users in the logs, there is a possibility that the survey sample might be overshadowed by the data or that the sample is indeed representative of the

broader audience. Analysing user engagement with the website and search systems can help identify the information users seek, how they navigate to it, and reveal any find-ability or user experience issues.

The main goal of this chapter is to analyse the server logs of the NML website in order identify and gain a better understanding of the behaviours of all live user interactions by answering the following research question:

Phase 3 Research Question:

RQ5: How do user groups identified in the transaction logs compare to those identified in the literature and survey results?

Data collection for this phase involved gathering the server log data from NML. The data was cleaned to ensure the removal of bots, crawlers and developer validation tools (see Chapter 3 Methodology, Section 3.4.3.3 and Section 3.4.3.4). After pre-processing and sessionisation the transaction log data resulted in a total dataset of 552,323 rows of page requests.

The remainder of the Chapter is organised as follows: Section 6.2 Analysing the Log File Data; Section 6.3 Clustering of the Transaction Logs; Section 6.4 Discussion and finally Section 6.4.1 Summary.

6.2 Analysing the Log File Data

Transaction logs are a source of both quantitative and qualitative information on user behaviour (Jansen, 2006) that can be gained by examining the participant's user demographics in terms of location, date and time as well as the query strings used when searching the website that could provide clues on search motivations and search strategy (see Appendix I.1).

The first stages of the analysis looked at the areas of location; date and time; Museum and Gallery Access; Page-level Access. The second stage looked at the search interactions, e.g., the terms used, the structure or complexity of the queries and the user sessions (Jones et al., 2000).

6.2.1 Location

Examining the logs in relation to the location of the user (determined by IP addresses) revealed that page requests came from 213 (see Table 6.1) different countries, although some 33,511 IP addresses were not able to be identified to a country. [1]. Underscoring the global reach of NML's digital content. Notably, the United Kingdom stands out with 303,174 requests and 54,509 queries, followed by the United States with 92,534 requests and 9,432 queries. These figures indicate the sites strong engagement from English-speaking populations, which could be attributed to museum location and that the content is created in English.

Beyond the predominantly English-speaking counties, the dataset encompasses a broad spectrum of global interactions. Countries such as Germany, Denmark, and France, despite having smaller numbers, demonstrate significant engagement, highlighting the website's appeal across different cultural and linguistic backgrounds. Interestingly, the presence of countries like India and China, with 3,455 and 3,072 requests respectively, signals an expanding interest from regions with vast internet user bases, suggesting the potential for further outreach and engagement strategies.

This extensive array of server log data not only showcases the wide-reaching interest in NML's offerings but also emphasises the need for understanding the diversity of its global audience. The data serves as a foundation for tailoring content and engagement strategies to cater to the varied needs and preferences of users worldwide, reflecting the universal appeal of NML's digital resources.

[1] Based on the `IP2Location` IP4 allocated IP address ranges; however, it is noted that the United Nations only identifies 195 countries

Table 6.1: Request by Country

Country	Requests	Queries	Country	Requests	Queries	Country	Requests	Queries
United Kingdom	303,174	54,509	Cyprus	131	22	Bolivia Plurinational State of	17	1
United States	92,534	9,432	Belize	124	8	Myanmar	16	1
Germany	15,810	4,081	Guyana	113	1	Senegal	15	0
Australia	13,992	571	Slovenia	108	10	Namibia	14	0
Ireland	11,773	5,082	Kenya	102	6	Paraguay	13	0
Denmark	11,143	6,538	Ghana	100	5	Faroe Islands	13	4
France	10,340	2,970	Saint Vincent and The Grenadines	99	1	Brunei Darussalam	12	0
Canada	5,415	834	Bahamas	90	1	Gabon	12	2
Russian Federation	5,298	1,436	Tunisia	90	5	Guam	11	0
Spain	3,518	815	Peru	89	16	Armenia	11	2
India	3,455	1,219	Antigua and Barbuda	88	11	Honduras	11	0
Netherlands	3,116	851	Sri Lanka	83	3	Libya	11	4
China	3,072	718	Bosnia and Herzegovina	79	0	Falkland Islands (Malvinas)	11	3
Italy	3,020	641	El Salvador	78	6	Gambia	10	2
Poland	1,913	465	Iceland	76	17	Guadeloupe	10	0
New Zealand	1,562	120	Grenada	73	0	Mozambique	10	6
Norway	1,368	88	Zimbabwe	73	5	Angola	10	0
Brazil	1,246	236	Martinique	67	2	Aruba	9	1
Korea Republic of	1,105	190	Lebanon	66	5	Uzbekistan	9	3
Belgium	936	226	Macedonia The Former Yugoslav Republic of	64	2	Nicaragua	8	3
Sweden	915	170	Venezuela Bolivarian Republic of	63	7	Benin	7	0
Jamaica	831	28	Belarus	58	13	Rwanda	7	0
Japan	820	158	Saint Lucia	58	3	Afghanistan	6	1
Hong Kong	703	254	Guernsey	57	5	Monaco	6	2
Portugal	655	92	Puerto Rico	56	2	Maldives	6	0
South Africa	642	78	Costa Rica	52	6	Mauritania	5	0

Table 6.1 continued from previous page

Country	uristem	uriquery	Country	uristem	uriquery	Country	uristem	uriquery
Philippines	638	26	Panama	52	7	Seychelles	5	2
Switzerland	638	108	Mauritius	51	0	Cabo Verde	5	2
Isle of Man	638	89	Kuwait	50	3	Somalia	5	0
Ukraine	635	194	Liberia	49	3	Yemen	5	1
Trinidad and Tobago	634	23	Anguilla	47	3	Suriname	5	1
Turkey	618	197	Georgia	47	5	Cuba	5	0
Greece	614	131	Iraq	45	5	Curacao	5	2
Czech Republic	532	99	Dominican Republic	45	2	Madagascar	5	0
Mexico	502	47	Nepal	44	7	Congo	5	0
Israel	475	174	Ecuador	40	4	Mali	4	2
Egypt	469	131	Cayman Islands	40	0	Fiji	4	0
Slovakia	465	206	Tanzania United Republic of	40	0	Northern Mariana Islands	4	0
Colombia	458	12	Gibraltar	39	3	Burkina Faso	3	0
Romania	456	34	Uruguay	38	4	Lesotho	3	0
Hungary	450	88	Albania	36	1	Andorra	3	1
Singapore	448	74	Azerbaijan	36	5	Lao People's Democratic Republic	3	2
Argentina	431	38	Togo	35	2	Guinea	3	0
Austria	430	56	Macao	35	10	Aland Islands	3	1
Viet Nam	423	124	Oman	35	2	Saint Martin (French Part)	3	0
Finland	378	74	Kazakhstan	35	11	Liechtenstein	3	0
Indonesia	369	66	Bermuda	35	5	Cook Islands	2	0
Malaysia	365	13	Bhutan	34	28	Montserrat	2	0
Nigeria	332	10	Virgin Islands British	33	0	Mayotte	2	0
Pakistan	329	9	Cambodia	33	1	Burundi	2	0
Serbia	328	17	Bahrain	32	7	Eswatini	2	0
United Arab Emirates	309	34	Dominica	31	0	Holy See	2	0
Luxembourg	303	156	Palestine State of	29	1	Sierra Leone	2	0

Table 6.1 continued from previous page

Country	uristem	uriquery	Country	uristem	uriquery	Country	uristem	uriquery
Croatia	280	30	Haiti	29	2	Kyrgyzstan	2	0
Malta	259	14	Reunion	27	3	Bonaire Sint Eustatius and Saba	2	0
Saudi Arabia	255	35	Cote D'ivoire	27	1	Timor-Leste	2	0
Thailand	233	58	Ethiopia	27	0	French Polynesia	1	0
Bulgaria	231	35	Virgin Islands U.S.	25	0	New Caledonia	1	0
Taiwan Province of China	230	37	Zambia	25	0	Greenland	1	0
Bangladesh	206	14	Moldova Republic of	24	3	American Samoa	1	0
Estonia	205	50	Turks and Caicos Islands	23	1	Equatorial Guinea	1	0
Barbados	201	10	Sint Maarten (Dutch Part)	22	6	Arizona	1	0
Iran Islamic Republic of	197	94	French Guiana	22	7	Niger	1	0
Algeria	190	3	Malawi	21	2	Guinea-Bissau	1	0
Jordan	186	10	Montenegro	21	4	Djibouti	1	0
Latvia	172	31	Mongolia	20	7	Tajikistan	1	0
Chile	172	11	Guatemala	20	3	Marshall Islands	1	0
Morocco	169	10	Uganda	19	0	San Marino	1	0
Jersey	169	22	Cameroon	18	0	Samoa	1	0
Lithuania	151	19	Syrian Arab Republic	18	0			
Saint Kitts and Nevis	140	3	Sudan	18	9	**Column Total**	**518,812**	**94,908**
Qatar	133	12	Botswana	17	0			

While the locations indicated by the log data and those of the survey respondents do not align precisely (see to Table 6.2), the fact that both datasets show a lot of interactions from the UK and other English-speaking regions suggests a connection. The difference in the nature of these data sets, with website requests potentially representing multiple interactions from a single user, in contrast to the survey responses where each response corresponds to an individual user, complicates direct statistical comparisons. However, the simultaneous collection of survey data and log information helps to suggest a clearer link. It makes sense that patterns of English-speaking bias can be seen because the website's content and the survey are presented in English, and the museum itself is situated in the UK. This could help to explain why so many website visitors and survey participants are from English-speaking regions.

Table 6.2: Website access vs Survey response top 10.

Country	**Requests (n)**	**Requests (%)**	**Survey count (n)**	**Survey count (%)**
United Kingdom	303,174	58.4	389	79.7
United States	92,534	17.8	39	8
Germany	15,810	3.0	1	0.2
Australia	13,992	2.7	1	0.4
Ireland	11,773	2.3	4	0.8
Denmark	11,143	2.1	2	0.4
France	10,340	2.0	2	0.4
Canada	5,415	1.0	6	1.2
Spain	3,518	0.7	7	1.4

6.2.2 Date and Time

When examining the logs in relation to the availability of the survey (Jan - March) (see Table 6.5) it was clear that February was the most accessed month across all sites (45.01%). While Tuesday (17.09%) and Wednesday (17.34%) are the most frequently accessed days, with Thursday (16.62%) also indicating high activity. The least website usage can be seen at the weekends, with Saturday (9.21%) showing the lowest activity. To ensure a precise temporal analysis, the server log times were normalised from UK Greenwich Mean Time (GMT) to the users local time to account for different time zones. This was achieved by utilising the IP2Location database to ascertain the country of origin for each visitor and the timezone shift

from GMT, thereby enabling an accurate calculation of the times of access. This normalisation allowed for a consistent temporal comparison aligned with the actual local time the user was accessing the website.

The survey revealed that the evening was the most likely time period (see Table 5.18) that the website was accessed. This finding aligns with the transaction log (see Table 6.3) post-normalisation, which indicates that the majority of activity occurred between 7.00 - 11.00 pm. This suggests that visitors were most likely to engage with the NML website during their local evening hours.

Table 6.3: Requests by time of day

Hour	**No of requests**	**Percentage of requests %**
0	4,829	0.87
1	1,310	0.24
2	1,101	0.20
3	1,047	0.19
4	1,319	0.24
5	2,938	0.53
6	6,548	1.19
7	7,568	1.37
8	13,941	2.52
9	23,133	4.19
10	22,171	4.01
11	26,083	4.72
12	26,284	4.76
13	26,394	4.78
14	30,688	5.56
15	28,731	5.20
16	30,075	5.45
17	28,401	5.14
18	28,635	5.18
19	39,160	7.09
20	45,143	8.17
21	52,121	9.44
22	56,895	10.30
23	47,808	8.66
	552,323	**100**

Compared to the survey findings in relation to frequency of visit (see Section 5.2.4.2) revealed that the majority of activity was in relation to a first time visit of the NML websites with a sharp decline to visiting a couple of times a year and a further decline to only 11% visiting monthly. While specific days of interacting with the website was not part of the survey, the logs show that Tuesday to Thursday are

the business days, but there is significant traffic on all days of the week (see Table 6.4).

It is thought that the increased access that can be seen in Figure 6.1 during February relates to collections that are ending and new ones being announced. This period is also linked to the local school holidays and could explain the increase in accesses to the general information pages. While a low weekend activity could be in response to the participants having other commitments to undertake, e.g., family life.

Table 6.4: Cleaned Log Visits Per Day (Jan-Mar 2017)

Day	**Mon**	**Tue**	**Wed**	**Thur**	**Fri**	**Sat**	**Sun**	**Total**
Requests	76,950	94,371	95,791	91,769	80,419	50,879	62,144	552,323
Percentage	13.93%	17.09%	17.34%	16.62%	14.56%	9.21%	11.25%	100%

Table 6.5: Cleaned Log Visits Per Month (Jan-Mar 2017)

Month	**Jan**	**Feb**	**Mar**	**Total**
Requests	140,134	248,615	163,574	552,323
Percentage	25.37%	45.01%	29.62%	100.00%

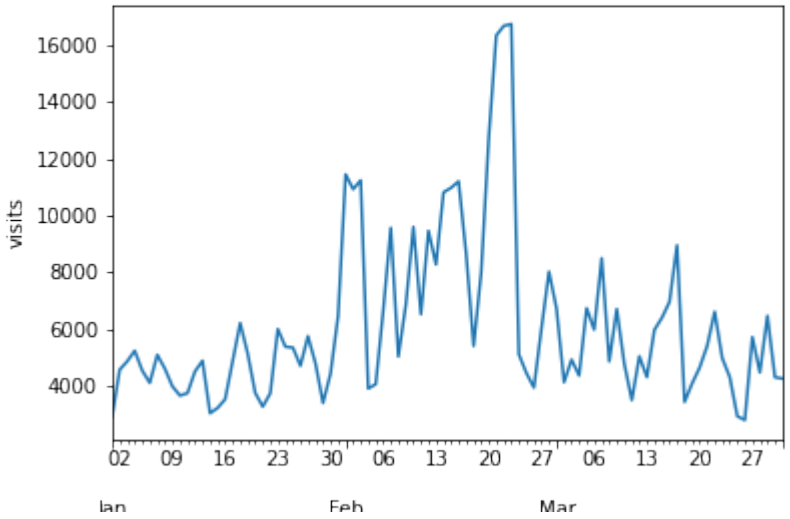


Figure 6.1: Total visits (Jan-Mar 2017)

6.2.3 Museum and Gallery Access

Examining the data logs for which museum and gallery were the focus of most attention could be a reflection of what is influencing some of the participants. The International Slavery Museum (ISM) is the most accessed museum site of the NML website with 95,384 (17.3%) page requests. This is followed by the World Museum Liverpool, Walker Art Gallery, Maritime Museum, Museum of Liverpool, Ladylever Art Gallery and lastly Sudley House. The overall request breakdown between the main sections of the website can be seen in Table 6.6.

Table 6.6: Total Requests by Museum Jan-Mar 2017

Museum	**Requests**	**Percentage**
ISM	95,384	17.3%
WML	83,558	16.9%
Walker	68,932	12.5%
Maritime	65,093	11.8%
MOL	49,070	8.8%
Ladylever	22,888	4.1%
Sudley House	8,056	1.5%
Museum Total	**329,981**	**71.1%**
Other pages	159,342	28.9%
Total	**552,323**	**100%**

The inability to identify the survey participants from the transaction logs resulted in the inclusion of the Museum of Liverpool, which was excluded from the survey by NML (see Phase 2). However, comparing the transaction logs with the captured referral URL of the survey (see Table 5.64), can be seen as validation of the survey as ISM, WML, Walker Art Gallery and the Maritime Museum are positioned in the top 4 of both data sets, although shown in a different order. The exception is the World Museum of Liverpool, which is ranked in second place in both data sets. The other interesting observation is the ranking of the Lady Lever Art Gallery and Sudley House, which are positioned in identical order at the bottom of both data sets. This validates the spread of users gathered from the different sections of the NML website for the survey as being representative of the users using the NML website at that time.

6.2.4 Page-level Access

Throughout the NML website(s), there are 8 main types of pages: **General** pages (G), which include the highest level NML home page, about page, visit pages, floor plans, policy pages and contact pages etc; the individual **Museum/gallery** home pages and the general overview pages (M); the **Collection overview** pages (C), which introduce the collections the museum is displaying on the site; the **Item** pages (I), which contain item/artefact detail pages for collections; the **Archive** pages (A) for the maritime museum only, separate to the collections but act as a more traditional archive of artefacts; the **Events** (E) pages, including 'Whats-on' pages regarding attractions visitors can view/book; the **Kids** (K) pages, which is a dedicated children's area; and **Other** pages: pages that do not fit the mentioned categories or occasional pages in new sections, such as Christmas e-cards or 'picture-of-the-month'. The general pages are the most requested (31%), followed by the item pages (24%) (see Table 6.7).

Each request has a URL for the request from which it was possible to ascertain the type of page the request was for. For example the following URL *"/ism/slavery/africa/capture_sale.aspx"* shows us that the "capture_sale" page is an item level page that sits within the "africa" collection, which sits within the "slavery" section which sits in the "ism" museum. So this request would be for a (I) Item page. Where as a request with a URL of "/walker/index.aspx" would be for a Museum general page.

Table 6.7: Page type usage (all requests)

Page Type	**Count**	**Percentage**
General	170,407	30.9%
Item	132,058	23.9%
Event	56,550	10.2%
Museum overview	55,275	10.0%
Other	43,105	7.8%
Kids	36,184	6.6%
Collection overview	30,543	5.5%
Shop	18,354	3.3%
Archive	9,847	1.8%
Total	552,323	100%

Comparing the transaction log data (TLD) with the data-set compiled from the

survey (see Table 5.65) similarities were presented in terms of engagement with the NML website. Although the identified pages have a different name, essentially some translation was made as follows: General page on TLD is the same as Landing, Talks, Visits and Contacts from the survey; Item from TLD has some traits with that of Collection and Curated; Events from TLD equates to Events and Exhibition; Other from TLD aligns to Education and Competition; TLD Kids page was not considered under the survey as the target audience was over 18 years of age, while shop was also excluded as this was not considered relevant to this research.

The data sets show that the majority of NML website users began their interaction at the General or Landing page, with 30.9% from TLD and a totalled 42.6% from the survey. The Item (TLD) shows some 23.9% of the requests made, which at first glance is significantly higher than the survey data set of 14.3%. However, if considered from a holistic approach then the order and allocation of the percentage breakdown of requests in the TLD data set compared to the survey results are reflective of each other, strengthening the validity of the survey.

Table 6.8: Top 20 pages accessed

URL	**Count**
/	41,601
/onlineshop/product-list.aspx	15,579
/wml/	9,217
/mol/	7,104
/walker/exhibitions/victorian-treasures/index.aspx	6,849
/about/sitesearch/search.aspx	6,768
/walker/	6,676
/maritime/visit/old_dock_tours.aspx	6,475
/mol/beatles-quiz/	6,258
/wml/events/displayevent.aspx	6,254
/wml/events/	5,528
/ladylever/events/displayevent.aspx	5,488
/ism/	5,355
/kids/games-quizzes/egypt-mummys/	5,281
/walker/events/displayevent.aspx	5,268
/walker/events/	5,238
/ism/slavery/triangle.aspx	4,936
/maritime/	4,662
/ism/slavery/middle_passage/	4,493
/wml/collections/antiquities/ancient-egypt/item.aspx	4,294
...	...

The most frequent (requested) page is the NML homepage (/) with 41,601 requests, followed by the online shop home page (/onlineshop/product-list.aspx with 15,579 requests, the next is the world museum home page (/wml/) with 9,217 requests. This is closely followed by the Museum of Liverpool home page (/mol/) with 7,104 requests. The next is the starting page of a walker art gallery exhibition of Victorian treasures (/wml/exhibitions/victorian-treasures/index.aspx) with 6,849 requests (see Table 6.8 for the top 20 pages).

With the page types extracted from the cs-uri-stem (uristem) it can be seen that the UK provides the most visitors to the website (although it is acknowledged that these visitors are not necessarily UK residents, as international holidaymakers could also be included, but there is no means of identifying these from the server logs) (see Table 6.9), although when compared to the data set compiled from the survey (see Table 5.66) 79.7% of respondents are UK based. It is also noted that the top two-page types are the General pages and the Item pages. Surprisingly, the Events pages are third, followed closely by the Museum overview pages (and less surprising is the Archive and Shop pages, which are the lowest accessed page types).

When exploring only the UK page requests, the general pages are still the most requested, with the item and events a close second and third (see Table 6.10). The events pages seem to be high for a number of the more local areas to Liverpool. This might be an indication of a possible physical visit and is comparable to the findings in the survey data set which shows that 35.9% of respondents are based in Merseyside with an additional 23.2% based in the Northwest (see Table 5.40). For those areas outside the Northwest UK, the pattern of requests focuses more on the general pages and the item pages (see Table 6.11). The NML website saw visits from, 3,828 of the UK's 49,178 cities, towns, villages, hamlets and suburban areas. [2]

[2] https://www.townslist.co.uk/

Table 6.9: Top 8 results listing page type by country with total requests and queries (see Appendix I Table I.2 for the complete results table).

Countries	Archive	Collection	Event	General	Item	Kids	Museum	Other	Shop	Requests	Queries
United Kingdom	6,005	17,438	43,061	111,138	47,842	7,979	40,666	21,865	9,096	305,090	55,514
United States	1,315	5,076	1,702	15,642	33,574	21,976	5,817	10,279	3,422	98,803	12,092
Denmark	354	1,278	585	5,591	17,180	1,744	791	900	3,107	31,530	20,368
Germany	463	1,219	341	8,302	3,096	115	886	1,120	268	15,810	4,081
Australia	210	309	48	1,089	10,902	1,524	267	1,336	86	15,771	571
Ireland	253	771	6,953	3,277	1,275	100	571	540	447	14,187	7,372
France	172	470	694	3,704	3,104	166	890	951	426	10,577	3,157
Canada	282	320	131	1,293	1,744	685	277	563	120	5,415	834
...	...	...	...	...	...	...	...	...	...	...	...
Column Total	**9,847**	**30,543**	**56,550**	**170,407**	**132,058**	**36,184**	**55,275**	**43,105**	**18,354**	**552,323**	**114,880**

Table 6.10: Page type by UK region with total requests and queries.

UK Region	Archive	Collection	Event	General	Item	Kids	Museum	Other	Shop	Requests	Queries
England	5,268	15,365	38,301	98,899	42,361	6,760	36,122	19,300	8,088	270,464	49,846
Wales	326	1,008	2,960	6,752	2,361	210	2,488	1,108	543	17,756	3,189
Scotland	309	831	1,351	4,159	2,559	785	1,550	1,220	346	13,110	1,887
Northern Ireland	102	234	449	1,326	561	224	505	237	119	3,757	592
Column Total	**6,005**	**17,438**	**43,061**	**111,136**	**47,842**	**7,979**	**40,665**	**21,865**	**9,096**	**305,087**	**55,514**

Table 6.11: Top 25 UK cities/places page types (in total there were requests from 3,828 places in the UK)

City/Place	Archive	Collection	Event	General	Item	Kids	Museum	Other	Shop	Requests	Queries
Manchester	691	2,263	5,890	15,185	6,707	625	5,458	2,726	1,375	40,920	7,033
Liverpool	625	2,108	7,372	15,188	4,402	120	4,680	1,973	1,241	37,709	9,687
London	474	1,643	2,531	8,456	5,695	1,188	3,119	1,971	750	25,827	3,543
Runcorn	165	677	1,987	3,737	1,211	47	1,369	560	275	10,028	1,917
Sheffield	137	333	737	2,451	1,292	198	938	849	207	7,142	1,100
Birmingham	140	401	831	2,515	1,199	246	1,046	564	198	7,140	1,054
Warrington	108	334	824	2,286	750	118	1,003	430	193	6,046	874
Chesterfield	57	266	680	2,051	653	69	929	250	137	5,092	1,122
Derby	68	300	424	2,288	536	326	296	448	213	4,899	2,420
Walton le Dale	75	198	623	1,825	555	134	759	267	116	4,552	649
Leeds	160	249	409	1,291	832	232	472	491	129	4,265	655
Chester	53	210	929	1,331	437	30	533	201	84	3,808	897
Reading	96	174	448	1,166	594	191	401	242	117	3,429	580
Edinburgh	63	209	281	853	579	226	378	308	75	2,972	385
Maidenhead	35	120	1,430	612	260	71	272	110	44	2,954	1,451
Cambridge	66	163	331	951	433	54	368	226	89	2,681	434
Daresbury	23	121	550	934	233	24	405	135	52	2,477	433
York	58	137	314	864	426	35	332	245	64	2,475	442
Exeter	86	180	293	863	351	20	362	190	69	2,414	398
Sunbury-on-Thames	34	165	222	807	552	46	274	202	52	2,354	333
Brighton	54	110	194	617	459	235	210	141	83	2,103	337
Glasgow	35	142	165	625	432	130	201	191	58	1,979	334
Bradford	38	90	243	577	316	30	216	248	34	1,792	290
Lancaster	39	100	217	590	298	32	260	133	30	1,699	286
Swindon	31	99	160	571	330	70	219	161	39	1,680	236
Top 25 Sub Total	**3,411**	**10,792**	**28,085**	**68,634**	**29,532**	**4,497**	**24,500**	**13,262**	**5,724**	**188,437**	**36,890**
...	...	...	...	...	...	...	...	...	...	...	...
Column Total	**6,005**	**17,438**	**43,061**	**111,136**	**47,842**	**7,979**	**40,665**	**21,865**	**9,096**	**305,087**	**55,514**

Table 6.12: Single page sessions compared to all sessions

Country	All Requests	Single Session requests	Percentage of all requests that are single page sessions	All Queries	Single Session Queries	Percentage of all queries that are single page sessions
United Kingdom	305,090	129,303	42%	55,514	19,859	36%
United States	98,803	37,496	38%	12,092	4,699	39%
Denmark	31,530	814	3%	20,368	34	0%
Germany	15,810	10,370	66%	4,081	3,534	87%
Australia	15,771	5,158	33%	571	255	45%
Ireland	14,187	2,141	15%	7,372	241	3%
France	10,577	3,869	37%	3,157	1,097	35%
Canada	5,415	2,646	49%	834	422	51%
...	...	...	...	...	...	...
Column Total	552,323	221,180	40%	114,880	34,764	30%

An extended list of the sessions by country and single page sessions by country can be seen in Appendix I.3.1 and I.3.2.

Table 6.13: Top 25 Single page session by UK city/place (sorted by physical distance to Liverpool - according to Google maps)

City/Place	Archive	Collection	Event	General	Item	Kids	Museum	Other	Shop	Requests	Queries	Distance from NML (Miles)
Liverpool	282	1,232	2,548	8,495	1,891	39	1,722	1,032	276	17,517	4,611	0
Runcorn	58	424	686	1,809	524	12	491	319	69	4,392	710	17.5
Daresbury	14	82	182	416	101	2	136	69	14	1,016	161	19.2
Warrington	59	218	325	1,257	364	39	439	241	58	3,000	370	19.5
Chester	17	131	281	638	169	3	172	101	21	1,533	299	27.6
Manchester	266	1,111	1,776	6,132	2,203	132	1,874	1,345	276	151,15	2,191	34.3
Walton le Dale	28	108	202	767	226	30	276	167	27	1,831	198	35.4
Bradford	15	42	63	273	108	8	99	168	15	791	126	67.8
Leeds	71	143	147	678	359	55	210	292	31	1,986	234	73.1
Chesterfield	29	134	243	1,049	219	2	330	107	41	2,154	502	78.2
Birmingham	68	199	252	1,087	437	50	371	317	45	2,826	380	99
York	19	78	89	392	195	15	128	154	19	1,089	147	102
Sheffield	84	174	261	1,052	491	56	358	588	60	3,124	481	102.1
Swindon	16	54	53	241	175	23	79	76	11	728	90	176.3
Newport	20	42	56	278	103	25	78	69	13	684	88	188
Cambridge	33	102	109	455	211	17	152	98	25	1,202	151	193.9
Reading	41	102	141	588	231	57	168	147	31	1,506	217	198.1
Maidenhead	18	59	58	286	105	14	113	73	4	730	65	198.7
Carluke	10	37	64	303	135	32	94	67	9	751	76	207.2
Glasgow	18	75	66	309	203	24	73	94	15	877	107	219.9
London	235	950	862	3,996	2,356	229	1,289	1,182	226	11,325	1,357	221.3
Sunbury-on-Thames	25	88	72	361	240	20	102	121	25	1,054	124	222
Edinburgh	39	118	91	411	240	59	149	195	18	1,320	137	230.3
Exeter	48	93	102	474	138	5	156	107	20	1,143	146	251.5
Bournemouth	24	61	78	301	95	7	109	59	13	747	83	265.3
...	...	...	...	...	...	...	...	...	...	...	...	
Column Total	2,910	9,764	13,831	51,341	19,584	2,013	15,361	12,187	2,312	129,303	19,859	

Findings also show that single page sessions/visits occur quite frequently. This is when users leave the site, without exploring more than the initial page. There are 552,323 total requests over the three months, with 221,180 (40%) of the requests being for single page accesses for the complete session (a breakdown of this can be seen in Table 6.12). It is not possible to establish if these single page sessions are bounces, or if the page provided the correct information in the right detail that the user felt like their information seeking goal was completed. Google defines a bounce as

> "A bounce is a single-page session on your site. In Analytics, a bounce is calculated specifically for a session that triggers only a single request to the Analytics server, such as when a user opens a single page on your site and then exits without triggering any other requests to the Analytics server during that session."

(Google Analytics Support, 2022)

However, the 40% single page access rate does correspond with the 48.2% of single page access indicated by the survey (see Table 5.46) which in turn suggests that the majority of interaction with the website is for establishing what is currently available to the public for viewing.

Page level access by Museum/Gallery can be seen in Table 6.14. These results show how each institutional areas of the site differ in the types of pages accessed in each.

Table 6.14: Page level access by Museum/Gallery

Page Type	WML	ISM	MOL	Maritime	LadyLever	Sudley	Walker	Undefined	Row Total
General	24,245	7,830	14,075	30,037	3,838	1,629	7,914	80,839	**170,407**
Item	24,350	55,910	4,706	5,765	5,435	2,229	25,781	7,882	**132,058**
Event	12,618	3,558	5,673	8,155	8,163	1,851	12,036	4,496	**56,550**
Museum	15,837	6,909	10,072	7,369	3,331	1,495	10,239	23	**55,275**
Other	1,843	13,047	11,777	827	1,218	278	3,070	11,045	**43,105**
Kids	0	0	0	0	0	0	0	36,184	**36,184**
Collection	4,665	8,130	2,767	3,093	903	574	9,892	519	**30,543**
Shop	0	0	0	0	0	0	0	18,354	**18,354**
Archive	0	0	0	9,847	0	0	0	0	**9,847**
Column Total	**83,558**	**95,384**	**49,070**	**65,093**	**22,888**	**8,056**	**68,932**	**159,342**	**552,323**

Table 6.14 indicates that the ISM received the majority of requests with WML and Walker in second and third place respectively, followed by Maritime, MOL, Lady Lever and Sudley. This aligns fairly closely to the survey results (see Table 5.64) with some changing of the order of places for the top 4).

6.2.5 Search Analysis

The TLD allowed for closer examination of the constructs that were being used to engage with the NML website. While the survey participants indicated their desired medium for interaction (see Section 5.3.6) the dissection of the Search queries presented an opportunity to examine the actual constructs that were being used to find the desired information.

6.2.5.1 Search Queries

The TLD revealed some 125,436 search queries issued during the three-month period. The search queries focused on the top level general pages such as the NML homepage, about page etc and the deepest level, item pages (see Figure 6.2). While the majority

of searches were made by visitors of the slavery museum in comparison to only 2,868 queries for maritime archive pages; 26,897 queries were for event pages; 14,354 queries for shop items; and 1,661 queries for collections and collection items. It appears as though there are over three times more requests using queries than there are for standard page requests. A possible explanation for this is that a number of the pages that are navigable by in page links or buttons are structured using the query terms.

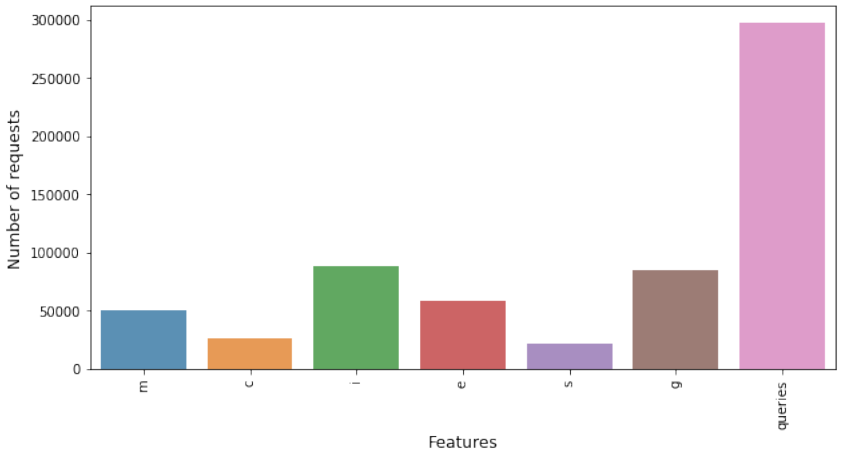


Figure 6.2: Page type counts

6.2.5.2 Query Terms

The TLD data-set also indicated that there were 6,755 searches made on the NML "free search" system. These were identified from the 114,091 rows in the server logs that did not contain blanks, by searching for the rows with the "ssterms" keywords in the uriquery that the NML search system appends to the page url upon a "free search" being made. The complete logs had 552,323 rows before the rows with blanks were removed.

There were 574 queries that contained a single or double quote in an attempt to SQL inject the NML database server (see Table 6.15). Some examples of these queries are:

1. 107 attempts using " slavery museum'0=A ", 90 attempts using " slavery museum' and 'x'='y ", 76 attempts using " slavery museum " and " x"="y " and 16 attempts to use " pre-raphaelite'||SLeeP(3)&&'1 " which are all generic

Table 6.15: Search term refinements

Filter	Identifier	Counts	Total search terms
Search terms before filtering			6,755
SQL injection attempts			
	single quote	99	
	double quotes	475	
	cOnVeRt	107	
	and 1>1	95	
	Total SQL Injection attempts	776	
After SQL Injections removed			5,979
Potential persona data			
	"@"	1	
	www	1	
After personal data removed			5,977
After abusive terms removed			5,974
After blanks removed			5,676

payload attempts to try and close the query with an apostrophe, then setting a variable. In essence, this is just a test to see if the system is vulnerable.

2. 78 attempts using " slavery museum'&&sLEEp(3)&&'1 " and 67 attempts at using " slavery museum' AnD sLeep(3) ANd '1 " both where the apostrophe escapes the query and tries to use the SQL sleep command in an attempt to block access to the database. The sleep command itself does nothing more than create a long running SELECT query but in reality it also puts a lock on the DB server until it has finished which in essence creates a denial of service on the database. [3]

3. Three is also a slightly more interesting remote code execution attempt. There were 107 attempts at using " `/**/cOnVeRt(int,(char(33)+char(126)+char(33)+(char(65)+char(66)+char(67)+char(49)+char(52)+char(53)+char(90)+char(81)+char(54)+char(50)+char(68)+char(87)+char(81)+char(65)+char(70)+char(80)+char(79)+char(73)+char(89)+char(67)+char(70)+char(68))+char(33)+char(126)+char(33)))` " which starts with a mixed case "cOnVeRt" in and attempt to obfuscate the basic string captures that look for 'convert', along with character encoding. Decoding the string reveals: " ! !ABC145ZQ62DWQAFPOIYCFD! ! ", which appears like a pretty

[3] `https://blog.pythian.com/mysql-injection-sleep/`

random string but a quick Google search shows it is likely a known attack from a SQL injection piece of hacking software called 'SQLI Dumper' that has a series of in build and automated SQL attack methods that anyone could run. [4]

After removing these SQL injection attempt rows there were 5,979 search terms. Further cleaning of the search terms saw 2 rows removed for containing personal data e.g., a person's email address and 3 more rows removed for containing an abusive term. This left 5,976 search terms to be used in the following analysis.

It was also found that the NML search system allowed the user to simply press the search button without entering a search term. Given that the majority of users (identified in the survey) are General public and Non-Professional users, it is safe to assume that they are not aware of what is hidden in the database behind the search box and so could just hit search to see if any results appear that could hint as to what is in the database. This resulted in 298 searches that were blank being removed, leaving a data-set of search terms of 5,676.

A closer examination of the 5,676 free text search queries used, found that 3,582 were unique. However, 2,094 were made more than once. Unfortunately, it is not possible to establish if these are duplicate searches made by the same person within the same session.

6.2.5.3 Common Occurring Search Terms

Examining the formulation of the search terms in the TLD revealed the identification of some fairly common and frequent words being used, with the occasional changes in spelling and use of capitalisation.

[4] https://www.threatx.com/blog/unexpected-web-application-attacks-when-not-to-trust-your-search-engine/

Table 6.16: List of all searches made more than 5 times over the 3 month period

Keywords (ssterms in uriquery)	Times searched for
slavery museum	760
slavery museum21211211212.1	90
Liverpool	54
jobs	46
Jobs	29
slavery	29
titanic	27
vacancies	23
terracotta	16
mucha	15
Dinosaurs	15
Slavery	15
parking	14
egypt	13
pre-raphaelite	13
Cafe	12
house+of+memories	12
planetarium	12
Terracotta+army	11
Titanic	11
middle+passage	11
Photography	11
tickets	11
opening+times	11
kofi	10
space	10
Dinosaurs+	10
kwame	10
turner	10
price	10
careers	10
Opening+times	10
dinosaurs	10
volunteering	9
terracotta+army	9
Lusitania	9
prices	9
games+and+quizzes	9
Parking	9
milroy	9
cafe	9
⋮	⋮

Keywords (ssterms in uriquery)	Times searched for
⋮	⋮
volunteer	8
games	8
Prices	8
Ancient+Egypt+gallery	8
slaves	8
terracotta+warriors	8
the+middle+passage	8
paintings	7
Parrot	7
burne+jones	7
lgbt	7
photography	7
hermaphrodite	7
echo and narcissus	7
atlantic+slave+trade	7
millais	7
victorian+treasures	7
Terracotta+warriors	7
lady+lever	7
stubbs	7
opening+hours	7
job+vacancies	7
Kids	7
ww1	7
terracota+warriors	6
youtube	6
vikings	6
Entry+fee	6
Space	6
map	6
lusitania	6
coal	6
ancient+egypt	6
puppet	6
wallpaper	6
Vacancies	6
Egypt	6
kids	6
⋮	⋮

Keywords (ssterms in uriquery)	Times searched for
⋮	⋮
beatles	6
slave+trade	6
oyeladun	6
slave+auctions	5
Cost	5
pomegranate	5
Titanic+	5
staff	5
Price	5
rubens	5
Monet	5
dress	5
virtual+tour	5
waterhouse	5
sankofa	5
abolition	5
chinese	5
faience artefacts	5
isabella	5
picasso	5
alfred h read	5
slave	5
Stone+age	5
monet	5
leighton	5
fish	5
blitz	5
Planetarium+	5
video	5
henry+viii	5
cascade	5
Middle+passage	5
disability	5
Subtotal (keywords searched 5 or more times)	**1,850**
Sub total searches 2 - 4 times	**244**
Total unique searches	**3,582**
Total searches	**5,676**

The majority (91.4%) of the searches made are very short, with only one, two or three words making up the search (see Table 6.17). The majority of these short tail keywords can be seen in Table 6.16 and are for:

1. General information pages: e.g. Opening times, Jobs, Tickets, Kids, Games, etc.
2. Facilities information: e.g. Cafe, Planetarium, Parking, etc.
3. Specific museums: e.g. Slavery museum, Lady Lever, etc.
4. Specific or potential collections or subject information: e.g. Egypt, Slavery, Dinosaurs, Titanic, Lusitania, Beatles, WW1, etc.
5. Specific events: e.g. Terracotta Warriors, Victorian treasures, house of memories, etc.
6. Specific artists or styles: e.g. Turner, Monet, Rubens, pre-raphaelite, etc.

Table 6.17: Length of keywords used

Number of Words	**Searches (n)**	**Searches (%)**
1	2,061	36.31
2	2,426	42.74
3	681	12.00
4	240	4.23
5	112	1.97
6	59	1.04
7	42	0.74
8	18	0.32
9	17	0.30
10	8	0.14
11	7	0.12
13	2	0.04
14	2	0.04
12	1	0.02
Column Total	5,676	100.00

The longer tailed keyphrases with more than 6 words were all used only once and tended to be structured as a question or as an attempt at a boolean query (see Table 6.18 for examples).

Table 6.18: Longtail keywords (longer than 6 words)

Longtail Keyphrase	Word Count	Time used
when is the new egyptian gallery opening	7	1
roles of enslave woman during british west indies	8	1
5 facts about liverpools role in the slave trade	9	1
Name the ways slaves were taken out of Africa	9	1
Pictures of troops arriving to Liverpool in the fourth	9	1
Portrait Strangers in a Strange Land by Albert Starling	9	1
Telescope, in case, by Abraham & Co. of Liverpool	9	1
The Liverpool and London War Risks Insurance Association Limited	9	1
Walker Art Gallery and the Craft and Design Gallery	9	1
Whish island are heavily with Africans that speak English	9	1
can I copy and use data from the site	9	1
do any of these places do sleepovers for cubs	9	1
how was the work divided on the sugar plantation	9	1
official abolition of slavery in the British West Indies	9	1
the type of lighting in the Walker art gallery	9	1
what role did Liverpool have in the slave trade	9	1
what music did the slaves bring to the Caribbean	9	1
when is the new Egyptian gallery going to reopen	9	1
What factors caused the rise of the Atlantic slave trade?	10	1
When did the south hear about the abolishment of slavery	10	1
how to acknowledge the use of data from the website	10	1
process of events that took place in the boiling house	10	1
what sort of conditions did slaves endure on the voyage	10	1
what was it like to be captured by the british	10	1
where are they taken , if they survive the voyage	10	1
A comb and jug from the post-emancipation period at the exhibition	11	1
what made the buyers and sellers of slavery chang their mind	11	1
when the colonists wrote a law who could reject the law	11	1
He said his family had a great heart but the wizard was	12	1
Destinations of were the english,spanish and french ports stop to let off slaves	13	1
What Happened To The Slaves When rived To They ArThe West Indies ?	13	1
between what years were millions of africans forced out of africa by europeans	13	1
Pamphlet: What Does Your Sugar Cost - A Cottage Conversation re British Negro slavery	14	1

6.2.5.4 Boolean Queries

Boolean searches allow for the combination of words and phrases using the words AND, OR, NOT (known as Boolean operators) to limit, broaden, or define a search. The TLD show 4 instances of long tail keywords where the search included the word "and" (see Table 6.18). However, there is no evidence of any actual boolean searches being made using the correct phrasing.

The NML search system at the time was not set up to provide boolean search results using the boolean phrases (AND, OR, NOT); but it is not believed that the users were aware of this, nor aware of how to use Boolean searches or at least

chose not to use them. The TLD would have captured the attempts as part of the uriquery, for which there is no evidence.

6.2.6 Sessions

The TLD produced a total of 321,174 sessions were identified when the 30 minute cut-off time limit was applied (265,776 sessions without applying the cut-off). Looking at sessions comprising single pages, it can be seen that the general pages (G) are accessed the most (see Table 6.13) closely followed by item pages (I). Bounces are probable in the case of NML item level pages when the visitor has acquired the information required or wanted. Furthermore, the item level pages do not advertise additional goods; instead, they highlight the collection to which the item belongs (but the links are not obvious).

6.2.6.1 Session entry pages:

Understanding visitors' entrance points can provide information on how a site is utilised as well as the amount of knowledge requested. According to the logs, the most frequently accessed pages are the NML group homepage (G) and the item pages (I). This is followed by the museum homepages, collection overview sites, event pages, kids pages, the store, and just a few archive pages (see Table 6.19).

6.2.6.2 Session exit pages:

Just as interesting, and potentially as informative as entry pages, are the exit pages. The closeness of the entry and exit results clearly highlight the amount of single page sessions (see Table 6.19). There is an indication of some movement around the site from the Museum overview pages and some of the undefined pages to all other page types, from which they then leave.

6.2.6.3 Length of Session

Results show that 246,643 sessions ended within 10 seconds, only 7,637 of these sessions ended within 10 seconds after visiting more than one page - a 43% drop-out within the 10 second period. Google Analytics reports a 52.6% bounce rate for the same period. The drop-out rates over time clearly identify that there are very few

Table 6.19: Session entry and exit pages

Page type	Entry	Exit
General	108,860	96,504
Item	58,618	59,466
Museum overview	23,153	42,045
Undefined	27,678	35,588
Collection overview	24,561	18,692
Event	31,584	20,109
Kids	13,662	13,757
Shop	5,399	8,759
Archive	7,126	5,721

Table 6.20: Session drop-out rates

Session Length (time)	Count	Percentage (%)
<0.1 sec	257,270	85.61
1 sec and below	257,858	85.81
10 secs and below	261,764	87.11
>10 secs and upto 20	3,172	1.06
>20 secs and upto 30	2,417	0.80
>30 secs and upto 60	5,509	1.83
>60 secs and upto 2 mins	6,317	2.10
>2 mins and upto 3 mins	3,871	1.29
>3 mins and upto 4 mins	2,685	0.89
>4 mins and upto 10 mins	7,207	2.40
>10 mins and upto 20 mins	4,535	1.51
>20 mins and upto 30 mins	3,037	1.01
Total sessions	**300,514**	**100.00**

users during this three month period who spend significant time on the site (see Table 6.20).

6.2.6.4 Known Item (Direct) Interactions

Item pages accessed as the first interaction of the session indicate known items. There were, 58,607 item pages accessed as an initial request of a session (see Table 6.21 for a list of the most requested initial pages and see Appendix I.3.3 for a list of all item initial session page requests).

Table 6.21: Session starting pages with Items (known items) - over 900 request.

Item URL	Count of Initial session hits
`/walker/exhibitions/victorian-treasures/index.aspx`	4,741
`/ism/slavery/africa/capture_sale.aspx`	2,722
`/walker/exhibitions/mucha/index.aspx`	2,657
`/ism/slavery/triangle.aspx`	1,939
`/ism/slavery/europe/liverpool.aspx`	1,048
`/ism/slavery/africa/effect.aspx`	998
`/ism/slavery/middle_passage/olaudah_equiano.aspx`	989

Table 6.21 continued from previous page

Item URL	Count of Initial session hits
/ism/slavery/europe/abolition.aspx	907
...	...
Total	**58,607**

6.2.6.5 Searches per Session

The TLD shows that there are only a fraction of the sessions where the user decided to undertake an actual search. The majority of sessions saw the user browsing.

If a user wished to search using keywords whilst using the NML site, they were taken to a dedicated search page (/about/sitesearch/search.aspx). But it is possible to infer the museum/gallery area of the site that the user was using by looking at that the previous uristem prior to the search row within the same session. This shows that the user of the Walker gallery conducted the most searches, followed closely by the World Museum (Wml) and then this was closely followed by the Maritime museum, then the Slavery museum (Ism), museum of Liverpool (Mol) and lastly the Ladylever gallery (see Table 6.22). However, it is important to also note that 63.6% (3,608) of the 5,676 total searches conducted were from sessions which started directly on the search form (`/about/sitesearch/search.aspx`) and not from after searching other areas of the NML website.

Table 6.22: Searches by museum

Museum/Gallery	Count of searches
Ism	328
Ladylever	160
Maritime	423
Mol	212
Walker	513
Wml	432
Sub-Total	**2,068**
Sessions started with a search	3,608
Total	**5,676**

95.7% (5,434 sessions) of all the 5,676 sessions with a search only saw the user conduct one or two searches. The most searches conducted in a single session was eight, and there were only three sessions where this happened (see Table 6.23).

Table 6.23: Searches in sessions

Number of searches	Session Count
1	4,875
2	559
3	152
4	50
5	15
7	12
6	10
8	3
Total sessions with a search	**5,676**

6.2.6.5.1 Query Usage Patterns

In the sessions where between five and eight searches were conducted, it is clear that there is some pagination through the initial queries results. In many cases this means that even though the logs show there were numerous searches conducted, the reality is that there were usually only one or two searches and the rest were paginations (see Table 6.24).

Table 6.24: A single session's search activity. This is one of the sessions that has eight rows of search activity.

Session	Date-time	Uristem	Query	Action
47700	2017-02-16 19:51:34+00:00	/about/sitesearch/search.aspx	ssterms=colliery&page=1	Search
47700	2017-02-16 19:51:57+00:00	/about/sitesearch/search.aspx	ssterms=mines&page=1	Search
47700	2017-02-16 19:54:17+00:00	/about/sitesearch/search.aspx	ssterms=coal&page=4	Search
47700	2017-02-16 19:54:43+00:00	/about/sitesearch/search.aspx	ssterms=coal&page=5	Pagination
47700	2017-02-16 19:55:10+00:00	/about/sitesearch/search.aspx	ssterms=coal&page=6	Pagination
47700	2017-02-16 19:55:34+00:00	/about/sitesearch/search.aspx	ssterms=coal&page=7	Pagination
47700	2017-02-16 19:57:21+00:00	/about/sitesearch/search.aspx	ssterms=coal&page=10	Pagination
47700	2017-02-16 19:57:42+00:00	/about/sitesearch/search.aspx	ssterms=coal&page=11	Pagination

Some of the sessions with the higher number of searches is because of spelling issues and the correction of the spellings. Table 6.25 shows a single session where the refinements happen over a very short time period and the search terms of the query are only adapted very slightly, sometime just by adding or removing a letter to correct spellings.

There are, however, some signs in a couple of the search sessions that show the user may have potentially been learning and applying this learning to subsequent searches. Table 6.26 shows an example of a single session with multiple searches

Table 6.25: A singe search session showing potential spelling corrections.

Session	Query	Duration from start of session	Action
152244	ssterms=darby+housse+%2b+caalport+ching+museum...	0 seconds	Search
152244	ssterms=arby+housses&page=1	48 seconds	Regfinement and spelling change
152244	ssterms=darby+housse+*&page=1	1 minute 31 seconds	Spelling change
152244	ssterms=drby+housse&page=1	2 minutes 11 seconds	Spelling change
152244	ssterms=housse&page=1	2 minutes 24 seconds	Refinement
152244	ssterms=darby+housses&page=1	2 minutes 45 seconds	Refinement - step back to previous search

that shows that the user is potentially learning from the materials returned from the search, to allow them to refine their search by changing the search term and in doing so either narrowing or widening the scope of the results returned. The time between the searches shows that the new searches are not simply unconsidered refinements.

Table 6.26: A single search session showing potential learning.

Session	Query	Duration from start of session	Action
283821	ssterms=Pottery&page=1	32 seconds	Search
283821	ssterms=Staffordshire%20pottery&page=2	1minte 45 seconds	More specfic Search
283821	ssterms=Staffordshire%20pottery&page=4	2 minutes 17 seconds	Pagination
283821	ssterms=English+ceramics&page=1	2 minutes 46 seconds	Widen search scope
283821	ssterms=vases&page=1	3 minutes 59 seconds	Narrow search
283821	ssterms=Staffordshire&page=1	4 minutes 45 seconds	Widen search scope
283821	ssterms=Classifying+ceramics&page=1	5 minute 50 seconds	Redefine search
283821	ssterms=Staffordshire&page=2	6 minutes 7 seconds	Step back and Pagination

6.3 Clustering of the Transaction Logs

To extend the initial stages of the TLD analysis and identify potential user group labels, cluster analysis was used on features derived from the server logs and performed on a re-structured version of TLD where each row represents users and features that capture interactions averaged across user sessions.

Previous studies have clustered user sessions Russell-Rose and Clough (2016); Bogaard et al. (2019); however, in this stage it is intended to establish if potential user groups could be identified from the logs. By identifying, all sessions containing the same IP address and User-Agent were grouped together and features derived from the logs were used to infer user groups (see Table 6.27). K-means clustering was applied using the features that are likely to distinguish user groups[5].

[5]Alternative algorithms such as k-modes (k-prototypes) and DBScan were also tested, but no

Table 6.27: Mapping user group characteristics to log data

User group characteristic	**Log data**
Motivation	Starting level page
Domain / CH knowledge	Page type and queries
Task	Page type and possibly queries
Location	IP (reversed) identifying country, region and city.
Frequency of visits	Repeat visits (sessions), queries, length of session

Features used to characterise users were based on the findings of prior work Walsh et al. (2017, 2020) and can be divided into 3 main categories: *Locational* (country, region and city); *Behavioural* (#sessions; total page requests; #page requests for museum overview, collection overview, item, event, shop and general information page; #museums accessed; average interactions/session); and *Knowledge* (total queries; average queries/session). Clustering was also attempted using the location and session count columns; however, results for both features in all clusters did not result in clear clusters.

As previously stated, rather than depending on pre-defined categories, cluster analysis was utilised to construct user groupings from data. The widely utilised k-means technique, which is ideally suited to numerical data, was applied. Prior to clustering variable values were scaled and to choose the appropriate number of clusters k, the elbow method was used Bholowalia and Kumar (2014). This identified the suitable number of clusters between 6 and 8.

To ensure the number of clusters was valid and stable, the cluster analysis was run repeatedly with different values of k (between 5-9), with $k = 7$ producing the most stable results. The resulting clusters were then analysed for stability and the clusters were evaluated.

6.3.1 Clustering Analysis

The logs provide a data set of 225,796 total users, which reduces to 225,730 when reducing the noise of potential uncaught bots by removing rows with over 100 requests in a session in the columns of Museum overview pages (M), Collection overview pages (C), artefact Item pages (I), Event pages (E), Shop pages (S), General information pages (G), and queries.

stable clusters emerged.

Table 6.28: Cluster results with assigned labels

Cluster	# Users	M	C	I	E	S	G	Queries	Label
1	172,692 (76.6%)	0.17	0.09	0.21	0.15	0.04	0.00	1.07	Single page viewers
2	46 (0.02%)	2.02	4.33	31.48	9.91	26.85	11.04	56.91	High all round searchers
3	4,162 (2.1%)	0.80	0.30	0.34	3.77	0.15	0.57	3.39	Event visitors
4	45,282 (20%)	0.20	0.05	0.07	0.08	0.02	1.29	1.15	Single query general page visitors
5	292 (0.1%)	1.10	3.48	23.35	4.73	7.48	5.90	21.84	Deep level browsers
6	290 (0.1%)	4.96	1.69	2.02	1.68	0.76	17.58	3.78	General museum visitors
7	2,966 (0.8%)	0.42	0.63	4.57	0.37	1.11	0.74	5.25	Known-item searchers

6.3.1.1 Describing the Clusters

As discussed previously, 7 groups of users were identified from the cluster analysis and assigned labels, based on their characteristics (see Table 6.28):

Cluster 1: The cluster labelled "Single page viewers" is by far the largest grouping (76.6%). This cluster consists of users who look at only one page and then leave. The pages are at various levels (general to detailed item pages) and on average the users issue 1-2 queries.

Cluster 2: The smallest cluster (0.02%) we label as "High all round searchers" as they view high numbers of all level pages and also issue many queries. Hobbyists have been identified as undertaking research for personal reasons, but are very knowledgeable in their particular areas of interest and as such are able to use search successfully. They are also interested in both the overview and the deeper item level pages Skov and Ingwersen (2008); Skov (2009); Elsweiler et al. (2011).

Cluster 3: This cluster (2.1% of users) are labelled as "Event visitors" and reflect users who predominantly visit event pages and tend to query for them.

Cluster 4: The second-biggest cluster (20%) of users has been labelled as "Single query general page visitors" with the pages mostly being viewed being general and museum overview pages. This aligns with Booth's Booth (1998) findings of the general user - someone seeking general information about opening hours, prices, facilities, etc.

Cluster 5: This cluster (0.1% of users) we label as "Deep level browsers" as they commonly use queries, view many item level pages, but very few general or

museum overview pages. Similarities can be drawn with Vilar and Šauperl (2014) who define professional users (museum staff) as those who have good knowledge of the task, are trained in the systems and have deep understanding of the context.

Cluster 6: This cluster (0.1% of users) is labelled as "General museum visitors" who mostly view general and museum overview pages, with the occasional "deep dive". This aligns with Booth's (1998) general user.

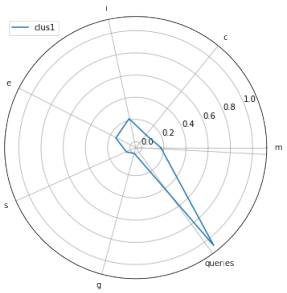
Cluster 7: This cluster (0.8% of users) we label as "Known item searchers" who view only item level pages and query frequently. This behaviour may arise as the majority of item level pages in the NML site(s) can only be accessed by the search feature. This behaviour aligns with Marchionini et al.'s (2003) professional user (including academics) who are described as being highly motivated, having high domain knowledge, system knowledge, time available and focus.

The differences in cluster profiles can be clearly seen from the example polar/spider charts shown in Figure 6.3. The differences in the shapes and the scales are some of the most striking differences.

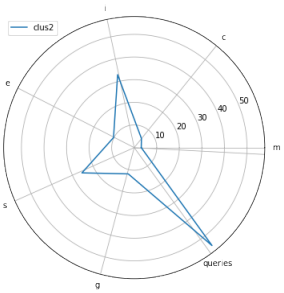
Some of the clusters produced as part of this study show some potential as being candidates for mapping to the known user groups. The potential candidates are shown in Table 6.29.

Table 6.29: Clusters mapped to potential user groups

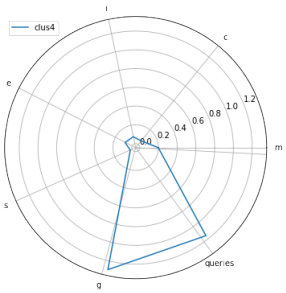
Cluster	**# Users**	**Label**	**Potential user group**
1	172,692	Single page viewers	Currently undocumented user-group called "Bouncers"
2	46	High all round searchers	Non-professionals (hobbyists)
3	4,162	Event visitors	Teachers / General Public
4	45,282	Single query general page visitors	General public (Pre-visit) / Teachers
5	292	Deep level browsers	Museum Staff
6	290	General museum visitors	General public / Students
7	2,966	Known item searchers	Academics (experts) / Non-professionals



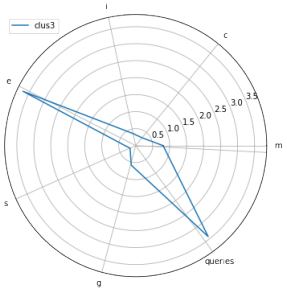
(a) cluster 1



(b) cluster 2

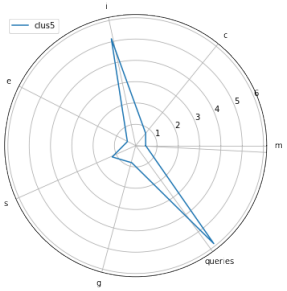


(c) cluster 4



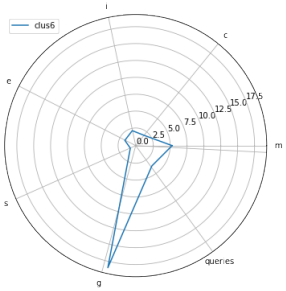
(d) Cluster 3

Event visitors



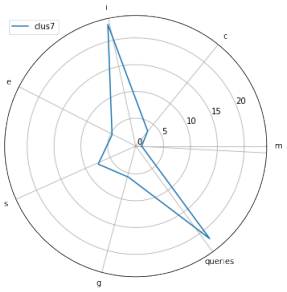
(e) Cluster 5

Deep level browsers



(f) Cluster 6

General museum visitors



(g) cluster 7

Figure 6.3: Example cluster polar charts

6.4 Discussion

In order to address RQ5: How do user groups identified in the transaction logs compare to those identified in the literature and survey results?, the interrogation of NML website transaction logs was made in an attempt to extend the findings from Phase 1 and Phase 2 and provide additional clarity, where possible, regarding the definition and characteristics of a user group label. Analysis of the TLD was undertaken by first looking at the available demographic data (Location, date and time); the access points to the NML website (Museum and Gallery access pages) and finally the Search interactions (terms used, structure or complexity of the queries) and the user sessions. A cluster analysis was also used on features derived from the server logs and performed on a re-structured version of the TLD where each row represents users and features that capture interactions averaged across the user sessions in order to enhance the overall analysis and findings of this captured data set.

Examination of the TLD revealed that the NML websites are being accessed from all over the world (see Table 6.10). When the IP addresses are geo-located, it became evident that the largest number of visits are from the UK (52.3%), which while this is not as high as the survey results from Phase 2 (Table 5.40) (Walsh et al., 2020) indicated (75%), it is still the largest proportion of requests. The majority (12.7%) of these UK users are local to the Liverpool area (Table 6.11) and therefore are within a reasonable distance to physically visit the museums/galleries. This also aligns with findings of Phase 2, with location being a good indicator of a user group and in this instance, the high number of local users could indicate general public (pre-visit) users.

The results also show that the access times of NML website activity tends to fall in accordance with GMT, generally in the evening, which when coupled with the high number of single page access sessions and subsequent high bounce rate, again points towards a local user who is preparing for a physical visit. This high number of single page sessions (see Tables 6.12 and 6.13) may indicate that users are being provided with the correct information to meet their needs (e.g., finding a known-item). These numbers also align with Phase 2 (Walsh et al., 2020), where

the majority of users (especially the general public and non-professional users) were visiting for the first time. Linking to the number of sessions that end within 10 seconds (Table 6.20) and little movement around the site indicated by session entry and exit pages (Table 6.19) highlights limited engagement with the site(s).

A closer examination of the majority of pages accessed by users fall under the categories of general and item (see Table 6.7). This aligns with the findings identified in Phase 2 (Walsh et al., 2020) for understanding users' purposes for using the website, which showed that the major reason for visiting by the general public and non-professional users was to prepare for a visit.

Inspection of the rareness of search results (only 7,121 from 586,868 requests) also aligns with the Phase 2 (Walsh et al., 2020) study where the survey identified that the use of navigational links was the preferred method of access by the majority of users (general public and non-professionals). Taking into consideration their higher CH and domain knowledge levels, this could indicate that the searchers of known-items or collections are from these user groups.

Performing a cluster analysis upon a restructured version of TLD was an attempt at establishing if potential user groups could be identified from the types of pages visited. This analysis produced seven clusters, some of which bore similarities to user label categories previously discovered in the literature (see Chapter 4) and Phase 5. However, the vagueness offered by the cluster analysis (see Table 6.28) does not assist in the clarity of the established user group labels, rather it adds to the non-operative state of said labels.

Therefore, in reaction to the findings of the Transaction Log Data analysis, minor amendments/extensions can be made to the following group descriptions:

General Public (GP) - Local to the area. Are more likely to interact with the web-pages through the use of navigational links and access the General, Item and Events pages of the website in preparation of a physical visit. Access to the website is likely to be undertaken in the evening consisting of short visits, no browsing and usually leaving once the relative information has been found.

Non-Professional (NP) - Are more likely to interact with the web-pages through the use of navigational links and access the General, Item or specific known item pages of the website in preparation of a physical visit. Access to the website is likely to be undertaken in the evening consisting of short visits, no browsing and

usually leaving once the relative information has been found.

Museum Staff - Deep level browsers.

Therefore, interrogation of the TLD proved invaluable in terms of validating the data derived from the survey and corroborating the existence of certain groups, e.g., general public, Non-professional Museum Staff. However, what is also apparent is the continued vagueness regarding what a user group label should look like. Therefore, in an attempt to clarify and expand user group label definitions, further exploration of the NML survey results will be undertaken from a data-driven approach.

6.4.1 Summary

The aim of Phase 3 was to extend the work of Phase 1 and Phase 2 through the use of identifying information seeking behaviours based on interactions with the NML website, which are captured in the website's transaction logs. Transaction logs are automatically generated and data captured through the use of technology, usually in the form of a computer, laptop, tablet or phone;and the subsequent interaction with the website. The data captured by transaction logs is not as nuanced as the survey questions but does support the validation of the survey e.g., examination of demographic data such as location, date and time as well as the formulation of the query strings that are used when searching the NML websites.

Preparation of the log data consisted of a cleaning code to remove artificial interactions e.g., bots, crawlers and developer tools. The code itself required adjustments following manual examination of the data logs after each code/programme sweep, therefore removing the impact of some, 34,281 artificial interactions from the data set.

Examination of the logs in respect of the limited demographic data found strong similarities between the log data and the survey responses in that the majority of activity took place in February, was UK based and local to the museum. Time access also tended to be in the evening and while the survey did not capture specific days of the week the logs pointed towards mid-week activity e.g., Tuesday, Wednesday and Thursday.

Accessing the NML website pages, according to the transaction logs, generally

revealed a high percentage of single page access requests that focused on primarily on the General, Item and Event pages, (see Table 6.12) possibly in preparation of a physical visit. Starting points of the NML website pages aligned to the survey findings and mainly focused on the main 4 museums' and Galleries e.g., Walker, WML, Maritime, ISM, although minor reordering could be seen in the repositioning of the ISM from fourth in the survey to first in the transaction logs. The amount of time spent previewing the website pages also revealed relatively limited interaction, e.g., 10 seconds or less, which could be an indication that information was readily found but could also suggest that the NML website was difficult to navigate or tedious for the users.

Inspection of the query constructs was less revealing. The majority of searches consisted of either one, two or three words. The use of intentional Boolean operators was not evident with the use of "AND" to form the creation of a question (see Table 6.17).

Transaction log analysis identified a number of documented activities in the server logs that align with known user group activities and behaviours. Using only the types of pages visited as features for cluster analysis (see Table 6.28) the resulting groups are similar to user categories previously discovered in the literature e.g., Non-professionals (Hobbyists), General Public, Teachers, Museum Staff, Academics (experts), Students. However, because of the multiple categories that each of the user groups could be allocated to, no specific determination can be made.

However, Grimes et al. (2007) discusses how the query log alone is not enough to be able to fully understand the users as they only allow the understanding of how users do what they do and what the users do, but they are unable to identify why users are doing what they are. In order to establish why users do what they do, Grimes et al. (2007) performed formal usability tests (field/lab studies or instrument panels; explore and question tests) to aid this missing understanding.

For the purposes of this research, the why was explored through the use of the survey (Phase 2) instead of usability tests. The reasons for adopting the survey over the usability tests was due to the fact that the users of the NML are from all over the globe and also are being drawn to the NML websites to undertake an unknown vast range of tasks for a range of unknown reasons. This creates difficulty in recruiting remote participants from all different times zones and also conducting usability tests

remotely with the global audiences, as well as creating valid usability tests with the unknown tasks and reasons the audience are visiting with. During the survey, questions were asked about the users' reasons for their visit, tasks and experiences; however, in Phase 2, it was found that the groups often were not operational as they often overlapped with other groups, and the definitions and labels used were not unique or descriptive enough. The sample in some of the groups also showed that for many of the questions there were inconsistencies, so much so that in some questions the same had some users claim to be at opposite ends of the scales to others in the same group. In this phase, whilst, it was possible to extend three of the group definitions slightly, it is not felt that these extensions have aided in separating the groups to a point that they become operational. Therefore, maybe trying to define these labelled groups is actually the wrong approach.

In Phase 4 a different bottom-up approach will be adopted and instead of focusing on the labelled groups that have been used throughout the literature and that the survey users self categorised into, the focus will be given to all the other questions and the categorisation labels ignored. By undertaking a clustering approach and establishing what groups naturally emerge from the survey results (excluding Q22) more realistic and operational groups may appear.

Chapter 7

Phase 4: CLUSTERING USERS FROM NML SURVEY RESPONSES

7.1 Introduction

Phases 1-3 aimed to identify and characterise user group labels established in previous studies. The literature review and Phase 1 emphasised the limited scope of previous website user studies, focusing primarily on specific user categories e.g., professionals and hobbyists, and aspects like motivation or experience relevant to them. Phase 2 identified NML website user categories through an online survey, revealing a prevalence of General Public and Non-Professional users, which were largely understudied in previous literature. However, descriptions for all user groups were either sparse or contained gaps, rendering them incomparable and impractical for industry use or future research contextualisation. Phase 3 attempted to identify user groups within the web-server log data and expand the user group label definitions through transaction log analysis. However, while it was possible to marginally extend three of the established group definitions e.g., General Public, Non-Professional, Museum Staff, it was not to the point of being able to operationalise such characterisations. Although, despite the lack of actionable or contextual descriptions for the user groups, this study has instilled confidence that the survey data is representative of the overall user population.

The goal of this phase is to continue the exploration of user groups and establishing characteristics, both currently held and naturally occurring, through the adop-

tion of data-driven approach that exams the accuracy of self-classification through the use of unsupervised machine learning cluster analysis by answering the following research question:

Phase 4 Research Questions:

RQ6: What are the patterns and characteristics of user groups as determined by cluster analysis of survey data?

Following the methodological framework established in Chapter 3, this chapter delves into the practical execution of these methods, offering a comprehensive view of their operationalisation. The selection of PAM and Agglomerative Hierarchical (Hclust) as primary clustering algorithms for this phase was based on empirical evidence, specifically Silhouette and Dunn Index scores, providing a robust basis for determination (see Chapter 3 Methodology, Section 3.4.4). This dual approach demonstrates their respective advantages through the actual performance metrics, the rationale for employing PAM for its deterministic clustering capabilities and Hierarchical (Hclust) for its insights into structures becomes evident, reflecting the nuanced and methodical decision-making process in the survey data's cluster analysis.

The remainder of the Chapter is organised as follows: Section 7.2 Multivariate analysis; Section 7.2.2 Cluster Analysis; Section 7.3 Results and Analysis; Section 7.5 Discussion and finally Section 7.6 Summary.

7.2 Multivariate Analysis

To set the stage for effective cluster analysis, multivariate analysis is performed as a crucial preliminary step. Multivariate analysis offers a reduction in bias and provides a result that could be considered closest to reality (D, 2021). Several techniques, such as Correspondence Analysis or Multiple Correspondence Analysis (MCA), were therefore considered as a means to manipulate this data. According to Abdi and Valentin (2007) MCA is an extension of correspondence analysis or a generalisation of principal component analysis (PCA) that allows for patterns of relationships between several categorical dependent variables, while Kassambara (2017a) points

out that MCA is generally used to analyse a data set from a survey and that the goals are to identify:

1. A group of individuals with similar profiles in their answers to the questions.

2. The associations or relationships between variable categories.

The group data (Q22) column was not processed in the MCA as it was deemed potentially biased and supplementary. It was, therefore, only used to colour the individuals by groups and to divide the self-selected General Public from non-general public to enable further exploration of the General Public group.

Prior to further analysis, dimensionality reduction using MCA was run with categorical variables. This reduction is justified as it simplifies the dataset by reducing the number of variables while retaining the essential relationships, making the subsequent cluster analysis more efficient and interpretable.

7.2.1 Running MCA (Multiple Correspondence Analysis)

MCA was used to identify potential relationships between variables, as well as lower the number of dimensions that can represent the variability in the data-set without losing important information (Brickey et al., 2010). The analysis was performed on 487 individuals, described by 8 variables. This generated 25 dimensions ranging from a percentage variance of 0.98 to 8.68.

Using Dimension 1 and Dimension 2, as these had the highest level of percentage variance, a plot of the variable categories was undertaken (see Figure 7.1).

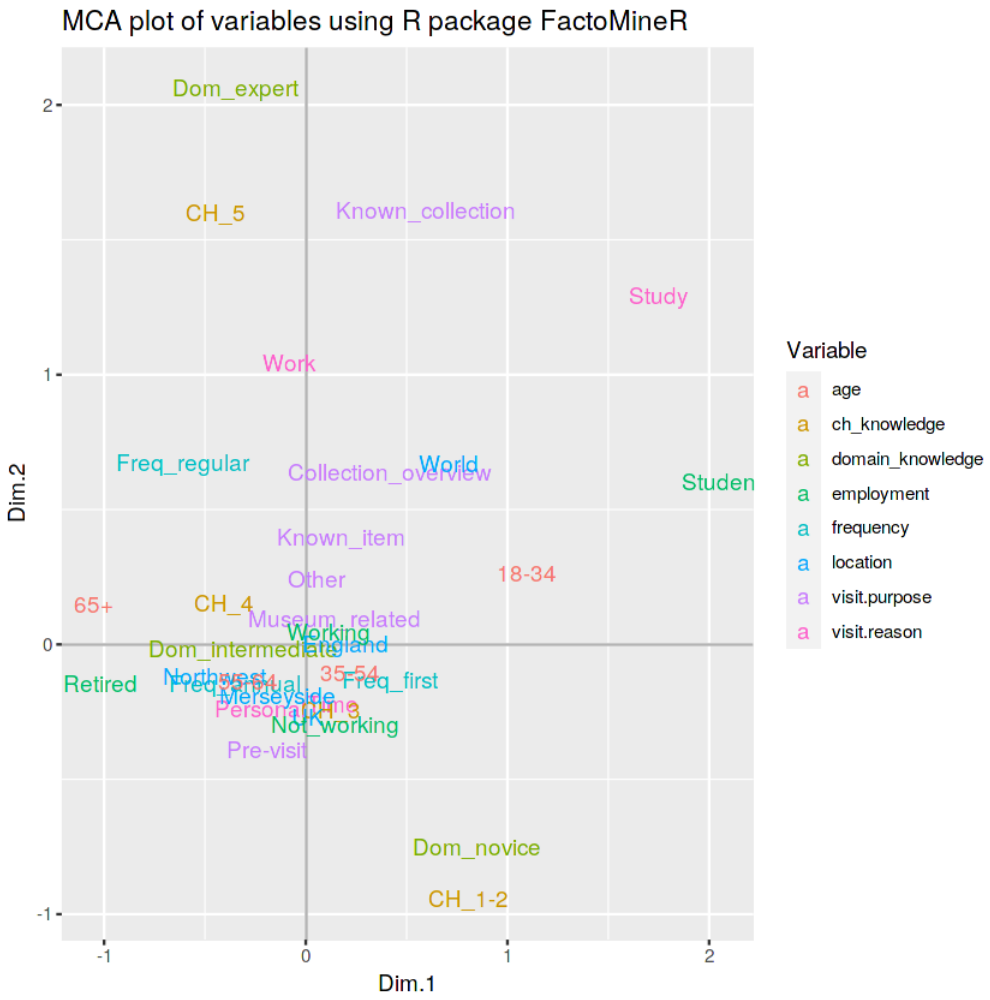


Figure 7.1: MCA plot showing grouping of individual variable categories on first 2 dimensions with the highest variance.

Figure 7.1 shows how each variable relates to the first two dimensions. Dimension 1 appears to be more defined by a novice level of domain knowledge and also a low level of general CH knowledge. Dimension 2, however, appears to be described by the polar opposite, with an expert level of domain knowledge and the highest level of general CH knowledge.

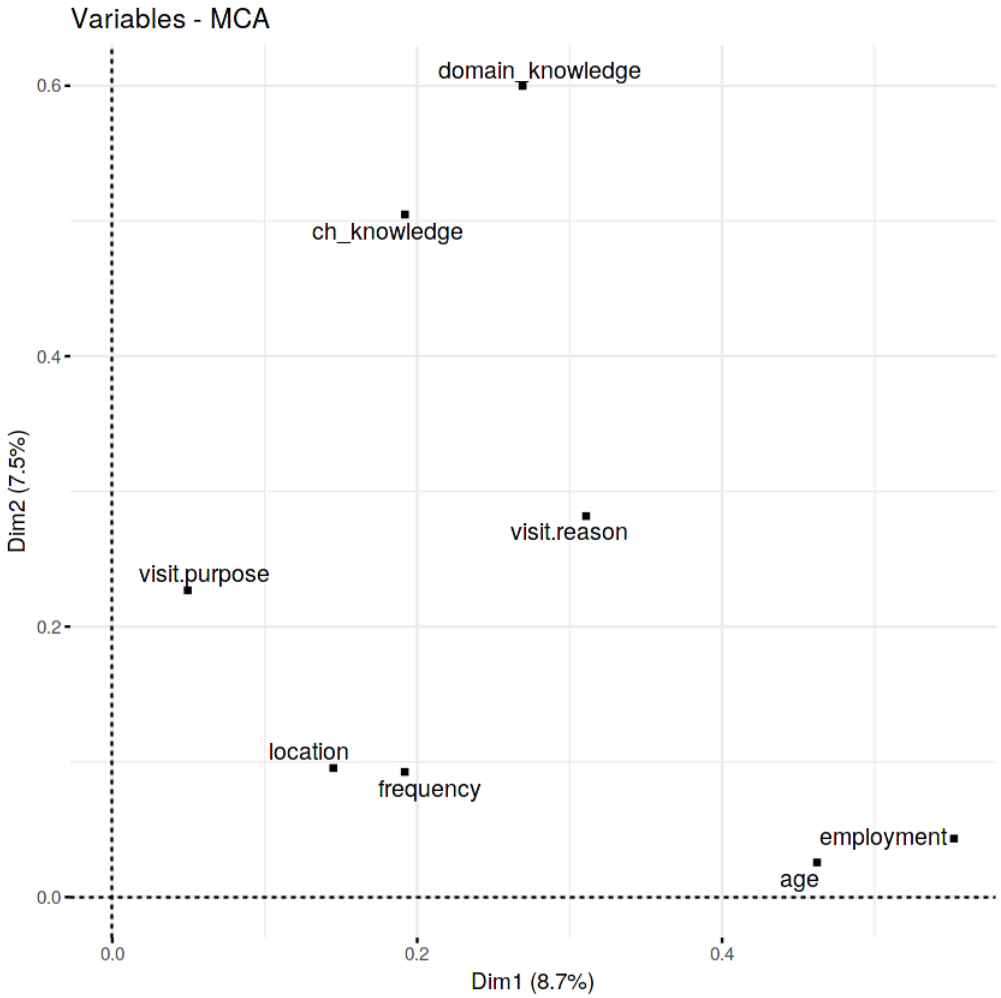


Figure 7.2: MCA plot showing the grouping of individual variable categories on first 2 dimensions with the factors shown.

Figure 7.2 shows that the factors of Age and Employment are most associated with Dimension 1. Whilst the factors of Domain knowledge and CH knowledge are more associated with Dimension 2. The factors of location, frequency, visit reason and visit purpose appear to be pretty neutral.

A factor map was then plotted (see Figure 7.3) to show the variables projected onto the principal components. The plot helped to identify variables that are most correlated with each dimension. The squared correlations between the variables and the dimensions were used as coordinates.

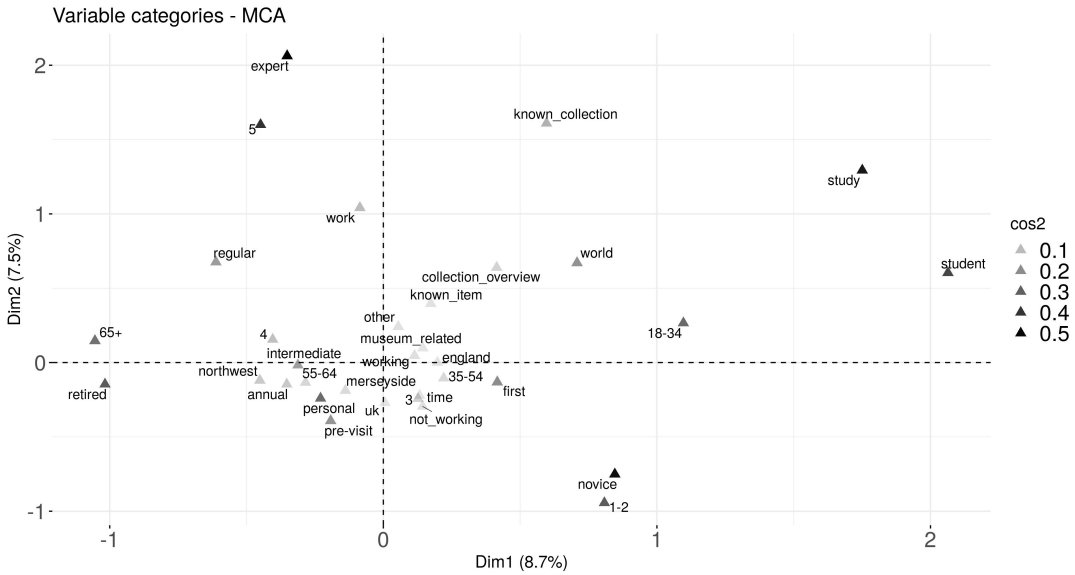


Figure 7.3: MCA plot showing grouping of individual variable categories on first 2 dimensions

The first 5 dimensions account for 34.8% of the variance in the data. Figure 7.3 shows an MCA plot for individual variable categories on the first 2 dimensions (representing 16.2% of variance), with the shading of the points representing their squared cosine (cos2) score - this measures the degree of association between variable categories and a particular axis. The plot confirms that variable categories with a similar profile are grouped (e.g., e.g. 65+ and retired), that some variables are well represented on the dimensions (e.g., student, study, retired) and that some variables are negatively correlated and positioned in opposing quadrants (e.g., expert and 5 vs. novice and 1-2).

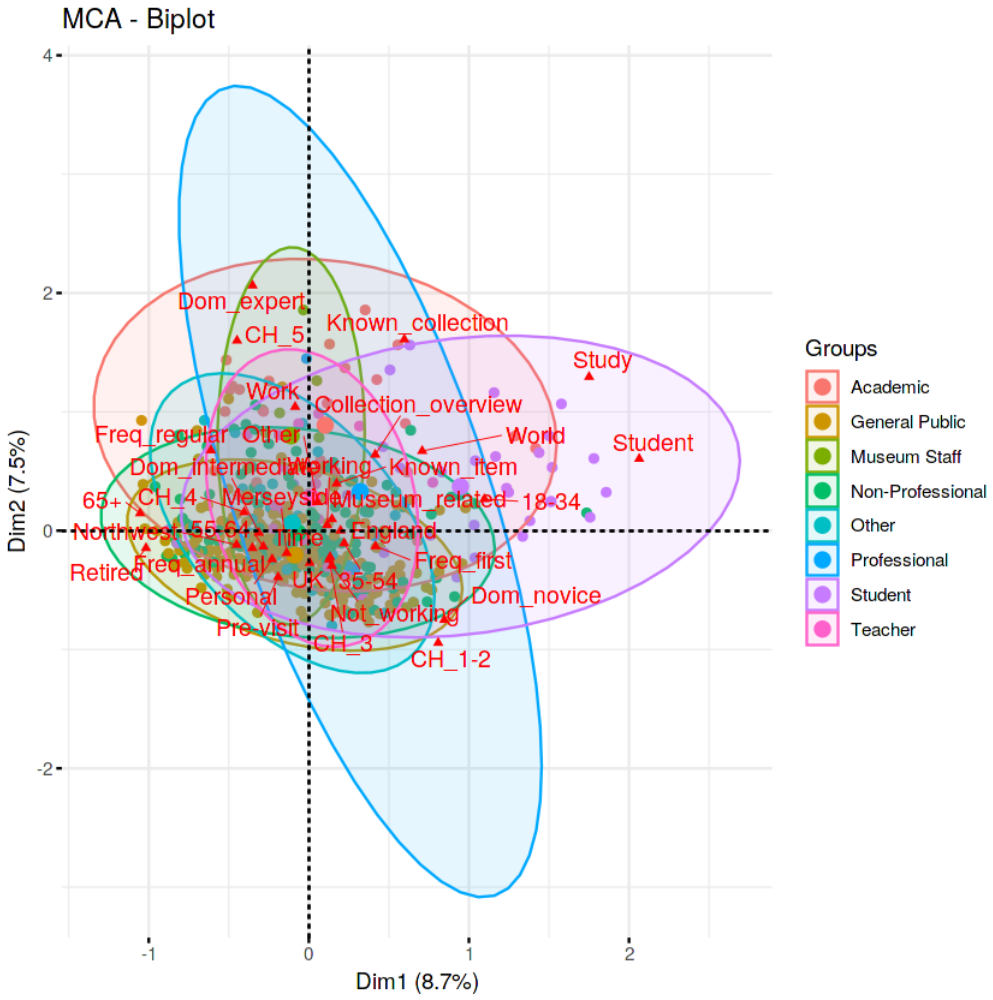


Figure 7.4: MCA plot showing the grouping of individual variable categories on first 2 dimensions, highlighted by self-selected groups.

When applying an overlay of the self-selected groupings (see Figure 7.4, it becomes clear that there are no clear clusters and variable groupings. The plot shows that all the groups overlap each other considerably.

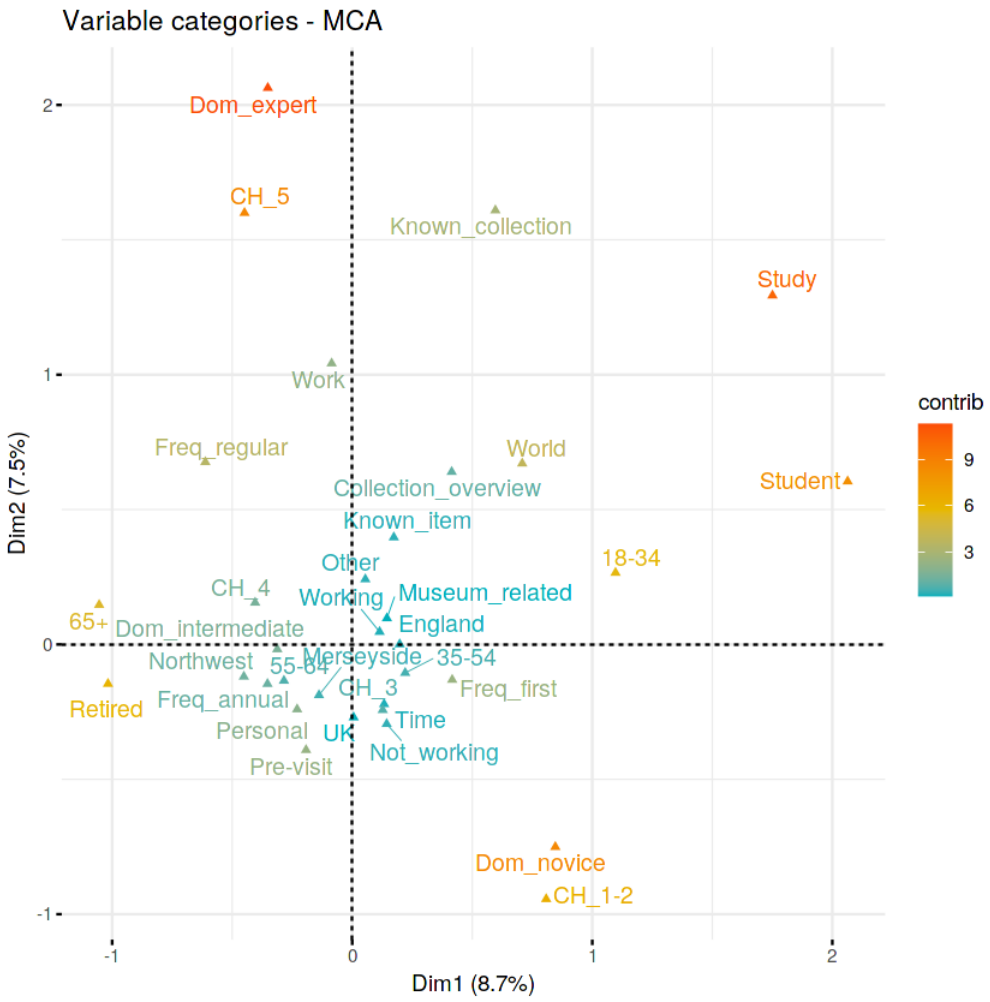


Figure 7.5: MCA plot showing the grouping of individual variable categories on first 2 dimensions, highlighted by contribution using a colour scale from blue to red.

Figure 7.5 shows a variation of the same diagram but with colour-coded variables against a scale. The scale ranges from blue to red. Blue means that the variable is not really contributing to the dimension, while those at the other end of the scale, the variables that are orange and red, are very significant to the dimension. The plot shows that the domain knowledge novice and CH knowledge levels of 1 and 2 are contributing significantly in a negative way to Dimension 1 whilst Study and Student are contributing to Dimension 1 significantly in a positive way. It also shows that domain knowledge of expert and CH knowledge of level 5 are positively contributing in a significant way to Dimension 2.

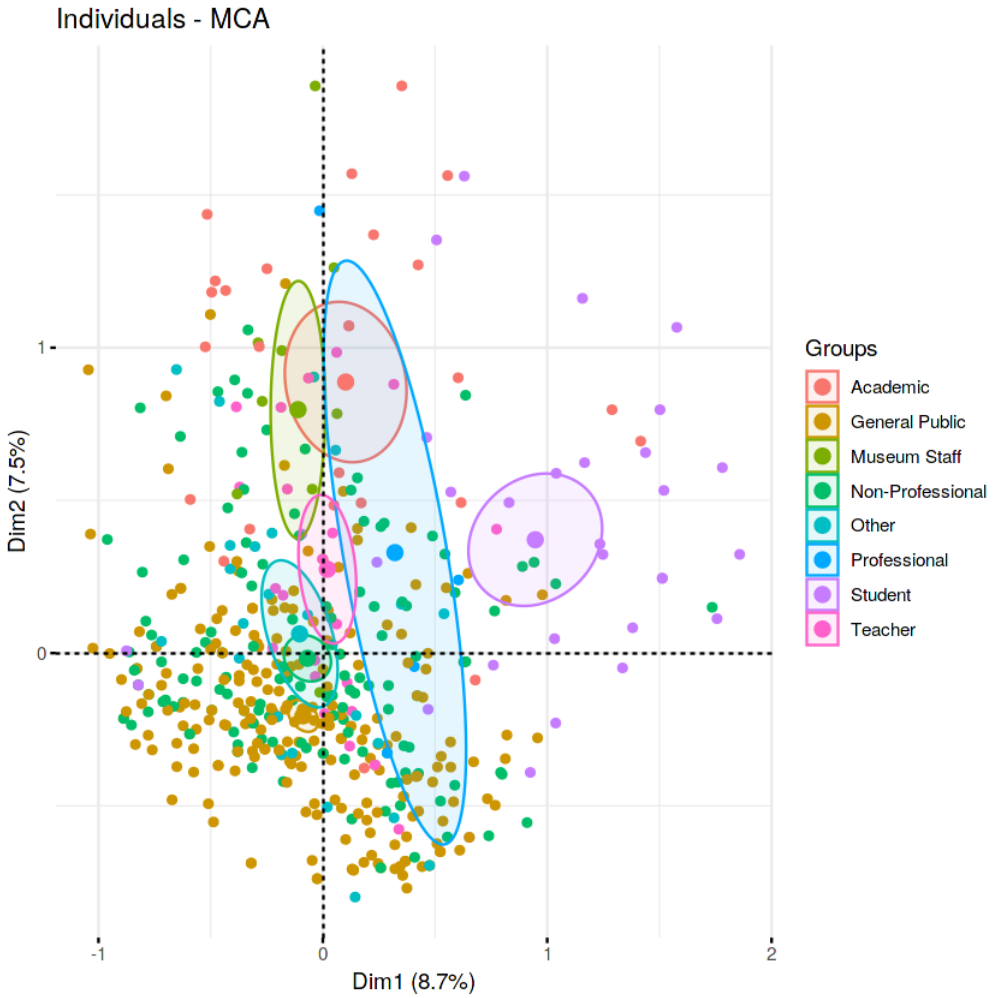


Figure 7.6: MCA plot showing the grouping of individual variable categories on first 2 dimensions, highlighted by self-selected groups with the addition of confidence ellipses.

Colouring each individual data point in alignment with the corresponding Q22 answers allows us to see if there are any groupings that could be based on the labels. The eclipses in Figure 7.6 show that the student group is significantly different from all other groups. It also shows that there are some levels of overlap between all the other groups. For example, The academic group is mostly covered by the professional group and partly covered by the Museum staff group. The Teacher groups stand semi-significantly, with only a slight overlap with the professional, non-professional and museum staff groups. There does not seem to be much difference between the Other and Non-Professional groups. This suggests that the choice to omit the labels as a clustering facor was a wise choice, as there seems to be a vast spread of points away from the centroid and the distance between the points represents the degree

of association between the variables.

7.2.2 Cluster Analysis Overview

This section will discuss the clustering process now the MCA has been conducted.

7.2.2.1 Assessing Cluster Tendency

Following the completion of the MCA, an initial check was made to establish whether the data could be clustered through the use of a dissimilarity matrix.

Table 7.1: Most similar pair in Gower matrix

Row	Visit Reason	Visit Purpose	Frequency	Domain Knowledge	CH Knowledge	Location	Age	Employment	Group
475	Study	Museum_related	Freq_first	Dom_novice	CH_3	World	18-34	Working	Student
1	Study	Museum_related	Freq_first	Dom_novice	CH_3	World	35-54	Working	Non-Professional

Table 7.2: Most dissimilar pair in Gower matrix

Row	Visit Reason	Visit Purpose	Frequency	Domain Knowledge	CH Knowledge	Location	Age	Employment	Group
35	Study	Museum_related	Freq_first	Dom_novice	CH_3	Merseyside	18-34	Student	Student
20	Personal	Pre-visit	Freq_regular	Dom_expert	CH_5	World	65+	Working	Academic

Table 7.3: Summary of results from the dissimilarity matrix.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0000	0.3750	0.4792	0.4866	0.6042	1.0000

Table 7.1 shows the most similar pair of data rows in the Gower dissimilarity matrix, with factors for group labels Student and Non-Professionals being tightly aligned in all but 1 of the variables, namely age. Table 7.2 shows a pair of rows of data in the Gower matrix that are the most dissimilar across the majority of variables, except for domain knowledge. Table 7.3 shows the results from the dissimilarity matrix with all 487 objects. All variables are treated as nominal (N) and ordinal (O) as expected, and the distribution, whilst very slightly positively skewed, is so slight it is negligible.

An attempt was also made to measure the probability that the data set is generated by a uniform data distribution. This can be achieved using statistical tests for spatial randomness. Hopkins Statistic was used to illustrate this idea. The Hopkins Statistic is a spatial statistic that tests the spatial randomness of a variable as

distributed in a space. If H < 0.5 then it can be concluded that the distribution is non-uniform (Han et al., 2012) e.g., 0.1827 the score is well below 0.5 which suggests there may be clusters in the data set. Undertaking a visual inspection was then made.

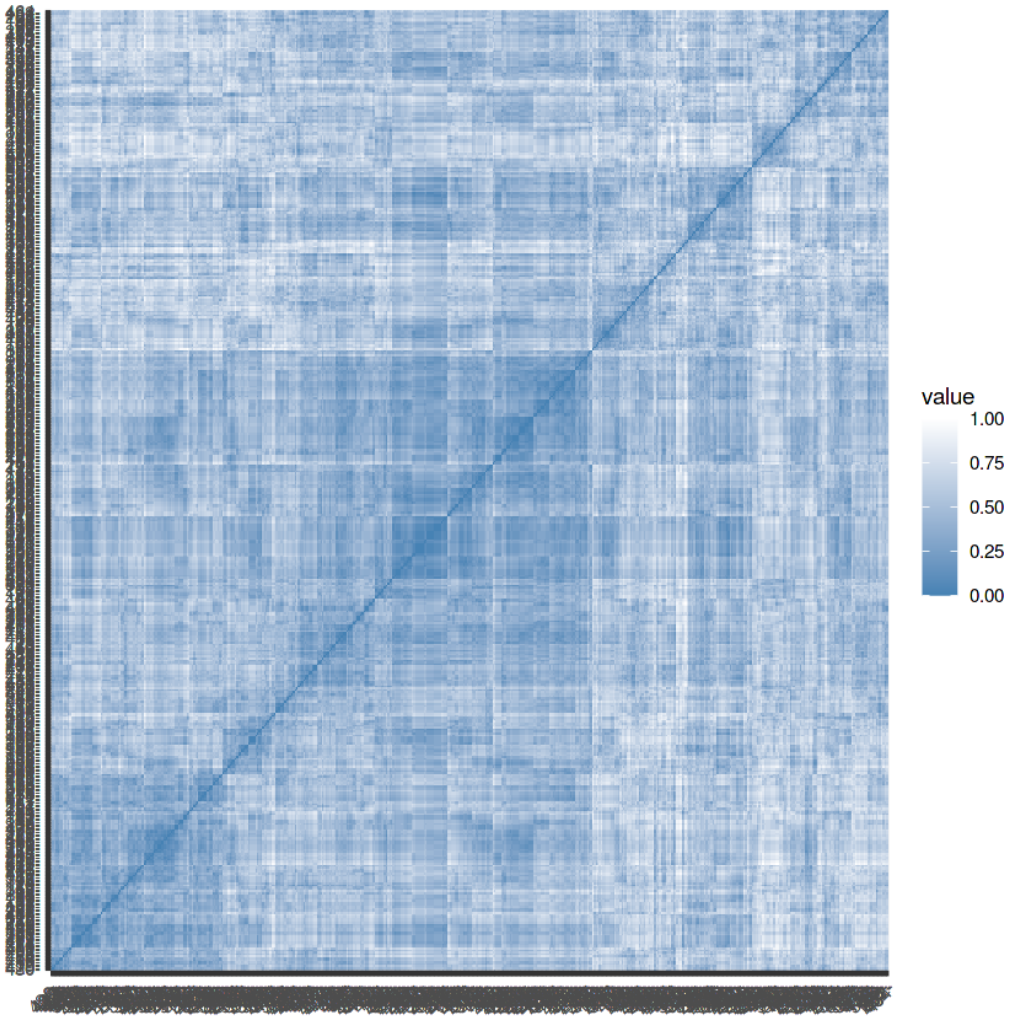


Figure 7.7: Visual assessment of cluster tendency.

Figure 7.7 shows a visual representation of the dissimilarity matrix. The plot has each data point represented on both the left side and across the bottom. In the plot, where each pair of points meets is a coloured square. The colour of the square is dependent on the level of similarity between the pair. The more similar the darker the blue colour, the less similar the whiter the colour. A horizontal Dark Blue line moving across the plot showing where each object is paired with itself. There are a few darker rectangles that appear in the plot which suggests that there may be some cluster structure in the data, but because it is not a dense square of blue the

clusters may not exist. Therefore, it may be worth attempting an alternate method.

Another visualisation technique that was performed in order to establish whether high dimensional data exhibits enough clustering was to use Non-linear dimensionality reduction using t-Distribution Stochastic Neighbour Embedding (t-SNE).

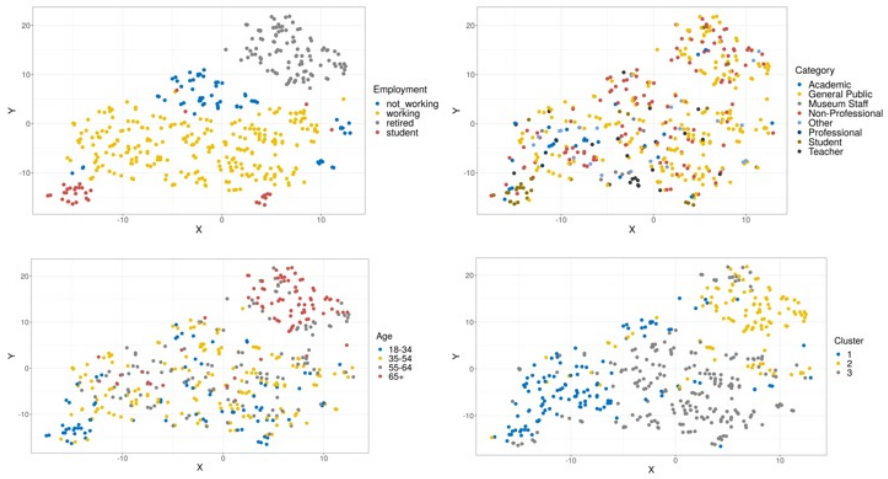


Figure 7.8: 2D t-SNE plots showing users for selected categories, together with self-assigned groups (top right) and PAM clusters for all users (bottom right)

This allowed for the visualisation of data in a lower dimensional space, such as 2D, to identify patterns and trends Van der Maaten and Hinton (2008). Figure 7.8 shows various t-SNE plots for selected categories (employment, age, visit purpose and location). The plot clearly seems to suggest that at least 2 groups may exist within the data. The group in the top right would seem to be clearly representative of the retired users (see the plot in the top right of Figure 7.8, where points are coloured by employment).

7.2.2.1.1 Clusterability Evaluation

The most similar pair is very close, the least similar pair is vastly different, and the Hopkins statistic is less than 0.5, all suggesting that the data is clusterable. However, the visualisation of the dissimilarity matrix was not very conclusive, as there were only very faint blocks visible. The t-SNE plots, however, did suggest that there could be some clusters within the data. Therefore, the results suggest that clustering would be possible.

7.2.2.2 Running the Algorithms

This section explores how clustering algorithms were utilised to discover meaningful patterns/user-groups in the survey dataset. Both the K-Medoids (PAM) and Hierarchical algorithms showed promise in the initial analysis (see Chapter 3), which therefore prompted further investigation into the user groups they generate.

7.2.2.2.1 Partitioning Around Medoids (PAM) Clustering

A common K-Medoids partitioning algorithm, such as PAM (Partitioning Around Medoids) (Kaufman and Rousseeuw, 2009), is a partition-based algorithm that operates similarly to k-means clustering, but cluster centres are constrained to being the observations themselves. (e.g., medoids). In comparison to k-means, the technique is more resistant to noise and outliers, and it also has the advantage of using one observation as the example for each cluster, making cluster interpretation easier.

Stage 1: Establish Optimal Number of Clusters

In order to establish the optimum number of clusters, approaches including Elbow, Silhouette and Gap, were used to determine the number as well as evaluating the quality of said clusters.

Utilising the Elbow method (see Figure 7.9), the inertia is a decreasing function of the number of clusters k. However, its rate of decrease changes noticeably around the optimal number of clusters. The bend, or "elbow", suggests that additional clusters beyond the third have diminishing value. Although this method indicates 3 as the optimal number of clusters, it can be somewhat subjective, as the elbow point might appear at different locations or not be clearly defined. Other possible but less likely cluster numbers include 5 and 9.

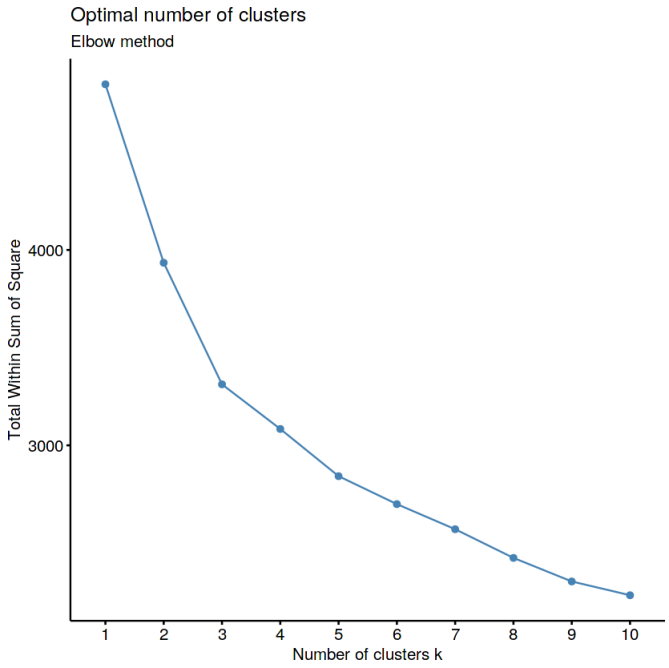


Figure 7.9: PAM elbow method to determine the optimal number of clusters.

The optimal number of clusters found using the Silhouette method (see Figure 7.10) is performed using K-Means clustering over a range of k and finds the optimal K that produces the largest silhouette coefficient and assigns data points to clusters based on the optimised K. In this instance, Figure 7.10 indicated the optimal number of clusters in the data as 3.

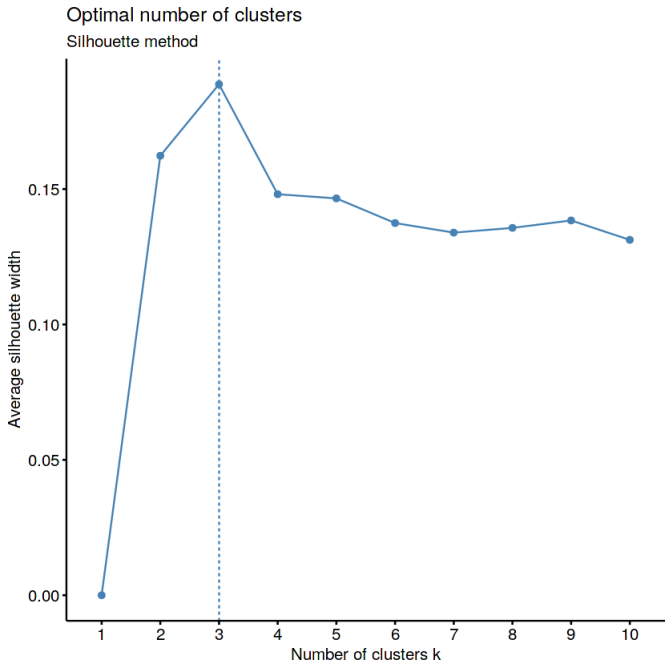


Figure 7.10: PAM silhouette method to determine the optimal number of clusters.

The optimal number of clusters determined by the GAP statistic approach (see Figure 7.11) compares the total inside intra-cluster variation for different values of k with their predicted values under the data's null reference distribution. The value that maximises the gap statistic will be used to estimate the best clusters. (i.e., that yields the largest gap statistic). This signifies that the clustering structure deviates significantly from a random, uniform distribution of points.

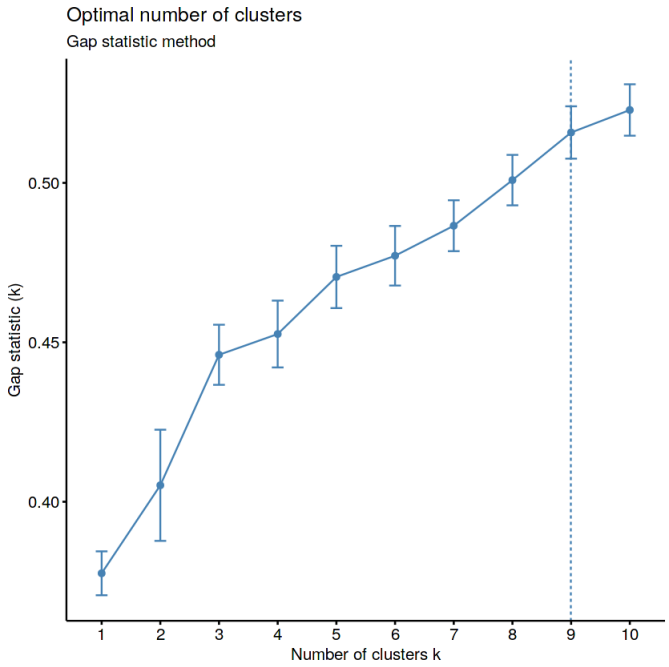


Figure 7.11: PAM gap statistic method to determine the optimal number of clusters.

The gap stats plot shows the statistics by the number of clusters (k) with standard errors drawn with vertical segments and the optimal value of k marked with a vertical dashed blue line. According to this observation, $k = 9$ is the optimal number of clusters in the data. While the elbow plot suggests 3 as the optimal number of clusters due to the noticeable bend, it also shows a less pronounced but significant change at 4 clusters. The gap statistic method, however, provides a more granular optimal point at 9 clusters. This discrepancy indicates that while 3 clusters offer a simpler segmentation, the data may benefit from a more detailed partition into 9 clusters, capturing finer distinctions among the data points. Additionally, both methods show agreement at 3 clusters as a significant change point, suggesting its robustness as an option.

Therefore, with two of the three methods suggesting that 3 clusters were optimal, it was prudent to proceed with the 3 clusters.

Stage 2: PAM Clustering Implementation

The results of applying a PAM clustering algorithm with three clusters to the dataset provided the following results:

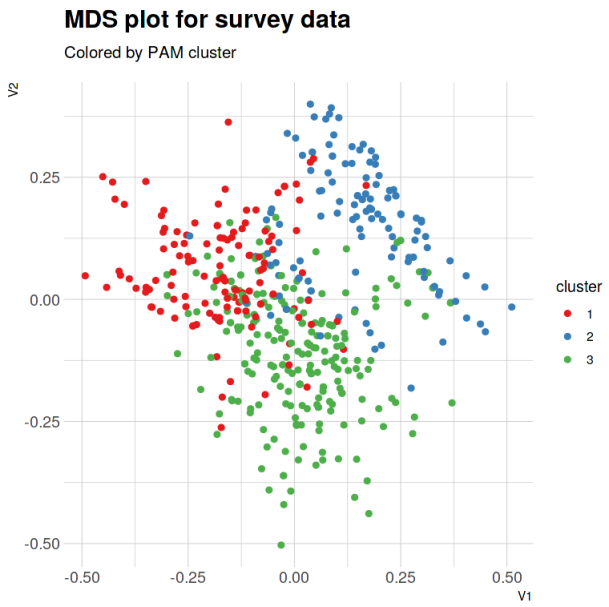


Figure 7.12: PAM MDS plot showing clusters in 2D.

The MDS plot in Figure 7.12, gives a good visual representation that the data clusters, but in the 2D plot the separation is not clean cut. Even though the clusters appear to be overlapping, this does not mean that the clusters are when in a higher-dimensional space.

Table 7.4: PAM cluster results information.

Cluster	Size	Max_diss	Av_diss	Diameter	Separation
1	125	0.6458333	0.3150000	0.9166667	0.04166667
2	126	0.5833333	0.2693452	0.8958333	0.04166667
3	236	0.6666667	0.2693326	0.9583333	0.04166667

Table 7.4 shows that clusters are split into reasonably sized clusters, with cluster 1 and 2 being of a very similar size and cluster 3 being the largest. The Max distance parameter, which specifies the maximum dissimilarity between

any two medoids in a cluster, is quite low, meaning that the clusters are more similar to each other. Values closer to 0 indicate more similar data points, and values closer to 1 indicate more dissimilar data points. The av_diss parameter refers to the average dissimilarity between medoids in a cluster. The values returned here are all less than 0.4, indicating that the quality of the clusters are higher as the data points have a low dissimilarity. All the diameter results are good (over 0.895), indicating that the data points in each of the clusters are closely related to the other data points in that cluster and have a low level of dissimilarity. The separation scores are all 0.041, indicating that the data points in the cluster are closely related and have a low level of dissimilarity. Despite the overlap in the MDS plot (Figure 7.12, These cluster results suggest potentially suitable clusters.

Table 7.5: Medoid representative rows of each cluster in PAM results.

Cluster	Medoid (Row)	Visit Reason	Visit Purpose	Frequency	Domain_Knowledge	CH_Knowledge	Location	Age	Employment	Group
1	483	Personal	Other	Freq_first	Dom_intermediate	CH_3	World	35-54	Working	Non-Professional
2	154	Personal	Pre-visit	Freq_first	Dom_intermediate	CH_4	Northwest	65+	Retired	Non-Professional
3	28	Personal	Pre-visit	Freq_annual	Dom_intermediate	CH_3	Merseyside	35-54	Working	Non-Professional

Table 7.5 shows the rows of data for the medoid data point for each cluster. There appears to be a number of factors that are the same in each cluster, suggesting that those factors (Visit reason, domain knowledge) are not distinguishing factors in this clustering. However, location, age, employment CH knowledge, and frequency appear to be. The group column was not a factor used for clustering but has been added after to see how the self-classification aligned with the cluster results. In this case, it is clear that the clusters do not map to the self-groupings at all.

Stage 3: PAM Clustering Evaluation

Table 7.6: PAM Clusters mapped against the self-assigned groups.

Cluster	**1**	**2**	**3**
Academic	11	5	9
General Public	40	72	124
Museum Staff	1	0	9
Non-Professionals	41	42	51
Other	8	6	10
Professional	4	0	0
Student	13	1	17
Teacher	7	0	16

The cluster medoids in Table 7.5 gave a clear indication that the PAM clusters did not map well to the self grouping of the participants of the survey made in Q22. Table 7.6 shows the sum of all self-grouping answers for all data points across all the self-assigned groups. Unlike the medoid, these holistic results suggest that whilst most groups are represented in all clusters, there could be a slightly more weighted cluster for some types of groups. e.g. cluster 2 has less professional and work style groups, consisting mostly of General Public and Non-Professional. Whereas clusters 1 and 3 are much more of a general mix of all groups.

This discrepancy is to be expected as self-classifications often differ from algorithmic clusters due to subjective perceptions and diverse criteria used by participants in self-assignment.

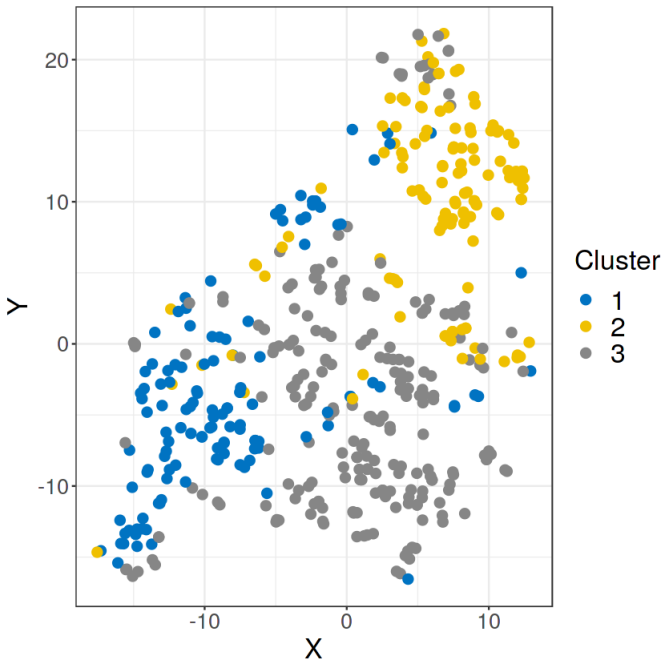


Figure 7.13: PAM clustering t-SNE plot.

The t-SNE plot (Figure 7.13) shows the cluster results as a 2D plot similar to the MDS plot (Figure 7.12). The difference between the two is the way they measure similarity between data points. MDS uses a metric of distances to measure similarity, whereas t-SNE uses a probabilistic approach to measure similarity. The t-SNE plot suggests something similar to the MDS plot in that the clusters do all have some overlap when plotted into the 2D space. The t-SNE, however, suggests that cluster 3 overlaps a little more with the other 2 clusters and, therefore, may contain some outliers. It could also suggest that with higher numbers of members representing all the self-assigned groups this is what we are seeing in this cluster, and so when plotted, it would spread across the whole plot.

Table 7.7: PAM clustering silhouette.

Cluster	Size	Ave.sil.width
1	125	0.12
2	126	0.2
3	236	0.15

Table 7.7 shows the cluster size and average silhouette score for each cluster. The average silhouette score is a measure of how well-defined a cluster of data

points is. It is calculated by taking the average silhouette value of all the data points in a cluster. A score of 0 indicates overlapping clusters and a negative score indicates that data points may have been assigned to the wrong clusters. A higher average silhouette score indicates that the data points in a cluster are more similar to each other and are more distinct from the data points in the other clusters. Generally, a score between 0.4 and 0.7 is considered good and a score above 0.7 is considered excellent. A score below 0.4 is considered to be poor. However, what is considered a "good" or "bad" silhouette score depends on the specific context and the nature of the data being clustered. As can be seen in Table 7.7, the scores of all clusters are above 0. The result for cluster 2 is slightly better than the score for clusters 1 and 3. This indicates that the data points in the clusters are not very distinct from the data points from the other clusters. This could mean that the data points in the cluster are too close together and/or that the data points from the other clusters are too similar to the data points in the cluster. In other words, the distance between the clusters is not significant. However, all clusters have positive average silhouette widths, indicating that the clustering solution is reasonable.

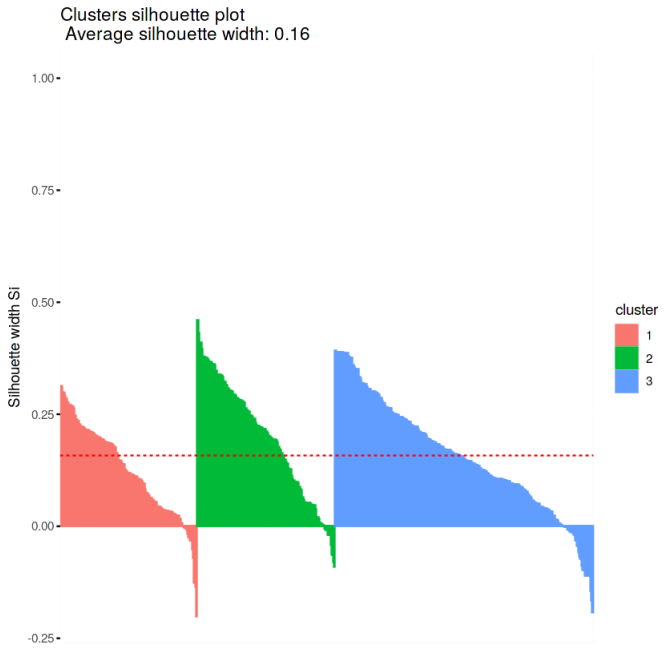


Figure 7.14: PAM clustering results silhouette.

Figure 7.14 provides a visual representation of the silhouette scores, and the

little tails (the negative Si) in each suggest the amount of data points possibly placed in the wrong cluster.

Table 7.8: All 3 PAM clusters values.

Cluster	**1**	**2**	**3**
Visit Reason			
Personal	65	107	175
Study	22	4	17
Time	21	10	27
Work	17	5	17
Visit Purpose			
Collection overview	12	9	13
Known Collection	10	4	5
Known Item	12	10	7
Museum Related	21	13	38
Other	58	20	16
Pre-Visit	12	70	157
Frequency			
Regular	5	14	63
Annual	14	34	106
First time	106	78	67
Domain Knowledge			
Novice	41	20	72
Intermediate	73	95	135
Expert	11	11	29
CH Knowledge			
1 or 2	23	16	42
3	49	35	107
4	39	59	55
5	14	16	32
Location			
England	29	20	40
Merseyside	3	21	147
Northwest	17	67	30
UK	8	7	17
World	68	11	2
Age			
18-34	29	5	53
35-54	54	13	108
55-64	38	38	58
65+	4	70	17
Employment			
Not Working	18	14	29
Working	87	19	174
Retired	5	89	17
Student	15	4	16

Table 7.8 provides a tabular breakdown of how the clusters are made up. It is clear that all clusters contain some values for every option from every question.

7.2.2.2.1.1 Classification In order to validate the clustering results, a Random Forest classifier was used for classification (Boehmke and Greenwell, 2020).

The Random Forest classifier, applied for classification in this study, harnesses the power of multiple decision trees, leveraging bootstrapped subsets of the data to construct an ensemble of trees (Boehmke and Greenwell, 2020; Breiman, 2001). The bagging technique integrates the trees' outcomes, often producing models with impressive predictive capabilities and requiring minimal hyperparameter adjustments.

A 70:30 partition of the dataset into training and test sets was undertaken for model training. Additionally, a 5-fold cross-validation, which segments the training data into five distinct portions to cross-validate the model, was employed and executed threefold. This approach aimed to refine the validation process, mitigating the risk of overestimating model performance and ensuring the robustness of the predictive accuracy across varied data segments. The model's hyperparameters of '*mtry*' and '*min.node.size*' were tuned via grid search, seeking an equilibrium to prevent overfitting and underfitting, thereby sharpening the model's predictive accuracy. The '*num.trees*' hyperparameter, delineating the count of trees within the forest, was investigated to determine its ideal value for both computational efficiency and prediction precision. Model evaluation was based on Accuracy and Kappa statistics, with the observation that additional trees beyond 200 did not markedly improve model performance. The Accuracy and Kappa peaked at 50, 200 and 400, with Accuracy and Kappa registering at 0.9722 and 0.9558, respectively (see Table 7.9). Notably, the model's performance showed minimal fluctuations up to 500 trees, underscoring the model's resilience against overfitting. The chosen hyperparameter of 400 trees effectively ensures model stability while conserving computational resources. The exploration for the number of trees commenced at a baseline, which was recommended as tenfold the number of dataset features, with incremental adjustments leading to the final choice of 400. This decision was based on the analysis revealing that additional trees beyond this point yielded no improvement in accuracy. This strategy aligns with the Random Forest methodology, where an excessive number of trees may not significantly enhance model performance and can increase

computational load (Breiman, 2001).

Table 7.9: Random Forest Classifier Number of Trees and Corresponding Accuracy

Number of Trees	**Accuracy**	**Kappa**
50	**0.9722222**	**0.9558011**
100	0.9652778	0.9451679
150	0.9652778	0.9448909
200	**0.9722222**	**0.9558011**
250	0.9652778	0.9446111
300	0.9652778	0.9451679
350	0.9652778	0.9448909
400	**0.9722222**	**0.9558011**
450	0.9652778	0.9448909
500	0.9652778	0.9448909
...	...	...

7.2.2.2.2 Hierarchical Clustering

In contrast to K-medoids (PAM), hierarchical clustering creates a hierarchical cluster structure where the number of clusters does not have to be pre-specified. This can be calculated from either a top-down (divisive) splitting of the clusters (one to many), or bottom-up (agglomerative) merging of the clusters (many to one).

The output is in the form of a hierarchy, which may help identify sub-groups within clusters.

Stage 1: Establish if clusterable and how many clusters (k) would be optimal

Similar to undertaking PAM clustering, the elbow, silhouette, and gap approaches were used to establish the optimum number of clusters for hierarchical clustering.

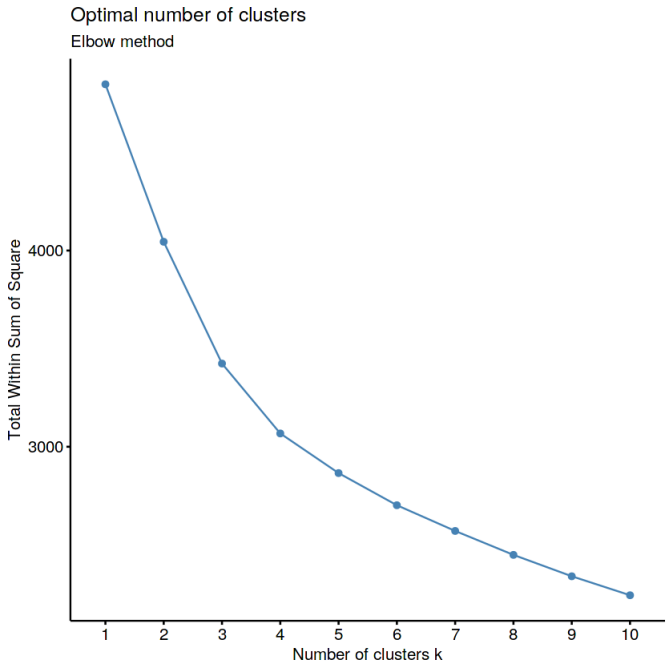


Figure 7.15: Hierarchical elbow method to determine the optimal number of clusters.

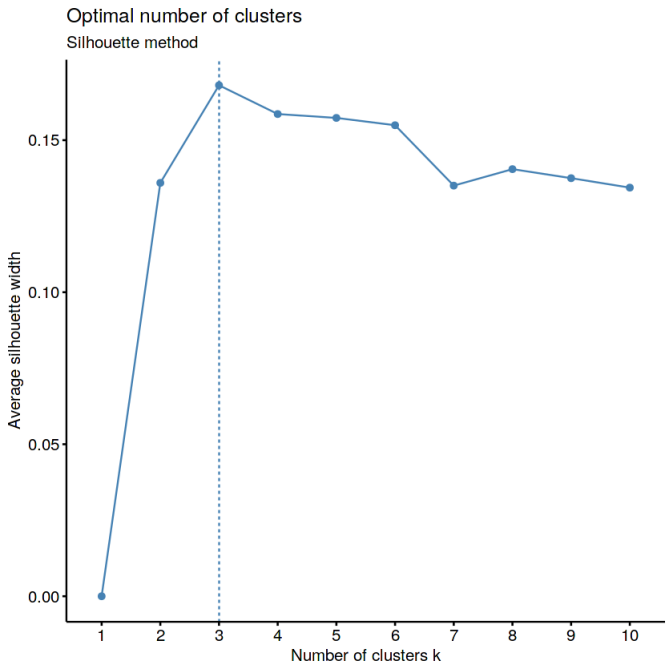


Figure 7.16: Hierarchical silhouette method to determine the optimal number of clusters.

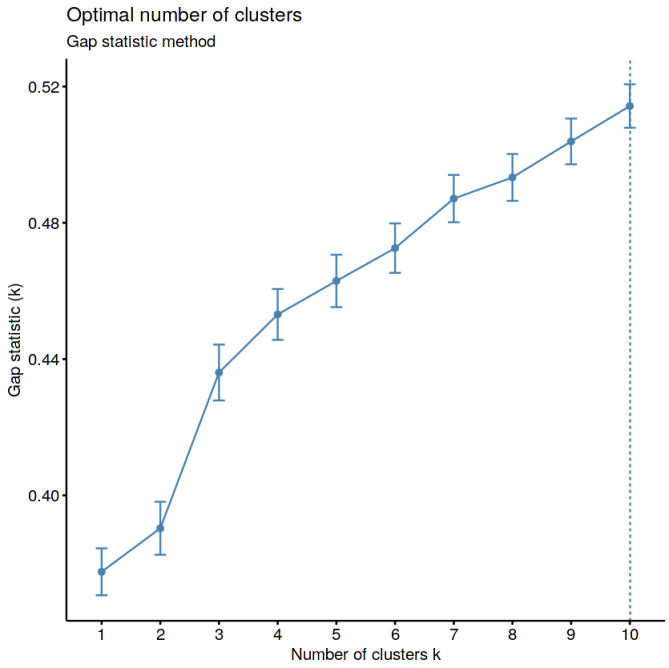


Figure 7.17: Hierarchical gap method to determine the optimal number of clusters.

The Elbow (Figure 7.15), Silhouette (Figure 7.16) and GAP statistic (Figure 7.17) methods all suggest that 3 clusters is the optimum number of cluster for this dataset.

Stage 2: Hierarchical Clustering

An advantage of hierarchical clustering over PAM (Partitioning Around Medoids) is that hierarchical clustering is able to identify clusters of different sizes and shapes, whereas PAM is limited to partitioning the data into clusters of equal sizes and equal densities. Additionally, hierarchical clustering is able to produce a dendrogram that can be used to visualise the relationships between clusters. Creating dendrograms for each type of linkage can also help with identifying the correct linkage method. A good dendrogram is one that clearly shows clusters of similar data points, with each cluster being distinct from the other clusters with the boundaries between the clusters clearly visible. A bad dendrogram is one where the clusters are not distinct or well-defined or where the boundaries between the clusters are not clearly visible. Additionally, a bad dendrogram may indicate that the linkage used to create it is not the best one for the dataset.

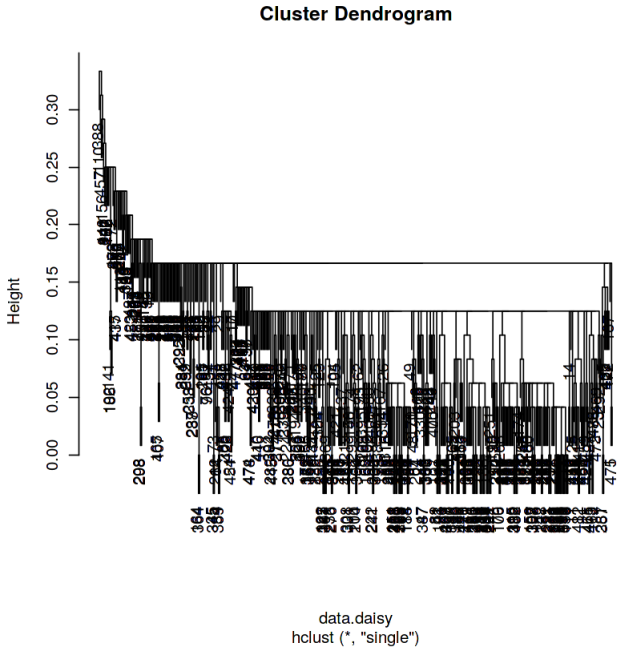


Figure 7.18: Hierarchical (hclust) single connectivity method.

Figure 7.18 shows a very tightly bound dendrogram for agglomerative hierarchical clustering using the single linkage (measures the similarity of the most similar pair or nearest neighbour (Clements, 2019)) connectivity. The dendrogram shows no clear clusters. Therefore, this may not be the correct linkage for this dataset.

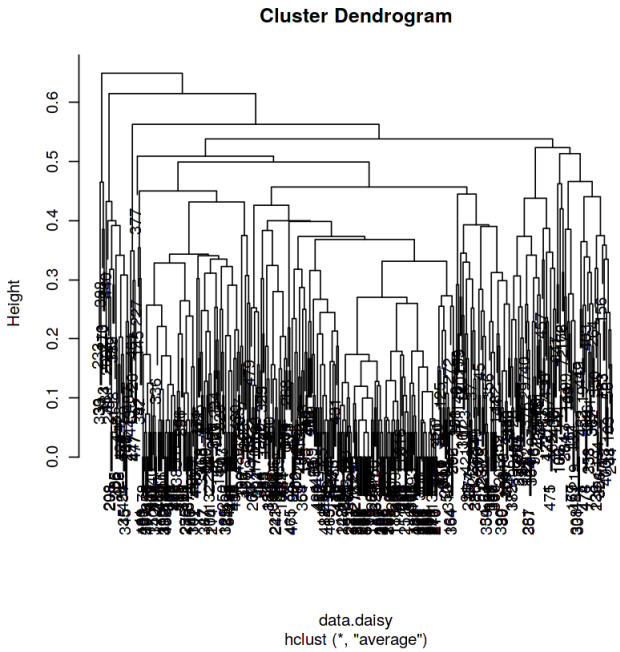


Figure 7.19: Hierarchical (hclust) average connectivity method.

Figure 7.19 shows a dendrogram for the average distance linkage. Clements (2019) states that the "average-linkage is where the distance between each pair of observations in each cluster are added up and divided by the number of pairs to get an average inter-cluster distance." This dendrogram is a less tightly bound dendrogram than that of the single linkage. There are also signs of clear clusters, but for 3 clusters that are suggested by the Elbow, Silhouette and GAP, this would mean that 2 of the 3 clusters would have very few data points (members), whilst cluster 3 would have the majority. Therefore, this may also not be the correct linkage for this dataset.

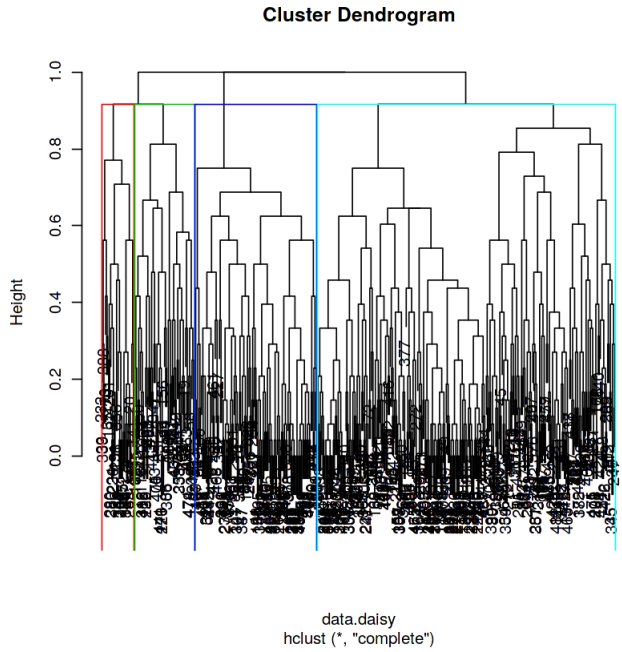


Figure 7.20: Hierarchical (hclust) complete connectivity method.

Figure 7.20 shows a dendrogram for agglomerative hierarchical clustering using the complete linkage. In complete linkage, the distance is measured between the farthest pair of observations in two clusters (Clements, 2019). The dendrogram for this method shows clear clusters. The 3 clusters are identified in the diagram by the coloured boxes. However, the height between branches in the dendrogram are all close together. The height represents the distance between two clusters. The greater the distance between two clusters, the higher the branch that separates them in the dendrogram. With the branches all being close together, it suggests that the clusters produced and not that distinct

from the other clusters.

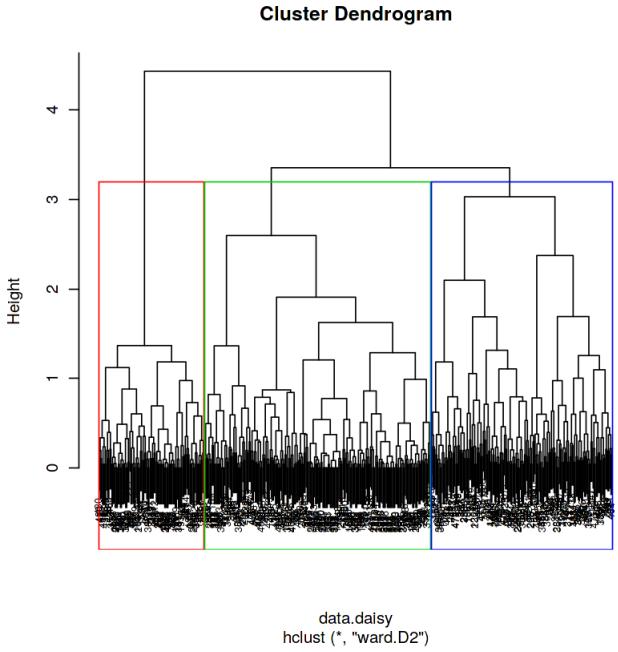


Figure 7.21: Hierarchical (hclust) ward connectivity method.

Figure 7.21 shows a dendrogram that has loose upper branches and also good height between the upper branches. The ward linkage is measured differently from single or complete linkage methods in that it uses an error sum of squares (ESS) criterion to determine the distance between clusters. This means that Ward linkage is based on the differences in the variance of the points in the clusters, whereas single and complete linkage is based on the distances between points in the clusters. Ward linkage tends to produce clusters with more homogeneous members, which can be useful when looking for clusters with similar characteristics (Yuxuan Hu and Meng, 2018).

The dendrogram shows a number of clear clusters and when having the tree cut at 3 clusters (represented by the coloured boxes), all the clusters appear to be of reasonable/comparable sizes and the branches have a good distance between them indicating the clusters are more distinct from one another than any of the other linkage methods. Therefore, this is the most suitable linkage method for this dataset.

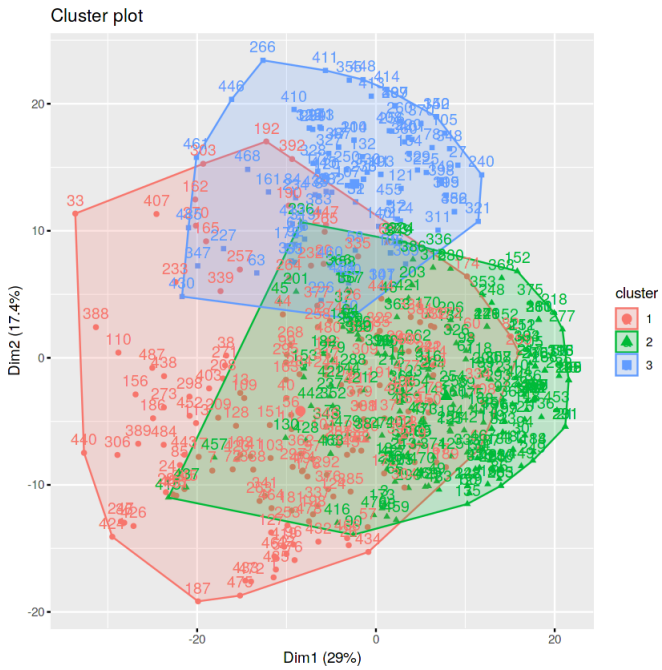


Figure 7.22: Hierarchical (hclust) factor map plot.

The above factor map plot (see Figure 7.22) Shows a 2D plot of the three clusters and also all data points, coloured by their cluster membership. It is clear that the clusters still have a lot of overlap when placed on a 2D plot, which is similar to the PAM method. So maybe the dataset does not allow for the creation of clean-cut and isolated clusters that would be ideal, but as the dataset is representative of people, maybe this is the correct representation that should be expected.

Table 7.10: Medoid representative rows of each cluster in Hierarchical results.

Cluster	Medoid (row)	Visit Reason	Visit Purpose	Frequency	Domain_Knowledge	CH_Knowledge	Location	Age	Employment	Group
1	239	Time	Other	Freq_annual	Dom_intermediate	CH_4	Merseyside	35-54	Working	General Public
2	21	Personal	Pre-visit	Freq_first	Dom_intermediate	CH_3	Merseyside	35-54	Working	General Public
3	105	Personal	Pre-visit	Freq_annual	Dom_intermediate	CH_3	Merseyside	65+	Retired	General Public

Table 7.10 shows the medoids for the 3 hierarchical clusters produced using the ward method. This set of medoids suggests that domain knowledge and location do not appear to be contributing factors to this clustering method. However, there are more factors here contributing than in the PAM method. The self-assigned groups are all for the same group, suggesting again that maybe the self-assignment is not a good separation method.

Stage 3: Hierarchical Clustering Evaluation

Table 7.11: Hierarchical Clusters mapped against the self-assigned groups.

Cluster	**1**	**2**	**3**
Academic	19	4	2
General Public	62	116	58
Museum Staff	10	0	0
Non-Professional	31	67	36
Other	10	12	2
Professional	3	1	0
Student	22	7	2
Teacher	15	8	0

Table 7.11 shows the 3 hierarchical clusters mapped against the self-classified groupings from Q22. Whilst, there is no clear mapping of a group to a cluster. Which is a similar result to the PAM method. However, this hierarchical method does allow for a little clearer interpretation of the clusters. It is possible to identify that cluster 3 has very little representation of professional style groups, such Museum staff, Professional, Teacher, Academic or Student. But instead is heavily dominant on General Public and Non-Professional. Cluster 2 has a similar makeup, but with some teachers and students. Cluster 1, whilst having the majority of the Academics, student and teachers as well as all the Museum staff, and 3 of the 4 Professionals it also has a good mix of the General Public and Non-Professional, and therefore it is not possible to define this group based on the self-assignment of group names.

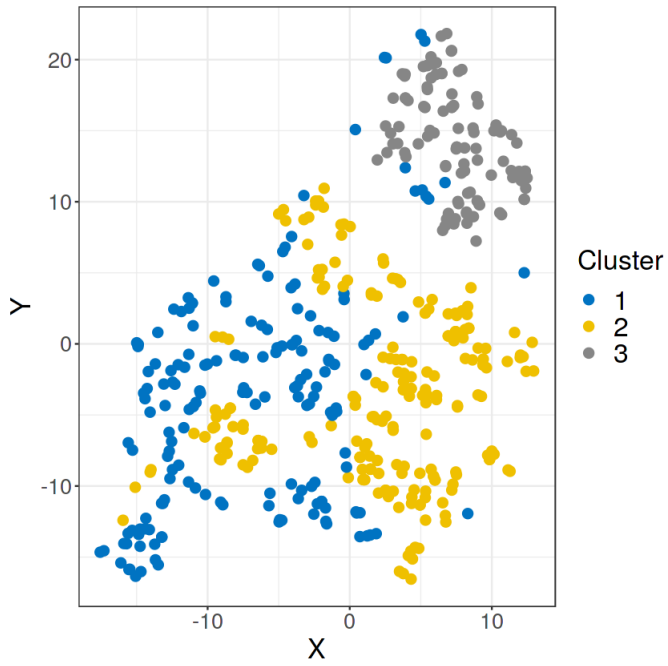


Figure 7.23: Hierarchical (hclust) clustering t-SNE plot.

Figure 7.23 shows that there appears to be 3 clusters with cluster 3 being more isolated and less overlapped than the other 2, but it appears cluster 1 does overlap cluster 2 more than cluster 3 within the 2D plot.

Table 7.12: Hclust members of clusters.

Cluster	Size	Ave.sil.width
1	172	-0.04
2	215	0.23
3	100	0.34

The negative silhouette width for cluster one, as shown in Table 7.12, suggests that there is no clear distinction for 40% of the cluster compared to one or both of the other classes.

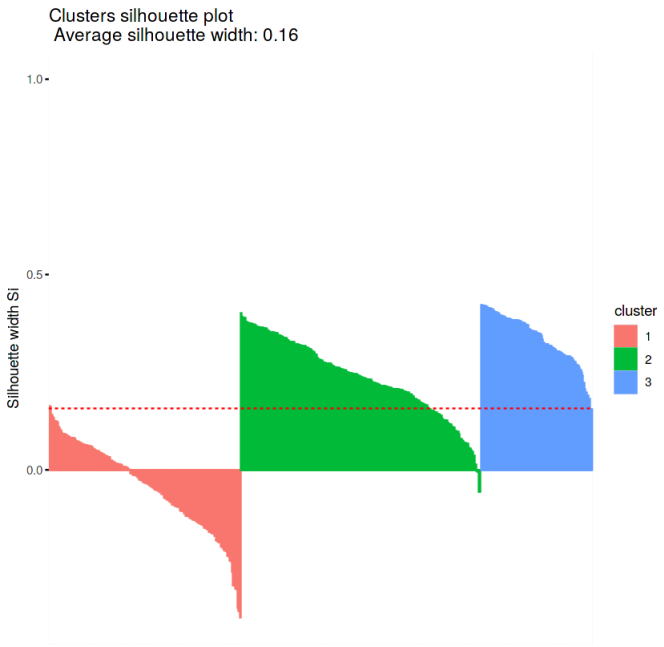


Figure 7.24: Hierarchical (hclust) silhouette.

The silhouette plot in Figure 7.24 shows that there are fewer data points that have possibly been miss-assigned, compared to the silhouette plot of PAM (Figure 7.14). This suggests that this hierarchical Ward linkage method, which minimises the total within-cluster variance, is a potentially better clustering method for this dataset than PAM.

Table 7.13: All 3 Hierarchical clusters values.

Cluster	**1**	**2**	**3**
Visit Reason			
Personal	45	208	94
Study	38	3	2
Time	52	4	2
Work	37	0	2
Visit Purpose			
Collection overview	18	8	8
Known Collection	10	6	3
Known Item	11	13	5
Museum Related	48	12	12
Other	40	32	22
Pre-Visit	45	144	50
Frequency			
Regular	45	19	18
Annual	51	71	32
First time	76	125	50
Domain Knowledge			
Novice	36	82	15
Intermediate	97	126	80
Expert	39	7	5
CH Knowledge			
1 or 2	18	49	14
3	59	92	40
4	56	61	36
5	39	13	10
Location			
England	31	41	17
Merseyside	63	74	34
Northwest	37	46	31
UK	9	18	5
World	32	36	13
Age			
18-34	47	40	0
35-54	63	111	1
55-64	43	60	31
65+	19	4	68
Employment			
Not Working	10	51	0
Working	127	153	0
Retired	11	0	100
Student	24	11	0

Table 7.13 shows the detailed breakdown of cluster members against all factors and all options.

7.3 Results and Analysis

This section will look at the cluster results for the self-assigned groupings, all users and lastly those that self-classified as General Public.

7.3.1 Analysing the Self-assigned Groups

The data was first analysed to determine whether the self-assigned groups are separable. By inspecting the top right t-SNE plot in Figure 7.8 it can seen that the self-assigned groups, on the whole, tend to spread across the plot, suggesting high overlap (including the General Public group). To further test this, the output of PAM clustering (with $k = 8$) was compared against the self-assigned groups using the Adjusted Rand Index (ARI). The result is a very low score of 0.021 indicating almost no overlap. Classifying the users by their self-assigned group results in an overall accuracy of 0.5315 (No Information Rate, NIR = 0.4895) on the test data, which is not that high. However, since a significant fraction is correctly classified, it is hypothesised that by using the PAM clustering method it should be possible to create a potentially, smaller set of more distinguishable user groups, which would be easier to cater to.

7.3.2 Clustering and Classifying All Users

The Hopkins statistic was initially used to test for cluster tendency. A score of 0.1826 is substantially below 0.5, indicating that the data is clusterable and thus eligible for cluster analysis. Following that, a cluster analysis was run on all users to see what groups emerged from the data. To determine the optimal number of clusters for PAM, the total WSS, average silhouette width, and gap statistic were estimated using the *fviz_ nbclust* package. The metrics recommend three, three, and nine cluster options, respectively. PAM was ran with $k = 3$, opting for the majority answer. Table 7.14 shows the average silhouette width of 0.16 and cluster medoids. The representatives reflect the mode value for each of the categories and therefore

Table 7.14: Cluster representatives for PAM clustering ($k = 3$) of all users

Cluster	Visit Reason	Visit Purpose	Freq	Domain Know	CH Know	Location	Age	Emp Status
1	personal	other	first	intermediate	3	world	35-54	working
2	personal	pre-visit	first	intermediate	4	northwest	65+	retired
3	personal	pre-visit	annual	intermediate	3	merseyside	35-54	working

hide some of the variation within the groups.

However, inspecting the distribution of individual categories and exemplars within the clusters, the clusters can be summarised as follows:

- **Cluster 1 - online researchers**: part and full-time workers (including students) visiting the website for a wide mix of reasons (including work or study), mainly seeking known items or collections and information about the museum, often first-time visitors with a range of domain knowledge (mostly intermediate) but higher CH knowledge, mostly aged between 18 and 54 and from outside the UK, therefore less likely to visit the museum in person (125 users).

- **Cluster 2 - CH enthusiasts**: mostly first-time and annual visitors to the website for personal reasons, perhaps preparing for a visit but also a range of other museum-related and other purposes, generally intermediate levels of domain and CH knowledge, predominately working or retired and aged 55+ and located in the Northwest of England and Merseyside (126 users).

- **Cluster 3 - visiting workers**: Typically regular users visiting the website for personal reasons and to pass time (although in a working capacity), mostly preparing for a visit to the museum, generally lower level of domain knowledge but intermediate level of CH knowledge, mainly in the 35-64 age range and working from the local Merseyside area (236 users).

Inspecting the t-SNE plot for the PAM clustering in Figure 7.8 (bottom right) would suggest that the clustering forms clear groups - the top right set of points is clearly representing Cluster 2, which seems to map onto mostly the retired and CH enthusiasts user group. Cluster 1 at the bottom left includes the student user group (amongst others). To check the stability of the clusters, bootstrapping was applied using the R *clusterboot* package. This runs PAM multiple times on samples

Table 7.15: Cluster representatives for PAM clustering ($k = 3$)

Cluster	Visit Reason	Visit Purpose	Freq	Domain Know	CH Know	Location	Age	Emp Status
1	personal	pre-visit	annual	intermediate	4	merseyside	35-54	working
2	personal	pre-visit	first	intermediate	3	northwest	65+	retired
3	personal	pre-visit	first	novice	3	england	35-54	working

of the data and compares the cluster outputs to determine how much points remain in the sample. The mean scores (1=all points remain in the same cluster) for the 3 clusters are: 0.7432, 0.6770 and 0.8044 suggesting the first and third clusters are the most stable.

Training the Random Forest classifier on clusters from PAM, an accuracy of 0.9306 (NIR=0.4861) on the test data (0.9086 on the training data) was obtained. Assessing global feature importance, variables are ranked as follows (by impurity): location (100), age (66.4), visit purpose (58.6), frequency (52.6), employment (52.5), CH knowledge (14.8), visit reason (6.2) and domain knowledge (0).

7.3.3 Clustering and Classifying General Public Users

In this section, the users who have identified themselves as General Public (or General Users) for the purposes of the survey are focused on. As this is a dominant group for NML (and DCH more generally Villaespesa (2019); Booth (1998)), there is a need to establish the homogeneity of this group, any sub-clusters and their defining characteristics. In prior work Walsh et al. (2020), it was found that general users could be distinguished as using the museum for personal use, often visiting for the first time, novice/intermediate domain knowledge, medium levels of CH knowledge, mainly from Merseyside/Northwest and generally in mid-life age range.

In Section 7.3.1, it was found that when classifying based on all self-assigned groups, the overall classification accuracy is low (0.5315). However, when inspecting the classification accuracy for the General Public class only, an accuracy of 0.8 (on test data) is obtained. Furthermore, performing binary classification for General Public vs. Other, an overall accuracy of 0.6966 (NIR=0.5172) on the test data is returned. Again, inspecting the General Public class only, there is an accuracy of 0.83. Therefore, although the General Public group shares similarities with other

groups (see the t-SNE plot in Figure 7.8), there are still differences that can be used to automatically distinguish this group, suggesting that the group is fairly homogeneous.

Prior to performing clustering to identify potential sub-groups within the general public users (236 users), a check for clusterability using the Hopkins statistic (0.20) and visual inspection of a visualisation of the dissimilarity matrix must first be run. It can be concluded that this sub-group may contain clusters. Similarly to clustering all users with PAM, a dissimilarity matrix using the Gower distance (using *daisy*) should be computed. To identify the optimum number of clusters, the total within-cluster sum of squares (WSS), average silhouette score and gap statistic were calculated. This time, all metrics output $k = 3$, and so this was adopted for the clustering with PAM. The resulting clustering has an average silhouette width of 0.18. The cluster medoids are shown in Table 7.15. The groups can be summarised as follows:

- **Cluster 1 - regular visit local workers**: generally, users mainly visit the website on a regular or annual basis for personal reasons, including preparing for a visit and seeking museum-related information, mostly intermediate levels of domain and CH knowledge, working, aged 35-64 and local to the Merseyside area (99 users).

- **Cluster 2 - local enthusiasts**: also mainly using the site for personal use (and pass time) and preparing for a visit; however, mostly first-time and annual website visitors with intermediate levels of domain and CH knowledge, mostly 55+, retired and from the Northwest and Merseyside (64 users).

- **Cluster 3 - first-time non-local workers**: mostly first-time users of the website using the website for personal use and to pass time, mostly in preparation for a visit; generally working with lower levels of domain and CH knowledge, mostly middle-aged 35-64 and from England but outside the Merseyside area (73 users).

The Random Forest classifier is trained with the target variable being the cluster number from PAM. Using a similar experimental setup as before, an accuracy of 0.8841 (NIR=0.4203) on the test data (0.8980 on training data) was obtained. Again,

using impurity to calculate global feature importance, the variables are ranked as follows: employment (100), frequency (68.9), location (62.6), domain knowledge (50.9), age (50.4), CH knowledge (24.4), visit purpose (7.2) and visit reason (0).

7.4 Clustering Comparisons

Comparing the Silhouette scores is one way to compare if the algorithms produce similar clusters statistically. The average silhouette score for hierarchical (see Figure 7.24) was 0.16 and for PAM it was 0.156. The difference between the silhouette scores is relatively small (0.016), suggesting that the resulting clusters may not be significantly different and that the resulting clusters are comparable.

Relying on a single method to compare clusters, such as the silhouette score alone, is not advised, so other metrics could be considered (Wong, 2022). The Rand index measures the similarity between two clusterings by comparing the number of agreements and disagreements in their assignments.

The Rand index is a measure of the similarity between two clustering solutions. It measures the number of pairs of data points that are assigned to the same or different clusters in both solutions. A Rand index of 0.164 was produced between the PAM and Hierarchical results, which suggests that there is a low level of similarity between the two clustering solutions being compared. However, when applying the Adjusted Rand Index (ARI), which also accounts for the clustering agreement expected by chance, produced a score of 0.1769772, suggesting that the similarity between the two clustering results is moderate.

Another measure that was used to calculate the similarity between the cluster results was the Normalized Mutual Information (NMI) score. This showed a score of 0.1784046, indicating that the two clusterings have a moderate degree of similarity.

In addition, the Normalized Variation of Information (NVI) calculation was run, which showed a score of 0.9013522, indicating that there is an amount of information shared between the two clusterings but that the two clusterings have a considerable degree of dissimilarity.

An alternative method undertaken was to compare the cluster assignments of each method against each other using a contingency table (see Table 7.16). This showed that whilst none of the clusters was completely the same (which is to be

expected), significant proportions of clusters aligned. For example, 78% of cluster 2 from hierarchical mapped to cluster 1 from PAM; 67.2% of hierarchical cluster 3 mapped to PAM's cluster 2, while 54.4% from hierarchical cluster 1 mapped to PAM cluster 1 (see Table 7.16).

Table 7.16: Hierarchical cluster results mapped against PAM cluster results.

	HC 1	**HC 2**	**HC 3**	**Row Total**
PAM 1	**54.4 %**	11.1 %	34.4 %	100.0 %
PAM 2	21.1 %	11.6 %	**67.3 %**	100.0 %
PAM 3	3.1 %	**78.1 %**	18.8 %	100.0 %

Having between 54% and 78% of each PAM cluster map to a single Hierarchical cluster does suggest a good level of comparability between the clusters, as there is a significant overlap between the clusters identified by the two clustering methods.

7.5 Discussion

This phase investigated clustering through two clustering algorithms (PAM and Hclust) to identify user groups from the data.

The initial MCA analysis (see Figure 7.1) showed that self-selection labels are not good predictors, as variable categories with similar profiles were grouped together, and some variables were negatively correlated and positioned in opposing quadrants. This suggests the need to cluster survey data excluding the self-classified label question to accurately identify user groups.

In comparison, clustering-based groupings are smaller and more distinct, resulting in considerably higher total classification scores. From the overall data three main categories can be identified: online researchers, CH enthusiasts and visiting workers. These distinct groups could help NML better tailor their services and resources.

Cluster analysis was performed on the General Public self-classified data, which is a dominant and relatively homogeneous user category for NML. Potential sub-groups within the General Public user group were identified, which have been labelled as: regular visit local workers, local enthusiasts, and first-time non-local workers. The first group may represent mostly teachers from the local area preparing for a visit;

the second group is mainly hobbyists and enthusiasts (e.g., those interested in local history and genealogy). Finally, the third group may represent groups such as teachers, but this time from outside the local area. However, these groups, although labelled as general users, are likely to have different needs and should therefore be catered for separately.

The equivalence of cluster results, as indicated by the silhouette score, Adjusted Rand Index (ARI), and Normalised Mutual Information (NMI), suggests that both the PAM and Hierarchical algorithms can be adopted, albeit with slight differences. Given the timing differences between these algorithms, it is advisable to develop a tool that incorporates both, enabling users to select their preferred algorithm based on their specific needs and preferences.

However, one of the main issues of using the PAM algorithm is the time required to perform the exploratory clustering process. GLAM employees will likely want to experiment with the ISCP outputs of the system quickly, and waiting for PAM to cluster each time the number of clusters is adjusted is not ideal. In a system where users may experiment with the number of clusters, running PAM each time could become frustrating and inefficient.

For example, considering the data gathered for this phase and using powerful cloud servers, the average time for the PAM clustering on the NML survey data was 0.14 seconds, while hierarchical clustering took just 0.03 seconds. Although a 0.1-second difference may seem trivial, it becomes significant when experimenting with different numbers of clusters. Each run of PAM consistently takes 0.14 seconds, whereas hierarchical clustering only takes 0.002 seconds for each subsequent experiment. This is because hierarchical clustering does not need to be re-run entirely; it simply cuts the hierarchical tree, created during the initial clustering, at a different level.

Similarly, if the clusters are experimented with, just 50 times to explore different numbers, PAM would take 7 seconds, while hierarchical clustering would only take 0.128 seconds. With larger datasets, the time difference between the two algorithms would increase significantly. However, a comparison of both algorithms revealed that the generated clusters are very similar, so the speed advantage of hierarchical clustering is unlikely to compromise the quality of the clusters.

Overall, the results on the clusters are far cleaner than the self-assigned groups,

although the General Public user group can be distinguished using classification. This group therefore, could be automatically identified and cluster analysis applied to further segment the group if desired.

7.6 Summary

The aim of Phase 4 was to extend the user group definitions and characteristics identified in previous phases through the use of unsupervised machine learning cluster analysis of survey data from National Museums Liverpool (NML). Previous phases had identified broad user categories but lacked detailed, actionable descriptions.

This phase used clustering techniques to provide a more comprehensive understanding of user behaviours and characteristics. Partitioning Around Medoids (PAM) and Agglomerative Hierarchical Clustering were employed, using performance metrics like Silhouette and Dunn Index scores to identify and validate naturally occurring user groups from survey data, moving beyond the limitations of self-reported categories. However, in terms of effectiveness, while PAM clustering is accurate, hierarchical clustering offers a faster and equally effective alternative, making it more suitable for exploratory analysis.

The application of PAM Clustering resulted in the identification of three optimal clusters with distinct characteristics, evaluated based on parameters like average dissimilarity and cluster diameter. Similarly, Hierarchical Clustering produced similar findings with three optimal clusters, with dendrograms providing a clear visual representation of the cluster structure. The following three main user groups identified by both clustering approaches are:

- Online Researchers: Users visiting for study or work, typically first-time visitors with intermediate to high CH knowledge.
- CH Enthusiasts: Mostly older, retired users from the Northwest of England, visiting for personal reasons and possessing intermediate CH knowledge.
- Visiting Workers: Regular users with intermediate domain and CH knowledge, mainly from the local area.

Clustering within the General Public group revealed three sub-groups:

- Regular Visit Local Workers: Regular or annual visitors for personal reasons, often preparing for a visit.
- Local Enthusiasts: First-time and annual visitors, mostly retired and aged 55+, from the Northwest and Merseyside.
- First-time Non-local Workers: First-time users from outside Merseyside, mostly working with lower CH knowledge.

The analysis demonstrates that clustering algorithms can effectively identify distinct user groups that are more meaningful than self-assigned labels. This approach provides valuable insights into user behaviour and that can guide GLAM (Galleries, Libraries, Archives, and Museums) institutions in better catering to their diverse user base.

The next chapter will discuss further practical applications of these clustering results, focusing on the development of the ISCP.

Chapter 8

PHASE 5: AUTOMATICALLY DESCRIBING USER GROUPS

8.1 Introduction

The creation of an organisation's digital presence can be undertaken by an individual or a diverse team, such as designers, developers, product managers, and end users if the team is working in a User Centred Design way. These teams are responsible for creating and improving digital products, including websites and applications. The priority of this group is to understand the needs, preferences, and behaviours of users in order to design and build products that are user-friendly, accessible, and effective.

However, diverse teams can mean an imbalance in technical ability and skill levels when dealing with the translation of human behaviour and technology. Drawing on academic research may be beneficial, but the focus is more practical and applied, with an emphasis on creating usable and operational digital products that meet business or museum goals and user needs. This is particularly crucial for museum teams, which are often smaller than typical development teams and operate with limited funds. As Green et al. (2020) pointed out, "what we often see is small teams with limited funds trying to do too much – too many projects, too much content – for 'all audiences'."

Therefore, to aid the design and development teams, a visual artefact based on the information-seeking behaviours captured by the survey has been created. The

Information Seeking Cluster Profiles (ISCPs) is an attempt to create a pre-persona quickly and economically that is based on the usually excluded information need and information behaviour data. By focusing on specific user needs and behaviours, museum teams can better allocate their limited resources and ensure their digital products are both effective and aligned with their audience's requirements.

Not to be confused with the already established Persona, this phase of work is not trying to replicate or enhance the automated persona, as they have access to many different data sources, which for most GLAM institutions would not be feasible to have access to. Neither is this work trying to recreate personas, as this is a well developed approach, despite their criticisms.

ISCPs can counter a number of the known issues with Personas. By undertaking the clustering and automatically extracting the ISCPs/groups, biases introduced by manual persona creation are omitted. However, it is important to note that biases introduced by the survey questions, sampling methods, and other factors still remain.

The goal of this phase is to create user group profiles, known as ISCPs, based on a survey that collects information on the audience's information seeking habits, needs, behaviours, preferences, and demographics (see Chapter 3, Section 3.4.5) by addressing the following research question:

Phase 5 Research Question:

RQ7: How can comprehensive and usable user group profiles be generated from the clustering data?

The ISCPs are developed by clustering the survey results and applying heuristics to these clusters (see Chapter 3, Section 3.4.5). Each cluster generates its own ISCP, forming a comprehensive set that represents the museum website's users. Similar to personas, each ISCP represents a segment of the user base and is depicted as a fictional person with a name and image to foster empathy. These ISCPs present detailed group information about each group in a clear, concise, and communicable format, using short text, tables, or charts.

The remainder of this Chapter is organised as follows: Section 8.2 Presenting Data-Driven Results; Section 3.4.5.2 Validation of ISCPs; Section 3.4.5.3 Qualitative Analysis; Section 3.4.5.4 Thematic Analysis Process; Section 8.2.1 Stage 5

- Thematic Analysis Insights and Supporting Quotes; Section 8.3 Discussion and Section 8.4 Summary.

8.2 Information Seeking Cluster Profiles

Information seeking cluster profiles (ISCPs) are representations of the different ways in which people go about seeking and gathering information that is relevant to their needs and goals. These profiles are based on the results of cluster analysis and heuristic evaluation of the clusters, utilising data gathered from surveys (see Chapter 3). They can be used to understand and design for the information-seeking behaviours of different groups of users. Figure 8.1 shows how the detailed user group description data is used in a more visual format to create a persona style poster (ISCP) that is more informative and useful to the museum practitioner.

The ISCP is structured into several sections, each based on topics within the survey:

1. **Overview** The overview highlights the sample size of this cluster. A name and relevant avatar are assigned, as well as a short textual description of the cluster.

2. **Demographics:** This section displays essential demographic information, including age distribution, gender, employment status, and primary locations of the user group. It visually presents the proportion of users within different age brackets, their gender split, various employment categories (such as full-time, part-time, and students), and geographical distribution.

3. **Goals and Motivations:** This section outlines the reasons for visiting museums, categorising them into personal reasons, studying, passing time, and other motivations. It also lists the specific museum and gallery website section the users are completing the survey from, providing insights into their preferred destinations. This section also highlights common activities and motivations of the users, such as pre-visit planning, seeking museum overviews, and looking for specific items within the collections.

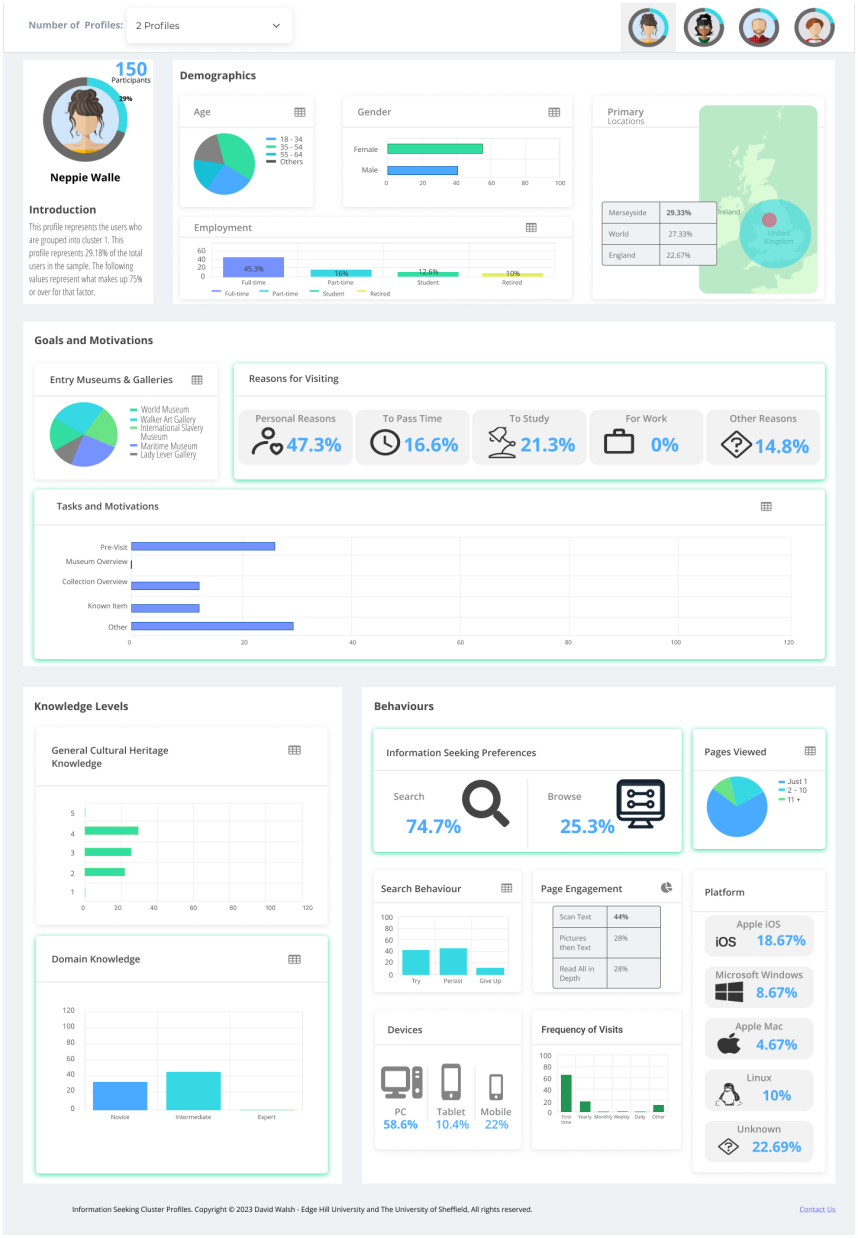


Figure 8.1: An example of a persona style Information Seeking Cluster Profile for cluster 1 (see Appendix J.1.2 for larger, more readable version)

4. **Knowledge Levels:** Bar charts illustrate the general cultural heritage knowledge and domain-specific expertise of the user group, showing levels from novice to expert.

5. **Behaviours:** This section captures information-seeking preferences, such as the preference for searching over browsing, and provides additional insights into search behaviour, page engagement, device usage, and visit frequency.

The ISCP layout remains consistent across different profiles, ensuring a uniform approach to presenting user data. The different profiles can be reviewed by clicking on the icons on the top right. The green outlines serve as a visual cue to identify significant factors within each section, allowing practitioners to quickly grasp the most critical insights relevant to each user group. This structured approach facilitates a better understanding of user behaviours and preferences, aiding in the effective design and delivery of digital content.

The development team can then experiment with the numbers by utilising the dropdown on the top to explore which number provides the cleanest and most suitable groups for the use of the current project.

8.2.1 Stage 5 - Thematic Analysis Insights and Supporting Quotes

The ISCPs were validated as described earlier in Section 3.4.5.2, and here are the results. The following thematic analysis is based on Table 8.1. The themes break down into 5 categories which align closely with the initial codes identified: Participant Background Knowledge and Experience, Perceptions of Traditional User Groupings, Thoughts and Applications of ISCP Profiles and Tool, Integration Challenges, and Improvement Opportunities.

Table 8.1: Codebook generated from Nvivo thematic analysis of the interviews.

Code	Description	Files	References
Awareness of Data-Driven Personas	Mentions of understanding or awareness of data-driven personas.	5	6
Perceived Value of Data-Driven Personas	Assessing the perceived benefits and effectiveness of data-driven personas.	2	2
Change in UX Workflow	Thoughts on how the new tool might alter standard UX practices.	6	7
Integration Challenges	Challenges anticipated in integrating the new tool into existing workflows.	1	1
Familiarity with Personas	References to previous knowledge or use of personas in UX design.	10	41
Depth of Use	Exploring the extent and manner in which personas are utilised in projects.	8	14
Evolving Attitudes towards UX-UI	Changes in perception and prioritisation of UX/UI within organisations.	6	9
Personas not worked well		2	3
Industry Sector	The industry sector of the participant's current working environment.	5	6
ISCP improvement suggestion		9	15
Perceived value of generated profiles	Thoughts on the generated profiles.	10	26
Perception of New Tool	Opinions, beliefs, or attitudes towards the newly introduced UX tool	10	40
Strengths		8	23
Unclarities or Concerns about the new tool	Specific concerns or reservations about the new tool.	5	6
Weaknesses		4	6
Perceptions of Personas		8	17
Personas perceived as useful		6	8
Personas perceived not useful		6	9
Potential Uses of ISCP	Perceived benefits and specific scenarios where ISCP profiles might be useful.	7	11
ISCP as a Time-Saving Tool	Views on how the ISCP tool could streamline UX processes.	3	4
Professional Background	Details about the participant's professional background.	10	26

Table 8.1 continued from previous page

Code	Description	Files	References
Specific Tasks - Responsibilities - Expertise	The specific tasks, responsibilities, or areas of expertise that they conduct.	10	14
Relevance to Study	Information that highlights how the participant's role makes their input relevant.	6	7
Experience with Personas	The level and depth of experience the participant has with using personas.	3	3
Willingness to Adopt	Indications of whether the practitioner would consider using the new tool.	10	16
Personal Willingness vs. Anticipated Resistance	Willingness to adopt the ISCP tool personally contrasted with perceived resistance from others.	4	5
Preconditions for Adoption	Conditions or factors that would influence the decision to adopt the new tool.	4	4

8.2.1.1 Participant Background Knowledge and Experience

The suitability of the participants of this study was captured through a collective lens of the following three themes: Industry Sector, Professional Background, and Relevance to Study.

8.2.1.1.1 Theme 1: Industry Sector

The industry sector theme captures the diverse range of businesses from where the participants come, emphasising a wide range of experiences in museum-related projects and beyond. Six participants have previously or are currently working on DCH (Museum) projects directly, with one previously being the NML web team manager at the time the survey data for this project was captured. Five of these six also have experience working on projects outside cultural heritage. The other four participants all have experience external to DCH.

The participants contribute experiences from various sectors (Arts, Entertainment and Recreation (Museums), Service (Funeral Care, Travel and tourism), Education (University), UK government, Logistics and distribution, Technology and

consultancy, Media, and Retail) offering a rich tapestry of insights pertinent to understanding museum website users. It is this diversity that highlights the wide range of expertise contributing to the validation of the ISCP tool and profiles.

8.2.1.1.2 Theme 2: Professional Background

This theme focuses on the Professional Backgrounds of participants, illustrating the variety of roles within the UX and web design industry. This diverse range of perspectives forms a solid foundation for examining the effectiveness and applicability of the ISCP profiles and tool in different professional contexts for example:

Participant 1 talked about the dual nature of the roles in this field, blending design with research but also highlighting the hands-on technical work also involved: *"My current role is a UX designer. That's what I've been employed as, but I am more of a UX researcher than a designer"* and *"Is me sat on Figma trying to make sense of designs that I'm creating for like an ancient system, a legacy system that needs to be brought into alignment with the modern aesthetic of the web"* (Participant 1).

The duality of responsibility was also shared by Participant 10 who balanced managerial tasks with practical, hands-on duties. Similarly, Participant 6 combined strategic oversight with technical expertise, highlighting the dynamic and project-based nature of their responsibilities. This further emphasises the multifaceted responsibilities of leaders in this sector.

"Web design management. I'm currently a leader with a team of web designers and design and develop my websites myself as well" (Participant 10).

and

"Title is technical lead and digital media and publishing team, that's. Manager of the development side of the digital media team" and *"we work in fortnightly sprints and lots of that would be working on deploying releases to either websites...and then discussing future projects"* (Participant 6).

The versatility of roles, as indicated by Participant's 7 and 9 also demonstrates substantial evolution, adaptability and continuous development that is closely linked to the strategic and visionary aspects of UX leadership:

"I'm a service designer...my job title changes year by year because the actual work that I do kind of is evolving all the time" (Participant 7).

and

"I'm currently the Creative Director at [Company], a digital agency focused on innovative user experiences. My role involves overseeing the UX design process, from research and conceptualisation to prototyping and user testing" (Participant 9).

Participants from various roles, spanning from UX designers and researchers to technical leads, service designers, and creative directors, showcase the diverse skill sets and professional backgrounds within the study. Each participant contributes a unique perspective, enhancing the analysis of the ISCP profiles and the tool's applicability and relevance in the dynamic field of UX and web design.

8.2.1.1.3 Theme 3: Relevance to Study

This theme captures the diverse experiences and backgrounds of the participants, emphasising their suitability for evaluating user profiles and the ISCP tool. Each participant provides invaluable insights gained from direct user engagement, the challenges they encountered, and their sector-specific knowledge, all of which are essential for understanding nuanced user needs and behaviours. For example:

"We were designing an app for the Gallery and Museum, so I spent a lot of time in there, observing users, talking to users, getting a feel for what they want from an app" (Participant 1), highlighting the importance of observing and engaging with users in their natural environments to gather authentic insights.

"I have used personas in the past, so you'd always have personas based on the type of drivers organised their parcels in quite different ways" (Participant 2), illustrating the practical use of personas to differentiate user behaviours and needs, while also acknowledging their limitations.

"They merged these things together and they ended up with 10 websites, 8 databases and a whole bunch of contents and data that didn't really talk to each other" (Participant 7), showcasing the complexities involved in consolidating digital assets to enhance user accessibility.

"working on the main website at the museum, [museum].ac.uk and then the collections website at the museum collections.[company].ac.uk" (Participant 6), demonstrating a deep understanding of the digital needs within cultural heritage institutions.

The diverse range of experiences among participants adds a rich layer of perspectives to the use and value of user profiles. The shared commonalities between the participants, highlight the samples exceptional suitability for providing insightful feedback on user profiles and the ISCP tool. Their direct engagement with users, critical evaluation of user-centric design tools, practical problem-solving across various sectors, and profound understanding of organisational challenges guarantee a thorough and nuanced analysis of the tool's effectiveness and applicability.

8.2.1.2 Perceptions of Traditional User Groupings

This category consists of three themes: Familiarity with Personas, Perceptions of Personas and Awareness of Data-Driven Personas. The objective is to understand the participants' usage experiences, perceptions, and awareness regarding the evolving field of data-driven personas.

8.2.1.2.1 Theme 4: Familiarity with Personas

This theme explores participants' past experiences when having worked with Personas. All participants had utilised personas in their professional endeavours, although the extent and consistency of usage varied. Some noted that they used personas in certain projects but not in others. For example, Participant 1 mentioned using them for a side project but not for their main ongoing project, where user characteristics were deemed more apparent. Additionally, certain participants highlighted that the adoption of Personas, and occasionally user research overall, was at times neglected due to financial limitations. This suggests a shifting perspective, where teams are increasingly adaptable and mindful of resources in their approaches. While they aim to incorporate personas whenever possible, they are also willing to forgo them when facing significant constraints. For example:

> *"We will do [[use personas]] for the higher-end clients that come in, the ones that have spent a lot more money that afford us a lot more time in the UX process. The ones that we don't get that kind of [[budget for]]...we don't. ... So if the UX hours are there, we will use personas."* (Participant 10)

It also became apparent that not all participants are directly involved in creating the personas, with some depending on user researchers for this task. While this approach implies the development of well-researched personas, it does suggest potential discrepancies in the perceived usefulness and acceptance of personas within the team. Additionally, the significance of organisational culture and attitudes within both the business and UX teams directly impacts participants' experiences with personas. Participants operate in a dynamic field where methodologies like persona adoption are shaped not only by practical or resource constraints but also by the values and priorities of the organisation and its members. For example, Participant 1 discussed different UX teams from different departments working in silos, each creating their own personas for the same projects:

> *"within the university, there are UX teams. So marketing have a UX team with about three people ... and then IT services ... UX team. So it seems like we're all doing different things. We're not really working together that well currently."* (Participant 1)

> *"I work very heavily with the user researcher, and I think if we were, I think if he was saying let's use personas, I think they'd be useful. I think I would be led by him because he's the kind of person who we think of as being closest to our audience."* (Participant 2)

What becomes evident from this theme, and is crucial to clarify for the remainder of this study, is that all participants possess familiarity with and have employed personas. However, the degree of involvement in creating or utilising personas varies among participants. Several factors, such as resources/budget constraints, attitudes towards personas, including personal perspectives, opinions of team members, and the broader corporate culture, influence the selection of workflow methodologies.

8.2.1.2.2 Theme 5: Perceptions of Personas

This theme delves into the participants' viewpoints on personas and their importance in the UX/web development workflow. While all participants have employed personas, there are varied opinions concerning their effectiveness and perceived limitations within the UX/UI workflow. Some acknowledge the value of personas and

utilise them as guiding and reference tools throughout the design process, for example

> *"We use personas for nearly all our projects. They're invaluable for keeping the user at the centre of our design process."* (Participant 9)

and

> *"The web designer will create the personas... I do anyway, at least as a designer, and I'll always push my designers to do the same where they keep referring back to who they're targeting as they work through designs."* (Participant 10)

Here, Participant 10 refers to trying to influence the culture and practices by pushing the use of Personas. While Participant 7 expands on the use not only within the group of designers, as suggested by Participant 10, but notes that personas can be particularly useful in large organisations for communicating user needs and characteristics across different teams or departments, enhancing the overall understanding of the user base. Several participants also emphasised the greater utility of personas when based on data rather than assumptions. Overall, there seemed to be a shared view among participants that personas constitute a vital aspect of the UX process, although their usefulness or value might not be fully recognised or appreciated by others in their teams.

However, not all participants perceive personas as beneficial. Those predominantly operating within a Service Design workflow (Participants 4, 7 and 8) voiced reservations about the concept of personas. Concerns ranged from the quality of the data to the rationale behind using the tool, and even the scope of the personas themselves, for example;

> *".. become a kind of a check box, kind of a tick box exercise that some organisations felt we need to keep but when you dig deeper into why they don't actually have an answer as to why they need them, they just want them."* (Participant 7)

Throughout the interviews, there was a recurring sense of discretion regarding the use of personas. This reinforces the idea that personas are sometimes perceived

more as a mandatory checklist item rather than a genuinely valuable tool to be consistently employed throughout the design and development phases. Additionally, some concerns were raised about teams adopting personas primarily for their convenience rather than their necessity, and that limitations such as personas not sufficiently addressing user needs and system functionality, and instead, they would opt for the alternative tool of using user stories.

> *".. personas can sometimes be the wrong tool to use in situations where we have people with a lack of digital experience because they can sometimes latch on to that a single persona. That single idea of a person....And that could be detrimental to what we ultimately try to design or navigate through. So instead we would lean towards user needs instead."* (Participant 4)

and

> *".. for the smaller projects we would fall back to personas as there are not really a lot of features on small projects to lead the user stories generation."* (Participant 9)

Additional concerns were raised regarding the scope of the personas and their effectiveness in capturing user needs, particularly in cases of complexity. Participants also provided clear examples of instances where personas were deemed ineffective. These shortcomings were often associated with the lack of detail, whether fabricated or derived from personas, resulting in a narrow scope that does not accurately represent broader audiences:

> *".. occasionally we don't always use them. And the reason for that is a lot of the time we can find them quite limiting. "* (Participant 8)

and

> *"You don't need to create a physical board that shows like the persona and the goal and stuff, so the act of grouping your users into specific persona types with key goals and things is a good practice, but outputting that as a actual physical sheet of paper with a picture on it and goal*

thing, I find that practiced sometimes is not really necessary and that's why I chuckle, because then personas are something that whenever anyone brings them up, people don't know what to do with the joke about it or they argue it's not necessary or they argue that it is necessary. It isn't. There is no consensus on personas." (Participant 7)

This theme uncovers a nuanced perception of personas among UX/UI professionals. While acknowledged for their effectiveness in guiding design processes and enhancing organisational understanding, personas also encounter critique for their constraints in specific scenarios and perceived limited applicability beyond the design phase. The discretionary use and adoption of alternative methods by practitioners raise concerns about the comprehension and actual validity of personas. This analysis highlights the significance of critically and knowledgeably employing personas, indicating the necessity for continuous assessment of their role and efficacy within UX/UI workflows.

8.2.1.2.3 Theme 6: Awareness and Evaluation of Data-Driven Approaches

This theme highlights that while all participants had a traditional understanding of personas, they were unaware of current research into data-driven personas by academics and practitioners, such as Jansen (2021); McGinn and Kotamraju (2008); Salminen et al. (2019). Many participants considered personas to be based on qualitative user research and not made up from assumptions as data-driven. After an explanation, many were interested in the concepts, but some questioned what the differences would be between their manually created personas and data-driven personas.

"all personas we create are driven by some form of data, right?" (Participant 7)

and

"I'd be curious to see, uh, a comparison of someone like me doing it the usual way and then getting a system to generate a persona using the method you just mentioned and seeing where the differences are, where the similarities are, that would be quite interesting." (Participant 2)

Overall, this theme highlighted that the interviewed practitioners were unfamiliar with the data-driven approach to persona creation. While they recognised the significance of incorporating real user qualitative data insights from interviews and observations into personas, they were unaware of utilising various data sources, such as web analytics, social media data, and customer databases, combined with algorithms and grounded in large-scale data analysis. This approach offers a more objective and scalable means of understanding user behaviours, needs, and preferences, a practice they had not previously heard of.

8.2.1.3 Thoughts and Applications of ISCP Profiles and Tool

This category encapsulates four themes: Theme 7 - The perceived value of generated profiles, Theme 8 - Perception of ISCP tool (strengths, weaknesses, unclarities), Theme 9 - Uses of ISCP, Theme 10 - Willingness to adopt (personal willingness vs anticipated resistance, pre-conditions for adoption). The objective of these themes is to validate the ISCP tool and the profiles it generates for the NML data.

8.2.1.3.1 Theme 7: Perceived value of Generated Profiles

The participant interviews reveal several key insights into the perceived value of generated user profiles, highlighting the potential benefits and applications of such profiles in various contexts. For example, all the participants saw the value the profiles could provide them if run on their own projects. For example:

> *"The way you've approached it looks great. The interface makes it really easy to understand. You know, obviously, there's a lot going on behind the scenes in terms of the way that that Data is crunched and worked out. There's obviously a lot of hard work going on there, but the way you've presented it makes it really straightforward to understand."* (Participant 3)

Indeed, many found the presentation of the profiles both appealing and useful in the museum context, particularly Participant 2, who was part of the NML team when the survey data used was captured:

> *"I can also see that this [[the profiles and ISCP tool]] would be really valuable for small museum-based teams where there is no data analyst or*

UX team member. When I was at the [Museum], we used personas, but they were less detailed than these and more wordy. I can see that these profiles based on data and in a visual format could be really useful." (Participant 2)

and

"I like the way you've presented it here, where you can see the differences between the profiles." (Participant 2)

Other participants who either work in other major UK museums or as part of a DCH project also saw the value the profiles provided:

"I think it's nice as a reference, and it gives us ... a rolling view of the audience in a way that again without great resource every time, it's not possible to Commission every, year for the website. ... It doesn't answer all questions, but then none of them [[meaning personas]] do, and it answers a different I think set of questions or gives. Information on a different area, you know. Which? In a lighter light way. I think is quite nice." (Participant 6)

and

"I mean, in terms of visual information, this is really good. So you know, you clearly at a glance see key pieces of information." (Participant 7)

and

"I thought it was like very detailed immediately and one of my first thoughts was this is a lot of information that a lot of it that wouldn't be included in a persona ... that helps us figure out all our assumptions were wrong." (Participant 8)

Participants recognised the considerable usefulness of the generated profiles for comprehending user behaviours, preferences, and technological usage. For instance, Participant 1 highlights the potential for verifying the accuracy of generated profiles

through direct observation in settings such as museums, demonstrating a practical real-world application. Additionally, Participant 4 proposed employing profiles in physical spaces, such as job centres or libraries, to gain better insights into visitor motivations and needs, showcasing the versatility of these profiles beyond digital contexts.

> *"I could use something like this to get an idea of what the users are, and then I'll probably go into the museum then after that, just to observe people, to see if they match what the cluster profiles that have been pulled out."* (Participant 1)

and

> *"I think it could be really useful in a physical space to help. Help me as a researcher and as a design team understand why people are going into these things and what their needs and wants are."* (Participant 4)

Several participants appreciate the data-driven nature of the profiles, with Participant 10 expressing enjoyment in the data-driven approach and finding the profiles useful and diverse. This indicates a perceived value in the depth and variety of insights the profiles provide. Furthermore, Participant 2 and Participant 3 commend the presentation and clarity of the profiles, suggesting that the visualisation of data makes it simple to understand and differentiate between user groups.

> *"I really like it. It looks really good .... I mean it looks varied more varied than you'd kind of expect. I'd find this really useful."* (Participant 10)

Participant 8 appreciates the detailed information provided by the profiles, which can help correct assumptions about user demographics, goals, and motivations before conducting interviews, thereby guiding research direction more effectively.

> *"If you go in with the assumptions of what people are doing and you get a whole load of others, then you're kind of like, oh, that helps us figure out all our assumptions were wrong."* (Participant 8)

Participant 7 compares the generated profiles to services like YouGov, suggesting that if the tool can offer similar demographic insights in a more targeted or cost-effective manner, it would be highly valuable for organisations looking to save time and resources.

> *"It actually reminds me quite strongly of I don't have been on you, Gov. … So yeah, if a solution exists out there that provides that kind of thing and actually more targeted to a particular. A topic or an area like culture or travel or something that if you even if you'd like almost tailor this solution to those different areas… And it would save an organization like ours time, which we're always wanting."* (Participant 7)

The profiles are viewed as enriching the persona development process by providing tangible data to substantiate assumptions regarding age, gender, and user behaviour. Participant 8 specifically highlights the significance of this feature for validating and fine-tuning personas:

> *"I think it's helpful to help define some of these things about personas, like particularly around like age and gender, because I feel like sometimes that's just plucked out of thin air… I think it, although it's a small thing, I think it's actually quite a nice thing to be able to properly validate."* (Participant 8)

These observations indicate a positive perception of the generated user profiles in terms of their usefulness, effectiveness and contribution to design and research endeavours. The profiles are valued for their data-driven approach, comprehensive insights, and capacity to enhance user comprehension across digital and physical contexts. Such feedback suggests that the profiles could be immensely relevant and valuable for research initiatives aimed at validating automated persona-style user profile and their associated tools, laying a strong foundation for further development and application in practical settings.

8.2.1.3.2 Theme 8: Perceptions of ISCP Tool

This theme reflects the participants' reactions to the ISCP tool in terms of its strengths, weaknesses, lack of clarity and concerns. There were notably (see Table

8.1) more references to strengths (8 participants, 23 references) than to weaknesses (4 participants, 6 references) or uncertainties (5 participants, 6 references).

8.2.1.3.2.1 Strengths: The strengths of the ISCP tool encompass various aspects, including its visual appeal, user-friendliness, customisation options, data-based robustness, and efficiency compared to traditional methods, for example:

- **Visual appeal and variety of data**: Participants appreciated the visual appeal and perceived variety offered by the ISCP tool. As one participant noted, *"It looks really good, though. I mean it looks varied more varied than you'd kind of expect."* (Participant 10)

- **Data-based robustness**: The foundation of the tool on quantitative survey data and its robustness were appreciated, distinguishing it from other analytics tools: *"I like the fact that it's a kind of got that robustness of being solidly based on data."* (Participant 2)

- **User engagement, understandability and encapsulation**: The tool's ease of use by hiding all of the complex clustering and analysis and the potential of the interface to engage users effectively was seen as particularly valuable: *"The way you've approached it looks great. The interface makes it really easy to understand. You know obviously there's a lot going on behind the scenes in terms of the way that that Data's crunched and worked out. There's obviously a lot of hard work going on there, but the way you've presented it makes it really straightforward to understand."* (Participant 3)

- **Depth of Insights and Customisation**: The depth of insights provided by the tool and its potential for extensive customization were seen as strengths. *"It gives you a level of detail thing and that obviously we could keep refining endlessly the audience for the music."* (Participant 6)

- **Speed compared to traditional methods**: The speed of generating ISCP profiles compared to traditional methods of user research highlights the efficiency of the tool in providing insights more quickly than conventional approaches. *"I mean, yeah, the, the speed, as you say of doing that versus actually*

organising in person, you know that's months more or less before we get any like a report." (Participant 6)

Participant 9 succinctly captures this sub-theme of strengths in a single quote:

> *"It seems to harness the power of data analytics to segment users into meaningful groups of users, which could potentially streamline the persona creation process."* (Participant 9)

8.2.1.3.2.2 Weaknesses: Few weaknesses of the ISCP tool were identified, with only four participants commenting on potential issues. Two participants raised concerns not directly related to the tool, profiles, or methods of analysis, but rather regarding the survey tool for data collection. They expressed concerns about potential biases in the data due to missed audiences caused by the length of time a survey ran or the attitude of people towards surveys in general e.g., some purposefully never complete them or they remember after the closing date. Also, as noted by Participant 8, people (survey respondents) can often self-inflate or deflate their abilities; however, they also went on to answer their own issue with

> *"further interviews would be really important to like validate or disprove"* (Participant 8).

Other weaknesses identified were in relation to the interface, but only in having a particular place on the dashboard where an overview of all main differences of all profiles in a summary would be useful e.g., Participant 2.

Considering the concept behind the ISCP tool is that initiation is undertaken early in the project, it is anticipated that additional qualitative user research would be conducted regardless.

8.2.1.3.2.3 Lack of clarity or concerns: There were inquiries regarding the potential biases introduced by the survey tool. Other concerns pertained to how the tool would manage questions, which differed from those used in this study. Additionally, a question was raised about the number of survey questions required to gather the data shown in the profiles. Participant 1 also posed an interesting question about institutions' interest in understanding their users and the necessity for a tool like this.

"I just think that certain museums, certain institutions are more receptive to doing these kind of things, the XXX are very old fashioned and very traditional, they don't even have counts of people that have come into the building. ... So yeah, I'm just a bit suspect about how certain institutions would want to do something like this. Also how interested are they in the user?" (Participant 1).

However, this study may not provide a definitive answer, as each organisation ultimately makes the decision.

8.2.1.3.3 Theme 9: Potential uses of ISCP

The ISCP tool is proposed to potentially provide a less biased approach to understanding user groups or personas compared to traditional manual methods. This could prove particularly relevant and valuable in resolving internal corporate disputes or when seeking consensus on user characteristics, for example:

"I mean, I'm sure there's bias probably in everything that's ever done isn't there, but that somehow it feels like because it's based on that data, it feels like it's potentially less biased." (Participant 2)

There was also perceived value for small teams without dedicated data analysis or user experience (UX) specialists. The ISCP profiles are viewed as a practical, data-driven alternative for less elaborate, more descriptive [[qualitative]] personas:

"I can also see that this would be really valuable for small museum based teams where there is no data analyst or UX team member." (Participant 2)

The ISCP tool and profiles are acknowledged for their effectiveness in improving product design and development, extending beyond the museum domain. They provide valuable insights into the preferences and requirements of various user groups, which can significantly influence the development process:

"Having those profiles to understand, you know, is this a real kind of travel savvy group of people, or is it someone that just wants to go on holiday for have an extended holiday or something like that..." (Participant 3)

The speed and efficiency of the ISCP tool and early use in the UX process of a project in collecting and analysing data compared to traditional methods are highlighted, with an emphasis on the immediacy of insights it can provide:

> *"I mean, yeah, the speed, ... of doing that versus actually organising in person [[interviews or observation data gathering and analysing]], you know that's months more or less before we get any like a report."* (Participant 6)

Participant 7 also commented on the potential for the ISCP tool to save time and resources, particularly when contrasted with costly alternatives such as YouGov surveys. This feature is particularly attractive to organisations seeking to understand their audiences without incurring substantial costs:

> *"if we can create a solution that allows an organisation like ours to send out a questionnaire which could then come back with a whole host of demographic data that's automatically collated and put together in a graph setup like that [[meaning eh ISCP profile]], I'd use it with bells on ..."* (Participant 7)

What also became evident is that the ISCP is not necessarily perceived as a replacement for persona or user story tools. Rather, it is viewed as an augmentation to the UX process, enriching traditional qualitative research practices used in creating personas or user stories. The tool is praised for offering valuable pre-research insights, aiding teams in focusing their efforts more effectively. This helps them avoid the common pitfall of realising too late in the research process that they should have targeted different user groups:

> *"It's really cool to be able to get this amount of information before you've even done any research like that in depth research, cause yeah, like I said, I think it helps focus."*(Participant 8)

Additionally, one participant recognised that the application for the ISCP tool (survey) extended beyond only gathering data in the digital spaces. They highlighted its potential use on kiosk-style screens in physical spaces such as job centres or

museums. This application is seen as beneficial for collecting metrics/data on visitor behaviour and needs, thereby assisting in designing and providing services in these environments:

> *"I'm thinking about how I would apply this to Universal Credit, right...I feel like this could work well in a physical space and what I mean by that is like it could go into job centres [[as a touch screen kiosk]] to ask people why are they there that day. Is it information seeking? Is it ... for an appointment? Is it to apply for something? ... So I think it could be good ... in a physical space where someone's having to be there and interact with something. I think it could be really beneficial in that space. And you could probably do that for museums too, right? You know, like people could fill this in on the way in."* (Participant 4)

8.2.1.3.4 Theme 10: Willingness to adopt

The ISCP tool was viewed by the participants as a valuable complement to traditional user research methods, providing a data-driven approach to understanding user groups prior to further qualitative investigation. Every participant expressed their desire to use a tool like this if it were available to them(although most participants viewed its adoption as an additional supportive enhancement to product design rather than a replacement for their current practices).

> *"Oh, absolutely, yeah, it'll be complimentary. I could use something like this to get an idea of what the users are..."* (Participant 1)

Participants particularly appreciated the ISCP tool for its potential to streamline the initial phases of user research by identifying key user groups for engagement with further qualitative methods.

> *"OK, so you can see this being ... used then to profile who you're going to go and actually do the qualitative data with..."* (Participant 5)

They also saw the potential to provide user data quickly and cost-effectively, especially compared to traditional methods that are time-consuming and expensive.

"if we can create a solution that allows an organization like ours to send out a questionnaire which could then come back with a whole host of demographic data that's automatically collated and put together in a graph setup like that, I'd use it with bells on..." (Participant 7)

The tool's ability to provide detailed user profiles is seen as beneficial for informing product design and development, especially in tailoring products to specific user needs.

"Having those profiles to understand...would really kind of affect the way we build out products..." (Participant 3)

Also, the tool's usefulness in heritage and cultural projects is specifically highlighted, where audience understanding is vital yet often challenging due to manual classification errors.

"I think that it would be super helpful to use. Like I would definitely advocate for that because on the project that we did with heritage work..." (Participant 8)

While recognising the usefulness of the ISCP tool, it is also acknowledged that its relevance may vary depending on the specific target audience and the context of the project.

"I think this is a useful tool. I think it very much depends on which company and which people it was targeting..." (Participant 2).

"I think this is a useful tool. I think its adoption would very much depend on the project and needs of the project as well as which users it was targeting..." (Participant 9).

However, Participant 10 expressed concerns about the time required to conduct the survey. They mentioned that on some of their projects, user research is not conducted due to budgetary and time constraints. They saw these constraints as potential barriers to adopting the ISCP tool in such projects.

"The one thing I think would stop us using this. Is that the time it would take to collect any data..." (Participant 10).

Overall, these observations illustrate a general willingness among participants to integrate the ISCP tool into their research methodologies. However, this enthusiasm is tempered by factors such as time, effort, and the specific contexts of their projects. Participants recognise the tool's potential to improve understanding of user groups, streamlining research processes, and provide cost-effective data analysis, indicating a positive disposition towards adopting the tool in diverse professional settings.

8.2.1.4 Integration Challenges

The ISCP tool should be employed in the early stages of a project to establish a solid understanding of the audience and form the foundation for further development. Implementing the ISCP tool requires a shift in conventional UX workflows. This section will explore the challenges associated with altering UX workflows and integrating new tools into existing processes.

8.2.1.4.1 Theme 11: Changes to UX workflow

The proposed adoption of the ISCP tool and early survey for some was not a significant departure from current practices, as according to the participants, many teams already conduct user research early in the process. However, while they typically engage in the more time-consuming qualitative user research, they recognise the potential benefits of an early survey tool. This tool could then inform their customary qualitative processes, aiding in the creation of complete personas or the investigation of user stories and issues:

> *"... that could inform your requirements gathering stage when you're actually speaking to your customer. I'm getting because you have data that you're able to play with before the actual persona part. So this is almost like pre-research as we would see at our company. And yeah, I can see that would be very useful."* (Participant 5)

and

> *"I mean, yeah, the speed ... of doing that [Meaning the survey and running the tool] versus actually organising in person [interviews], you know that's months more or less before we get any ... report."* (Participant 6)

Participant 8 appreciated the agility and speed provided by the ISCPs, enabling some customisation of their interview questions and potentially helping them resolve problems quicker:

> *"[[It's a]] quick way to see immediately that you're like ohh, our assumptions were wrong about this and you don't have to wait until the interviews to get to that point and then you can use the interviews, validate and understand a bit more about what people are doing."* (Participant 8)

Participants also recognise the profiles as supportive tools for design and research. For instance, Participant 5 sees them as beneficial for the requirements-gathering stage, suggesting that they provide a solid foundation for understanding user needs early in the design process.

> *"You know that could inform your requirements gathering stage when you're actually speaking to your customer. I'm getting because you have data that you're able to play with before the actual persona part."* (Participant 5)

8.2.1.5 Improvement Opportunities

This theme examines participant suggestions for enhancing and developing the ISCP tool through suggested modifications and improvements.

8.2.1.5.1 Theme 12: ISCP improvement suggestions

The participant commentary in this category encompasses suggestions for tailoring the tool/survey to specific industry needs or client bases. This includes modifying, adding, or adjusting questions to ensure relevance across various domains, catering to the specific interests and needs of different businesses or contexts. Additionally, integrating existing data, whether historically collected or from analytics/data management systems, would enhance the applicability of the ISCP tool and profile to a broader range of users:

> *"I think different places will have different things that are of interest to them and being able to kind of customise ... the questions ... for different businesses potentially"* (Participant 2)

and

> *"the reasons thing and that would have to be tailored specifically towards the kind of industry or client or user base"* (Participant 5)

While some participants find the demographic information extremely useful, there's an emphasis on ensuring that the tool's features, such as knowledge levels, are pertinent and suitable for the specific audience or project. There is also concern about potential data duplication and the possibility of linking it with other data sources. However, it is argued that capturing data simultaneously from the same sample may not provide a comprehensive profile.

The importance of the tool seamlessly integrating with existing tools and systems was voiced by Participant 6.

> *"if it was being sold as a product to our user research or our product manager, I guess the. The question is that like the integration with existing systems, you know? How easy is it to integrate with analytics? Yeah, I guess that's all I would like. That's probably a thing that would come up. How does it relate to the other systems we have? You know, how does it tie in with other ways we have of doing this to make it part of the sort of the steps of the project."* (Participant 6)

It was also noted that if customisability were implemented, clear instructions would be necessary on how to specify the type of question. This would enable the system to apply the appropriate analysis and generate accurate graphs:

> *"It seems that there would need to be some parameters put in place or you know like kind of containers around the questions asked. So anyone can create their own questionnaire and upload it."* (Participant 7)

A few of the suggestions, voiced by only one or two participants, comprised of the following:

- To improve the description on the profile to highlight the uniqueness of that profile. Highlighting the factors in text as opposed to just changing the colour of the borders of the factor, which is currently the case in the profiles. (Participant 1).

- Questions about how the tool would handle data over time suggest an interest in longitudinal studies and the ability to compare clusters or profiles as they evolve, indicating a need for the tool to support ongoing learning and adaptation (Participant 6).

- Inclusion of a shorthand persona style label for the profiles so discussions in the team would allow for easy identification between the profiles. (Participant 6).

- Providing users with help information, tooltips, and guides on how to effectively use the system, including what questions to ask and how to interpret the data, is seen as crucial for user engagement and effectiveness (Participant 7).

Overall, the thematic analysis explored the nuanced perceptions and applications of an ISCP's related tool within the context of Digital Cultural Heritage (DCH). Through a comprehensive examination of participant insights, the study revealed a broad appreciation for the data-driven nature of ISCPs, highlighting their potential to offer more objective, varied, and earlier-than-usual detailed user profiles compared to traditional methods. Participants recognised the value of these profiles in enhancing understanding of user behaviours, preferences, and technological usage across both digital and physical environments. The analysis demonstrates the potential of ISCPs and the associated tool to improve user understanding, facilitate efficient design and research processes, and offer valuable insights for organisations seeking to deepen their engagement with audiences without incurring high costs. This feedback suggests a strong foundation for further development and application of automated, data-driven user profiles like the ISCP in real-world settings, offering a new paradigm in user research that complements traditional qualitative methods with scalable, data-informed approaches.

8.3 Discussion

The evaluation of the ISCPs showed that an ISCP was seen as a useful tool for designers and researchers who are trying to understand and design for the information

seeking behaviours of different groups of users. They can provide insight into the types of information that users are looking for, the sources they rely on, and the challenges they face when seeking information. This can help designers to create solutions that better meet the needs of their users.

While the generated visual profile (see Figure: 8.1) resembles a Persona, ISCPs focus solely on basic demographics and information-seeking characteristics differing in their level of detail and angle of approach.

The inclusion of industry experienced participants with diverse capabilities offered varied perspectives on the utility of user profiles. Despite this diversity, their shared commonalities highlight their exceptional ability to provide insightful feedback on user profiles and the ISCP tool. Their direct user interaction, critical assessment of design tools, broad problem-solving skills, and deep organisational understanding ensured a thorough analysis of the tool's effectiveness.

Exploring participants' perceptions of traditional user groupings highlighted that while all participants have experience with personas, their level of involvement in creating or using them varies. Factors like budget constraints, attitudes toward personas, personal perspectives, team opinions, and corporate culture influence workflow methodology selection. What also became apparent was the nuanced view of personas among UX/UI professionals. While recognised for guiding design processes and enhancing organisational understanding, personas face criticism for their constraints and perceived limited applicability. The discretionary use of alternative methods raises concerns about the understanding and validity of personas. This emphasises the importance of critically and knowledgeably using personas and underlines the need for ongoing assessment of their role and effectiveness in UX/UI workflows.

Additionally, while the participants acknowledged the importance of integrating real user qualitative data insights into personas, they themselves lacked awareness of utilising various data sources, algorithms, and large-scale data analysis for this purpose. Therefore, applying an ISCP would provide a more objective and scalable understanding of user behaviours, needs, and preferences, a practice they had not encountered before.

ISCPs are designed as a precursor model that ISCPs could be integrated seamlessly into the standard UX process, typically conducted at the project outset. The

development of detailed a questionnaire, focussing on users' information-seeking behaviour (e.g., reasons for site visits, knowledge of content, preferred navigation methods, search persistence, visit frequency, content preferences) and demographics, template could be provided to institutions such as DCH for self-administration among their users. The ISCP tool could then utilise the survey results to identify user groups via clustering, minimising bias and avoiding generic user groupings.

Overall the participant interviews unveiled crucial insights into the value of the ISCP's, emphasising their potential benefits and applications in diverse contexts. These observations reflect a positive perception in terms of their usefulness, effectiveness, and contribution to design and research efforts. Appreciation was also noted for the data-driven approach, comprehensive insights, and ability to improve user understanding across digital and physical environments. The feedback provided by the participants overall implies that the profiles could be highly relevant and valuable for research initiatives focused on validating automated persona-style user-profiles and associated tools, paving the way for further development and application in real-world settings.

8.4 Summary

The chapter contributes new knowledge to the field through several key areas:

Development of ISCPs: Introduction of Information Seeking Cluster Profiles (ISCPs) as a novel method for profiling museum visitors based on their information-seeking behaviours. The structure of the ISCP into five detailed sections (Overview, Demographics, Goals and Motivations, Knowledge Levels, and Behaviours) provides a comprehensive framework for understanding user groups.

Application of Clustering Techniques: Utilisation of clustering analysis on survey results to create data-driven profiles, is a methodological innovation in the context of DCH. The clustering approach provides a more objective and scalable way to segment user groups compared to traditional qualitative methods.

Insights from Thematic Analysis: Identification of participant perceptions regarding traditional user groupings and data-driven personas. Highlighting the gap in awareness and the potential interest in data-driven approaches among museum professionals.

Practical Tool for Museum Teams: The development and application of the ISCP tool as a practical resource for museum teams allow for the effective allocation of resources and design user-centric digital products. Feedback from participants suggested the tool's utility in streamlining early-stage research and providing quick insights.

Integration Challenges and Improvement Opportunities: Identification of the challenges in integrating the ISCP tool into traditional UX workflows. Suggestions for tool improvement, including customisation options, better integration with existing systems, and clearer user guidance.

Shift to Bottom-Up Approach: Proposal to adopt a bottom-up approach in the next phase, focusing on natural groupings from survey results rather than predefined categories, aiming to identify more realistic and operational user groups.

These contributions represent new knowledge in the fields of digital heritage, user experience design, and data-driven profiling, providing both theoretical and practical advancements.

Chapter 9

DISCUSSION, CONCLUSION, CONTRIBUTION AND RECOMMENDATIONS FOR FUTURE WORK

The preceding chapters of the thesis have each contributed to the exploration of identifying the information-seeking characteristics of different user groups for museum websites. This chapter will first summarise in Section 9.1 the thesis and presents the answers to the research questions in order to meet the aim. Secondly, the core contributions will be outlined in Section 9.3 and finally, Section 9.4 will suggest any future work based on the results of this study.

9.1 Conclusions

Phase 1 established that while there where many separate studies of DCH users (see Table 4.1) many only covered one or two targeted groups and that attempting to establish useable descriptions was limited due to the lack of detail resulting in a misrepresentation of said groups. To corroborate the groups found in the literature (Phase 1), a comprehensive survey was undertaken with an exemplar DCH website (NML) (Phase2). The findings from Phase 2 both corroborate and contradict the findings of Phase 1 by confirming the presence of a range of roles, abilities, and

knowledge as well as there being a range of users. However, Phase 2 revealed that contrary to the literature, the main user groups were categorised as "General Public" and "Non-professional" as opposed to "Experts" and "Novices" thereby suggesting that the wrong audience had been the focus of previous literature and proposed search systems resulting in website high bounce rates and unadopted technological solutions.

The questionnaire used in Phase 2 and the associated clustering code was packaged as a tool that could be downloaded, surveys run on their audiences, and results processed via the code resulting in the relevant descriptions outputted. The automated approach was taken further to make the work more useful to a broader audience. This new approach allowed a tool/script using a clustering algorithm that enhanced the ability to experiment with the number of clusters to be created so that they could be used in conjunction with the questionnaire. The museums can use the questionnaire to gather data from their website users and then run the results through the script and then experiment with the numbers of groups without having to have the knowledge of how to create the script or clustering. The output from the script provides a set of user group cluster definitions that are complete human-readable descriptions in the form of a more visual Information Seeking Cluster Profile (ISCP), a pre-persona template that the UX/practitioner team could enhance with their additional user research to convert to a persona. The results of this phase showed that the hierarchical clustering algorithm in a script could provide a fast enough tool to be helpful to practitioners and researchers while the outputs of ISCP results in a format that enhances both fields and, according to the interviewees, has an even wider scope in the general UX field. For example the resulting groups from the NML clustering presented three detailed human-readable definitions of groups of users that can be shared in research and also set the standard for the user research that should be expected in future research in DCH, making the user groups comparable and actionable.

The following sections will demonstrate how each research question has been met, resulting in the overall aim of "identifying the information-seeking characteristics of different user groups for museum websites." being realised.

9.1.1 RQ1: What are the main user group categories present in the Digital Cultural Heritage literature?

The literature revealed that there have been many separate studies of the user groups that make up all the museum virtual visitors, but most studies are conducted in isolation of other groups. Evidence presented from the literature review suggested that no comprehensive approach to the identification and use of a group label was undertaken by any previous author; rather the label that was used was seen as appropriate to what was being investigated.

Phase 1 attempted to identify and draw together the labels and the limited descriptions that have been used through the exploration of how they have been used in terms of participant context and research theme. This produced an extensive list of labels that, upon closer examination, could be reduced to a more manageable listing by merging variations of labels (plurals of a singular name) and duplicates (provided characteristics were the same), along with disregarding complete outliers. A hierarchy could then be created through the determination of possible overarching labels from those that were available and that could also encapsulate other labels of a similar genre (see Figure 4.5) e.g., General Public, Educational Visitor and Researcher.

9.1.2 RQ2: What are the similarities, shared characteristics, and differences of the identified groups?

Similarly to RQ1, attempting to define a label through the literature via its characteristics proved difficult due to the isolation and individuality of the studies and the practices of the authors. While it became apparent that the focus for the majority of research centred on motivation, task domain and technical knowledge, it was the lack of detail relating to the definition of a label that added unnecessary complexity. The characteristics assigned to a label could be fluid in the extreme with not even consensus being achieved between authors as to what constitutes a novice or expert.

Examining the definitions that could be attributed from the publications, the context of the publication, the participants that were being used, consideration being given to knowledge areas and motivations for engaging with DCH, did not

produce the clarity or rationale for why a named label was being used which in turn left significant knowledge gaps about the other characteristics of the group and presented a rather more confused and non-operational interpretation.

9.1.3 RQ3: How are user groups defined based on the survey responses? and RQ4: How do the aspects of the user groups defined in the survey compare to those in the literature?

The literature review and Phase 1 established the need for a comprehensive investigation into capturing the characteristics of the DCH website user, leading to Phase 2, which involved gathering data from a range of visitors to the NML website through a survey consisting of 22 questions.

The analysis of the data focused on identifying the user groups as determined by the participants, using cross-tabulation exercises and chi testing to reduce the overall number of valid surveys to 488 and the group number to 7. The results of the survey revealed that the main user groups visiting the NML website were the "general public" and "non-professional" visitors, who make up nearly 80% of all visitors. In addition the survey also allowed the definition of the criteria that distinguish these two groups from each other and from other groups, such as motivation, task, engagement, domain knowledge, and location.

The differences between these two groups and those groups that have been studied more frequently (academics, students, and professionals) were also highlighted, primarily relating to why they are visiting NML's website and how much CH and domain knowledge/experience they have. Other significant characteristics of these groups revolve around the lower levels of CH knowledge and expertise that may result in the need for more domain knowledge support in their interactions. The findings of the survey also suggest a challenge to previous research, in which the domain knowledge of the non-professionals and experts was found to be similar; and that non-professionals have closer levels of domain knowledge to the General Public as opposed to the experts.

Overall, the findings of the survey highlighted the need for further investigation

into how the different user groups interact with the website and how user interactions can be better supported across a range of tasks and goals.

9.1.4 RQ5: How do user groups identified in the transaction logs compare to those identified in the literature and survey results?

Definitions of the user groups of an exemplar DCH museum website were captured and documented in Phase 2. However in order to prove these groups, which are based on voluntary survey respondents, are not biased or only representative of those that are likely to complete a survey, a study to try and identify the previously devised groups within weblogs was conducted.

In Phase 3, the aim of the study was to compare the user groups identified in the literature and survey results with those identified in the transaction logs of the NML website. The log data was automatically captured from website interactions that took place in February e.g., demographic data and search queries, and a cleaning code was used to remove artificial (Bots) interactions. The analysis found that there were strong similarities between the log data and the survey responses. The majority of activity that took place was UK-based and local to the museum. The transaction logs showed a high percentage of single-page access requests focused on General, Item, and Event pages, possibly in preparation for a physical visit. The cluster analysis of the types of pages visited revealed user groups that were similar to those previously discovered in the literature, but no specific determination could be made due to overlapping categories.

However, the study also revealed that focusing on labelled user groups may not be the most effective approach, as the multiple categories that each user group could be allocated to made it difficult to make specific determinations as well as the limitations of using transaction logs alone to fully understand user behaviour.

Nevertheless, Phase 3 was able to compare user groups identified in the transaction logs to those identified in the literature and survey results and found that the transaction logs provided valuable insights into user behaviour on the website. The study also highlighted the need for a different bottom-up approach to establish more realistic and operational user groups, which was explored in Phase 4. The research

question was met to a significant extent in Phase 3 of the study.

9.1.5 RQ6: What user groups can be identified from the survey responses?

Using transaction log analysis, Phase 3, found that the self-assigned groups were not separable, with the majority of users identifying as General Public or Non-professional. Therefore in Phase 4, the PAM clustering method was used to create a set of more distinguishable user groups. The resulting cluster medoids were summarised as Online Researchers, CH enthusiasts, and visiting workers.

The limitations of using self-assigned user groups were highlighted in that the General Public user group might contain potential sub-groups with different needs or be nothing more than a state of mind of the user at that point in time. The study concludes that a smaller set of more distinct groups may be easier to cater for than self-assigned groups that commonly overlap and share characteristics, demonstrating that the research question of identifying user groups from survey responses has been met.

9.1.6 RQ7: How can comprehensive and usable user group profiles be generated from the clustering data?

Phase 1 demonstrated that it was not possible to create detailed and comprehensive user group definitions from the literature. The conducting of the survey in Phase 2 and the manual analysis of this showed that improved definitions could be drafted. However, the results were time-consuming to produce and showed a significant overlap between groups and the resulting definitions were still vague in places, lacking comparability. In Phase 4 multiple clustering algorithms were run on the survey question responses from Phase 2 excluding the self-classification question (Q22). The result was that the PAM and Agglomerative hierarchical algorithms proved to be more useful in this survey data, producing a more comprehensive sets of group definitions for clusters. The definitions were based on the medoid of the resulting clusters but were found to not be 100% representative of all the cluster members, which meant that the description was lacking in comparative detail. In an attempt

to combat this and also to produce a tool that would be usable by museum professionals as well as DCH researchers, a set of heuristics was created that were applied to columns of the cluster results and provided much greater detailed, more comparable and actionable user group definitions, which are far more representative of all members. The use of the more flexible hierarchical algorithm is also recommended as this produces comparable clusters to the PAM algorithm, but being able to cut the tree multiple times (to explore the numbers of groups) with only one run of the clustering saves time.

The results are presented in a number of formats, from detailed textual descriptions to summaries to tabular and in a visual pre-Persona (ISCP) version which enables the results to be usable by either the researcher or the museum web team. A set of example ISCPs for NML are presented for comparability to future researchers (see Appendix J.1). These may be adoptable by similar museums on a wider scale however, future work of conducting the survey and using the tool on different institutions is advised first. The ISCPs that are conducted by museum practitioners could easily be incorporated into the standard UX practices and built upon to create full Personas and compared to the definitions created in this work.

9.2 Limitations

The original concept for this research stemmed from my interest in developing usable user interfaces, especially in areas with unique customer bases. The CH/DCH community has only been partially explored, leading to a narrow identification of its user base. Given the numerous CH institutions, it was strategically impossible to investigate each one, so a representative exemplar, NML, was used. NML includes several museums, making it a suitable focus for this study. While the study's context was limited, the results could be applicable to the wider DCH community. Involving multiple institutions would have provided more responses and firmer conclusions about the applicability of the user profiles for DCH museums and galleries. However, as NML comprises diverse museums and galleries, it is reasonable to conclude that the results may be adopted more broadly, though future work should confirm this.

Distribution of the survey was through a Pop-up menu on the NML website,

as per NML requirements. Self-selected sampling via popups will always introduce bias, however the inclusion of validation means that bias is as limited as possible. Serving the survey via a pop-up could mean that users who use a pop-up blocker would not have seen it. Also, hosting anything on the internet provides an invitation to the non-human or Bots. While automated and manual processes were undertaken to capture these bots, there could possibly still be some responses that were not captured. If there are any bots left in the data, these will be minimal and should not affect the results of analysis.

What could however have had a detrimental effect on the results is that of the limited sample size; the larger the number of participants the more robust the results; although while the sample is only a fraction of the whole website user base, it was proven to be representative in the log data and as such should not affect the results. Participants who chose to complete the survey could be the type of people who always complete surveys, and the sample misses those that would not normally choose to complete them; people cannot be forced to complete a survey; meaning that the sample may not fully represent the diversity of users who engage with DCH. That being said the sample did cover a full range of users and that if there was a slight bias it is unlikely that it would be any different to the biases in all other survey-based research that has been undertaken and published.

Similarly, the qualitative study data for Phase 5 was coded and themes identified by a single person, myself, followed by discussions with supervisors. While ensuring methodological consistency, this approach lacked the input from multiple coders with differing expertise and viewpoints. Also, while real-time/in-situ member-checking was conducted with interviewees validation steps were undertaken to ensure limited bias.

9.3 Core Contributions

The thesis contributes to the field of Digital Cultural Heritage (DCH) research in several ways. Firstly, it highlights the limited and ambiguous descriptions of user groups for DCH in the literature that causes confusion and assumptions due to the lack of clear labels and comprehensive definitions. Secondly, it identifies that the main audience for DCH is the General Public/Non-Professional, which is contrary to

most literature and research systems that are designed for experts/academics/professionals. Thirdly, it shows that domain knowledge is not a driver for search, as all user groups prefer browsing.

The thesis also provides an understanding of user behaviour on DCH websites, with evidence that some users visit a DCH website with others as a social activity, which was previously believed to be a solitary task. The thesis highlights the importance of detailed descriptions of user group labels, as users' and researchers' understanding of these labels is not always reliable. Moreover, it shows that groups from literature cannot be easily identified in transaction logs, but user groups and detailed definitions can be generated from the cluster analysis.

To support this research, the thesis presents an ISCP generation tool that was created to aid both experienced and non-data-science researchers, which can be used efficiently and effectively by researchers and museum web team professionals. The thesis also provides detailed ISCP presentations for each audience, as well as a set of NML user groups and definitions that can be published and compared with other research because they are presented in this detailed way [1]. Lastly, the thesis contributes a large data-set of self-reported metrics for a large number of DCH users. Overall, the contributions of this thesis advance our understanding of user behaviour and characteristics on DCH websites and provide valuable insights for researchers and professionals in the field.

9.4 Future Work

9.4.1 Running the Survey at Different Institution(s)

For future work, it would be valuable to re-run the survey on the audience of another institution(s). This would allow confirmation of the strengths and weaknesses of the survey, as well as a comparison and contrast of the results. By repeating the survey in a different context, it would be possible to assess the generalisability of the findings and identify any potential differences in user behaviour and characteristics between institutions.

Additionally, by comparing and contrasting the results from different institutions,

[1] `https://sweor1-dave-walsh.shinyapps.io/ISCPs/`

it would be possible to identify any commonalities and differences in user behaviour and characteristics. This information could be used to develop more targeted and effective strategies for engaging with users on DCH websites, tailored to the specific needs and preferences of different user groups.

Moreover, comparing the results from different institutions would also provide insights into the effectiveness of different website design and content strategies in engaging users. This information could be used to develop best practices and guidelines for DCH website design and content, aimed at optimising user engagement and satisfaction. Overall, rerunning the survey on another institution(s) audience would provide valuable insights into user behaviour and characteristics on DCH websites and help to advance our understanding of how to effectively engage with users in this context.

9.4.2 Validating the Profiles and Tool

In this study, the generated ISCP profiles were validated by a wide range of DCH, UX, and web professionals for their appearance, usefulness, and accuracy. However, validating on multiple projects the user profiles generated through the data-driven user clustering tool is crucial to ensure their accuracy and effectiveness. To achieve this, a possible future work would be to recruit a series of museum website development teams and DCH researchers to conduct the survey on their own users. After obtaining the data, the museum teams can then utilise the data-driven user clustering tool to create their own set of ISCPs. This step would help to establish how useful the resulting ISCPs are for the museum web teams and fellow researchers and how easy the tool was to use. Moreover, this step would help to identify any further improvements that the museum web teams would like to see in the tool. This process can help to refine the tool and make it more efficient for future use.

Overall, validating the user profiles and testing the effectiveness of the ISCP tool in a real-world scenario are essential future activities that could help to enhance the tool's reliability and usefulness.

9.4.3 Additional Features

A further area of potential future work would be to extend the ISCP tool to incorporate the server log data, in order to gain a deeper understanding of user behaviour and technical usage patterns. This would enable the creation of more detailed user stories, which would be based on actual user behaviour and would allow for more accurate and effective design decisions to be made. By incorporating this data into the ISCP tool, it could be moved closer to a fully-fledged persona tool, which would be extremely valuable for designers and developers in the digital cultural heritage sector.

The use of log data in combination with the ISCP tool would enable a more comprehensive analysis of user behaviour and preferences, as it would allow for the identification of patterns and trends that may not be immediately apparent from survey data alone. This would provide a more holistic view of user needs and preferences, which could then be used to create more effective and tailored digital experiences.

In addition to aiding the creation of user stories and Personas, the incorporation of log data into the ISCP tool would also have the potential to inform the design of new features and functionality on DCH websites. By identifying which areas of the website are most commonly used and which features are most popular, designers and developers would be better equipped to make informed decisions about which areas to prioritise for development.

Overall, the integration of server log data into the ISCP tool would be a valuable area of future work, as it would allow for a more detailed and accurate understanding of user behaviour and preferences and could provide a more effective basis for the creation of user stories and Personas.

9.4.4 Usable Search & Browse Hybrid Museum Interfaces

Lastly, as future work, there is a need to further investigate and develop an integrated hybrid interfaces for DCH websites. The current research into search systems, Rich Prospect Browsers (RPB), and Generous interfaces appears to be focused solely on either searching or browsing. It does not address the issue of how users can fully understand the entire collection and what is offered.

The research conducted on DCH users has revealed that the current interfaces and features available to users are not very usable. By developing an integrated hybrid interface, it is possible to combine the strengths of search systems, RPBs, and Generous interfaces, while minimising their weaknesses. An integrated hybrid interface could provide users with a comprehensive understanding of the collection, allowing them to explore and discover new items while still offering traditional search and browsing functions.

Further research could be conducted to investigate what elements of search systems, RPBs, and Generous interfaces should be included in an integrated hybrid interface. It would also be important to consider the technical requirements and challenges associated with developing such an interface, as well as conduct user testing to evaluate the effectiveness and usability of the interface.

Overall, the development of an integrated hybrid interface for DCH websites could significantly enhance the user experience, enabling users to better understand and explore the collection and ultimately increase engagement and interest in cultural heritage.

Bibliography

Abdi, H. and Valentin, D. (2007). Multiple correspondence analysis. *Encyclopedia of measurement and statistics*, 2(4):651–657.

Adams, A., Blandford, A., and Lunt, P. (2005). Social empowerment and exclusion: a case study on digital libraries. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 12(2):174–200.

Adlin, T. and Pruitt, J. (2010). *The essential persona lifecycle: Your guide to building and using personas*. Morgan Kaufmann.

Agirre, E., Aletras, N., Clough, P., Fernando, S., Goodale, P., Hall, M., Soroa, A., and Stevenson, M. (2013). Paths: A system for accessing cultural heritage collections. In *ACL (Conference System Demonstrations)*, pages 151–156. Citeseer.

Ahmed, Z. A., Qaed, F., and Almurbati, N. (2020). Enhancing museums' sustainability through digitalization. In *2020 Second International Sustainability and Resilience Conference: Technology and Innovation in Building Designs (51154)*, pages 1–4. IEEE.

Allen, R. B. (1997). Chapter 3 - mental models and user models. In Helander, M. G., Landauer, T. K., and Prabhu, P. V., editors, *Handbook of Human-Computer Interaction (Second Edition)*, pages 49–63. North-Holland, Amsterdam, second edition edition.

Amin, A., Hardman, L., and van Ossenbruggen, J. (2007). Searching in the cultural heritage domain: capturing cultural heritage expert information seeking needs. *Information Systems [INS]*, R0701(R0701).

Amin, A., Van Ossenbruggen, J., Hardman, L., and van Nispen, A. (2008). Understanding cultural heritage experts' information seeking needs. In *Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries*, pages 39–47. ACM.

An, J., Cho, H., Kwak, H., Hassen, M. Z., and Jansen, B. J. (2016). Towards automatic persona generation using social media. In *2016 IEEE 4th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW)*, pages 206–211. IEEE.

Andrews, D., Nonnecke, B., and Preece, J. (2003). Electronic survey methodology: A case study in reaching hard-to-involve internet users. *International journal of human-computer interaction*, 16(2):185–210.

Antoniou, A. and Lepouras, G. (2010). Modeling visitors' profiles: A study to investigate adaptation aspects for museum learning technologies. *Journal on Computing and Cultural Heritage (JOCCH)*, 3(2):1–19.

Ardissono, L., Kuflik, T., and Petrelli, D. (2012). Personalization in cultural heritage: the road travelled and the one ahead. *User modeling and user-adapted interaction*, 22(1-2):73–99.

Assadi, H., Beauvisage, T., Lupovici, C., and Cloarec, T. (2003). Users and uses of online digital libraries in france. In *International Conference on Theory and Practice of Digital Libraries*, pages 1–12. Springer.

Baeza-Yates, R. and Maarek, Y. (2011). Web retrieval: the role of users. In *Sigir*, pages 1303–1304. Citeseer.

Bagnall, P., Dewsbury, G., and Sommerville, I. (2005). The limits of personas. In *Proceedings of the 5th Annual DIRC Research Conference*, volume 39, page 40. Citeseer.

Bailey, M. (2022). A dark winter looms for uk museums as energy costs and fewer visitors continue to hit revenue. `https://www.theartnewspaper.com/2022/10/05/a-dark-winter-looms-for-british-museums-as-covid-energy-costs-fewer-visitors-and-an-elusive-workforce-continue-to-hit-revenue`. (Accessed on 03/06/2023).

Bartoli, F., Lisanti, G., Seidenari, L., Karaman, S., and Del Bimbo, A. (2015). Museumvisitors: a dataset for pedestrian and group detection, gaze estimation and behavior understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 19–27.

Baskas, H. (2013). *Hidden Treasures: What Museums Can't or Won't Show You*. Globe Pequot.

Bastanlar, Y. (2007). User behaviour in web-based interactive virtual tours. In *2007 29th International Conference on Information Technology Interfaces*, pages 221–226. Ieee.

Bates, M. J. (1979). Information search tactics. *Journal of the American Society for information Science*, 30(4):205–214.

Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Information Review*, 13(5):407–424.

Beaudoin, J. E. (2012). Context and its role in the digital preservation of cultural objects. *D-Lib Magazine*, 18(11):1.

Belkin, N. J., Bierig, R., Buscher, G., van Elst, L., Gwizdka, J., Jose, J., and Teevan, J. (2009). Understanding the user-logging and interpreting user interactions in information search and retrieval (uiir-2009). *Understanding the User-Logging and Interpreting User Interactions in Information Search and Retrieval (UIIR-2009)*.

Belkin, N. J., Oddy, R. N., and Brooks, H. M. (1982a). Ask for information retrieval: Part i. background and theory. *Journal of documentation*, 38(2):61–71.

Belkin, N. J., Oddy, R. N., and Brooks, H. M. (1982b). Ask for information retrieval: Part ii. results of a design study. *Journal of documentation*, 38(3):145–164.

Bell, C. (2022). The origin of customer journey mapping - cxmtoday. `https://cxmtoday.com/leadership/the-origin-of-customer-journey-mapping/`. (Accessed on 03/11/2023).

Benouaret, I. and Lenne, D. (2015). Personalizing the museum experience through context-aware recommendations. In *2015 IEEE International Conference on Systems, Man, and Cybernetics*, pages 743–748. Ieee.

Bergstrom, D. M. (2010). Personas catalogue v.3 an update. Technical report, Europeana.

Bertacchini, E. and Morando, F. (2011). The future of museums in the digital age: New models of access and use of digital collections. EBLA Working Papers 201105, University of Turin.

Bezdek, J. C. and Hathaway, R. J. (2002). Vat: A tool for visual assessment of (cluster) tendency. In *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02*, volume 3, pages 2225–2230. IEEE.

Bholowalia, P. and Kumar, A. (2014). Ebk-means: A clustering technique based on elbow method and k-means in wsn. *International Journal of Computer Applications*, 105(9).

Birt, L., Scott, S., Cavers, D., Campbell, C., and Walter, F. (2016). Member checking: a tool to enhance trustworthiness or merely a nod to validation? *Qualitative health research*, 26(13):1802–1811.

Blandford, A. and Attfield, S. (2010). Interacting with information. *Synthesis Lectures on Human-Centered Informatics*, 3(1):1–99.

Blandford, A., Green, T. R., Furniss, D., and Makri, S. (2008). Evaluating system utility and conceptual fit using cassm. *International Journal of Human-Computer Studies*, 66(6):393–409.

Boehmke, B. and Greenwell, B. (2020). Chapter 11 random forests | hands-on machine learning with r. `https://bradleyboehmke.github.io/HOML/random-forest.html`. (Accessed on 04/10/2023).

Bogaard, T., Hollink, L., Wielemaker, J., Hardman, L., and van Ossenbruggen, J. (2019). Searching for old news: User interests and behavior within a national collection. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*, pages 113–121. ACM.

Bohnert, F., Zukerman, I., and Laures, J. (2012). Geckommender: Personalised theme and tour recommendations for museums. In *International Conference on User Modeling, Adaptation, and Personalization*, pages 26–37. Springer, Springer.

Bolchini, D. and Mylopoulos, J. (2003). From task-oriented to goal-oriented web requirements analysis. In *Proceedings of the Fourth International Conference on Web Information Systems Engineering, 2003. WISE 2003.*, pages 166–175. Ieee.

Bonfigli, M. E., Cabri, G., Leonardi, L., and Zambonelli, F. (2004). Virtual visits to cultural heritage supported by web-agents. *Information and Software Technology*, 46(3):173–184.

Booth, B. (1998). Understanding the information needs of visitors to museums. *Museum Management and Curatorship*, 17(2):139–157.

Borgman, C. L. (1996). Why are online catalogs still hard to use? *Journal of the American Society for Information Science (1986-1998)*, 47(7):493.

Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development*. sage.

Boyce, R. D., Ragueneau-Majlessi, I., Yu, J., Tay-Sontheimer, J., Kinsella, C., Chou, E., Brochhausen, M., Judkins, J., Gufford, B. T., Pinkleton, B. E., et al. (2018). Developing user personas to aid in the design of a user-centered natural product-drug interaction information resource for researchers. In *AMIA Annual Symposium Proceedings*, volume 2018, page 279. American Medical Informatics Association.

Bradley, C., Oliveira, L., Birrell, S., and Cain, R. (2021). A new perspective on personas and customer journey maps: Proposing systemic ux. *International Journal of Human Computer Studies*, 148(October 2019).

Braun, V. and Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101.

Braun, V. and Clarke, V. (2021). One size fits all? what counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology*, 18(3):328–352.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

Brickey, J., Walczak, S., and Burgess, T. (2010). A comparative analysis of persona clustering methods. In *Amcis*, page 217.

British Design Council (2005). The design process: The 'double diamond' design process model. *http://www.designcouncil.org.uk/about-design/how-designers-work/the-design-process/*, 11(12):2013.

Broder, A. (2002). A taxonomy of web search. In *ACM Sigir forum*, volume 36, pages 3–10. ACM.

Byrne, D. (2022). A worked example of braun and clarke's approach to reflexive thematic analysis. *Quality & quantity*, 56(3):1391–1412.

Cairns, S. (2013). Mutualizing museum knowledge: Folksonomies and the changing shape of expertise. *Curator: The Museum Journal*, 56(1):107–119.

Cambridge Dictionary (2020). General public – cambridge dictionary. Accessed on 20 April 2020.

Candy, P. C. (1989). Constructivism and the study of self-direction in adult learning. *Studies in the Education of Adults*, 21(2):95–116.

Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P. S., and Sun, L. (2023). A comprehensive survey of ai-generated content (aigc): A history of generative ai from gan to chatgpt. *arXiv preprint arXiv:2303.04226*.

Carmel, D., Josifovski, V., and Maarek, Y. (2011). User modeling for web applications. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 7–8.

Case, D. O. and Given, L. M. (2016). *Looking for information: A survey of research on information seeking, needs, and behavior*. Emerald Group Publishing.

Chang, Y.-n., Lim, Y.-k., and Stolterman, E. (2008). Personas: from theory to practices. In *Proceedings of the 5th Nordic conference on Human-computer interaction: building bridges*, pages 439–442.

Chapman, C. N. and Milham, R. P. (2006). The personas' new clothes: methodological and practical arguments against a popular method. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 50, pages 634–636. SAGE Publications Sage CA: Los Angeles, CA.

Charles, M. C. and Mertler, C. A. (2002). *Introduction to educational research*. Allyn and Bacon.

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. sage.

Chatman, E. A. (1999). A theory of life in the round. *Journal of the American Society for Information Science*, 50(3):207–217.

Chatterji, M. (2004). Evidence on "what works": An argument for extended-term mixed-method (etmm) evaluation designs. *Educational Researcher*, 33(9):3–13.

Chen, H.-M. and Cooper, M. D. (2001). Using clustering techniques to detect usage patterns in a web-based information system. *Journal of the American Society for Information Science and Technology*, 52(11):888–904.

Cherryholmes, C. H. (1994). More notes on pragmatism. *Educational researcher*, 23(1):16–18.

Chowdhury, G. (2010). From digital libraries to digital preservation research: the importance of users and context. *Journal of documentation*.

Chowdhury, S. (2015). *Cultural Heritage Information Access and Management*, chapter Cultural heritage information and usability. Facet Publishing.

Ciber (2013). Europeana 2012-2013: usage and performance update. Technical report, CIBER Research.

Cifter, A. S. and Dong, H. (2008). User characteristics: Professional vs. lay users.

Clarke, V. and Braun, V. (2013). Successful qualitative research: A practical guide for beginners. *Successful qualitative research*, pages 1–400.

Clements, J. (2019). Introduction to hierarchical clustering | by john clements | towards data science. `https://towardsdatascience.com/introduction-hierarchical-clustering-d3066c6b560e`. (Accessed on 01/09/2023).

Coburn, J. (2016). I don't know what i'm looking for: Better understanding public usage and behaviours with tyne & wear archives & museums online collections

mw2016: Museums and the web 2016. http://mw2016.museumsandtheweb.com/paper/i-dont-know-what-im-looking-for-better-understanding-public-usage-and-behaviours-with-tyne-wear-archives-museums-online-collections/. (Accessed on 03/11/2017).

Cohen, L., Manion, L., and Morrison, K. (2018). *Research methods in education*. Routledge, 8th edition.

Connaway, L. S., White, D., and Lanclos, D. (2011). Visitors and residents: What motivates engagement with the digital information environment? *Proceedings of the American Society for Information Science and Technology*, 48(1):1–7.

Cooley, M. (1989). Human-centred systems. In *Designing human-centred technology*, pages 133–143. Springer.

Cooley, M. (1996). On human-machine symbiosis. In *Human machine symbiosis*, pages 69–100. Springer.

Cooper, A. (2004). *The inmates are running the asylum:[Why high-tech products drive us crazy and how to restore the sanity]*. Sams Indianapolis.

Coppola, D. (2021). Online purchasing penetration by age great britain 2012-2020.

Corbin, J. and Strauss, A. (2008). *Basics of Qualitative Research (3rd ed.): Techniques and Procedures for Developing Grounded Theory*. SAGE Publications, Inc., Thousand Oaks, CA.

Corona, L. (2021). Museums and communication: the case of the louvre museum at the covid-19 age. *Humanities and Social Science Research*, 4(1):p15–p15.

Crabtree, B. F. and Miller, W. F. (1992). A template approach to text analysis: Developing and using codebooks. In Crabtree, B. F. and Miller, W. L., editors, *Doing qualitative research*, pages 93–109. Sage Publications, Inc.

Creswell, J. W. (1998). *Qualitative inquiry and research design: Choosing among five traditions*. Thousand Oaks, CA: Sage.

Creswell, J. W. (2002). *Educational research: Planning, conducting, and evaluating quantitative*. Merrill.

Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches*. SAGE Publications, 2 edition.

Creswell, J. W. (2013). *Qualitative inquiry and research design : choosing among five approaches*. SAGE Publications, Los Angeles, 3rd ed. edition.

Creswell, J. W. (2014). *Qualitative, quantitative and mixed methods approaches*. Sage.

Cunliffe, D., Kritou, E., and Tudhope, D. (2001). Usability evaluation for museum web sites. *Museum Management and Curatorship*, 19(3):229–252.

D, V. (2021). Multivariate analysis: an overview - students 4 best evidence. `https://s4be.cochrane.org/blog/2021/09/09/multivariate-analysis-an-overview/`. (Accessed on 01/12/2023).

Denscombe, M. (2008). Communities of practice: A research paradigm for the mixed methods approach. *Journal of mixed methods research*, 2(3):270–283.

Dervin, B. (1983). *An overview of sense-making research: Concepts, methods, and results to date*. The Author.

Dev, H. and Liu, Z. (2017). Identifying frequent user tasks from application logs. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces*, pages 263–273. ACM.

Devine, C. (2015). The museum digital experience: Considering the visitor's journey. In *MWA2015: Museums and the Web Asia 2015*.

Dewing, M. (2012). *Social Media: Who Uses Them?* Library of Parliament Ottawa, ON, Canada.

Dierking, L. D. and Falk, J. H. (1994). Family behavior and learning in informal science settings: A review of the research. *Science Education*, 78(1):57–72.

Dobreva, M., O'Dwyer, A., and Feliciati, P. (2012). *User studies for digital library development*. Facet.

Dolnicar, S. (2002). A review of data-driven market segmentation in tourism. *Journal of Travel & Tourism Marketing*, 12(1):1–22.

Dolnicar, S., Grün, B., and Leisch, F. (2016). Increasing sample size compensates for data problems in segmentation studies. *Journal of Business Research*, 69(2):992–999.

Dolnicar, S., Grün, B., Leisch, F., and Schmidt, K. (2014). Required sample sizes for data-driven market segmentation analyses in tourism. *Journal of Travel Research*, 53(3):296–306.

Drivas, I., Kouis, D., Kyriaki-Manessi, D., and Giannakopoulos, G. (2021). Content management systems performance and compliance assessment based on a data-driven search engine optimization methodology. *Information*, 12(7):259.

Duff, W. M. and Johnson, C. A. (2002). Accidentally found on purpose: Information-seeking behavior of historians in archives. *The Library Quarterly*, 72(4):472–496.

Duff, W. M., Monks-Leeson, E., Galey, A., et al. (2012). Contexts built and found: a pilot study on the process of archival meaning-making. *Archival Science*, 12(1):69–92.

Durbin, G. (2008). Just do it! web 2.0 as task, not technology. In Trant, J. and Bearman, D., editors, *Museums and the Web 2008: Proceedings*, Toronto. Archives & Museum Informatics. Accessed on 26/2/23.

Eklund, J. L. (2011). Cultural objects digitization planning: Metadata overview. *Visual Resources Association Bulletin*, 38(1).

Ellis, D. (1989). A behavioural approach to information retrieval system design. *Journal of documentation*, 45(3):171–212.

Ellis, D., Cox, D., and Hall, K. (1993). A comparison of the information seeking patterns of researchers in the physical and social sciences. *Journal of documentation*.

Elo, S. and Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, 62(1):107–115.

Elsweiler, D., Wilson, M. L., and Lunn, B. K. (2011). Chapter 9 understanding casual-leisure information behaviour. *New Directions in Information Behaviour*

(Library and Information Science, Volume 1) Emerald Group Publishing Limited, 1:211–241.

Enge, E., Spencer, S., Stricchiola, J., and Fishkin, R. (2012). *The art of SEO*. " O'Reilly Media, Inc.".

europeana (2014). Results of the europeana user survey 2014 - europeana professional. `http://pro.europeana.eu/blogpost/results-of-the-europeana-user-survey-2014`. (Accessed on 03/14/2016).

Everitt, B. S., Landau, S., Leese, M., and Stahl, D. (2011). *Cluster Analysis*, volume 5th. John Wiley.

Falk, J. H. (2009). *Identity and the museum visitor experience*. Left Coast Press.

Fangi, G. (2019). Aleppo before and after the war 2010–2018. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42:449–456.

Fantoni, S. F., Stein, R., and Bowman, G. (2012). Exploring the relationship between visitor motivation and engagement in online museum audiences. In *Museums and the Web*.

Farber, A. and Radensky, P. (2008). Living museum: Supporting the creation of quality user-generated content. In Trant, J. and Bearman, D., editors, *Museums and the Web 2008: Proceedings*, Toronto. Archives & Museum Informatics. Accessed 26/2/23.

Filaire, T. (2018). Clustering on mixed type data. a proposed approach using r | by thomas filaire | towards data science. `https://towardsdatascience.com/clustering-on-mixed-type-data-8bbd0a2569c3`. (Accessed on 04/10/2022).

Filippini-Fantoni, S. and Bowen, J. (2007). Bookmarking in museums: Extending the museum experience beyond the visit? In Trant, J. and Bearman, D., editors, *Museums and the Web 2007: Proceedings*, Toronto. Archives & Museum Informatics. Accessed 26/2/23.

Fiorucci, M., Khoroshiltseva, M., Pontil, M., Traviglia, A., Del Bue, A., and James, S. (2020). Machine learning for cultural heritage: A survey. *Pattern Recognition Letters*, 133:102–108.

Fisher, M. and Twiss-Garrity, B. A. (2007). Remixing exhibits: Constructing participatory narratives with on-line tools to augment museum experiences. In Trant, J. and Bearman, D., editors, *Museums and the Web 2007: Proceedings*, Toronto. Archives & Museum Informatics. Accessed on 26/02/23.

Fleming, D. (2007). Are museums about stories or objects? - interpretation conference london, 24 september 2007. `http://www.bms.edu.lv/index.php?option=com_content&task=view&id=18&Itemid=32`. Accessed August 15, 2019.

Foo, B. (2016). Nypl-publicdomain/pd-visualization: A visualization and interface that helps users explore 187k digital items from nypl in the public domain. `https://github.com/NYPL-publicdomain/pd-visualization`. (Accessed on 03/20/2023).

Formann, A. K. (1984). *The latent class analysis - Introduction to theory and application.* Beltz.

French, A. (2016). Service design thinking for museums: Technology in contexts. *MW2016: Museums and the Web.*

Frias-Martinez, E., Chen, S. Y., and Liu, X. (2008). Investigation of behavior and perception of digital library users: A cognitive style perspective. *International Journal of Information Management*, 28(5):355–365.

Fuentetaja, I. G. and Economou, M. (2008). Online access to digital collections-design and use of museum databases. *Proceedings of VSMM 2008 (Limassol, Cyprus 20-25 October 2008).*

Fuentetaja, I. G. and Economou, M. (2009). Analysis of users' access to museums websites-comparison of weblogs. In *2009 15th International Conference on Virtual Systems and Multimedia*, pages 123–128. Ieee.

Gaede, M., Bogers, T., Hall, M. M., Koolen, M., Petras, V., and Larsen, B. (2024). The impact of chiir publications: A study of eight years of chiir. In *Proceedings*

of the 2024 Conference on Human Information Interaction and Retrieval, CHIIR '24, page 34–44, New York, NY, USA. Association for Computing Machinery.

Gage, N. L. (1989). The paradigm wars and their aftermath a "historical" sketch of research on teaching since 1989. *Educational researcher*, 18(7):4–10.

Giacometti, A. (2009). The textiles browser: An experiment in rich-prospect browsing for text collections. Master's thesis, University of Alberta.

Gibbons, S. (2018). Empathy mapping: The first step in design thinking. `https://www.nngroup.com/articles/empathy-mapping/`. (Accessed on 03/11/2023).

Gilliland-Swetland, A. and White, L. (2004). Museum information professionals as providers and users of online resources. *Bulletin of the American Society for Information Science and Technology*, 30(5):23–26.

Glaser, B. and Strauss, A. (2017). *Discovery of grounded theory: Strategies for qualitative research*. Routledge.

Goldgar, A. (2000). The british museum and the virtual representation of culture in the eighteenth century. *Albion: A Quarterly Journal Concerned with British Studies*, 32(2):195–231.

Goldman, K. H., Schaller, D. T., Adventures, E. W., et al. (2004). Exploring motivational factors and visitor satisfaction in on-line museum visits. In *In D. Bearman & J. Trant (Eds.), Museums and the Web 2004*. Citeseer.

Goltz, S. (2014). A closer look at personas: What they are and how they work. `https://www.smashingmagazine.com/2014/08/a-closer-look-at-personas-part-1/`. (Accessed on 09/02/2021).

Goodale, P., Hall, M., Archer, P., Tavanti, F., and Clough, P. (2011). D 1.1 user requirements analysis.

Goodwin, K. (2005). Perfecting your personas – ux articles by uie. `https://articles.uie.com/perfecting_personas/`. (Accessed on 11/05/2021).

Goodwin, K. (2008). Getting from research to personas : Harnessing the power of data. https://articles.uie.com/research_to_personas/. (Accessed on 31/08/2021).

Google (2023). Bounce rate - analytics help. https://support.google.com/analytics/answer/1009409?hl=en. (Accessed on 03/06/2023).

Google Analytics Support (2022). bounce rate - analytics help.

Górny, M. and Mazurek, J. (2012). Key users of polish digital libraries. *The Electronic Library*, 30(4):543–556.

Gov UK, O. (2019). Internet banking, by age group, great britain, 2019.

Graham, S., Milligan, I., Weingart, S. B., and Martin, K. (2016). *Exploring big historical data: the historian's macroscope*. World Scientific.

Grant, M. J. and Booth, A. (2009). A typology of reviews: an analysis of 14 review types and associated methodologies. *Health information & libraries journal*, 26(2):91–108.

Green, L., Webb, A., and Mann, L. (2020). Museums making money from online content: Might our beliefs be holding us back? https://www.franklygreenwebb.com/museums-making-money-from-online-content-might-our-beliefs-be-holding-us-back/. (Accessed on 27/05/2022).

Greene, J. C., Caracelli, V. J., and Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational evaluation and policy analysis*, 11(3):255–274.

Grimes, C., Tang, D., and Russell, D. (2007). Query logs alone are not enough. In *Proceedings of the Workshop on Query Log Analysis: Social and Technology Challenges at the 16th International World Wide Web Conference, Banff, Canada*.

Guba, E. G., Lincoln, Y. S., et al. (1994). Competing paradigms in qualitative research. *Handbook of qualitative research*, 2(163-194):105.

Guest, G., Bunce, A., and Johnson, L. (2006). How many interviews are enough?: An experiment with data saturation and variability. *Field methods*, 18(1):59–82.

Guest, G., MacQueen, K., and Namey, E. (2014). *Applied Thematic Analysis*. SAGE Publications, Inc., Thousand Oaks.

Guldbæk Rasmussen, K., Iversen, R., Petersen, G., and Primdahl Sørensen, S. (2009). M3. 2.3 personas catalogue. Technical report, Europeana Connect.

Hadley, W. (2020). Covid-19 impact museum sector research findings. [Accessed March 23rd, 2021].

Hall, M., Goodale, P., Clough, P., and Stevenson, M. (2012a). The paths system for exploring digital cultural heritage'. In *Clare Mills, Michael Pidd and Esther Ward. Proceedings of the Digital Humanities Congress*.

Hall, M. and Walsh, D. (2021). Exploring digital cultural heritage through browsing. In Golub, K. and Liu, Y.-H., editors, *Information Organization in Digital Humanities: Global Perspectives*, Digital Research in the Arts and Humanities, pages 261–284, United Kingdom. Routledge.

Hall, M. M., Clough, P. D., de Lacalle, O. L., Soroa, A., and Agirre, E. (2012b). Enabling the discovery of digital cultural heritage objects through wikipedia. In *Proceedings of the 6th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, pages 94–100. Association for Computational Linguistics.

Ham, S. H. (2013). *Interpretation: Making a difference on purpose*. Fulcrum publishing.

Hampson, C., Agosti, M., Orio, N., Bailey, E., Lawless, S., Conlan, O., and Wade, V. (2012). The cultura project: supporting next generation interaction with digital cultural heritage collections. In *Euro-Mediterranean Conference*, pages 668–675. Springer.

Han, J., Kamber, M., and Pei, J. (2012). Front matter. In *Data Mining (Third Edition)*, The Morgan Kaufmann Series in Data Management Systems, pages i–v. Morgan Kaufmann, Boston, third edition edition.

Han, J., Pei, J., and Tong, H. (2011). *Data mining: concepts and techniques*. Morgan kaufmann.

Harmouch, M. (2021). 17 clustering algorithms used in data science and mining | towards data science. https://towardsdatascience.com/17-clustering-algorithms-used-in-data-science-mining-49dbfa5bf69a. (Accessed on 12/16/2022).

He, D. and Göker, A. (2000). Detecting session boundaries from web user logs. In *Proceedings of the BCS-IRSG 22nd annual colloquium on information retrieval research*, pages 57–66.

Hedberg, B. (1980). The design and impact of real-time computer systems. *Bjorn-Anderson, N., Hedberg, B., Mercer, D., Mumford, E., and A. Sole,(eds.), The Impact of Systems Change in Organizations, Alphen: Sijtoff and Noordhoff.*

Hedberg, B. and Mumford, E. (1975). The design of computer systems: Man's vision of man as an integral part of the system design process. *Human choice and computers*, 31:59.

Hein, G. E. (1998). *Learning in the Museum*, volume 47. Routledge.

Hennes, T. (2002). Rethinking the visitor experience: Transforming obstacle into purpose. *Curator: The Museum Journal*, 45(2):109–121.

Heo, M., Song, J.-S., and Seol, M.-W. (2013). User needs of digital service web portals: A case study. *The Journal of Educational Research*, 106(6):469–477.

Holden, R. J., Kulanthaivel, A., Purkayastha, S., Goggins, K. M., and Kripalani, S. (2017). Know thy ehealth user: Development of biopsychosocial personas from a study of older adults with heart failure. *International journal of medical informatics*, 108:158–167.

Hood, M. G. (2004). Staying away: Why people choose. *Reinventing the museum: historical and contemporary perspectives on the paradigm shift*, page 150.

Hornecker, E. and Stifter, M. (2006). Learning from interactive museum installations about interaction design for public settings. In *Proceedings of the 18th Australia conference on computer-human interaction: design: activities, Artefacts and Environments*, pages 135–142.

Hsieh-Yee, I. (1993). Effects of search experience and subject knowledge on the search tactics of novice and experienced searchers. *Journal of the American Society for Information Science*, 44(3):161–174.

Hughes, L. M. (2004). *Digitizing collections: strategic issues for the information manager*, volume 2. Facet Publishing.

Hunt, T. (2019). Should museums return their colonial artefacts? *The Guardian*, 29.

Hutchinson, H. B., Rose, A., Bederson, B. B., Weeks, A. C., and Druin, A. (2005). The international children's digital library: a case study in designing for a multilingual, multicultural, multigenerational audience. *Information Technology and Libraries*, 24(1):4–12.

Ibrahim, N., Mohamad Ali, N., and Mohd Yatim, N. F. (2015). Factors facilitating cultural learning in virtual architectural heritage environments: End user perspective. *Journal on Computing and Cultural Heritage (JOCCH)*, 8(2):8.

IDEO (2015). Design kit: The human-centered design toolkit | ideo.com. `https://www.ideo.com/post/design-kit`. (Accessed on 02/24/2023).

Ingwersen, P. and Järvelin, K. (2005). Information retrieval in context: Irix. In *ACM SIGIR Forum*, volume 39, pages 31–39. ACM.

Isard, A., Oberlander, J., Matheson, C., and Androutsopoulos, I. (2003). Speaking the users' languages. *IEEE Intelligent Systems*, 18(1):40–45.

Ismail, M. N. (2007). *A data model for multimedia database of malaysian cultural heritage artefacts management*. PhD thesis, Universiti Teknologi Mara.

I.T.R.C. (2013). Groundwater statistics and monitoring compliance, statistical tools for the project life cycle.

Jackson, R., Bazley, M., Patten, D., and Martin, K. (1998). Using the web to change the relation between a museum and its users. *Museums and the Web 1998*.

Jafari, M. and Ansari-Pour, N. (2019). Why, when and how to adjust your p values? *Cell journal*, 20(4).

Jansen, A., Van Mechelen, M., and Slegers, K. (2017). Personas and behavioral theories: A case study using self-determination theory to construct overweight personas. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 2127–2136.

Jansen, B., Zhang, M., Booth, D., Park, D., Zhang, Y., Kathuria, A., and Bonner, P. (2009). To what degree can log data profile a web searcher? *Proceedings of the American Society for Information Science and Technology*, 46(1):1–19.

Jansen, B. J. (2006). Search log analysis: What it is, what's been done, how to do it. *Library & information science research*, 28(3):407–432.

Jansen, B. J., Jung, S.-G., Salminen, J., Guan, K. W., and Nielsen, L. (2021). Strengths and weaknesses of persona creation methods: Guidelines and opportunities for digital innovations. In *Proceedings of the Annual Hawaii International Conference on System Sciences*. Hawaii International Conference on System Sciences.

Jansen, B. J., Spink, A., and Koshman, S. (2007). Web searcher interaction with the dogpile. com metasearch engine. *Journal of the American Society for Information Science and Technology*, 58(5):744–755.

Jansen, B. J., Spink, A., and Pedersen, J. (2005). A temporal comparison of altavista web searching. *Journal of the American Society for Information Science and Technology*, 56(6):559–570.

Jansen, J. (2021). How are personas used in real organizations? `https://persona.qcri.org/blog/how-are-personas-used-in-real-organizations/`. (Accessed on 06/02/2023).

Jean-François, G., Laetitia, R., and Stefan, D. (2013). Is the coverage of google scholar enough to be used alone for systematic reviews. *BMC medical informatics and decision making*, 13(1):1.

Jin, X. and Han, J. (2010). K-medoids clustering. In Sammut, C. and Webb, G. I., editors, *Encyclopedia of Machine Learning*, pages 564–565. Springer US, Boston, MA.

Joffe, H. (2012). Thematic analysis. In Harper, D. and Thompson, A., editors, *Qualitative Research Methods in Mental Health and Psychotherapy: A Guide for Students and Practitioners*, pages 209–223. Wiley, Chichester.

Johnson, A. (2008). Users, use and context: supporting interaction between users and digital archives. *What Are Archives?: Cultural and Theoretical Perspectives: A Reader*, pages 145–64.

Johnson, R. B. and Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational researcher*, 33(7):14–26.

Jones, S., Cunningham, S. J., McNab, R., and Boddie, S. (2000). A transaction log analysis of a digital library. *International Journal on Digital Libraries*, 3(2):152–169.

Kachhadiya, B. C. and Patel, B. (2018). A survey on sequential pattern mining algorithm for web log pattern data. In *2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI)*, pages 1269–1273. Ieee.

Kansal, T., Bahuguna, S., Singh, V., and Choudhury, T. (2018). Customer segmentation using k-means clustering. In *2018 international conference on computational techniques, electronics and mechanical systems (CTEMS)*, pages 135–139. IEEE.

Kassambara, A. (2017a). Mca - multiple correspondence analysis in r: Essentials. `http://www.sthda.com/english/articles/31-principal-component-methods-in-r-practical-guide/114-mca-multiple-correspondence-analysis-in-r-essentials/`. (Accessed on 06/22/2021).

Kassambara, A. (2017b). *Practical guide to cluster analysis in R: Unsupervised machine learning*, volume 1. Sthda.

Kaufman, L. and Rousseeuw, P. J. (2009). *Finding groups in data: an introduction to cluster analysis*, volume 344. John Wiley & Sons.

Kelly, L. and Russo, A. (2008). From ladders of participation to networks of participation: Social media and museum audiences. In *Museums and the Web*.

Kim, H. M. and Wiggins, J. (2016). A factor analysis approach to persona development using survey data. In *Proceedings of the 2016 Library Assessment Conference*, volume 11.

King, E., Smith, M. P., Wilson, P. F., and Williams, M. A. (2021). Digital responses of uk museum exhibitions to the covid-19 crisis, march–june 2020. *Curator: The Museum Journal*.

Kling, R. (1977). The organizational context of user-centered software designs. *MIS quarterly*, pages 41–52.

Knemeyer, D. (2015). Design thinking and ux: Two sides of the same coin. *Interactions*, 22(5):66–68.

Koch, T., Golub, K., and Ardö, A. (2006). Users browsing behaviour in a ddc-based web service: a log analysis. *Cataloging & classification quarterly*, 42(3-4):163–186.

Koleva, B., Egglestone, S. R., Schnädelbach, H., Glover, K., Greenhalgh, C., Rodden, T., and Dade-Robertson, M. (2009). Supporting the creation of hybrid museum experiences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1973–1982.

Kolko, J. (2015). Design thinking comes of age. *Harvard Business Review*, 2015(September):pp.66–71.

Komarac, T., Ozretic-Dosen, D., and Skare, V. (2020). Managing edutainment and perceived authenticity of museum visitor experience: insights from qualitative study. *Museum Management and Curatorship*, 35(2):160–181.

Koushik, M., Lee, E. J., Pieroni, L., Sun, E., and Yeh, C.-W. (2010). Re-envisioning the museum experience: combining new technology with social-networking. In *International Conference on Entertainment Computing*, pages 248–253. Springer.

Koutrika, G. and Simitsis, A. (2006). An enhanced search interface for information discovery from digital libraries. In *International Conference on Theory and Practice of Digital Libraries*, pages 87–98. Springer.

Krantz, A., Korn, R., and Menninger, M. (2009). Rethinking museum visitors: Using k-means cluster analysis to explore a museum's audience. *Curator: The Museum Journal*, 52(4):363–374.

Kravchyna, V. and Hastings, S. K. (2002). Informational value of museum web sites. *First Monday*, 7(2).

Kuhlthau, C. C. (1989). Information search process: A summary of research and implications for school library media programs. *School Library Media Quarterly*, 18(1):19–25.

Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user's perspective. *Journal of the American society for information science*, 42(5):361–371.

Kujala, S. and Kauppinen, M. (2004). Identifying and selecting users for user-centered design. In *Proceedings of the Third Nordic Conference on Human-Computer Interaction*, NordiCHI '04, page 297–303, New York, NY, USA. Association for Computing Machinery.

Kuncheva, L. I., Hadjitodorov, S. T., and Todorova, L. P. (2006). Experimental comparison of cluster ensemble methods. In *Information Fusion, 2006 9th International Conference on*, pages 1–7. Ieee.

Kusnick, J., Mayr, E., Seirafi, K., Beck, S., Liem, J., and Windhager, F. (2024). Every thing can be a hero! narrative visualization of person, object, and other biographies. *Informatics*, 11(2):26.

Lang, B., Liu, X., and Li, W. (2013). The next-generation search engine: Challenges and key technologies. In Gaol, F. L., editor, *Recent Progress in Data Engineering and Internet Technology*, pages 239–248, Berlin, Heidelberg. Springer Berlin Heidelberg.

Lapatovska, I., O'Brien, H. L., Rieh, S. Y., and Wildemuth, B. (2011). Capturing the complexity of information interactions: Measurement and evaluation issues. *Proceedings of the ASIST Annual Meeting*, 48.

Lau, T. and Horvitz, E. (1999). Patterns of search: analyzing and modeling web query refinement. In *UM99 User Modeling*, pages 119–128. Springer.

Lee, D. (1988). Developing effective communications (1988). Accessed 18/2/2023.

Lee, M., Kwahk, J., Han, S. H., Jeong, D., Park, K., Oh, S., and Chae, G. (2020). Developing personas & use cases with user survey data: A study on the millennials' media usage. *Journal of Retailing and Consumer Services*, 54:102051.

Lehmann, J., Lalmas, M., Yom-Tov, E., and Dupret, G. (2012). Models of user engagement. In *User Modeling, Adaptation, and Personalization*, pages 164–175. Springer.

Lepkowska-White, E. and Imboden, K. (2013). Effective design for usability and interaction: the case of art museum websites. *Journal of Internet commerce*, 12(3):284–305.

Leporini, B. and Norscia, I. (2009). Translating museum visual contents into descriptions for blind users: A multidisciplinary approach. In *Computer-Aided Design of User Interfaces VI*, pages 47–57. Springer.

Li, C., Bernoff, J., Fiorentino, R., and Glass, S. (2007). Social technographics. mapping participation in activities forms the foundation of a social strategy. *Forrester Research Inc. Published in April*, 19.

Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., and Pechenkina, E. (2023). Generative ai and the future of education: Ragnarök or reformation? a paradoxical perspective from management educators. *The International Journal of Management Education*, 21(2):100790.

Lin, A. C. and Gregor, S. D. (2006). Designing websites for learning and enjoyment: A study of museum experiences. *The International Review of Research in Open and Distributed Learning*, 7(3).

Lin, F. S. and Cassidy, T. (2008). Affective textile and costume museum website design. *International Journal of Fashion Design, Technology and Education*, 1(1):23–33.

Lončarić, D., Prodan, M. P., and Ribarić, I. (2016). The influence of a visitor's perceptions of a museum's website design on behavioural intentions. *Ekonomski vjesnik/Econviews-Review of Contemporary Business, Entrepreneurship and Economic Issues*, 29(1):65–79.

Lopatovska, I., Bierlein, I., Lember, H., and Meyer, E. (2013). Exploring requirements for online art collections. *Proceedings of the American Society for Information Science and Technology*, 50(1):1–4.

Lotina, L. (2014). Reviewing museum participation in online channels in latvia. *Museum Management and Curatorship*, 29(3):280–292.

Lourdi, I. and Nikolaidou, M. (2009). Guidelines for developing digital cultural collections. In *Handbook of research on digital libraries: design, development, and impact*, pages 198–205. IGI Global.

Lucas, H. C. (1975). *Why information systems fail*. Columbia University Press.

Lucas Jr, H. C. (1971). A user-oriented approach to systems design. In *Proceedings of the 1971 26th annual conference*, pages 325–338.

Mansourian, Y. (2008). Coping strategies in web searching. *Program*.

Marchionini, G. (1989). Information-seeking strategies of novices using a full-text electronic encyclopedia. *Journal of the American Society for Information Science*, 40(1):54–66.

Marchionini, G. (1995). *Information seeking in electronic environments*. Cambridge university press.

Marchionini, G. (2006). Exploratory search: from finding to understanding. *Communications of the ACM*, 49(4):41–46.

Marchionini, G., Plaisant, C., and Komlodi, A. (2003). The people in digital libraries: Multifaceted approaches to assessing needs and impact. *Digital library use: Social practice in design and evaluation*, pages 119–160.

Martella, C., Miraglia, A., Frost, J., Cattani, M., and van Steen, M. (2016). Visualizing, clustering, and predicting the behavior of museum visitors. *Pervasive and Mobile Computing*, 38:430–443.

Martins, C. (2012). Museum audio guides as an accessibility enhancer. *Accesibilidad en la nueva era de las comunicaciones Profesionales y universidad: un diálogo imprescindible*, pages 101–115.

Marty, P. F. (2004). The evolving roles of information professionals in museums. *Bulletin of the American Society for Information Science and Technology*, 30(5):20–23.

Marty, P. F. (2006). Meeting user needs in the modern museum: Profiles of the new museum information professional. *Library & information science research*, 28(1):128–144.

Marty, P. F. (2007a). The changing nature of information work in museums. *Journal of the American Society for Information Science and Technology*, 58(1):97–107.

Marty, P. F. (2007b). Museum websites and museum visitors: Before and after the museum visit. *Museum Management and Curatorship*, 22(4):337–360.

Marty, P. F. (2008). Museum websites and museum visitors: Digital museum resources and their use. *Museum Management and Curatorship*, 23(1):81–99.

Mason, D. D. and McCarthy, C. (2008). Museums and the culture of new media: an empirical model of new zealand museum websites. *Museum Management and Curatorship*, 23(1):63–80.

Mathis, L. J. (1999). *Comparison of Nurse Practitioner Exercise Prescription: Mississippi versus Retrospective International Internet Data*. PhD thesis, Mississippi University for Women.

Matthews, Bob & Ross, L. (2010). *Research Methods: A Practical Guide for the Social Sciences*. Pearson.

Matthews, T., Judge, T., and Whittaker, S. (2012). How do designers and user experience professionals actually perceive and use personas? In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1219–1228.

Mayr, E., Federico, P., Miksch, S., Schreder, G., Smuc, M., and Windhager, F. (2016). Visualization of cultural heritage data for casual users. In *Proceedings of the 1st Workshop for Visualization for the Digital Humanities, Baltimore, MD*.

McGinn, J. J. and Kotamraju, N. (2008). Data-driven persona development. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1521–1524. ACM.

Medić, S., Pivac, T., Ivkov-Džigurski, A., Košić, K., and Pavlović, N. (2015). Leisure time and museums-motives of visits. *Journal of the Geographical Institute "Jovan Cvijić" SASA*, 65(3):391–406.

Mertens, D. M. (2014). *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods*. Sage publications.

Meyerson, J., Galloway, P., and Bias, R. (2012). Improving the user experience of professional researchers: Applying a user-centered design framework in archival repositories. *Proceedings of the American Society for Information Science and Technology*, 49(1):1–7.

Miaskiewicz, T. and Kozar, K. A. (2011). Personas and user-centered design: How can personas benefit product design processes? *Design studies*, 32(5):417–430.

Michalakis, K. and Caridakis, G. (2022). Context awareness in cultural heritage applications: a survey. *ACM Journal on Computing and Cultural Heritage (JOCCH)*, 15(2):1–31.

Mijač, T., Jadrić, M., and Ćukušić, M. (2018). The potential and issues in data-driven development of web personas. In *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pages 1237–1242. IEEE.

Miles, M. B. and Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Sage.

Miller, D. and Slater, D. (2000). *The Internet: an ethnographic approach*. Routledge.

Moher, D., Liberati, A., Tetzlaff, J., and Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the prisma statement. *Annals of internal medicine*, 151(4):264–269.

Morris, M. R. and Teevan, J. (2012). Exploring the complementary roles of social networks and search engines. In *Human-Computer Interaction Consortium Workshop (HCIC)*, pages 1–10.

Morse, C., Niess, J., Lallemand, C., Wieneke, L., and Koenig, V. (2021). Casual leisure in rich-prospect: Advancing visual information behavior for digital museum collections. *J. Comput. Cult. Herit.*, 14(3).

Morse, J. (2006). Biased reflections: principles of sampling and analysis in qualitative inquiry. *Moving beyond effectiveness in evidence synthesis: Methodological issues in the synthesis of diverse sources of evidence*, pages 53–60.

Mulder, S. and Yaar, Z. (2006). *The user is always right: A practical guide to creating and using personas for the web*. New Riders.

Nauta, G. J. and van den Heuvel, W. (2015). Survey report on digitisation in european cultural heritage institutions 2015. *Den Haag: DEN Foundation*.

Ndayizigamiye, P. (2022). Implementation of mhealth interventions for public healthcare in east africa. *Advances in Healthcare Information Systems and Administration*, pages 106–124.

Nemo (2020). Survey on the impact of the covid-19 situation on museums in europe. [Accessed June 29th, 2021].

Newman, D. A. (2014). Missing data: Five practical guidelines. *Organizational Research Methods*, 17(4):372–411.

Newman, K. and Tourle, P. (2011). The impact of cuts on uk museums. *London: Museums Association*.

News, B. (2013). Syria clashes destroy ancient aleppo minaret - bbc news. `https://www.bbc.co.uk/news/world-middle-east-22283746`. (Accessed on 02/27/2023).

Nicholas, D., Huntington, P., and Jamali, H. R. (2007). The use, users, and role of abstracts in the digital scholarly environment. *The Journal of Academic Librarianship*, 33(4):446–453.

Nicholas, D., Huntington, P., Jamali, H. R., and Watkinson, A. (2006). The information seeking behaviour of the users of digital scholarly journals. *Information Processing & Management*, 42(5):1345–1365.

Nielsen, L., Salminen, J., Jansen, B. J., An, J., Kwak, H., and Jung, S.-G. (2019). Automatic persona generation for online content creators: Conceptual rationale and a research agenda. *Personas-User focused design*, pages 135–160.

Norman, D. A. (1988). *The psychology of everyday things*. Basic books.

Norman, D. A. (2002). *The design of everyday things*. Basic books.

Norman, D. A. and Draper, S. W. (1986). *User Centered System Design: New Perspectives on Human-computer Interaction*. CRC Press.

Normore, L. (2003). Studying special collections and the web: An analysis of practice. *First Monday*, 8(10).

Normore, L. F. (2008). Characterizing a digital library's users: steps towards a nuanced view of the user. *Proceedings of the American Society for Information Science and Technology*, 45(1):1–7.

Noyes, J. and Baber, C. (1999). *User-centred design of systems*. Springer Science & Business Media.

Oakley, A. (1999). Paradigm wars: some thoughts on a personal and public trajectory. *International Journal of Social Research Methodology*, 2(3):247–254.

O'Brien, H. L. and Toms, E. G. (2008). What is user engagement? a conceptual framework for defining user engagement with technology. *Journal of the American Society for Information Science and Technology*, 59(6):938–955.

on Longevity, S. C. (2016). The sightlines project: seeing our way to living long & living well in 21st century america. Accessed November 12, 2016.

Oppenheim, A. N. (1992). *Questionnaire design, interviewing and attitude measurement*. Pinter Publishing.

Osman, E. E. M. (2016). Relevant museum experiences: A proposed visitor categorization matrix. Master's thesis, University of Cincinnati.

Ottaviano, S. and Allegra, M. (2013). Cultural heritage and educational web sites. In *The International Scientific Conference eLearning and Software for Education*, volume 1, page 622. " Carol I" National Defence University.

Padmaja, S. and Sheshasaayee, A. (2016). Clustering of user behaviour based on web log data using improved k-means clustering algorithm. *International Journal of Engineering and Technology (IJET)*, 8(1):305–310.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., and Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, page n71.

Pantano, E. (2011). Virtual cultural heritage consumption: a 3d learning experience. *International Journal of Technology Enhanced Learning*, 3(5):482–495.

Parker, C., Scott, S., and Geddes, A. (2019). Snowball sampling. *SAGE research methods foundations*.

Parry, R., Eikhof, D. R., Barnes, S.-A., and Kispeter, E. (2018). Mapping the museum digital skills ecosystem - phase one report. Technical report, University of Leicester AHRC project 'One by One'.

Paterno, F. and Mancini, C. (1999). Designing web user interfaces adaptable to different types of use. In *Proceedings of Museums and the Web, New Orleans, LA, USA, Available online at http://www. archimuse. com/mw99/papers/paterno/paterno. html*.

Paternò, F. and Mancini, C. (2000). Effective levels of adaptation to different types of users in interactive museum systems. *Journal of the American Society for Information Science*, 51(1):5–13.

Pavlik, J. V. (2023). Collaborating with chatgpt: Considering the implications of generative artificial intelligence for journalism and media education. *Journalism & Mass Communication Educator*, 78(1):84–93.

Payne, E. and Whittaker, L. (2006). *Developing essential study skills*. Pearson Education.

Peacock, D. and Brownbill, J. (2007). Audiences, visitors, users: Reconceptualising users of museum on-line content and services. In *Museums and the Web*. Citeseer.

Pirolli, P. and Card, S. (1999). Information foraging. *Psychological review*, 106(4):643.

Pirolli, P. and Card, S. (2005). The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of international conference on intelligence analysis*, volume 5, pages 2–4.

Pisoni, G., Díaz-Rodríguez, N., Gijlers, H., and Tonolli, L. (2021). Human-centered artificial intelligence for designing accessible cultural heritage. *Applied Sciences*, 11(2).

Pitkow, J. and Recker, M. (1994). Results from the first world-wide web user survey. *Computer Networks and ISDN Systems*.

Porter, S. R., Whitcomb, M. E., and Weitzer, W. H. (2004). Multiple surveys of students and survey fatigue. *New Directions for Institutional Research*, 2004(121):63–73.

Pousman, Z., Stasko, J., and Mateas, M. (2007). Casual information visualization: Depictions of data in everyday life. *IEEE transactions on visualization and computer graphics*, 13(6).

Power, C., Freire, A., Petrie, H., and Swallow, D. (2012). Guidelines are only half of the story: accessibility problems encountered by blind users on the web. In

Proceedings of the SIGCHI conference on human factors in computing systems, pages 433–442.

Prescott, L. and Erway, R. (2011). *Single search: the quest for the Holy Grail*. OCLC Research Dublinˆ eOhio Ohio.

Pretorius, M. and Sangham, I. (2016). Personas informed by user research for the western cape government. In *Proceedings of the 16th European Conference on EGovernment. Ljubljana, Slovenia*, pages 162–169.

Proudfoot, K. (2023). Inductive/deductive hybrid thematic analysis in mixed methods research. *Journal of Mixed Methods Research*, 17(3):308–326.

Pruitt, J. and Adlin, T. (2010). *The persona lifecycle: keeping people in mind throughout product design*. Elsevier.

Pruitt, J. and Grudin, J. (2003). Personas: practice and theory. In *Proceedings of the 2003 conference on Designing for user experiences*, pages 1–15. ACM.

Qiu, W. and Joe, H. (2006). Generation of random clusters with specified degree of separation. *Journal of Classification*, 23(2):315–334.

Radjenović, D., Heričko, M., Torkar, R., and Živkovič, A. (2013). Software fault prediction metrics: A systematic literature review. *Information and software technology*, 55(8):1397–1418.

Rasmussen, K. and Petersen, G. (2012). Personas. In Dobreva, M., O'Dwyer, A., and Feliciati, P., editors, *User studies for digital library development*, chapter 10, pages 105–113. Facet.

Redish, J. and Wixon, J. (2003). Task analysis. In Jacko, J. A. and Sears, A., editors, *The human-computer interaction handbook: fundamentals, evolving technologies, and emerging applications*, page 922–940. Lawrence Erlbaum Associates.

Regmi, P. R., Waithaka, E., Paudyal, A., Simkhada, P., and Teijlingen, E. V. (2017). Guide to the design and application of online questionnaire surveys. *Nepal Journal of Epidemiology*, 6(4):640–644.

Rex, B. (2020). Which museums to fund? examining local government decision-making in austerity. *Local Government Studies*, 46(2):186–205.

Richards, L. and Morse, J. M. (2012). *README FIRST for a Users Guide to Qualitative Methods*. Sage publications.

Rivas, R., Sadah, S. A., Guo, Y., and Hristidis, V. (2020). Classification of health-related social media posts: Evaluation of post content–classifier models and analysis of user demographics. *JMIR Public Health Surveill*, 6(2):e14952.

Rivas, S. (2006). *Coverage Error: The Achilles' Heel of Web-based Surveys*. Center for Demography and Ecology, University of Wisconsin.

Roberto, F. (2008). Exploring museum collections on-line: The quantitative method. In Trant, J. and Bearman, D., editors, *Museums and the Web 2008: Proceedings*, Toronto. Archives & Museum Informatics. Accessed on 15/02/23.

Rodríguez-Vázquez, C., Castellanos-García, P., and Martínez-Fernández, V. A. (2023). Cultural tourism in a post-covid-19 scenario: The french way of saint james in spain from the perspective of promotional communication. *Societies*, 13(1).

Rogina, J. (2022). New vs returning visitors in google analytics | clearpath online. `https://clearpath.online/resources/new-vs-returning-visitors-google-analytics/`. (Accessed on 08/14/2022).

Rönkkö, K., Hellman, M., Kilander, B., and Dittrich, Y. (2004). Personas is not applicable: local remedies interpreted in a wider context. In *Proceedings of the eighth conference on Participatory design: Artful integration: interweaving media, materials and practices-Volume 1*, pages 112–120.

Roppola, T. (2013). *Designing for the museum visitor experience*. Routledge.

Roussou, M., Katifori, A., Pujol, L., Vayanou, M., and Rennick-Egglestone, S. J. (2013). A life of their own: museum visitor personas penetrating the design lifecycle of a mobile experience. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, pages 547–552. ACM.

Ruecker, S., Radzikowska, M., and Sinclair, S. (2011). *Visual interface design for digital cultural heritage: A guide to rich-prospect browsing*. Ashgate Publishing, Ltd.

Russell-Rose, T. and Clough, P. (2016). Mining search logs for usage patterns. *Text Mining and Visualization: Case Studies using Open-Source Tools*, 40:153.

Russell-Rose, T. and Tate, T. (2012). *Designing the search experience: The information architecture of discovery*. Newnes.

Rutledge, L., Aroyo, L., and Stash, N. (2006). Determining user interests about museum collections. In *Proceedings of the 15th international conference on World Wide Web*, pages 855–856. ACM.

Ryan, J. (2023). A planet without google search - cnet. `https://www.cnet.com/science/features/a-planet-without-google-search/`. (Accessed on 03/20/2023).

Salgado, M. (2008). Breaking apart participation in museums. In Trant, J. and Bearman, D., editors, *Museums and the Web 2008: Proceedings*, Toronto. Archives & Museum Informatics. Accessed on 15/02/23.

Salminen, J., Guan, K., Jung, S.-G., Chowdhury, S. A., and Jansen, B. J. (2020). A literature review of quantitative persona creation. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–14.

Salminen, J., gyo Jung, S., Nielsen, L., Şengün, S., and Jansen, B. J. (2022a). How does varying the number of personas affect user perceptions and behavior? challenging the 'small personas' hypothesis! *International Journal of Human-Computer Studies*, 168:102915.

Salminen, J., Jansen, B., and Jung, S.-G. (2022b). Survey2persona: Rendering survey responses as personas. In *Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*, UMAP '22 Adjunct, page 67–73, New York, NY, USA. Association for Computing Machinery.

Salminen, J., Jansen, B. J., An, J., Kwak, H., and Jung, S.-G. (2019). *Automatic Persona Generation for Online Content Creators: Conceptual Rationale and a Research Agenda*, pages 135–160. Springer London, London.

Salminen, J., SENGUN, S., Kwak, H., Jansen, B. J., An, J., JUNG, S.-g., Vieweg, S., and Harrell, D. F. (2018). From 2,772 segments to five personas: Summarizing a diverse online audience by generating culturally adapted personas. *First Monday*, 23(6).

Samis, P. (2008). Who has the responsibility for saying what we see? mashing up museum, artist, and visitor voices, on-site and on-line. In Trant, J. and Bearman, D., editors, *Museums and the Web 2008: Proceedings*, Toronto. Archives & Museum Informatics. Accessed on 15/02/23.

Sarraf, S. (1999). A survey of museums on the web: Who uses museum websites? *Curator: The Museum Journal*, 42(3):231–243.

Sarstedt, M., Mooi, E., et al. (2014). A concise guide to market research. *The Process, Data, and*.

Schäfer, K., Rasche, P., Bröhl, C., Theis, S., Barton, L., Brandl, C., Wille, M., Nitsch, V., and Mertens, A. (2019). Survey-based personas for a target-group-specific consideration of elderly end users of information and communication systems in the german health-care sector. *International journal of medical informatics*, 132:103924.

Schaller, D. and Goldman, K. H. (2004). Exploring motivational factors and visitor satisfaction in on-line museum visits. *Museums and the Web 2004*, pages 1–17.

Schatz, R. and Chen, H. (1999). Digital libraries: technological advances and social impacts. *Computer*, 32(2):45–50.

Schneiderman, B. and Plaisant, C. (2010). *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Pearson Education.

Schouten, F. (1995). Improving visitor care in heritage attractions. *Tourism Management*, 16(4):259–261.

Schweibenz, W. (2004). Virtual museums. *The Development of Virtual Museums„,ICOM News Magazine*, 3(3):3.

Shaheen, K. (2015). Isis blows up arch of triumph in 2,000-year-old city of palmyra | islamic state | the guardian. `https://www.theguardian.com/world/2015/oct/05/isis-blows-up-another-monument-in-2000-year-old-city-of-palmyra`. (Accessed on 02/27/2023).

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423.

Shneiderman, B. (1998). *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Addison Weseley Longman.

Skov, M. (2009). *The reinvented museum: Exploring information seeking behaviour in a digital museum context*. PhD thesis, Københavns Universitet'Københavns Universitet', Faculty of Humanities, School of Library and Information Science, Royal School of Library and Information Science. unpublished thesis.

Skov, M. (2013). Hobby-related information-seeking behaviour of highly dedicated online museum visitors. *Information Research*, 18(4).

Skov, M. and Ingwersen, P. (2008). Exploring information seeking behaviour in a digital museum context. In *Proceedings of the Second International Symposium on Information Interaction in Context*, IIiX '08, pages 110–115, New York, NY, USA. ACM, ACM.

Skov, M. and Ingwersen, P. (2014). Museum web search behavior of special interest visitors. *Library & Information Science Research*, 36(2):91–98.

Skov, M., Pedersen, H., Larsen, B., and Ingwersen, P. (2004). Testing the principle of polyrepresentation. In *Proceedings of ACM SIGIR 2004 Workshop on" Information Retrieval in Context*, pages 47–49.

Soloman, H. (2021). just-eat-case-study. Accessed December 11, 2022.

Soren, B. J. and Network, C. H. I. (2005). Best practices in creating quality online experiences for museum users. *Museum management and curatorship*, 20(2):131–148.

Speakman, R., Hall, M. M., and Walsh, D. (2018). User engagement with generous interfaces for digital cultural heritage. In *Digital Libraries for Open Knowledge: 22nd International Conference on Theory and Practice of Digital Libraries, TPDL 2018, Porto, Portugal, September 10–13, 2018, Proceedings 22*, pages 186–191. Springer.

Spellerberg, M., Granata, E., and Wambold, S. (2016). Visitor-first, mobile-first: Designing a visitor-centric mobile experience. In *Museums and the Web*.

Spennemann, D. H. (2024). Will artificial intelligence affect how cultural heritage will be managed in the future? responses generated by four genai models. *Heritage*, 7(3):1453–1471.

Spink, A., Bateman, J., and Jansen, B. J. (1999). Searching the web: A survey of excite users. *Internet research*, 9(2):117–128.

Spink, A., Wolfram, D., Jansen, M. B., and Saracevic, T. (2001). Searching the web: The public and their queries. *Journal of the American society for information science and technology*, 52(3):226–234.

Srinivasan, R., Boast, R., Becvar, K. M., and Furner, J. (2009). Blobgects: Digital museum catalogs and diverse user communities. *Journal of the American Society for Information Science and Technology*, 60(4):666–678.

Stebbins, R. A. (2007). *Serious leisure: A perspective for our time*, volume 95. Transaction Publishers.

Steinerová, J. and Šušol, J. (2005). Library users in human information behaviour. *Online Information Review*.

Stenmark, D. (2008). Identifying clusters of user behavior in intranet search engine log files. *Journal of the American Society for Information Science and Technology*, 59(14):2232–2243.

Stickdorn, M. and Schneider, J. (2011). *This is Service Design Thinking: Basics, Tools, Cases*. BIS Publishers.

Stiller, J. (2014). *From curation to collaboration*. PhD thesis, Humboldt-Universität zu Berlin, Philosophische Fakultät I.

Straughan, C. (2019). The changing face of museum tours - museumnext. https://www.museumnext.com/article/the-changing-face-of-museum-tours/. (Accessed on 11/20/2021).

Sullivan, A. M. (2016). Cultural heritage & new media: A future for the past. *J. Marshall Rev. Intell. Prop. L.*, 15:604–645.

Suter, L. E. (2005). Multiple methods: research methods in education projects at nsf. *International Journal of Research & Method in Education*, 28(2):171–181.

Swain, J. (2018). A hybrid approach to thematic analysis in qualitative research: Using a practical example. *Sage research methods*.

Sweetnam, M., Siochru, M., Agosti, M., Manfioletti, M., Orio, N., and Ponchia, C. (2013). Stereotype or spectrum: Designing for a user continuum. In *the Proceedings of the First Workshop on the Exploration, Navigation and Retrieval of Information in Cultural Heritage, ENRICH*.

Sweetnam, M. S., Agosti, M., Orio, N., Ponchia, C., Steiner, C. M., Hillemann, E.-C., Siochrú, M. Ó., and Lawless, S. (2012). User needs for enhanced engagement with cultural heritage collections. In *International Conference on Theory and Practice of Digital Libraries*, pages 64–75. Springer.

Sylaiou, S., Mania, K., Karoulis, A., and White, M. (2010). Exploring the relationship between presence and enjoyment in a virtual museum. *International Journal of Human-Computer Studies*, 68(5):243–253.

Talbot, J. F., Kaplan, R., Kuo, F. E., and Kaplan, S. (1993). Factors that enhance effectiveness of visitor maps. *Environment and Behavior*, 25(6):743–760.

Tashakkori, A. and Teddlie, C. (2003). *Handbook of Mixed Methods in Social & Behavioral Research*. SAGE Publications.

Tashakkori, A., Teddlie, C., and Teddlie, C. B. (1998). *Mixed methodology: Combining qualitative and quantitative approaches*. SAGE Publications.

Taylor, E. (2015). The national archive, uk government web archive discovery project: User survey. An unpublished technical report.

Taylor, R. S. (1982). Value-added processes in the information life cycle. *Journal of the American Society for Information Science*, 33(5):341–346.

Templeton, C. A. (2011). Museum visitor engagement through resonant, rich and interactive experiences. Master's thesis, School of Design, Carnegie Mellon University.

Terras, M. (2008). Digital heritage: Applying digital imaging to cultural heritage.lindsay macdonald (ed.). *Literary and Linguistic Computing*, 23(2):244–246.

Thoma, V. and Williams, B. (2009). Developing and validating personas in e-commerce: A heuristic approach. In *IFIP Conference on Human-Computer Interaction*, pages 524–527. Springer.

Torres de Souza, M., Hörding, O., and Karol, S. (2019). The story of spotify personas | spotify design. `https://spotify.design/article/the-story-of-spotify-personas`. (Accessed on 04/06/2023).

Tranfield, D., Denyer, D., and Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British journal of management*, 14(3):207–222.

UK, G. (2017). Learning about users and their needs - service manual - gov.uk. `https://www.gov.uk/service-manual/user-research/start-by-learning-user-needs`. (Accessed on 07/13/2022).

UNESCO (2009). Unesco framework for cultural statistics 2009. [Accessed 2022].

UNESCO (2020). Museums around the world: In the face of covid-19. [Accessed June 29th, 2021].

University, C. (nd). A brief history of oa | open research. `https://osc.cam.ac.uk/open-access/brief-history-oa`. (Accessed on 03/18/2024).

U.S. General Services Administration (GSA) Technology Transformation Service (2019). Personas. `https://www.usability.gov/how-to-and-tools/methods/personas.html`. (Accessed on 12/20/2022).

Van der Maaten, L. and Hinton, G. (2008). Visualizing data using t-sne. *Journal of machine learning research*, 9(11).

Van Hooland, S. (2006). From spectator to annotator: possibilities offered by user-generated metadata for digital cultural heritage collections. In *Immaculate Catalogues: Taxonomy, Metadata and Resource Discovery in the 21st Century, Proceedings of CILIP Conference*.

Vane, O. (2020). *Timeline design for visualising cultural heritage data*. PhD thesis, Royal College of Art.

Vilar, P. and Šauperl, A. (2014). Archival literacy: Different users, different information needs, behaviour and skills. In *Information Literacy. Lifelong Learning and Digital Citizenship in the 21st Century*, pages 149–159. Springer.

Villa, R., Clough, P. D., Hall, M. M., and Rutter, S. A. (2013). Search or browse? casual information access to a cultural heritage collection. In *EuroHCIR*, volume 1033, pages 19–22. Citeseer.

Villaespesa, E. (2019). Museum collections and online users: Development of a segmentation model for the metropolitan museum of art. *Visitor Studies*, 22(2):233–252.

Villaespesa, E., Tate, U., and Stack, J. (2015). Finding the motivation behind a click: Definition and implementation of a website audience segmentation. *Museums and the Web 2015: Proceedings*.

Voorbij, H. (2010). The use of web statistics in cultural heritage institutions. *Performance Measurement and Metrics*, 11(3):266–279.

Walliman, N. (2006). *Research strategies and design*. SAGE Publications.

Walsh, D., Clough, P., and Foster, J. (2016). User categories for digital cultural heritage. In Clough, P., Goodale, P., Agosti, M., and Lawless, S., editors, *Proceedings of the First International Workshop on Accessing Cultural Heritage at Scale co-located with Joint Conference on Digital Libraries 2016 (JCDL 2016)*, volume 1611 of *CEUR Workshop Proceedings*. CEUR-WS.org.

Walsh, D., Clough, P., Hall, M. M., Hopfgartner, F., and Foster, J. (2021). Clustering and classifying users from the national museums liverpool website. In Berget, G., Hall, M. M., Brenn, D., and Kumpulainen, S., editors, *Linking Theory and Practice of Digital Libraries*, pages 202–214, Cham. Springer International Publishing.

Walsh, D., Clough, P., Hall, M. M., Hopfgartner, F., Foster, J., and Kontonatsios, G. (2019). Analysis of transaction logs from national museums liverpool. In Doucet, A., Isaac, A., Golub, K., Aalberg, T., and Jatowt, A., editors, *Digital Libraries for Open Knowledge*, pages 84–98, Cham. Springer International Publishing.

Walsh, D. and Hall, M. (2015). Just looking around: Supporting casual users initial encounters with digital cultural heritage. In Gade, M., Hall, M., Huurdeman, H., Kamps, J., Koolen, M., Skov, M., Toms, E., and Walsh, D., editors, *Proceedings of the First International Workshop on Supporting Complex Search Tasks co-located with the 37th European Conference on Information Retrieval (ECIR 2015)*, volume 1338 of *CEUR Workshop Proceedings*. CEUR-WS.org.

Walsh, D., Hall, M., Clough, P., and Foster, J. (2017). The ghost in the museum website: Investigating the general public's interactions with museum websites. In Kamps, J., Tsakonas, G., Manolopoulos, Y., Iliadis, L., and Karydis, I., editors, *Research and Advanced Technology for Digital Libraries*, pages 434–445, Cham. Springer, Springer International Publishing.

Walsh, D., Hall, Mark M.and Clough, P., and Foster, J. (2020). Characterising online museum users: a study of the national museums liverpool museum website. *International Journal on Digital Libraries*, 21(1):75–87.

Wang, G., Zhang, X., Tang, S., Zheng, H., and Zhao, B. Y. (2016). Unsupervised clickstream clustering for user behavior analysis. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 225–236. ACM.

Wang, Y., Aroyo, L. M., Stash, N., and Rutledge, L. (2007). Interactive user modeling for personalized access to museum collections: The rijksmuseum case study. In *International Conference on User Modeling*, pages 385–389. Springer.

Warwick, C. (2012). Studying users in digital humanities. *Digital Humanities in Practice*, pages 1–21.

Warwick, C., Terras, M., Galina, I., Huntington, P., and Pappa, N. (2008). Library and information resources and users of digital resources in the humanities. *Program*, 42(1):5–27.

Watters, A. (2022). Top 25 information technology statistics & facts for 2023 | comptia. `https://connect.comptia.org/blog/information-technology-stats-facts`. (Accessed on 02/27/2023).

White, R. W. (2016). *Interactions with search systems*. Cambridge University Press.

White, R. W., Kules, B., and Bederson, B. (2005). Exploratory search interfaces: categorization, clustering and beyond: report on the xsi 2005 workshop at the human-computer interaction laboratory, university of maryland. In *ACM SIGIR Forum*, volume 39, pages 52–56. ACM.

Whitelaw, M. (2015). Generous interfaces for digital cultural collections. *Digital Humanities Quarterly*, 9(1):38.

Wilson, M. L. (2017). The tetris model of resolving information needs within the information seeking process. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, pages 147–154.

Wilson, T. D. (1981). On user studies and information needs. *Journal of documentation*, 37(1):3–15.

Wilson, T. D. (1999). Models in information behaviour research. *Journal of documentation*, 55(3):249–270.

Wilson, T. D. (2000). Human information behavior. *Informing science*, 3:49.

Wixon, D. and Wilson, C. (1997). The usability engineering framework for product design and evaluation. In *Handbook of human-computer interaction*, pages 653–688. Elsevier.

Wong, K. J. (2022). 7 evaluation metrics for clustering algorithms | by kay jan wong | towards data science. https://towardsdatascience.com/7-evaluation-metrics-for-clustering-algorithms-bdc537ff54d2. (Accessed on 04/03/2023).

Woodruff, A., Aoki, P. M., Hurst, A., and Szymanski, M. H. (2001). Electronic guidebooks and visitor attention. In *Ichim (1)*, pages 437–454.

Woolfolk, A. (2001). *Educational psychology*. Allyn and Bacon, 8th edition.

Xie, H. (2006). Evaluation of digital libraries: Criteria and problems from users' perspectives. *Library and Information Science Research*, 28(3):433–452.

Yang, G. (2009). A study on the user-centered interface design for virtual museums. In *2009 IEEE 10th International Conference on Computer-Aided Industrial Design & Conceptual Design*, pages 1647–1651. Ieee.

Yu, J. X., Ou, Y., Zhang, C., and Zhang, S. (2005). Identifying interesting visitors through web log classification. *IEEE Intelligent Systems*, 20(3):55–59.

Yuxuan Hu, K. L. and Meng, A. (2018). Agglomertive hierarchical clustering using ward linkage. https://jbhender.github.io/Stats506/F18/GP/Group10.html. (Accessed on 01/09/2023).

Yvonne Feilzer, M. (2010). Doing mixed methods research pragmatically: Implications for the rediscovery of pragmatism as a research paradigm. *Journal of mixed methods research*, 4(1):6–16.

Zahidi, Z., Lim, Y. P., and Woods, P. C. (2013). User experience for digitization and preservation of cultural heritage. In *2013 International Conference on Informatics and Creative Multimedia*, pages 13–16. IEEE.

Zahidi, Z., Lim, Y. P., and Woods, P. C. (2014). Understanding the user experience (ux) factors that influence user satisfaction in digital culture heritage online collections for non-expert users. In *2014 Science and Information Conference*, pages 57–63. IEEE.

Zairul, M. (2021). Can member check be verified in real time? introducing arc (asking, record, confirm) for member checking validation strategy in qualitative research. *Engineering Journal*, 25(1):245–251.

Zakraoui, J., Saleh, M., and Ja'am, J. A. (2019). Text-to-picture tools, systems, and approaches: a survey. *Multimedia Tools and Applications*, 78(16):22833–22859.

Zhang, J. and Kamps, J. (2010). Search log analysis of user stereotypes, information seeking behavior, and contextual evaluation. In *Proceedings of the third symposium on Information interaction in context*, pages 245–254, New York, NY, USA. ACM, ACM.

Appendices

Appendix A

Publications included in the systematic literature review

A.1 Starting list of publications

Table A.1: Table of publications identified for possible use in the systematic literature review

Year of publication	**Publications**	**Excluded at Screening**	**Excluded at Eligibility**	**Included in final set**
1981	Wilson, On user studies and information needs	x		
1993	Hsieh-Yee, Effects of search experience and subject knowledge on the search tactics of novice and experienced searchers			x
1995	Schouten, Improving visitor care in heritage attractions			x
1998	Booth, Understanding the information needs of visitors to museums			x
1998	Jackson et al., Using the Web to Change the Relation Between a Museum and its Users	x		
1999	Paterno and Mancini, Designing web user interfaces adaptable to different types of use			x
1999	Sarraf, A survey of museums on the web: Who uses museum websites?			x
1999	Schatz and Chen, Digital libraries: technological advances and social impacts	x		
2000	Paternò and Mancini, Effective levels of adaptation to different types of users in interactive museum systems			x
2001	Cunliffe et al., Usability evaluation for museum web sites			x
2002	Duff and Johnson, Accidentally found on purpose: Information-seeking behavior of historians in archives			x
2002	Kravchyna and Hastings, Informational value of museum web sites			x
2003	Assadi et al., Users and uses of online digital libraries in France			x
2003	Bolchini and Mylopoulos, From task-oriented to goal-oriented Web requirements analysis			x
2003	Isard et al., Speaking the users' languages			x
2003	Marchionini et al., The people in digital libraries: Multifaceted approaches to assessing needs and impact			x
2003	Normore, Studying special collections and the Web: An analysis of practice			x

Table A.1 continued from previous page

Year of publication	Publications	Excluded at Screening	Excluded at Eligibility	Included in final set
2004	Bonfigli et al., Virtual visits to cultural heritage supported by web-agents	x		
2004	Gilliland-Swetland and White, Museum information professionals as providers and users of online resources			x
2004	Goldman et al., Exploring motivational factors and visitor satisfaction in on-line museum visits			x
2004	Marty, The evolving roles of information professionals in museums	x		
2004	Schweibenz, Virtual museums			x
2004	White et al., Exploratory search interfaces: categorization, clustering and beyond: report on the XSI 2005 workshop at the Human-Computer Interaction Laboratory, University of Maryland	x		
2005	Soren and Network, Best practices in creating quality online experiences for museum users			x
2005	Steinerová and Šušol, Library users in human information behaviour		x	
2005	Hutchinson et al., The international children's digital library: a case study in designing for a multilingual, multicultural, multigenerational audience	x		
2006	Hornecker and Stifter, Learning from interactive museum installations about interaction design for public settings	x		
2006	Koutrika and Simitsis, An enhanced search interface for information discovery from digital libraries	x		
2006	Lin and Gregor, Designing websites for learning and enjoyment: A study of museum experiences			x
2006	Marty, Meeting user needs in the modern museum: Profiles of the new museum information professional			x
2006	Rutledge et al., Determining user interests about museum collections			x
2006	Nicholas et al., The information seeking behaviour of the users of digital scholarly journals	x		
2006	Xie, Evaluation of digital libraries: Criteria and problems from users' perspectives	x		

Table A.1 continued from previous page

Year of publication	Publications	Excluded at Screening	Excluded at Eligibility	Included in final set
2007	Amin et al., Searching in the cultural heritage domain: capturing cultural heritage expert information seeking needs			x
2007	Bastanlar, User behaviour in web-based interactive virtual tours			x
2007	Marty, Museum websites and museum visitors: Before and after the museum visit			x
2007	Nicholas et al., The use, users, and role of abstracts in the digital scholarly environment			x
2007	Li et al., Social technographics. mapping participation in activities forms the foundation of A social strategy			x
2007	Peacock and Brownbill, Audiences, visitors, users: Reconceptualising users of museum on-line content and services	x		
2007	Stebbins, Serious leisure: A perspective for our time			x
2007	Wang et al., Interactive user modeling for personalized access to museum collections: The rijksmuseum case study		x	
2008	Frias-Martinez et al., Investigation of behavior and perception of digital library users: A cognitive style perspective			x
2008	Amin et al., Understanding cultural heritage experts' information seeking needs			x
2008	Blandford et al., Evaluating system utility and conceptual fit using CASSM			x
2008	Fuentetaja and Economou, Online access to digital collections-Design and use of museum databases		x	
2008	Johnson, Users, use and context: supporting interaction between users and digital archives			x
2008	Lin and Cassidy, Affective textile and costume museum website design		x	
2008	Marty, Museum websites and museum visitors: Digital museum resources and their use			x
2008	Mason and McCarthy, Museums and the culture of new media: an empirical model of New Zealand museum websites			x
2008	Normore, Characterizing a digital library's users: steps towards a nuanced view of the user			x
2008	Skov and Ingwersen, Exploring Information Seeking Behaviour in a Digital Museum Context			x

Table A.1 continued from previous page

Year of publication	Publications	Excluded at Screening	Excluded at Eligibility	Included in final set
2008	Warwick et al., Library and information resources and users of digital resources in the humanities			x
2008	Cifter and Dong, User characteristics: Professional vs. lay users			x
2009	Falk, Identity and the museum visitor experience			x
2009	Fuentetaja and Economou, Analysis of Users' Access to Museums Websites-Comparison of Weblogs			x
2009	Leporini and Norscia, Translating museum visual contents into descriptions for blind users: A multidisciplinary approach			x
2009	Guldbæk Rasmussen et al., M3. 2.3 Personas Catalogue		x	
2009	Srinivasan et al., Blobgects: Digital museum catalogs and diverse user communities			x
2009	Skov, The reinvented museum: Exploring information seeking behaviour in a digital museum context			x
2009	Koleva et al., Supporting the creation of hybrid museum experiences		x	
2009	Yang, A study on the user-centered interface design for virtual museums		x	
2010	Sylaiou et al., Exploring the relationship between presence and enjoyment in a virtual museum		x	
2010	Antoniou and Lepouras, Modeling visitors' profiles: A study to investigate adaptation aspects for museum learning technologies	x		
2010	Chowdhury, From digital libraries to digital preservation research: the importance of users and context	x		
2011	Baeza-Yates and Maarek, Web retrieval: the role of users.	x		
2011	Bertacchini and Morando, The Future of Museums in the Digital Age: New Models of Access and Use of Digital Collections			x
2011	Connaway et al., Visitors and residents: What motivates engagement with the digital information environment?		x	
2011	Templeton, Museum visitor engagement through resonant, rich and interactive experiences			x
2011	Goodale et al., D 1.1 User Requirements Analysis			x

Table A.1 continued from previous page

Year of publication	Publications	Excluded at Screening	Excluded at Eligibility	Included in final set
2011	Carmel et al., User modeling for web applications	x		
2012	Ardissono et al., Personalization in cultural heritage: the road travelled and the one ahead			x
2012	Duff et al., Contexts built and found: a pilot study on the process of archival meaning-making			x
2012	Dobreva et al., User studies for digital library development		x	
2012	Fantoni et al., Exploring the relationship between visitor motivation and engagement in online museum audiences			x
2012	Górny and Mazurek, Key users of Polish digital libraries			x
2012	Hampson et al., The CULTURA project: supporting next generation interaction with digital cultural heritage collections			x
2012	Power et al., Guidelines are only half of the story: accessibility problems encountered by blind users on the web		x	
2012	Sweetnam et al., User needs for enhanced engagement with cultural heritage collections			x
2012	Warwick, Studying users in digital humanities			x
2012	Koushik et al., Re-envisioning the museum experience: combining new technology with social-networking	x		
2013	Cairns, Mutualizing museum knowledge: Folksonomies and the changing shape of expertise			x
2013	Ottaviano and Allegra, Cultural heritage and educational web sites			x
2013	Roussou et al., A life of their own: museum visitor personas penetrating the design lifecycle of a mobile experience		x	
2013	Skov, Hobby-related information-seeking behaviour of highly dedicated online museum visitors			x
2013	Heo et al., User Needs of Digital Service Web Portals: A Case Study	x		
2013	Lopatovska et al., Exploring requirements for online art collections		x	
2013	Lepkowska-White and Imboden, Effective design for usability and interaction: the case of art museum websites		x	
2014	Skov and Ingwersen, Museum Web search behavior of special interest visitors			x

Table A.1 continued from previous page

Year of publication	Publications	Excluded at Screening	Excluded at Eligibility	Included in final set
2014	Lotina, Reviewing museum participation in online channels in Latvia		x	
2015	Devine, The Museum Digital Experience: Considering the Visitor's Journey	x		
2015	Ibrahim et al., Factors Facilitating Cultural Learning in Virtual Architectural Heritage Environments: End User Perspective			x
2015	Bartoli et al., Museumvisitors: a dataset for pedestrian and group detection, gaze estimation and behavior understanding	x		
2015	Benouaret and Lenne, Personalizing the museum experience through context-aware recommendations	x		
2016	Osman, Relevant Museum Experiences: A Proposed Visitor Categorization Matrix			x
2016	Spellerberg et al., Visitor-first, mobile-first: Designing a visitor-centric mobile experience			x
2016	Martella et al., Visualizing, clustering, and predicting the behavior of museum visitors	x		
2016	Lončarić et al., The influence of a visitor's perceptions of a museum's website design on behavioural intentions	x		
2016	French, Service design thinking for museums: Technology in contexts	x		

A.2 Final list of publications after filtering

Table A.2: Table of publications included in the systematic literature review

Year of publication	**Publications**	**# of articles**
1993	Hsieh-Yee, Effects of search experience and subject knowledge on the search tactics of novice and experienced searchers	1
1995	Schouten, Improving visitor care in heritage attractions	1
1998	Booth, Understanding the information needs of visitors to museums	1
1999	Paterno and Mancini, Designing web user interfaces adaptable to different types of use Sarraf, A survey of museums on the web: Who uses museum websites?	2
2000	Paternò and Mancini, Effective levels of adaptation to different types of users in interactive museum systems	1
2001	Cunliffe et al., Usability evaluation for museum web sites	1
2002	Duff and Johnson, Accidentally found on purpose: Information-seeking behavior of historians in archives Kravchyna and Hastings, Informational value of museum web sites	2
2003	Assadi et al., Users and uses of online digital libraries in France Bolchini and Mylopoulos, From task-oriented to goal-oriented Web requirements analysis Isard et al., Speaking the users' languages Marchionini et al., The people in digital libraries: Multifaceted approaches to assessing needs and impact Normore, Studying special collections and the Web: An analysis of practice	5
2004	Gilliland-Swetland and White, Museum information professionals as providers and users of online resources Goldman et al., Exploring motivational factors and visitor satisfaction in on-line museum visits Schweibenz, Virtual museums	3
2005	Soren and Network, Best practices in creating quality online experiences for museum users	1

Table A.2 continued from previous page

Year of publication	Publications	# of articles
2006	Lin and Gregor, Designing websites for learning and enjoyment: A study of museum experiences Marty, Meeting user needs in the modern museum: Profiles of the new museum information professional Rutledge et al., Determining user interests about museum collections	3
2007	Amin et al., Searching in the cultural heritage domain: capturing cultural heritage expert information seeking needs Bastanlar, User behaviour in web-based interactive virtual tours Marty, Museum websites and museum visitors: Before and after the museum visit Nicholas et al., The use, users, and role of abstracts in the digital scholarly environment Li et al., Social technographics. mapping participation in activities forms the foundation of A social strategy Stebbins, Serious leisure: A perspective for our time	6
2008	Frias-Martinez et al., Investigation of behavior and perception of digital library users: A cognitive style perspective Amin et al., Understanding cultural heritage experts' information seeking needs Blandford et al., Evaluating system utility and conceptual fit using CASSM Johnson, Users, use and context: supporting interaction between users and digital archives Marty, Museum websites and museum visitors: Digital museum resources and their use Mason and McCarthy, Museums and the culture of new media: an empirical model of New Zealand museum websites Normore, Characterizing a digital library's users: steps towards a nuanced view of the user Skov and Ingwersen, Exploring Information Seeking Behaviour in a Digital Museum Context Warwick et al., Library and information resources and users of digital resources in the humanities Cifter and Dong, User characteristics: Professional vs. lay users	10
2009	Falk, Identity and the museum visitor experience Fuentetaja and Economou, Analysis of Users' Access to Museums Websites-Comparison of Weblogs Leporini and Norscia, Translating museum visual contents into descriptions for blind users: A multidisciplinary approach Srinivasan et al., Blobgects: Digital museum catalogs and diverse user communities Skov, The reinvented museum: Exploring information seeking behaviour in a digital museum context	5

Table A.2 continued from previous page

Year of publication	**Publications**	**# of articles**
2011	Bertacchini and Morando, The Future of Museums in the Digital Age: New Models of Access and Use of Digital Collections Templeton, Museum visitor engagement through resonant, rich and interactive experiences Goodale et al., D 1.1 User Requirements Analysis	3
2012	Ardissono et al., Personalization in cultural heritage: the road travelled and the one ahead Duff et al., Contexts built and found: a pilot study on the process of archival meaning-making Fantoni et al., Exploring the relationship between visitor motivation and engagement in online museum audiences Górny and Mazurek, Key users of Polish digital libraries Hampson et al., The CULTURA project: supporting next generation interaction with digital cultural heritage collections Sweetnam et al., User needs for enhanced engagement with cultural heritage collections, Warwick, Studying users in digital humanities	7
2013	Cairns, Mutualizing museum knowledge: Folksonomies and the changing shape of expertise Ottaviano and Allegra, Cultural heritage and educational web sites Skov, Hobby-related information-seeking behaviour of highly dedicated online museum visitors	3
2014	Skov and Ingwersen, Museum Web search behavior of special interest visitors	1
2015	Ibrahim et al., Factors Facilitating Cultural Learning in Virtual Architectural Heritage Environments: End User Perspective	1
2016	Osman, Relevant Museum Experiences: A Proposed Visitor Categorization Matrix Spellerberg et al., Visitor-first, mobile-first: Designing a visitor-centric mobile experience	2

A.3 2024 Paper Analysis

Due to the methodological error between April 2016 and 2024 the original data for the keyword searches has been lost. This analysis conducted in March 2024 demonstrates that the starting figure of 84 papers for the PRISMA systematic literature review is a viable figure.

A.3.1 Method

A list of all keyword combinations has been created, containing one option from each of the 4 groups. Groups of three and four word keyphrases were generated. 560 3-word combinations and 1820 4-word combinations were produced, totalling 2380 keyphrases.

Using SerpApi was used to query Google Scholar and retrieve the paper listings for the keyphrases. All 2380 keyphrases were searched one by one, with the following search parameters used: the upper year was set to 2016 (which is when the original survey was carried out and written up); also, the upper limit of 20 results a page was also set (this is because API returns a JSON list for each page of Google Scholar results, and so to keep this pagination searching to a minimum the maximum number of results per page was used). All searches were conducted as exact searches, meaning the keyphrases were encased in double-quotes.

Many of the terms returned with no results (only 13 returned any actual results). The zero results were randomly manually checked on Google Scholar directly to confirm it was not an issue with the script or API. All checks returned no results, confirming the script and API were working correctly.

The terms that did return results can be seen in table A.3.

Nearly all of the results that were returned came from queries was a combination of three phrases/words, with only one query returning a result from the four keyword combinations.

Table A.3: 2024 literature search using Google Scholar via SERPAPI.

Keyword	**Results**	**Results with PDF**
museum online users	3	2
museum online users searching	1	1
museum virtual users	1	1
museum virtual visitors	12	8
museum web users	7	5
museum web visitors	13	9
museum website user studies	1	0
museum website users	31	20
museum website visitors	34	25
online museum visitors	101	65
users information needs personas	1	0
virtual museum visitors	104	68
web users information needs	37	31
Total papers (to the end of 2016) with duplicates	**346**	**235**
Duplicates		127
Total papers without duplicates		**108**
Number of papers post April 2016		21
Total papers up to end April 2016		**87**

From all searches, 346 listings were returned, with 235 having a PDF resource link available. There were 127 papers which appeared in a number of query results, so duplicates were removed, leaving 108 papers to review. However, this search included papers that were published till the end of 2016, and the original survey was conducted up to and including April 2016 (gathering all papers available at that time) these were printed and analysed in late April/May 2016, so any results which had papers for 2016 listed (25 papers) were manually checked to see if they were published post-April and if they were, they were excluded. Removing another 21 papers. This leaves 87 papers available.

The original PRISMA had 84 papers available from the initial searches of the DBs. All were returned from Google Scholar, and no additional were added during the manual searches of the individual DBs. Whilst 87 is close but not exactly 84, the addition of time may account for the differences, 2016 did not have as many authors self-publishing their works as they do today; there may also have been papers which

were published and listed in Google Scholar but were embargoed and did not have a PDF then, but the PDF has since been released. Also, these results still include all papers returned till the end of April 2016, but the original search could not have had all of these due to the search concluding in mid-April.

The full list of papers that started the PRISMA process can be seen in table A.1.

In this study, 87 papers, equating to 25% of the Search Engine Results Page (SERP) listings, were identified as having an available PDF. This proportion initially appears low; however, a recent study conducted in 2024 by Gaede et al. reveals that only 38% of all Conference on Human Information Interaction and Retrieval (CHIIR) papers had a downloadable version accessible (Gaede et al., 2024). Considering that the UK's Research Excellence Framework only announced the policy for open access in 2014 and the policy came into effect in April 2016 (University, nd), the figure of 25% of papers having a PDF does not seem overly surprising.

Appendix B

All user groups with characteristics

Table B.1: All user groups with characteristics

Year	**Author**	**User Group Label**	**Characteristics**
2012	Warwick	Academic	Higher education use of educational portals
2008	Warwick	Academic Support Person In HE	No definition in paper
2011	Bertacchini	Academic/Commercial Publisher	No definition in paper
2016	Falk & Dierking (2013) in Osman	Affinity Seekers	Visitors who are motivated to visit a museum because it speaks to their sense of heritage and their bigger identity.
2008	Cifter	Amateur	Might be as advanced as professionals in terms of both use and innovativeness.
2012	Sweetnam	Apprentice Investigator	Students at an advanced undergraduate and post graduate level. Some knowledge of historical period and/or cultural context addressed by the resource.
2003	Assadi	Book Lover	Gallica served as a catalogue prior to making a purchase
2015	Ham (2013) in Ibrahim	Captive User	Those who learn because they are required to
2015	Ibrahim	Casual User	Those who voluntarily seek cultural information due to intrinsic motivation such as to satisfy personal curiosity; but will not spend a lot of time dealing with uninteresting information. Not domain or system experts but could have some background of heritage but not to level of expert
2013	Ottaviano	Children	seeks educational resources.
2007	Li	Collectors	Focussed on building: developing and maintaining private collections.
2013	Skov	Collectors	Build and maintain personal archives on private collection - pursue technical information on the commercial, social and physical circumstances in which items are acquired along with knowledge, thereby providing a broad understanding of the items historical and contemporary production and use.

Table B.1 continued from previous page

Year	Author	User Group Label	Characteristics
2011	(as in Dallas Museum of Art (DMA) Templeton	Committed Enthusiast	Knowledgeable and connected to art. Enjoy learning from experts.
2007	Li	Creators	Create and upload assets online such as web pages and videos.
2007	Li	Critics	Comments on blogs and other sites.
1995	Schouten	Curator	Consider words and letters the only medium to transfer and idea
2011	(as in Dallas Museum of Art (DMA) Templeton	Curious Participants	Enjoy art and museums
2008	Johnson	Digital Archive User	Does not possess knowledge or experience to locate archival sources. Will often become frustrated early in the interaction process.
2013	Ottaviano	Disabled	No definition in paper
2011	(as in Dallas Museum of Art (DMA) Templeton	Discerning Independents	Knowledgeable and connected to art. Enjoy developing their own views on art
1998	Booth	Educational Visitor	Requires information on opening hours, prices, the museums facilities, what's on, notable exhibits and navigation aids and more detailed information to help plan visits. . . and project based information
2008	Skov & Ingwersen - (Described by Booth, 1998)	Educational Visitor	Requires information on opening hours; prices and facilities etc in addition more detailed information to help plan visits and project information
2015	Pantano (2011) in Ibrahim	Enjoyer	Enjoyers are those who appreciate the virtual exploration of the cultural heritage for personal pleasure
2009	Falk	Experience Seekers	Want to see the most renowned pieces and make memories.
2011	Templeton (using Falk 2009)	Experience Seekers	Want to see the most renowned pieces and make memories.
2012	Fantoni (using Falk 2009)	Experience Seekers	aspires to be exposed to the things and ideas that exemplify what is best and intellectually most important with in a culture or community.

Table B.1 continued from previous page

Year	Author	User Group Label	Characteristics
2016	Falk & Dierking (2013) in Osman	Experience Seekers	See museum as an important destination - they are collecting experiences
2016	Spellerberg (using Falk 2009)	Experience Seekers	No definition in paper
2008	Cifter	Experienced User	may have some previous experience with the product or the relevant task, but their knowledge of the task is much more limited when compared with professional user.
1999	Paterno	Expert	paper appears to be a pre-write-up of the 2000 version
2000	Paterno	Expert	Experts want to have full access to all the information available. They need minimal support to formulate their requests, and should be allowed to formulate such requests in a flexible way.
2003	Isard	Expert	Might be interested in a list of references to published articles that discusses the selected exhibits.
2004	Schweibenz	Expert	No definition in paper
2006	Rutledge	Expert	knowledge structure developed by experts such as art theoreticians and art historians.
2007	Amin	Expert	Professional in three roles: curators - registrars and researchers

Table B.1 continued from previous page

Year	Author	User Group Label	Characteristics
2008	Amin	Expert	five expert roles - Researchers - Curator -Registrars - Teachers and students - IT personnel - Experts daily search tasks are dominated by a range of different (relatively complex and high level) information gathering tasks, while the tools tend to be geared towards support for (relatively simple and low level) fact finding tasks. Second, many search tasks require experts to use and combine results from multiple sources, while the tools typically provide access to a single source. Third direct communication as a means for information transfer is greatly valued by experts, also trust in information source is an important aspect of experts search activities.
2012	Hampson	Expert	Expert users already have a deep knowledge of the content of the digital collections allowing their research to create new insights into these artefacts. Aware of the relevant search parameters to access the digital content and are motivated to interact with the system because of their research interest.
2013	Cairns	Expert	No definition in paper
2015	Pantano (2011) in Ibrahim	Expert	Experts are specialists in the field of cultural heritage.
2009	Srinivasan	Expert Communities	have a meaningful relationship to the collections.
2009	Falk	Explorer	Want to learn new information and understand new concepts. Motivated by a need to satisfy personal curiosity and interest in an intellectually challenging environment
2012	Fantoni (using Falk 2009)	Explorer	motivated by a need to satisfy personal curiosity and interest in an intellectually challenging environment.
2016	Falk & Dierking (2013) in Osman	Explorer	visitors who are driven by curiosity, they value learning but they are not necessarily experts.
2016	Spellerberg (using Falk 2009)	Explorer	No definition in paper

Table B.1 continued from previous page

Year	Author	User Group Label	Characteristics
2011	Templeton (using Falk 2009)	Explorers	Want to learn new information and understand new concepts.
2009	Falk	Facilitators	Want to ensure that their companions meet their visit goal
2011	Templeton (using Falk 2009)	Facilitators	Want to ensure that their companions meet their visit goals.
2012	Fantoni (using Falk 2009)	Facilitators	motivated by the wish to engage in a meaningful social experience with someone who they care about in a educationally supportive environment.
2016	Falk & Dierking (2013) in Osman	Facilitators	visitors who are socially motivated, they are focussed on enabling the learning and experience of others they care about.
2016	Spellerberg (using Falk 2009)	Facilitators	No definition in paper
2008	Frias-Martinez	Field Independence/Dependence	individuals exhibit more individualistic behaviours - do not need external referents to aid the processing of information vs more social orientation better at learning material with human content
2008	Skov & Ingwersen	Genealogist	Seek meaning and connections between documents in the archives
2012	Sweetnam	General Public	adults and children - a large diverse set of users that bring a very wide range of interests, technical abilities, contextual and/or historical awareness to collections
1999	Sarraf	General User	Mostly females that are college educated, work in museums.
2015	Pantano(2011) in Ibrahim	General User	General users are those who usually visit the website more for some passing curiosity rather than to retrieve information to improve their knowledge of cultural heritage
1998	Booth	General Visitor	Requires information on opening hours, prices, the museums facilities, what's on, notable exhibits and navigation aids
2008	Skov & Ingwersen	General Visitor	Requires information on opening hours; prices and facilities etc

Table B.1 continued from previous page

Year	Author	User Group Label	Characteristics
2011	Goodale	General Visitor	Can be a lifelong learner such as a hobbyist, Genealogist or amateur historian. Could also be a general tourist with no knowledge.
2002	Duff	Historian	an expert user of text
2003	Marchionini	Hobbyist	(e.g., genealogy, Civil War, railroads, other examples). high motivation, typically high domain knowledge, a range of library system knowledge, high focus, and high time allocations
2007	Stebbins	Hobbyist & Professionals	launch themselves on a (leisure) career centred on acquiring and expressing a combination of its special skills; knowledge and experience. Hobbyists are classified according to five categories: collectors, makers and tinkerers, activity participants (in non-competitive, rule-based, pursuits such as fishing and barbershop singing), players of sports and games (in competitive, rule-based activities with no professional counterparts like long-distance running and competitive swimming) and the enthusiasts of the liberal arts hobbies
2009	Falk	Hobbyist & Professionals	Want to see specific pieces or exhibits. Motivated by the desire to further specific intellectual needs in a setting with a specific subject matter focus
2009	Skov	Hobbyist & Professionals	A need to persevere. Leisure career that proceeds in stages: beginning; development; establishment; maintenance; and decline. Proactive knowledge. experience and skill acquisition. Durable benefits of personal and social rewards. Unique ethos or culture. Strong identification with the chosen pursuit
2011	Templeton (using Falk 2009)	Hobbyist & Professionals	Want to see and study specific pieces or exhibits.
2012	Fantoni (using Falk 2009)	Hobbyist & Professionals	posses the desire to further specific intellectual needs in a setting with a specific subject mater focus.

Table B.1 continued from previous page

Year	Author	User Group Label	Characteristics
2016	Falk & Dierking (2013) in Osman	Hobbyist & Professionals	Visitors who feel a close tie between the museum content and their professional or hobbyist passion.
2016	Spellerberg (using Falk 2009)	Hobbyist & Professionals	No definition in paper
2002	Duff	Humanities Scholars	willing to use special libraries and collections but need assistance when using the systems. Often turn to browsing to identify relevant sources.
2012	Warwick	Humanities Scholars	Higher education use of educational portals in a specific domain
2008	Frias-Martinez	Imager/Verbalizer	internal and passive vs external and stimulating
2007	Li	Inactives	do not participate in social computing activities
2012	Gorny	Incidental Users	Users who visit a digital library on an occasional basis, and do not plan to make systematic use of its resources.
2008	Warwick	Independent Researchers	No definition in paper
2015	Pantano (2011) in Ibrahim	Informationalist	Informationalists are those who have the intent of improving their knowledge
2012	Sweetnam et, al.,	Informed User	Researchers who are not professional academics but have knowledge of some aspects addressed by the resource
1999	Paterno	Intermediate	paper appears to be a pre-writeup of the 2000 version
2007	Li	Joiners	Primarily active on social networks.
2007	Nicholas	Junior Researcher	Conduct abstract only searches - no clear definition.
2008	Cifter	Lay Users	Lay users have limited or no training in a particular area, however they are likely to have personal interests or special needs in that area.
2004	Schweibenz	Laymen	No definition in paper
2008	Warwick	Lecturer/Academic	No definition in paper

Table B.1 continued from previous page

Year	Author	User Group Label	Characteristics
2003	Marchionini	Library Staff (Library of congress)	high motivation, medium to high domain knowledge, high library system knowledge, high focus, and limited time allocations
2008	Normore	Lifelong Learners Group	people who no longer enrolled in educational programmes but had a personal interest in information about local history - but who had no formal or professional need for that type of information
2012	Hampson	Mediators	mediators are scholar who act as go between of general public and the system.
2004	Gilliland	MIP (Museum Information Professional)	Users, producers and providers of digital content while understanding the needs of staff and public when producing online resources
2006	Marty	MIP (Museum Information Professional)	Specialises in the application of information science and technology in museums and whose interests lie in managing unique information resources found in museums
2006	Lin	Museum And Educational Experts	has educational, e-learning or informatics and communications backgrounds or experiences.
2009	Skov	Museum Curator	Museum curators have a dual interaction between museum objects (including their documentation) and the associated historic context. Their work tasks include: collection management, exhibition planning and design, research, and answering of inquiries to the museum from the public.
2011	Goodale	Museum Curator	Producer of content for heritage consumers. Creates exhibitions and promotes collections/exhibitions.
2007	Marty	Museum Professional (curators, educators, web developers)	Museum professionals need to understand what museum visitors are looking for before and after a museum visit and make this information available to them.

Table B.1 continued from previous page

Year	Author	User Group Label	Characteristics
2009	Skov	Museum Professional	Museum documentation is traditionally written for and by museum professionals. Often containing professional jargon. Digitizing cultural heritage collections involves opening up databases, previously the sole domain of the museum professionals.
2008	Mason	Museum Staff	Museum staff plan, implement and maintain museum websites. Often developing them in a haphazard fashion.
2009	Srinivasan	Museum Staff	elite experts of the collections held.

Table B.1 continued from previous page

Year	Author	User Group Label	Characteristics
2008	Marty	Museum Visitor	The typical online museum visitor completing this online survey, therefore, visits museums approximately four times a year, visits museum websites approximately once a week, and considers it very important for museums to have museum websites. Online museum visitors are likely to use most types of online museum resources, especially online images and research materials. Most online museum visitors have very different expectations for museums and museum websites. Online museum visitors have a strong positive relationship with museum information resources in that they are frequent users of information resources in museums, and they expect museum information resources to be easily accessible online. Online museum visitors see museums and museum websites as complementary, where one is not likely to replace the other as users search for and access information. Online museum visitors are likely to make frequent visits to museums and museum websites, including visits to museum websites independent of planning or returning from museum visits. They are also likely to use a wide variety of digital museum resources in their daily lives, including online images of artefacts and online research materials.
2016	Osman	Museum Visitor	literature based
2015	Ham (2013) in Ibrahim	Non-Captive User	User who have the option to ignore the information without baring any punishment
2012	Hampson	Non-Domain Users	need an improved user experience.
2013	Cairns	Non-Expert	No definition in paper

Table B.1 continued from previous page

Year	Author	User Group Label	Characteristics
2003	Assadi	Non-Professional Researcher	non professional researchers centres of interest were specific and well defined. Happy to utilise catalogues/indexes.
1999	Paterno	Novice	paper appears to be a pre-write-up of the 2000 version
1993	Hsieh-Yee	Novice (End User Searchers)	defined as non-professional searchers who have little or no search experience and have not taken courses on online searching or attended workshops provided by librarians or system vendors.
2006	Rutledge	Novice User	Can be hindered by lacking knowledge of rich and complex knowledge structures of museum collections as they are developed by experts.
2008	Cifter	Novice User	are new to the task or the product and usually they do not have enough information to perform the task with the device.
2009	Srinivasan	Novice User	Do not find systems easy to use because they are not familiar with the terms used for searches. Prefer content to be in language they would use and not the language of experts.
2012	Warwick	Novice User	May become confused and give up using them [applications] despite the rich possibilities they offer.
2003	Marchionini	Object Seeker	(e.g., some authors, CD-ROM/multimedia developers, TV/video producers, and instructional materials developers). high motivation, range of domain knowledge, low library system knowledge, high focus, and low to medium time allocations
2005	Soren	Online Visitor	based on knowledge of the participants

Table B.1 continued from previous page

Year	Author	User Group Label	Characteristics
2007	Marty	Online Visitor	Online museum visitors have specific and different needs and expectations of museum websites before and after museum visits. Prior to the visit, they require information such as hours of operation or driving instructions, while after the visit they are more likely to want information about future exhibits and special events. On the other hand, visitors are more likely to use online tours of galleries before visits than after visits, and more likely to use online images of artefacts after visits than before. Many online visitors prior to a visit have already decided to visit the museum. You can not assume that visitors using websites to plan their first visit will need the same information as visitors planning their second visit, or that they will be interested in the same exhibits, collections and so on.
2008	Warwick	Other - Retired Academics / Support Personnel	No definition in paper
2007	Nicholas	Phd-Student	No definition in paper
2008	Warwick	Post Doc Researcher	No definition in paper
2007	Nicholas	Postgraduate Student	No definition in paper
2001	Cunliffe	Potential Visitor	Looking to plan visit to museum
2004	Schweibenz	Potential Visitor	seeks information on the website about the museum, its collections and contact details.
2008	Blandford	Professional Information User	Uses of digital information resources in the context of their everyday life
2003	Marchionini	Professional Researcher,	e.g., picture researchers). high motivation, medium domain knowledge, average to high library system knowledge, very high focus, and medium time allocations,

Table B.1 continued from previous page

Year	Author	User Group Label	Characteristics
2012	Sweetnam	Professional Researcher	established academics, experienced in the general area covered by the resource, but not necessarily with the specific content of the resource
1993	Hsieh-Yee	Professional Searcher	Experienced searchers who have at least one year of experience and have taken courses on online searching or attended workshops provided by system vendors.
2008	Cifter	Professional User	a user who has a well defined goal for his/her activities, acting within the formal part of the economy - have good knowledge about the task that they perform with the product: they are trained and they may have previous experience with the product or the relevant task. - users have gone through extensive training to achieve particular knowledge which is valuable in a social or economical context.
2012	Gorny	Prospective Users	Users who already use a library (sometimes even quite frequently) and who clearly demonstrate an interest in its collections. Such users are limited on one hand by their capabilities and habits, and on the other hand do not always find what interests them in the library.
2009	Falk	Rechargers	Want to relax in a peaceful atmosphere
2011	Templeton (using Falk 2009)	Rechargers	Want to relax in a peaceful atmosphere.
2012	Fantoni (using Falk 2009)	Rechargers	motivated by the yearning to physically, emotionally, and intellectually recharge in a beautiful and refreshing environment.
2016	Falk & Dierking (2013) in Osman	Rechargers	seeking a restorative experience that allows them to reflect or rejuvenate
2016	Spellerberg (using Falk 2009)	Rechargers	No definition in paper

Table B.1 continued from previous page

Year	Author	User Group Label	Characteristics
2012	Gorny	Regular Users	A user who makes use of digital libraries systematically, with a defined frequency.
2003	Normore	Researcher (Independent Researcher, Professional Researcher)	the older and most educated of the studied groups. Uses secondary sources to point them to their most important information sources: letters, journals, diaries, historic newspapers and government and legal documents (e.g. deeds, wills and court records). Use digitised collections as filters which guides their research visit more productively.
2008	Normore	Researcher Group	people who do historical research - recruited from academic departments and historical societies in the local area. They held advanced degrees specializing in the history of the American South.
2008	Warwick	Researcher Post Graduate Student	No definition in paper
2016	Falk & Dierking (2013) in Osman	Respectful Pilgrims	Visitors who go to museums as a sense of duty to honour the memory of what represented as a memorial.
2003	Marchionini	Scholar	(e.g., historians, sociologists, anthropologists, authors). high motivation, high domain knowledge, high library system knowledge, high focus, and high time allocations
2008	Blandford	Scholar	how these users can be supported in broader work activities. Used as surrogate users - subject experts who can better assess features of a DL then the target user population, but who are not usability experts.
2011	Bertacchini	Scholar	No definition in paper
2012	Hampson	Scholar	Have access and are able to use unnormalised text of collections. Benefit from user friendly interfaces to enable easy access to relevant digital content.

Table B.1 continued from previous page

Year	Author	User Group Label	Characteristics
2012	Sweetnam	School Children	Challenging group to address that offers rewarding gains - Teacher access to primary sources has the potential to educate students about the specific historical context of artefacts
2001	Cunliffe	School Enquirer	Looking for information on projects and also to plan a visit.
2007	Nicholas	Senior Staff	No definition in paper
2014	Skov	Special Interest Museum Visitor	pursue a long standing interest or hobby.
1998	Booth	Specialist Visitor	Requires information on opening hours, prices, the museums facilities, what's on, notable exhibits and navigation aids, information concerning the Museums collections and access to its expertise, together with links to other sources of information
2008	Skov & Ingwersen	Specialist Visitor	most likely to engage in exploration and searching of online collection databases.
2007	LI	Spectators	Passive user.
2000	Paterno	Student (art student)	Students have a better knowledge of the application domain, so they want to access a wider range of topics, receiving more detailed information.
2003	Marchionini	Student (K-16)	low to medium motivation, low domain knowledge, low library system knowledge, low to medium focus, and low to medium time allocations.
2009	Srinivasan	Student	representative of the types of expert communities - interested in museum objects
2012	Duff	Student	No definition in paper
2008	Normore	Student Group	people who were enrolled in a graduate program for school media specialists. Not done extensive research for local or regional history sources.

Table B.1 continued from previous page

Year	Author	User Group Label	Characteristics
2003	Marchionini	Surfers	(e.g., those who are curious, those who bump into the NDL, etc.). low motivation, low domain knowledge, low library system knowledge (but may be high computing system knowledge), low focus, and very low time allocations
2008	Warwick	Taught Post-Doc Student	No definition in paper
2003	Marchionini	Teacher (K-16)	medium motivation, medium to high domain knowledge, low to medium library system knowledge, medium focus, and low time allocations
2011	Goodale	Teacher	uses the museum's material to create lessons/lectures or other learning resources. They design use of the website for homework assignments or to do project work for students. Could also use the material to support research publications.
2001	Cunliffe	Technical Enquirer	seeking answers to detailed enquiries.
2011	(as in Dallas Museum of Art (DMA) Templeton	Tentative Observers	Little experience with art
2000	Paterno	Tourist	Tourists are characterised by the need for basic general information, expressed and presented clearly. They like to access information by spacial representations (e.g. a museum or city map) because this gives implicitly information to them concerning how to organise a physical visit.
2003	Bolchini	Tourist	seek out locations, visiting hours, costs and start and end dates of exhibitions.
2015	Pantano(2011) in Ibrahim	Tourist	Tourists are those who visit the site to help organize their personal tours
2007	Nicholas	Undergraduate Student	No definition in paper
2007	Bastanlar	User	No definition in paper
2012	Ardissono	User	heterogeneous group of elderly members, parents and children.

Table B.1 continued from previous page

Year	Author	User Group Label	Characteristics
2002	Kravchyna	Virtual Museum Visitor	expect to get needed information about museum collections.they assume that the museum they visit online will offer digital image collections full of paintings, art works, historic illustrated manuscripts, drawings and museum objects (images of items that are not normally accessible.
2008	Skov & Ingwersen	Virtual Museum Visitor	average age is 46 years; 21%retired; 95% men; 11% novices with little domain knowledge. This indicates that mainly people with some (51%) or extensive (31% background knowledge use this specialized online collection.
2001	Cunliffe	Virtual Visitor	unable to visit in person.
2004	Schweibenz	Virtual Visitor	explores the museum's collections online.Depending on age, background and knowledge might seek different entry points to the website. Virtual visitors who might never be able to visit a certain museum in person.
2009	Fuentetaja	Virtual Visitor	browses the website over a concrete period of time.
1995	Schouten	Visitor	Visitors do not attend alone - visiting displays is a means to interact
2004	Goldman	Visitor	Linked to educational assignment or resource. Planning a physical visit or thought it maybe an interesting site to explore.
2011	Bertacchini	Visitor	No definition in paper
2009	Leporini	Visually Impaired User	No definition in paper
2011	Bertacchini	Web Users	No definition in paper
2012	Hampson	Wider -Public	No definition in paper

Appendix C

Lists of user groups

C.1 List of user groups after duplicates removed

Table C.1: List of user groups after duplicates removed

User Groups
Academic
Affinity Seekers
Amateur
Apprentice Investigator
Captive User
Casual User
Children (School Children)
Collectors
Commercial publisher
Committed enthusiast
Conceptualists
Creators
Critics
Curator
Curious participants
Disabled
Discerning independent
Educational visitor
Enjoyer
Experience seekers
Experienced User
Expert
Explorer
Facilitators
Families
Field dependent

Table C.1 continued from previous page

User Groups
Field independent
Genealogist
General public
General user
General visitor
Historian
Hobbyist
Hobbyist & Professionals
Humanities Scholars
Inactives
Incidental users
Independent Researchers
Inexperienced Searchers
Informationalist
Informed user
Intermediate
Joiners
Junior researcher
Lay users
Laymen
Library staff (Library of congress)
Mediators
MIP (Museum Information Professional)
Museum and Educational Experts
Museum curator
Museum Professional
Museum Professional (curators, educators, web developers)
Museum staff
Museum user
Non-captive User
Non-domain Users
Non-professional Researcher
Novice
Novice end user searchers
Novice User
Object seeker
Online visitor
Operationalist
PhD-student
Post doc researcher
Postgraduate student
Potential visitor
Professional Researcher
Professional searcher
Professional user
Prospective users

Table C.1 continued from previous page

User Groups
Rechargers
Regular users
Researcher (independent researcher, professional researcher)
Researcher post graduate student
Respectful pilgrims
Scholar
School Enquirer
Senior staff
Special interest museum visitor
Specialist visitor
Spectators
Student (art student)
Student(K-16)
Surfers
Taught post-doc student
Teacher(K-16)
Technical enquirer
Tentative observers
Tourist
Undergraduate student
Uni student
Virtual museum visitor
Virtual visitor
Visitor
Visually impaired user
Web users
Wider-public

C.2 List of user groups after similar grouped

Table C.2: Table of reduced user list based on grouping similars

Removed duplicates and similars	**Similars**	
Academic	Lecturer/Academic	
Academic support person in HE		
Academic/Commercial Publisher		
Affinity Seekers		
Amateur		
Apprentice Investigator		
Book Lover		
Captive User		
Casual User		
Children		
Collectors		
Committed Enthusiast		
Creators		
Critics		
Curator	Museum Curator	
Curious Participants		
Digital Archive User		
Disabled		
Discerning Independents		
Educational Visitor		
Enjoyer		
Experience Seekers		

Table C.2 continued from previous page

Removed duplicates and similars	Similars	
Experienced User		
Expert		
Explorer	Explorers	
Facilitators		
Field Independence/Dependence		
Genealogist		
General Public	General User	General Visitor
Historian		
Hobbyist & Hobbyist from Hobbyist & Professionals		
Imager/Verbalizer		
Inactives		
Incidental Users		
Independent Researchers		
Informationalist		
Informed User		
Intermediate		
Joiners		
Junior Researcher		
Lay Users	Laymen	
Library Staff		
Lifelong Learners Group		
Mediators		
MIP (Museum Information Professional)		
Museum and Educational Experts		
Museum Professional		
Museum Staff		

Table C.2 continued from previous page

Removed duplicates and similars	Similars	
Museum Visitor		
Non-captive User		
Non-domain Users		
Non-expert		
Non-professional Researcher		
Novice	Novice User	
Novice (End User Searchers)		
Object Seeker		
Other - Retired Academics / Support Personnel		
PhD-Student		
Post Doc Researcher		
Postgraduate Student	Researcher Post Graduate Student	Expert Communities
Potential Visitor		
Professional Information User		
Professional Researcher		
Professional Searcher		
Professional User	Professionals from Hobbyist & Professionals	
Prospective Users		
Rechargers		
Regular Users		
Researcher (Independent Researcher, Professional Researcher)	Researcher Group	
Respectful Pilgrims		
Scholar	Humanities Scholars	
School Children		
School Enquirer		
Senior Staff		

Table C.2 continued from previous page

Removed duplicates and similars	Similars	
Special Interest Museum Visitor		
Specialist Visitor		
Spectators		
Student	Student Group	
Surfers		
Taught Post-Doc Student		
Teacher		
Technical Enquirer		
Tentative Observers		
Tourist		
Undergraduate Student		
User		
Virtual Museum Visitor	Virtual Visitor	Online Visitor
Visitor		
Visually Impaired User		
Web Users		
Wider-public		

C.3 User group usage through time

Table C.3: User groups used and introduced over time

User Group Labels	1993	1995	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2011	2012	2013	2014	2015	2016	Grand Total
Academic																1					1
Academic support person in HE													1								1
Academic/Commercial publisher															1						1
Affinity seekers																				1	1
Amateur													1								1
Apprentice investigator																1					1
Book lover								1													1
Captive User																			1		1
Casual user																			1		1
Children																	1				1
Collectors												1					1				2
Committed enthusiast															1						1
Creators												1									1
Critics												1									1
Curator		1																			1

Curious participants															1						1
Digital archive user													1								1
Disabled																	1				1
Discerning Independents															1						1
Educational visitor			1										1								2
Enjoyer																			1		1
Experience seekers														1	1	1				2	5
Experienced User													1								1
Expert				1	1			1	1		1	1	1			1	1		1		10
Expert communities														1							1
Explorer														1		1				2	4
Explorers															1						1
Facilitators														1	1	1				2	5
Field Independence/Dependence													1								1
Genealogist													1								1
General public																1					1
General user				1															1		2
General visitor			1										1		1						3
Historian							1														1

Hobbyist								1													1
Hobbyist & Professionals												1		2	1	1				2	7
Humanities Scholars							1									1					2
Imager/Verbalizer													1								1
Inactives												1									1
Incidental users																1					1
Independent Researchers													1								1
Informationalist																			1		1
Informed user																1					1
Intermediate				1																	1
Joiners												1									1
Junior researcher												1									1
Lay users													1								1
Laymen									1												1
Lecturer/academic													1								1
Library staff								1													1
Lifelong learners group													1								1
Mediators																1					1
MIP (Museum Information Professional)									1		1										2

Museum and Educational Experts											1										1
Museum curator														1	1						2
Museum Professional												1		1							2
Museum staff													1	1							2
Museum Visitor													1							1	2
Non-captive User																			1		1
Non-domain users																1					1
Non-expert																	1				1
Non-professional researcher								1													1
Novice				1																	1
Novice (end user searchers)	1																				1
Novice User											1		1	1		1					4
Object seeker								1													1
Online visitor										1		1									2
Other - retired academics / support personnel													1								1

PhD-student												1									1
Post doc researcher													1								1
Postgraduate student												1									1
Potential visitor						1			1												2
Professional information user													1								1
Professional Researcher								1								1					2
Professional searcher	1																				1
Professional user													1								1
Prospective users																1					1
Rechargers														1	1	1				2	5
Regular users																1					1
Researcher (independent researcher, professional researcher)								1													1
Researcher group													1								1

Researcher post graduate student													1								1
Respectful pilgrims																				1	1
Scholar								1					1		1	1					4
School children																1					1
School Enquirer						1															1
Senior staff												1									1
Special interest museum visitor																		1			1
Specialist visitor			1										1								2
Spectators												1									1
Student					1			1						1		1					4
Student group													1								1
Surfers								1													1
Taught postdoc student													1								1
Teacher								1							1						2
Technical enquirer						1															1
Tentative observers															1						1
Tourist					1			1											1		3
Undergraduate student												1									1

User												1				1					2
Virtual museum visitor							1						1								2
Virtual visitor						1			1					1							3
Visitor		1							1						1						3
Visually impaired user														1							1
Web users															1						1
Wider -public																1					1
Grand Total	**2**	**2**	**3**	**4**	**3**	**4**	**3**	**13**	**6**	**1**	**4**	**16**	**28**	**14**	**16**	**23**	**5**	**1**	**8**	**13**	**169**

Appendix D

User groups, after being grouped by similar names and definitions

Table D.1: User groups, after being grouped by similar names and definitions

User Groups	**Sub User Groups**	**Sub Sub User Groups**	**Groups with Similar Definitions**
Academic			Academic, Humanities Scholars, Mediators, Scholar
Novice			Novice, Amateur, Digital Archive User, Novice (End User Searchers), Non-expert, Novice User
Expert			Expert, Expert Communities, Informationalist, Museum and Educational Experts, Professional Information User, Professional Searcher, Professional User
	Professional Researcher		Professional Researcher, Researcher (Independent Researcher, Professional Researcher), Researcher Group
		Museum Staff	Museum Staff, Museum Professional, Senior Staff
		Library Staff	Library Staff, Senior Staff
		Curator	Curator, Creators, Museum Curator, Senior Staff
		MIPS	MIP (Museum Information Professional), Senior Staff
	Post Doc Researcher		Post Doc Researcher, Researcher Post Graduate Student
	Historian		Historian
Commercial Publisher			Academic/Commercial Publisher

Table D.1 continued from previous page

User Groups	Sub User Groups	Sub Sub User Groups	Groups with Similar Definitions
General Public			General Public, General User, General Visitor, Experienced User, Laymen, Incidental Users, Potential Visitor, Spectators, Surfers, Tentative Observers, Wider-public
	Affinity Seekers		Affinity Seekers
	Experience Seeker		Experience Seekers, Curious Participants, Enjoyer
	Online Visitor		Online Visitor, Museum Visitor
	Rechargers		Rechargers
	Tourist		Tourist
Student			Student, Informationalist
	School Children		School Children, Captive User
	College Students		College Students, Captive User
	Uni Student		Junior Researcher
		Undergraduate Student	Undergraduate Student, Apprentice Investigator, Captive User, Imager/Verbalizer
		Taught-Post-Graduate Students	Taught Post-Doc Student, Apprentice Investigator, Captive User, Casual User, Field Independence/Dependence, Postgraduate Student
		PhD-Student	PhD-Student
Teacher			Teacher, Educational Visitor, School Enquirer
Hobbyist			Hobbyist, Committed Enthusiast, Discerning Independents, Lay Users, Hobbyist & Professionals, Explorer, Genealogist, Informationalist, Informed User, Lifelong Learners Group, Non-Professional Researcher, Object Seeker, Special Interest Museum Visitor, Specialist Visitor
	Collectors		Collectors
Facilitators			Facilitators
Respectful Pilgrims			Respectful Pilgrims
Virtual Visitor			Virtual Visitor, Virtual Museum Visitor, Web Users

Appendix E

Uses of user group labels

Table E.1: Uses of User group labels

Labels	**Count of Uses**
Academic	1
Academic support person in HE	1
Academic/Commercial publisher	1
Affinity seekers	1
Amateur	1
Apprentice investigator	1
Book lover	1
Captive User	1
Casual user	1
Children	1
Collectors	2
Committed enthusiast	1
Creators	1
Critics	1
Curator	1
Curious participants	1
Digital archive user	1
Disabled	1
Discerning Independents	1
Educational visitor	2
Enjoyer	1
Experience seekers	5
Experienced User	1
Expert	10
Expert communities	1
Explorer	4
Explorers	1
Facilitators	5
Field Independence/Dependence	1

Table E.1 continued from previous page

Labels	Count of Uses
Genealogist	1
General public	1
General user	2
General visitor	3
Historian	1
Hobbyist	1
Hobbyist & Professionals	7
Humanities Scholars	2
Imager/Verbalizer	1
Inactives	1
Incidental users	1
Independent Researchers	1
Informationalist	1
Informed user	1
Intermediate	1
Joiners	1
Junior resercher	1
Lay users	1
Laymen	1
Lecturer/academic	1
Library staff	1
Lifelong learners group	1
Mediators	1
MIP (Museum Information Professional)	2
Museum and Educational Experts	1
Museum curator	2
Museum Professional	2
Museum staff	2
Museum Visitor	2
Non-captive User	1
Non-domain users	1
Non-expert	1
Non-professional researcher	1
Novice	1
Novice (end user searchers)	1
Novice User	4
Object seeker	1
Online visitor	2
Other - retired academics / support personnel	1
PhD-student	1
Post doc researcher	1
Postgraduate student	1
Potential visitor	2
Professional information user	1

Table E.1 continued from previous page

Labels	Count of Uses
Professional Researcher	2
Professional searcher	1
Professional user	1
Prospective users	1
Rechargers	5
Regular users	1
Researcher (independent researcher, professional researcher)	1
Researcher group	1
Researcher post graduate student	1
Respectful pilgrims	1
Scholar	4
School children	1
School Enquirer	1
Senior staff	1
Special interest museum visitor	1
Specialist visitor	2
Spectators	1
Student	4
Student group	1
Surfers	1
Taught post-doc student	1
Teacher	2
Technical enquirer	1
Tentative observers	1
Tourist	3
Undergraduate student	1
User	2
Virtual museum visitor	2
Virtual visitor	3
Visitor	3
Visually impaired user	1
Web users	1
Wider -public	1

Appendix F

All user group definitions concatenated

Table F.1: All definitions concatenated to form one definition for the refined list of user groups

User group name	**Definition**
User group name	**Definition**
Academic	Higher education use of educational portals (e.g., historians, sociologists, anthropologists, authors). High motivation, high domain knowledge, high library system knowledge, high focus, and high time allocations. Have access and are able to use unnormalised text of collections. Benefit from user friendly interfaces to enable easy access to relevant digital content. Willing to use special libraries and collections but need assistance when using the systems. Often turn to browsing to identify relevant sources. Has educational, e-learning or informatics and communications backgrounds or experiences. Seeking answers to detailed enquiries.
Affinity seekers,	Visitors who are motivated to visit a museum because it speaks to their sense of heritage and their bigger identity.
Novice	Does not possess much search experience. Subject knowledge is their only asset when they start doing their own searches. Are new to the task or the product, and usually they do not have enough information to perform the task with the device. Can be hindered by lacking knowledge of rich and complex knowledge structures of museum collections. Does not possess knowledge or experience to locate archival sources. Will often become frustrated early in the interaction process. May become confused and give up using them [applications] despite the rich possibilities they offer. Do not find systems easy to use because they are not familiar with the terms used for searches. Prefer content to be in language they would use and not the language of experts. Might be as advanced as professionals in terms of both use and innovativeness. They can perform basic online searches but do not utilise sophisticated system features and great variability in search performance exists. Need an improved user experience (e.g., some authors, CD-ROM/multimedia developers, TV/video producers, and instructional materials developers). high motivation, range of domain knowledge, low library system knowledge, high focus, and low to medium time allocations
Expert	Might be interested in a list of references to published articles that discusses the selected exhibits. Museum knowledge structure developed by experts such as art theoreticians and art historians. Experts are specialists in the field of cultural heritage. Experts want to have full access to all the information available. They need minimal support to formulate their requests, and should be allowed to formulate such requests in a flexible way. Have a meaningful relationship to the collections. Expert users already have a deep knowledge of the content of the digital collections, allowing their research to create new insights into the artefacts. Aware of the relevant search parameters to access the digital content and are motivated to interact with the system because of their research interest. Mediators are scholar who act as go between of general public and the system. An expert user of text. Informationalists are those who have the intent of improving their knowledge. Uses system features to modify a search result without changing the concepts. A user who makes use of digital libraries systematically, with a defined frequency. Seeking answers to detailed enquiries.

Table F.1 continued from previous page

Professional researcher	want to see specific pieces or exhibits.Motivated by the desire to further specific intellectual needs in a setting with a specific subject matter focus. A need to persevere. Leisure career that proceeds in stages: beginning; development; establishment; maintenance; and decline. Proactive knowledge. experience and skill acquisition. Durable benefits of personal and social rewards. Unique ethos or culture.Strong identification with the chosen pursuit launch themselves on a (leisure) career centred on acquiring and expressing a combination of its special skills; knowledge and experience, Seek meaning and connections between documents in the archives
Museum staff,	elite experts of the collections held.Museum professionals need to understand what museum visitors are looking for before and after a museum visit, and make this information available to them. Create and upload assets online such as web pages and videos. Has educational, e-learning or informatics and communications backgrounds o experiences.
Library staff	high motivation, medium domain knowledge, high library system knowledge, high focus, and limited time allocations
Curators	Consider words and letters the only medium to transfer and idea
MIPS	Specialises in the application of information science and technology in museums and whose interests lie in managing unique information resources found in museums
Commercial publisher	

Table F.1 continued from previous page

General public	View exhibitions as places where they can learn something, but they rarely do. e.g. window shopping adults and children - a large diverse set of users that bring a very wide range of interests, technical abilities, contextual and/or historical awareness to collections. Mostly females that are college educated and work in museums. General users are those who usually visit the website more for some passing curiosity rather than to retrieve information to improve their knowledge of cultural heritage. Requires information on opening hours, prices, the museum's facilities, what's on, notable exhibits and navigation aids. Requires information on opening hours; prices and facilities etc. Lay users have limited or no training in a particular area, however they are likely to have personal interests or special needs in that area. Unable to visit in person. Browses the website over a concrete period of time. explores the museum's collections online.Depending on age, background and knowledge might seek different entry points to the website. Virtual visitors who might never be able to visit a certain museum in person.Enjoyers are those who appreciate the virtual exploration of the cultural heritage for personal pleasure. No definition in paper. No definition in paper. Linked to educational assignment or resource. Planning a physical visit or thought it may be an interesting site to explore.no definition in paper, expect to get needed information about museum collections. They assume that the museum they visit online will offer digital image collections full of paintings, art works, historic illustrated manuscripts, drawings and museum objects (images of items that are not normally accessible. Average age is 46 years; 21%retired; 95% men; 11% novices with little domain knowledge.Users who visit a digital library on an occasional basis, and do not plan to make systematic use of its resources. Online museum visitors have specific and different needs and expectations of museum websites before and after museum visits. Prior to the visit, they require information such as hours of operation or driving instructions, while after the visit they are more likely to want information about future exhibits and special events. On the other hand, visitors are more likely to use online tours of galleries before visits than after visits, and more likely to use online images of artefacts after visits than before. Many online visitors prior to a visit have already decided to visit the museum. You can not assume that visitors using websites to plan their first visit will need the same information as visitors planning their second visit, or that they will be interested in the same exhibits, collections and so on.Looking to plan a visit to the museum seeks information on the website about the museum, its collections and contact details (e.g., those who are curious, those who bump into the NDL, etc.). low motivation, low domain knowledge, low library system knowledge (but may be high computing system knowledge), low focus, and very low time allocations. Little experience with art

Table F.1 continued from previous page

Experience seeker	Want to see the most renowned pieces and make memories. Aspires to be exposed to the things and ideas that exemplify what is best and intellectually most important with in a culture or community.See museum as an important destination - they are collecting experiences. No definition in paper. Tourists are characterised by the need for basic general information, expressed and presented clearly. They like to access information by spacial representations (e.g. a museum of city map) because this gives implicitly information to them concerning how to organise a physical visit. Seek out locations, visiting hours, costs and start and end dates of exhibitions. Tourists are those who visit the site to help organize their personal tours
Online visitor	Online museum visitors have specific and different needs and expectations of museum websites before and after museum visits. Prior to the visit, they require information such as hours of operation or driving instructions, while after the visit they are more likely to want information about future exhibits and special events. On the other hand, visitors are more likely to use online tours of galleries before visits than after visits, and more likely to use online images of artefacts after visits than before. Many online visitors prior to a visit have already decided to visit the museum. You can not assume that visitors using websites to plan their first visit will need the same information as visitors planning their second visit, or that they will be interested in the same exhibits, collections and so on.
Rechargers	Want to relax in a peaceful atmosphere seeking a restorative experience that allows them to reflect or rejuvenate, motivated by the yearning to physically, emotionally, and intellectually recharge in a beautiful and refreshing environment. No definition in paper
Student	
School children	seek educational resources. Low to medium motivation, low domain knowledge, low library system knowledge, low to medium focus, and low to medium time allocations.Those who learn because they are required to
College Students,	Those who learn because they are required to
Uni Students	Students have a better knowledge of the application domain, so they want to access a wider range of topics, receiving more detailed information.Those who learn because they are required to. No definition in paper
Undergraduate students,	no definition in paper. Students at an advanced undergraduate and post graduate level. Some knowledge of historical period and/or cultural context addressed by the resource.Those who learn because they are required to. No definition in paper
Taught-Post-graduate students	Students at an advanced undergraduate and post graduate level. Some knowledge of historical period and/or cultural context addressed by the resource.no definition in paper
PostDoc researcher,	no definition in paper. No definition in paper. No definition in paper
PhD-student	

Table F.1 continued from previous page

Educator	Requires information on opening hours, prices, the museum's facilities, what's on, notable exhibits and navigation aids and more detailed information to help plan visits. . . and project based information. Medium motivation, medium to high domain knowledge, low to medium library system knowledge, medium focus, and low time allocations willing to use special libraries and collections but need assistance when using the systems. Often turn to browsing to identify relevant sources.Looking for information on projects and also to plan a visit. Seeking answers to detailed enquiries.
Hobbyist	(e.g., genealogy, Civil War, railroads, other examples). high motivation, typically high domain knowledge, a range of library system knowledge, high focus, and high time allocations. Want to see specific pieces or exhibits.Motivated by the desire to further specific intellectual needs in a setting with a specific subject matter focus, A need to persevere. Leisure career that proceeds in stages: beginning; development; establishment; maintenance; and decline. Proactive knowledge. experience and skill acquisition. Durable benefits of personal and social rewards.Unique ethos or culture. Strong identification with the chosen pursuit launch themselves on a (leisure) career centred on acquiring and expressing a combination of its special skills; knowledge and experience. No definition in paper. Knowledgeable and connected to art. Enjoy learning from experts.Researchers who are not professional academics but have knowledge of some aspects addressed by the resource. Non professional researchers, centres of interest were specific and well defined. Happy to utilise catalogues/indexes. Pursue a long standing interest or hobby.Requires information on opening hours, prices, the museum's facilities, what's on, notable exhibits and navigation aids, information concerning the Museum's collections and access to its expertise, together with links to other sources of information most likely to engage in exploration and searching of online collection databases.Want to learn new information and understand new concepts.Motivated by a need to satisfy personal curiosity and interest in an intellectually challenging environment visitors who are driven by curiosity, they value learning, but they are not necessarily experts. Motivated by a need to satisfy personal curiosity and interest in an intellectually challenging environment.no definition in paper. Those who voluntarily seek cultural information due to intrinsic motivation, such as to satisfy personal curiosity; but will not spend a lot of time dealing with uninteresting information. Not domain or system experts but could have some background of heritage but not to level of expert. Enjoy art and museums. Knowledgeable and connected to art. Enjoy developing their own views on art. Informationalists are those who have the intent of improving their knowledge. Lay users have limited or no training in a particular area, however they are likely to have personal interests or special needs in that area.User who have the option to ignore the information without baring any punishment (e.g., some authors, CD-ROM/multimedia developers, TV/video producers, and instructional materials developers). high motivation, range of domain knowledge, low library system knowledge, high focus, and low to medium time allocations seeking answers to detailed enquiries. May have some previous experience with the product or the relevant task, but their knowledge of the task is much more limited when compared with a professional user.Seek meaning and connections between documents in the archives.

Table F.1 continued from previous page

Collectors	Focussed on building: developing and maintaining private collections.Place primary emphasis on factual object related information.
Professional	a user who has a well defined goal for his/her activities, acting within the formal part of the economy - have good knowledge about the task that they perform with the product: they are trained, and they may have previous experience with the product or the relevant task. - users have gone through extensive training to achieve particular knowledge which is valuable in a social or economical context. A user who makes use of digital libraries systematically, with a defined frequency.
Researchers	no definition in paper. No definition in paper. e.g., picture researchers). High motivation, medium domain knowledge, average to high library system knowledge, very high focus, and medium time allocations, established academics, experienced in the general area covered by the resource, but not necessarily with the specific content of the resource. Experienced searchers who have at least one year of experience and have taken courses on online searching or attended workshops provided by system vendors. the older and most educated of the studied groups. Uses secondary sources to point them to their most important information sources: letters, journals, diaries, historic newspapers and government and legal documents (e.g. deeds, wills and court records). Use digitised collections as filters which guides their research visit more productively. No definition in paper. No definition in paper. No definition in paper. Informationalists are those who have the intent of improving their knowledge. Researchers who are not professional academics but have knowledge of some aspects addressed by the resource
Facilitators	Want to ensure that their companions meet their visit goal, motivated by the wish to engage in a meaningful social experience with someone who they care about in an educationally supportive environment. No definition in paper. Visitors who are socially motivated, they are focussed on enabling the learning and experience of others they care about.
Respectful pilgrims,	Visitors who go to museums as a sense of duty to honour the memory of what represented as a memorial.

F.1 Refined and coherent group definitions

The list below contains a set of user group labels and definitions that are more coherent than the table above.

- **Academic -** Academics are often used as surrogate users, but are usually subject experts who are highly educated and motivated by work/research tasks to use DCH websites and collections. Highly focused on completing tasks/research and allocate significant time to the tasks. Possess a high level of domain and library/collection system knowledge. Some assistance may be required using new interfaces, but will use browsing for relevant content over search (sourced from 8 publications).

- **Novice -** Novices are frequently defined as non-professionals or early career professionals who have limited domain knowledge or experience with search systems. They lack motivation and possess no formal training, resulting in quickly giving up when using search systems. Prefer content language in simplistic terms rather than that of an expert, which can also hinder the search process, as systems tend to be built by experts. Although, a Novice can be highly advanced in using generic systems and also highly innovative when achieving goals (sourced from 9 publications).

- **Expert -** Experts are specialists in their field and are motivated by improving their knowledge or achieving a task quickly, while expecting access to all the information that is available. They have extensive training and familiarity with search systems (minimum one year), often combining multiple systems. Possess high domain knowledge, requiring minimal support in formulating search requests, and can understand the complex knowledge structures of collections that allows them to create new insights into artefacts. Experts have been identified as researchers, curators, registrars, art theoreticians, historians, teachers, IT personal and some advanced students; all of which are linked to a formal career (sourced from 15 publications).

- **Professional Researcher -** The older and most educated of the studied groups, this user group label is often intertwined with academics and histor-

ical society personnel. They are highly motivated and focused on furthering personal knowledge in a specific historic field and/or to advance their research, although domain knowledge ranges from medium to high levels with medium levels of time allocations. Confident using digitised collections; often having a meaningful relationship with the collections; and search systems and frequently use secondary sources to find important sources: letters, journals, diaries, historical newspapers, and government legal documents (e.g. deeds, wills and court records) (sourced from 5 publications).

- **Museum Staff -** Museum staff need to understand what museum visitors are looking for before and after a museum visit and make this information available to them. Museum staff plan, implement and maintain museum websites, often developing them in a haphazard fashion. The digitisation of cultural heritage collections involves opening databases, containing documentation consisting of profession jargon; traditionally written for and by museum professionals (sourced from 5 publications).

- **Library Staff -** Library staff are highly motivated with medium domain knowledge but very high library system knowledge. They have a high focus level but limited time allocations (sourced from 2 publications).

- **Curator -** Museum curators have a dual interaction between museum objects and the associated historic context. Their work tasks include collection management, exhibition planning and design, research, and answering of inquiries to the museum from the public. They create exhibitions both physical and online; although preferring the written word, they produce content through the creation of assets such as videos and web-pages (sourced from 5 publications).

- **MIPS -** Museum Information Professionals (MIPs) specialise in the application of information science and technology in museums. They understand the needs of the users and produce and create digital content that both museum staff and museum users need. They manage the information resources found in museums (sourced from 3 publications).

- **Post-Doc Researcher -** No characteristics, traits or information in general

relating to the user label could be extracted from the publication (sourced from 2 publications).

- **Historian -** A historian is an expert user of text (sourced from 1 publication).
- **Academic Commercial Publisher -** No characteristics, traits or information in general relating to the user label could be extracted from the publication (sourced from 1 publication).
- **General Public -** General users consist of a large, diverse set of users that bring a very wide range of interests, technical abilities, contextual and/or historical awareness to collections/museums. They are those who usually visit the website more for some passing curiosity rather than to retrieve information to improve their knowledge of cultural heritage and are rather looking to plan a visit to a museum and require information on opening hours; prices and facilities etc adults and children. They can be considered a Passive user. (e.g., those who are curious, those who bump into the NDL, etc.) and possess low levels of motivation, domain knowledge, library system knowledge (but may be high computing system knowledge), focus, and with limited (low) time allocation levels (sourced from 12 publication).
- **Affinity Seeker -** Affinity seekers are motivated to visit a museum because it speaks to their sense of heritage and bigger identity (sourced from 1 publication).
- **Experience Seeker -** Experience seekers want to virtually explore the most renowned pieces and make memories. They enjoy art galleries and museums and aspire to be exposed to the things and ideas that exemplify what is best and intellectually most important within a culture or community. They see museums as important destinations and are collecting experiences (sourced from 7 publications).
- **Virtual Museum Visitor -** Virtual museum visitors (VMV) have specific and different needs and expectations of museum websites before and after museum visits. They visit museums approximately four times a year, and the museum website approximately once a week, and consider it very important

for museums to have a website; they see a museum website as complementary, and not likely to replace the physical museum presence as users search for and access information. Prior to the visit, the VMV requires information such as hours of operation or driving instructions, while after the visit they are more likely to want information about future exhibits and special events. They are more likely to use online tours of galleries before visits than after visits, and more likely to use online images of artefacts after visits than before, as they are likely to use a wide variety of digital museum resources in their daily lives e.g., online images of artefacts and online research materials (sourced from 9 publications).

- **Recharger -** Rechargers are motivated by the yearning to physically, emotionally and intellectually recharge in a peaceful, refreshing environment. They seek to have an experience that allows them to reflect or rejuvenate (sourced from 5 publications).

- **Tourist -** Tourists require basic general information such as location, visiting hours, pricing, exhibition details that will help them on a physical visit. They require the information presented clearly; ideally via spacial representations such as maps because this provides implicit information that they can use on a physical visit (sourced from 3 publications).

- **Student -** Students are those with an intent of improving their knowledge and were enrolled in an educational program - at graduate level, they could be considered representative of the types of expert communities. They have a better knowledge of the application domain, so require access a wider range of topics, receiving more detailed information. They can possess low to medium levels of focus, motivation and time allocations, combined with low domain knowledge and library system knowledge (sourced from 6 publications).

- **School Children -** School Children seek information, usually around a task set by a teacher (teachers tend to seek resources relative to specific historical artefacts in order to engage their classes), and use the system as part of their required learning (sourced from 3 publications).

- **College Student -** Mainly college educated females who learn because they are required to (sourced from 2 publications).

- **Uni Student -** No clear characteristics, traits or behaviours evident in the publication. The only comment was in relation to conducting abstract only searches (sourced from 1 publication).

- **Undergraduate Student -** Undergraduate students have some level of historical or cultural knowledge. They learn because they have to in order to pass the course (sourced from 4 publications).

- **Post-graduate Student -** Students at an advanced educational level, while not domain or system experts, possess some knowledge of historical period and/or cultural context. They tend to voluntarily seek cultural information due to intrinsic motivation, such as to satisfy personal curiosity; but will not spend a lot of time dealing with uninteresting information. They do not need external referents to aid the processing of information (sourced from 6 publications).

- **PhD Student -** PhD Students could be considered representative of the types of expert communities. They have a better knowledge of the application domain, so require access a wider range of topics, receiving more detailed information. (sourced from 1 publication)

- **Teacher -** Teachers are generally looking for information on opening hours, prices, the museum's facilities, what's on, notable exhibits and navigation aids and more detailed information to help plan visits. Teachers use the museum's material to create lessons/lectures or other learning resources. They use the website for homework assignments or to set project work for students. Occasionally they could also use the material to support research publications. As teachers, they possess medium levels of focus and motivation and low level time allocations, with medium to high levels of domain knowledge combined with low to medium levels of library system knowledge (sourced from 5 publications).

- **Hobbyist -** Hobbyists are motivated and driven by curiosity and interest in an

intellectually challenging environment with a specific subject. (e.g., genealogy, Civil War, railroads, other examples). Classified according to five categories: collectors, makers and tinkerers, activity participants, players of sports and games and enthusiasts of the liberal arts. They are highly focused and possess high levels of motivation and domain knowledge. Familiarity with a range of library systems ranges from low to high levels, with time allocation levels varying from low to high. Strong identification with the chosen pursuit and seek meaning and connections between documents in the archives and acquiring and expressing a combination of its special skills, knowledge and experience. They are not professional academics or enrolled in educational programmes, but have a personal interest and/or knowledge of some aspects addressed by the resource. Enjoy learning from experts and enjoy developing their own views. Want to learn new information and understand new concepts. Want to see and study specific pieces or exhibits (sourced from 20 publications).

- **Collectors -** Focused on building, developing and maintaining private collections. They place primary emphasis on factual object related information, thereby providing a broad understanding of the item's historical and contemporary production and use (sourced from 2 publications).

- **Facilitators -** Wish to engage in a meaningful social experience with someone who they care about in an educationally supportive environment. They are socially motivated and focused on enabling the learning and experience of others; ensuring that their companions meet their visit goals (sourced from 5 publications).

- **Respectful Pilgrims -** Visitors who go to museums as a sense of duty to honour the memory of what represented as a memorial (sourced from 1 publication).

Appendix G

Pilot survey responses

G.1 Pilot participant 1:

Well, I haven't got very far because I clicked the do not wish to participate box, submitted my response, and it won't allow me to go back in again. Here is the small amount of feedback I have from seeing two screens

Pop-up box - put in headings then reader can skip read - put START SURVEY in white box or something else to make it more visible. I didn't even see the START SURVEY BUTTON to start off with. I came to the end of the white box and thought that was it.

Consent - do not wish to participate, link goes to Survey Completed. You'll need different text here

G.1.1 second attempt

Comments on the questions. q10 why are you collecting student status when you already know this from previous Q. Could reduce this to just Student. q10. What is expert? q11 perhaps you've deliberately unordered this but I would put in order of knowledge q15 is task right word here. Think it is an information science term rather than a term a website visitor would use q18 are they visiting today? Could just be visiting the website Q22 – complicated wording Q23 do people get to the questionnaire from a web page. Will you know what page they have come from. Q24 what is navigation? Will people understand this? Do you mean menus?

G.2 Pilot participant 2:

Some comments/suggestions: **Consent form** I know you can't reduce the amount of info in the consent form. Could you reduce it a bit and link to a page with it on? I don't mind adding it to a web page on the NML website if that is acceptable. Can you put (NML) in brackets after the first use of National Museums Liverpool, just to help make it clear that NML is National Museums Liverpool?

The questions about education level - I'd be tempted to put these a bit later on in the questions, because I think it could put someone off if they are being asked this so quickly or make them feel like they aren't 'educated' enough to do the survey (might just be me being a bit overly sensitive tho!) Q10: Change NML to National Museums Liverpool Q22: On previous visits, how easy do you find content you are satisfied with when searching? I found this a bit awkwardly worded and had to read it a few times. People might not think you mean using the search, they might just think it means 'looking for' something. Could maybe change to: If you've used the 'search' box previously how useful did you find it? If you've used the 'searchb' previously, did you find what you were looking for? Or perhaps just asking this question after question 24 would make it a bit more obvious that it's referring specifically to search. Q23: I'm happy for you to drop question 23 to keep the survey a bit shorter.

G.3 Pilot participant 3:

Here are some comments: - What is the reasoning behind the age groups? - What is the reasoning behind the levels in Q11. Should those be formatted in some order? - Q18 Capitalisation inconsistencies

G.4 Pilot participant 4:

I started off writing a few quick notes but have ended up writing a lot! I hope you don't mind the detail. I did a survey as part of my PhD research and wish I'd done a lot of things differently, so I'm always very keen to offer feedback on other people's. You should feel free to disregard any or all of it! Anyway, here we go...

One tip on the information page. Maybe it's possible to include a short paragraph in bold at the top which briefly summarises the key information (anonymous, used for research, stop at any time). I'm not suggesting that people shouldn't read the rest of it, but...! Q6 - I'm at work AND at University. I'd assume I should select work, but you may wish to clarify ("at your place of study?"). Q10 - "visit to the NML" implies a physical visit (rather than visiting the website). Maybe "..for this visit to the NML website"? Q11 - "What is your knowledge level in relation to the reason for your current visit" I think the wording of this question is a little unclear (although reading the answer options I'm able to work out what you are asking). Q13 - I'm not sure what "this type of activity" refers to. Do you mean looking at the NML website, or undertaking the broader information seeking task that has led the user to the site? I think you could use the wording for Q14 ("...the type of activity you are using the National Museums Liverpool website for today") here. Q14 - Is midday really required? It seems incredibly specific. Maybe morning/lunchtime/afternoon would make more sense? Q15 - I think people might find this hard to answer with the level of detail you want. Also note that with your current options, the minimum choice is one day a week. What if I do this task much less frequently than that? It seems to me like the kind of question better answered with options like every day, a few times per week, a few times per month etc (as you offer for Q12). Q17 - Maybe it would be useful to know if someone is just starting, or if they have just finished? These seem like quite distinct states. But I guess you must have a reason for only giving these two options (i.e. not separating option two). Q18 - I feel like this question could be much earlier in the survey. Questions 11-16 all refer to the purpose of the visit,so I think this question (Q18) should come before them. Respondents will find Q11-16 easier to answer if you have already got them to understand what you mean by "activity". It will also help resolve the issues I mentioned with Q11and 13. Also are you certain that these options are mutually exclusive? If I am planning a visit to the museum then I might want to know what collections I can see - so I'd be unsure which option to select. Also I may have two distinct reasons for visiting (as an academic I might be looking for information on a specific object, and also seeing what other online collections there are). I suppose there are two solutions - either ask "what is your primary reason for visiting", or allow multiple answers ("select all that apply"). I think the second option is probably better. Oh and very

minor point - the first answer option starts with a capital letter, but the rest don't. You should probably capitalise them all to be consistent with the rest of the survey. Q19 - doesn't exist..I jump straight from 18 to 20 (and from page 7 to page 9). Is this based on an answer I've given,or just a numbering error? Q20 - URI may be technically correct, but I guarantee most people will think it's a typo! Q21 - To be pedantic, "Give up easily" should probably be "Give up immediately", since if you try more than once you will select option 2. In general this is always a tricky thing to ask people, and I wonder whether it's possible to make such a broad statement. Surely for most people the answer depends on what the task is? There are times when I will give up immediately (some meaningless bit of trivia) and times I will persist as long as necessary (the phone number of the hospital my wife has just been rushed to) . I think you should try and clarify exactly what you are asking people. Are you referring explicitly to the NML website? If so you should say that. Are you referring to the activity the person is currently engaged in? Again you should make that clear. Personally I think that without specifying a context the question is basically impossible to answer, and there results will be pretty meaningless. If you do want to keep the question general then maybe you could give time options (i.e. I try for 5 minutes 10 minutes, half an hour etc)? Q22 - The question should read "how easily" rather than "how easy", and the tense should be changed. So: "...how easily did you find content you were satisfied with..." To make clear the distinction between this Q and Q24, I would say "..when using the search box" rather than "when searching". Some people will equate searching with the general process of looking for information (i.e. using both search box and menus), so it will help to highlight that you are only asking about the search function. What if this is my first visit? I think you should include that as an option. "After a couple of queries on other pages but the first of results" - I'm afraid I can't work out what this answer option means! I think you should clarify what you mean by "standard navigation" - not everyone will understand the term. There is a difficulty with this question if on my previous visit I looked for more than one piece of information, and found one very easily but struggled with the other. What option should I choose then? Q23 I would find it very difficult to answer this question. The key question is whether you mean information that is useful for me, or just information in general. For example, if the last page I viewed was a set of search results, it might be that there was loads

of information (snippets, thumbnails etc), but none of it was relevant to what I was looking for.In this case should I answer “not enough information” or “just the right amount of information”? Also is the pop-up programmed to appear on certain pages, or can it appear at any time during a session? If the latter then I wonder whether you will get useful results from this question, since there will be huge variation in the pages viewed prior to the survey appearing. I suppose you may be able to track the page that the user was looking at, and then link the answers to that page type, which may be of use. Q24 - this seems to compliment Q22, so I would swap this Q with Q23. It should also read “easily” rather than “easy”, and you change to the past tense (“how easily did you find content you were satisfied with”, There is also the same issue for first time users of the site, and for people who did multiple tasks (see comments for Q22) I assume an aim here is to compare the ease of searching, with that of using navigation features. To that end I think you might find it helpful to construct questions in such a way that the data can be directly compared. For example you could use a 5 or 7 point scale for both questions, and ask “how easy was it to find information using the search box/navigation features, with a score for each.” You might also want to consider asking one or two general questions about perceptions of the website using a likert scale (“overall, how satisfied are you with the information you found on the NML webiste”, “overall, how easy is it to find information on the NML website). This will give you some useful quantitative data to work with, and you may be able to look for correlations between task type and experience with the website, and satisfaction / ease of use.

G.5 Pilot participant 5:

Sorry I do not have much to time at the moment but from a quick look I would say it seems on the right track. There appears to be a few capitalisation issues for example in Q18 the first option begins with a capital letter, but the rest are all lowercase.

G.6 Pilot participant 6:

For me I would have expected to see specific questions around the users' characteristics as discussed in the workshop paper. I thought part of the aim of the survey was to try and validate the proposed dimensions (i.e. domain experience and technical/search experience) and perhaps use other criteria that could then be used to classify users?

G.7 Pilot participant 7:

The language used in places seems very technical and on some of the options there are a couple of choices I could make because of the wording you have used, e.g. There does not seem to be a question 19.

G.8 Pilot participant 8:

Having so many demographic questions early may be a little off putting for some people. Could these be moved about? Try and get the important questions answered first. There are a few questions that seem to be worded either to technically or complex, for example Q18 and Q22.

G.9 Pilot participant 9:

Technically, the poll would work on our system fine. There are a number of typos and inconsistencies that need sorting before it was able to go live and represent the museum though. Is all of the ethics information required as this is very long?

G.10 Pilot participant 10:

The number of questions is a little too many. Could you try reducing them down? Q10, I would advise changing NML to National Museums Liverpool and may be best to do this everywhere. Also Question 18 has some capitalisation inconsistencies.

Appendix H

Phase 2 Analysis and Findings

H.1 Question 4 tables

Table H.1: A complete list of all content type combinations users selected for Question 4

Ranked Content Combination	No. selecting this combination
Text, Pictures, Videos, Audio (speech or music), Interactive / hands on content (Quizzes, virtual tours)	46
Pictures, Text, Videos, Audio (speech or music), Interactive / hands on content (Quizzes, virtual tours)	45
Pictures, Text, Videos, Interactive / hands on content (Quizzes, virtual tours), Audio (speech or music)	35
Pictures, Text, Interactive / hands on content (Quizzes, virtual tours), Videos, Audio (speech or music)	29
Text, Pictures, Videos, Interactive / hands on content (Quizzes, virtual tours), Audio (speech or music)	26
Text, Pictures, Audio (speech or music), Videos, Interactive / hands on content (Quizzes, virtual tours)	20
Pictures, Text, Audio (speech or music), Videos, Interactive / hands on content (Quizzes, virtual tours)	18
Text, Pictures, Interactive / hands on content (Quizzes, virtual tours), Videos, Audio (speech or music)	17
Pictures, Text, Interactive / hands on content (Quizzes, virtual tours), Audio (speech or music), Videos	15
Pictures, Text, Audio (speech or music), Interactive / hands on content (Quizzes, virtual tours), Videos	14
Pictures, Videos, Text, Audio (speech or music), Interactive / hands on content (Quizzes, virtual tours)	13
Pictures, Videos, Interactive / hands on content (Quizzes, virtual tours), Text, Audio (speech or music)	10
Pictures, Videos, Audio (speech or music), Text, Interactive / hands on content (Quizzes, virtual tours)	8

Pictures, Interactive / hands on content (Quizzes, virtual tours), Text, Videos, Audio (speech or music)	7
Text, Pictures, Interactive / hands on content (Quizzes, virtual tours), Audio (speech or music), Videos	7
Text, Interactive / hands on content (Quizzes, virtual tours), Pictures, Videos, Audio (speech or music)	7
Text, Videos, Audio (speech or music), Pictures, Interactive / hands on content (Quizzes, virtual tours)	7
Text, Videos, Pictures, Interactive / hands on content (Quizzes, virtual tours), Audio (speech or music)	7
Text, Videos, Pictures, Audio (speech or music), Interactive / hands on content (Quizzes, virtual tours)	7
Interactive / hands on content (Quizzes, virtual tours), Pictures, Text, Videos, Audio (speech or music)	6
Audio (speech or music), Text, Interactive / hands on content (Quizzes, virtual tours), Videos, Pictures	6
Interactive / hands on content (Quizzes, virtual tours), Videos, Pictures, Audio (speech or music), Text	6
Text, Pictures, Audio (speech or music), Interactive / hands on content (Quizzes, virtual tours), Videos	6
Pictures, Interactive / hands on content (Quizzes, virtual tours), Videos, Text, Audio (speech or music)	5
Pictures, Videos, Text, Interactive / hands on content (Quizzes, virtual tours), Audio (speech or music)	5
Videos, Text, Pictures, Audio (speech or music), Interactive / hands on content (Quizzes, virtual tours)	5
Interactive / hands on content (Quizzes, virtual tours), Pictures, Videos, Audio (speech or music), Text	5
Interactive / hands on content (Quizzes, virtual tours), Videos, Pictures, Text, Audio (speech or music)	5
Videos, Pictures, Audio (speech or music), Text, Interactive / hands on content (Quizzes, virtual tours)	4
Videos, Pictures, Audio (speech or music), Interactive / hands on content (Quizzes, virtual tours), Text	4
Videos, Interactive / hands on content (Quizzes, virtual tours), Pictures, Audio (speech or music), Text	4
Interactive / hands on content (Quizzes, virtual tours), Pictures, Videos, Text, Audio (speech or music)	4
Videos, Pictures, Text, Interactive / hands on content (Quizzes, virtual tours), Audio (speech or music)	4
Pictures, Interactive / hands on content (Quizzes, virtual tours), Videos, Audio (speech or music), Text	3
Pictures, Text, Audio (speech or music), Interactive / hands on content (Quizzes, virtual tours)	3
Audio (speech or music), Text, Interactive / hands on content (Quizzes, virtual tours), Pictures, Videos	3
Pictures, Audio (speech or music), Videos, Text, Interactive / hands on content (Quizzes, virtual tours)	3

Interactive / hands on content (Quizzes, virtual tours), Pictures, Audio (speech or music), Videos, Text	3
Pictures, Videos, Interactive / hands on content (Quizzes, virtual tours), Audio (speech or music), Text	3
Interactive / hands on content (Quizzes, virtual tours), Audio (speech or music), Pictures, Videos, Text	3
Pictures, Interactive / hands on content (Quizzes, virtual tours), Audio (speech or music), Text, Videos	3
Pictures, Videos, Audio (speech or music), Interactive / hands on content (Quizzes, virtual tours), Text	3
Text, Videos, Audio (speech or music), Interactive / hands on content (Quizzes, virtual tours), Pictures	3
Pictures, Interactive / hands on content (Quizzes, virtual tours), Text, Audio (speech or music), Videos	3
Pictures, Videos, Audio (speech or music), Text	3
Interactive / hands on content (Quizzes, virtual tours), Pictures, Text, Audio (speech or music), Videos	3
Videos, Interactive / hands on content (Quizzes, virtual tours), Audio (speech or music), Pictures, Text	3
Interactive / hands on content (Quizzes, virtual tours), Videos, Audio (speech or music), Pictures, Text	3
Pictures, Audio (speech or music), Interactive / hands on content (Quizzes, virtual tours), Text, Videos	3
Audio (speech or music), Text, Videos, Pictures, Interactive / hands on content (Quizzes, virtual tours)	3
Text, Audio (speech or music), Videos, Pictures, Interactive / hands on content (Quizzes, virtual tours)	3
Videos, Pictures, Text, Audio (speech or music), Interactive / hands on content (Quizzes, virtual tours)	3
Videos, Audio (speech or music), Pictures, Interactive / hands on content (Quizzes, virtual tours), Text	3
Pictures, Audio (speech or music), Interactive / hands on content (Quizzes, virtual tours), Videos, Text	2
Text, Pictures, Interactive / hands on content (Quizzes, virtual tours), Audio (speech or music)	2
Text, Audio (speech or music), Pictures, Videos, Interactive / hands on content (Quizzes, virtual tours)	2
Videos, Audio (speech or music), Text, Pictures, Interactive / hands on content (Quizzes, virtual tours)	2
Interactive / hands on content (Quizzes, virtual tours), Text, Audio (speech or music), Pictures, Videos	2
Text, Interactive / hands on content (Quizzes, virtual tours), Audio (speech or music), Videos, Pictures	2
Text, Videos, Interactive / hands on content (Quizzes, virtual tours), Audio (speech or music)	2
Pictures, Audio (speech or music), Text, Interactive / hands on content (Quizzes, virtual tours), Videos	2
Videos, Audio (speech or music), Interactive / hands on content (Quizzes, virtual tours), Text, Pictures	2

Interactive / hands on content (Quizzes, virtual tours), Text, Videos, Pictures, Audio (speech or music)	2
Audio (speech or music), Videos, Interactive / hands on content (Quizzes, virtual tours), Pictures, Text	2
Pictures, Audio (speech or music), Text, Videos, Interactive / hands on content (Quizzes, virtual tours)	2
Pictures, Interactive / hands on content (Quizzes, virtual tours), Audio (speech or music), Videos, Text	2
Interactive / hands on content (Quizzes, virtual tours), Pictures, Audio (speech or music), Text, Videos	2
Interactive / hands on content (Quizzes, virtual tours), Text, Pictures, Videos, Audio (speech or music)	2
Text, Videos, Audio (speech or music), Interactive / hands on content (Quizzes, virtual tours)	2
Audio (speech or music), Videos, Interactive / hands on content (Quizzes, virtual tours), Text, Pictures	2
Audio (speech or music), Pictures, Videos, Interactive / hands on content (Quizzes, virtual tours), Text	2
Audio (speech or music), Videos, Pictures, Interactive / hands on content (Quizzes, virtual tours), Text	2
Audio (speech or music), Pictures, Text, Interactive / hands on content (Quizzes, virtual tours), Videos	2
Interactive / hands on content (Quizzes, virtual tours), Audio (speech or music), Videos, Pictures, Text	2
Interactive / hands on content (Quizzes, virtual tours), Text, Pictures, Audio (speech or music), Videos	2
Audio (speech or music), Pictures, Interactive / hands on content (Quizzes, virtual tours), Videos, Text	2
Text, Interactive / hands on content (Quizzes, virtual tours), Videos, Pictures, Audio (speech or music)	2
Videos, Audio (speech or music), Pictures, Text, Interactive / hands on content (Quizzes, virtual tours)	2
Text, Videos, Interactive / hands on content (Quizzes, virtual tours), Pictures, Audio (speech or music)	2
Interactive / hands on content (Quizzes, virtual tours), Videos, Audio (speech or music), Text, Pictures	2
Interactive / hands on content (Quizzes, virtual tours), Text, Audio (speech or music), Videos, Pictures	2
Text, Audio (speech or music), Interactive / hands on content (Quizzes, virtual tours), Pictures, Videos	1
Audio (speech or music), Interactive / hands on content (Quizzes, virtual tours), Videos, Pictures, Text	1
Pictures, Text, Audio (speech or music), Videos	1
Audio (speech or music), Pictures, Text, Interactive / hands on content (Quizzes, virtual tours)	1
Text, Interactive / hands on content (Quizzes, virtual tours), Audio (speech or music), Pictures, Videos	1
Audio (speech or music), Pictures, Interactive / hands on content (Quizzes, virtual tours), Text, Videos	1

Text, Pictures, Interactive / hands on content (Quizzes, virtual tours), Videos	1
Pictures, Text, Interactive / hands on content (Quizzes, virtual tours), Audio (speech or music)	1
Audio (speech or music), Interactive / hands on content (Quizzes, virtual tours), Text, Pictures, Videos	1
Audio (speech or music), Videos, Pictures, Text, Interactive / hands on content (Quizzes, virtual tours)	1
Interactive / hands on content (Quizzes, virtual tours), Videos, Text, Audio (speech or music), Pictures	1
Videos, Text, Pictures, Interactive / hands on content (Quizzes, virtual tours), Audio (speech or music)	1
Audio (speech or music), Text, Videos, Interactive / hands on content (Quizzes, virtual tours), Pictures	1
Text, Interactive / hands on content (Quizzes, virtual tours), Pictures, Audio (speech or music), Videos	1
Text, Audio (speech or music), Interactive / hands on content (Quizzes, virtual tours), Videos, Pictures	1
Audio (speech or music), Interactive / hands on content (Quizzes, virtual tours), Pictures, Text, Videos	1
Videos, Interactive / hands on content (Quizzes, virtual tours), Text, Audio (speech or music), Pictures	1
Videos, Pictures, Interactive / hands on content (Quizzes, virtual tours), Audio (speech or music), Text	1
Pictures, Interactive / hands on content (Quizzes, virtual tours), Text, Videos	1
Pictures, Audio (speech or music), Videos, Interactive / hands on content (Quizzes, virtual tours)	1
Pictures, Audio (speech or music), Interactive / hands on content (Quizzes, virtual tours), Text	1
Text, Interactive / hands on content (Quizzes, virtual tours), Videos, Audio (speech or music), Pictures	1
Audio (speech or music), Interactive / hands on content (Quizzes, virtual tours), Videos, Text, Pictures	1
Audio (speech or music), Interactive / hands on content (Quizzes, virtual tours), Text, Videos, Pictures	1
Text, Audio (speech or music), Pictures, Interactive / hands on content (Quizzes, virtual tours), Videos	1
Videos, Interactive / hands on content (Quizzes, virtual tours), Pictures, Text, Audio (speech or music)	1
Interactive / hands on content (Quizzes, virtual tours), Audio (speech or music), Videos, Text, Pictures	1
Interactive / hands on content (Quizzes, virtual tours), Videos, Text, Pictures, Audio (speech or music)	1
Pictures, Text, Videos, Interactive / hands on content (Quizzes, virtual tours)	1
Audio (speech or music), Text, Pictures, Interactive / hands on content (Quizzes, virtual tours), Videos	1

Text, Pictures, Videos, Interactive / hands on content (Quizzes, virtual tours)	1
Pictures, Audio (speech or music), Videos, Text	1
Audio (speech or music), Interactive / hands on content (Quizzes, virtual tours), Pictures, Videos, Text	1
Text, Videos, Pictures, Interactive / hands on content (Quizzes, virtual tours)	1
Interactive / hands on content (Quizzes, virtual tours), Text, Videos, Audio (speech or music), Pictures	1
Audio (speech or music), Pictures, Videos, Text, Interactive / hands on content (Quizzes, virtual tours)	1
Pictures, Audio (speech or music), Videos, Interactive / hands on content (Quizzes, virtual tours), Text	1
Interactive / hands on content (Quizzes, virtual tours), Videos, Audio (speech or music), Pictures	1
Interactive / hands on content (Quizzes, virtual tours), Audio (speech or music), Pictures, Text	1
Pictures, Text, Videos, Audio (speech or music)	1
Pictures, Videos, Audio (speech or music), Interactive / hands on content (Quizzes, virtual tours)	1

H.2 Question 22 (Groups) tables

Table H.2: Self selected groups by respondents

Grouping	Count of respondents	Percentage
g	253	44.86%
np	137	24.29%
s	33	5.85%
o	26	4.61%
t	25	4.43%
a	25	4.43%
m	10	1.77%
p	5	0.89%
s/g	5	0.89%
p/a	4	0.71%
a/s	4	0.71%
a/s/g	3	0.53%
np/s/g	3	0.53%
p/a/s/m	2	0.35%
np/a/s/g	2	0.35%
t/g	2	0.35%
np/a/g	2	0.35%
np/a/s	2	0.35%
np/s	2	0.35%
p/m/g	2	0.35%
p/m	2	0.35%
a/s/m	2	0.35%
np/p	2	0.35%
t/s	1	0.18%
m/g	1	0.18%
p/t/m/g	1	0.18%
a/m	1	0.18%
np/p/a/s/m/g	1	0.18%
t/s/g	1	0.18%
a/g	1	0.18%
p/t/s	1	0.18%
s/m	1	0.18%
t/m	1	0.18%
np/a	1	0.18%

H.3 Referral Link Data

Table H.3: Table showing all permutations of referrer links and counts.

Referrer link	Count
http://www.liverpoolmuseums.org.uk/walker/exhibitions/victorian-treasures/index.aspx	55
http://www.liverpoolmuseums.org.uk/maritime/visit/old_dock_tours.aspx	27
http://www.liverpoolmuseums.org.uk/wml	26
http://www.liverpoolmuseums.org.uk/walker	24
http://www.liverpoolmuseums.org.uk/walker/events	22
http://www.liverpoolmuseums.org.uk/wml/events	14
http://www.liverpoolmuseums.org.uk/ism	14
http://www.liverpoolmuseums.org.uk/maritime	12
http://www.liverpoolmuseums.org.uk/ladylever/events	12
http://www.liverpoolmuseums.org.uk/ladylever	9
http://www.liverpoolmuseums.org.uk/walker/index.aspx	9
http://www.liverpoolmuseums.org.uk/ism/slavery/archaeology/caribbean	7
http://www.liverpoolmuseums.org.uk/wml/things-to-see	6
http://www.liverpoolmuseums.org.uk/maritime/index.aspx	6
http://www.liverpoolmuseums.org.uk/maritime/events	6
http://www.liverpoolmuseums.org.uk/sudley	6
http://www.liverpoolmuseums.org.uk/sudley/events	5
http://www.liverpoolmuseums.org.uk/walker/visit	5
http://www.liverpoolmuseums.org.uk/wml/index.aspx	5
http://www.liverpoolmuseums.org.uk/walker/exhibitions/preraphaelites	4
http://www.liverpoolmuseums.org.uk/sudley/index.aspx	4
http://www.liverpoolmuseums.org.uk/ladylever/index.aspx	4
http://www.liverpoolmuseums.org.uk/wml/exhibitions/animal-mummies/index.aspx	4
http://www.liverpoolmuseums.org.uk/wml/visit/index.aspx	4
http://www.liverpoolmuseums.org.uk/ism/slavery/africa/capture_sale.aspx	4
http://www.liverpoolmuseums.org.uk/ladylever/collections	3
http://www.liverpoolmuseums.org.uk/maritime/visit/index.aspx	3
http://www.liverpoolmuseums.org.uk/wml/visit/floor-plans/planetarium.aspx	3
http://www.liverpoolmuseums.org.uk/maritime/visit/floor-plan/life-at-sea/gaylife/polari.aspx	3
http://www.liverpoolmuseums.org.uk/wml/exhibitions/animal-mummies	3
http://www.liverpoolmuseums.org.uk/wml/visit	3
http://www.liverpoolmuseums.org.uk/ism/index.aspx	3
http://www.liverpoolmuseums.org.uk/walker/exhibitions/gay-fashion-icons/index.aspx	3

Table H.3 continued from previous page

Referrer link	Count
http://www.liverpoolmuseums.org.uk/sudley/exhibitions/glitz	3
http://www.liverpoolmuseums.org.uk/wml/collections/antiquities/ancient-egypt/index.aspx	3
http://www.liverpoolmuseums.org.uk/ladylever/events/displayevent.aspx?EventId	3
http://www.liverpoolmuseums.org.uk/maritime/archive	3
http://www.liverpoolmuseums.org.uk/maritime/exhibitions/liverpool-pilots	3
http://www.liverpoolmuseums.org.uk/wml/collections/antiquities/ancient-egypt	2
http://www.liverpoolmuseums.org.uk/walker/events/victorian-treasures-listings.aspx	2
http://www.liverpoolmuseums.org.uk/ism/resources/origins_chattel_slavery.aspx	2
http://www.liverpoolmuseums.org.uk/wml/collections/antiquities/roman/ince-blundell/index.aspx?utm_source=social	2
http://www.liverpoolmuseums.org.uk/walker/exhibitions/victorian-treasures/index.aspx?utm_source=NMLEmail&utm_campa	2
http://www.liverpoolmuseums.org.uk/ism/resources/slave_trade_ports.aspx	2
http://www.liverpoolmuseums.org.uk/ism/visit	2
http://www.liverpoolmuseums.org.uk/walker/collections	2
http://www.liverpoolmuseums.org.uk//ism/index.aspx	2
http://www.liverpoolmuseums.org.uk/walker/events/displayevent.aspx?EventId	2
http://www.liverpoolmuseums.org.uk/wml/visit/floor-plans/bug-house/index.aspx	2
http://www.liverpoolmuseums.org.uk/maritime/visit/floor-plan/titanic	2
http://www.liverpoolmuseums.org.uk/walker/events/displayevent.aspx?EventId	2
http://www.liverpoolmuseums.org.uk/walker/exhibitions/arts-council-collection/coming-out/index.aspx	2
http://www.liverpoolmuseums.org.uk/walker/exhibitions/arts-council-collection/looking-north	2
http://www.liverpoolmuseums.org.uk/maritime/visit/floor-plan/life-at-sea/derbyshire	2
http://www.liverpoolmuseums.org.uk/walker/johnmoores/history/previous-winners-list	2
http://www.liverpoolmuseums.org.uk/walker/events/acc-listings.aspx	2
http://www.liverpoolmuseums.org.uk/maritime/archive/category-sheets.aspx?catId	2
http://www.liverpoolmuseums.org.uk/wml/exhibitions/terracotta-warriors/terracotta-warriors.aspx	2
http://www.liverpoolmuseums.org.uk/ism/slavery/europe/liverpool.aspx	2

Table H.3 continued from previous page

Referrer link	Count
http://www.liverpoolmuseums.org.uk/ladylever/events/chinese-new-year-listings.aspx	2
http://www.liverpoolmuseums.org.uk/walker/contact/index.aspx	2
http://www.liverpoolmuseums.org.uk/maritime/events/displayevent.aspx?EventId	2
http://www.liverpoolmuseums.org.uk/ladylever/visit/getting-here.aspx	2
http://www.liverpoolmuseums.org.uk/walker/collections/paintings/19c/item-238802.aspx	2
http://www.liverpoolmuseums.org.uk/walker/exhibitions/matisse/poem-transcript-en.aspx	1
http://www.liverpoolmuseums.org.uk/walker/exhibitions/victorian-treasures/item-239059.aspx	1
http://www.liverpoolmuseums.org.uk/walker/exhibitions/peterchang/life_in_plastic.aspx	1
http://www.liverpoolmuseums.org.uk/walker/johnmoores/jm2016/index.aspx	1
http://www.liverpoolmuseums.org.uk/walker/exhibitions/victorian-treasures/item-235695.aspx	1
http://www.liverpoolmuseums.org.uk/walker/events/displayevent.aspx?EventId	1
http://www.liverpoolmuseums.org.uk/walker/exhibitions/mucha/index.aspx	1
http://www.liverpoolmuseums.org.uk/ladylever/collections/paintings/gallery3/maymorning.aspx	1
http://www.liverpoolmuseums.org.uk/ladylever/events/displayevent.aspx?EventId	1
http://www.liverpoolmuseums.org.uk/walker/exhibitions/arts-council-collection	1
http://www.liverpoolmuseums.org.uk/walker/events/displayevent.aspx?EventId	1
http://www.liverpoolmuseums.org.uk/walker/johnmoores	1
http://www.liverpoolmuseums.org.uk/ladylever/events/displayevent.aspx?EventId	1
http://www.liverpoolmuseums.org.uk/ladylever/exhibitions/model-image	1
http://www.liverpoolmuseums.org.uk/walker/exhibitions/hockney/gregory.aspx	1
http://www.liverpoolmuseums.org.uk/walker/events/displayevent.aspx?EventId	1
http://www.liverpoolmuseums.org.uk/walker/visit/getting-here.aspx	1
http://www.liverpoolmuseums.org.uk/walker/visit/index.aspx	1
http://www.liverpoolmuseums.org.uk/walker/events/displayevent.aspx?EventId	1
http://www.liverpoolmuseums.org.uk/maritime/exhibitions/child-migrants	1

Table H.3 continued from previous page

Referrer link	Count
http://www.liverpoolmuseums.org.uk/ladylever/collections/star-objects	1
http://www.liverpoolmuseums.org.uk/maritime/visit/directions.aspx	1
http://www.liverpoolmuseums.org.uk/walker/collections/paintings/19c/index.aspx	1
http://www.liverpoolmuseums.org.uk/ism/slavery/africa	1
http://www.liverpoolmuseums.org.uk/wml/collections/antiquities/ivories/related-material-92-1.aspx	1
http://www.liverpoolmuseums.org.uk/wml/collections/ethnology/research.aspx	1
http://www.liverpoolmuseums.org.uk/wml/collections/antiquities/roman/ince-blundell/item-442569.aspx	1
http://www.liverpoolmuseums.org.uk/ladylever/exhibitions/impressionists/timeline.aspx	1
http://www.liverpoolmuseums.org.uk/maritime/collections/lusitania/item-192227.aspx	1
http://www.liverpoolmuseums.org.uk/ladylever/collections/wedgwood	1
http://www.liverpoolmuseums.org.uk/walker/exhibitions/wrightofderby	1
http://www.liverpoolmuseums.org.uk/walker/collections/paintings/19c	1
http://www.liverpoolmuseums.org.uk/walker/exhibitions/henry/whitehall.aspx	1
http://www.liverpoolmuseums.org.uk/ism/slavery/archaeology/caribbean/plantations	1
http://www.liverpoolmuseums.org.uk/walker/collections/paintings/19c/item-242794.aspx	1
http://www.liverpoolmuseums.org.uk/ism/slavery/americas/abolition_americas.aspx	1
http://www.liverpoolmuseums.org.uk/ism/slavery/africa/effect.aspx	1
http://www.liverpoolmuseums.org.uk/ism/exhibitions/art-of-solidarity	1
http://www.liverpoolmuseums.org.uk/wml/collections/ethnology/americas.aspx	1
http://www.liverpoolmuseums.org.uk/maritime/visit/tourist-information.aspx	1
http://www.liverpoolmuseums.org.uk/ladylever/events/displayevent.aspx?EventId	1
http://www.liverpoolmuseums.org.uk/maritime/archive/collections/maritime-family-archives.aspx	1
http://www.liverpoolmuseums.org.uk/walker/collections/highlights	1
http://www.liverpoolmuseums.org.uk/walker/collections/A-Z-of-artists/A-Z-of-artists.aspx	1
http://www.liverpoolmuseums.org.uk/walker/johnmoores/history/fact_file/fact_file16.aspx	1
http://www.liverpoolmuseums.org.uk/walker/visit/kids.aspx	1

Table H.3 continued from previous page

Referrer link	Count
http://www.liverpoolmuseums.org.uk/walker/exhibitions/paultrevor/biography.aspx	1
http://www.liverpoolmuseums.org.uk/walker/collections/paintings/13c-16c/index.aspx	1
http://www.liverpoolmuseums.org.uk/maritime/visit/floor-plan/life-at-sea/gaylife	1
http://www.liverpoolmuseums.org.uk/walker/exhibitions/victorian-treasures/videos.aspx	1
http://www.liverpoolmuseums.org.uk/maritime/visit/floor-plan/life-at-sea/derbyshire/crew-list.aspx	1
http://www.liverpoolmuseums.org.uk/wml/collections/antiquities/ancient-egypt/item-295406.aspx	1
http://www.liverpoolmuseums.org.uk/walker/collections/paintings/17c/item-236479.aspx	1
http://www.liverpoolmuseums.org.uk/ism/visit/directions.aspx	1
http://www.liverpoolmuseums.org.uk/ism/slavery	1
http://www.liverpoolmuseums.org.uk/walker/exhibitions/arts-council-collection/coming-out	1
http://www.liverpoolmuseums.org.uk/ladylever/collections/paintings/gallery3/thenewfrock.aspx	1
http://www.liverpoolmuseums.org.uk/maritime/visit/floor-plan/seized/display/drug-suit.aspx	1
http://www.liverpoolmuseums.org.uk/sudley/collections/entrancehall/sunday_morning_alma-tadema.aspx	1
http://www.liverpoolmuseums.org.uk/maritime/visit	1
http://www.liverpoolmuseums.org.uk/walker/exhibitions/art-in-revolution/online-catalogue.aspx	1
http://www.liverpoolmuseums.org.uk/maritime/collections/boa/albert-dock.aspx	1
http://www.liverpoolmuseums.org.uk/sudley/collections/costumeroom/index.aspx	1
http://www.liverpoolmuseums.org.uk/ism/visit/floor-plan/index.aspx	1
http://www.liverpoolmuseums.org.uk/ism/slavery/middle_passage/olaudah_equiano.aspx	1
http://www.liverpoolmuseums.org.uk/maritime/collections/lusitania/related-person-8897-1.aspx	1
http://www.liverpoolmuseums.org.uk/ism/slavery/middle_passage	1
http://www.liverpoolmuseums.org.uk/wml/exhibitions/thebeatgoeson/thebeatgoesonline/sites/locating/recordshops.aspx	1
http://www.liverpoolmuseums.org.uk/ism/resources/representation_slavery_curacao.aspx	1
http://www.liverpoolmuseums.org.uk/walker/exhibitions/arts-council-collection/kaput/index.aspx	1
http://www.liverpoolmuseums.org.uk/maritime/collections/models	1
http://www.liverpoolmuseums.org.uk/sudley/events/displayevent.aspx?EventId	1

Table H.3 continued from previous page

Referrer link	Count
http://www.liverpoolmuseums.org.uk/ism/events/past-exhibitions.aspx	1
http://www.liverpoolmuseums.org.uk/wml/visit/floor-plans/aquarium	1
http://www.liverpoolmuseums.org.uk/wml/collections/ethnology/africa/selim-aga/item-489235.aspx	1
http://www.liverpoolmuseums.org.uk/ism/slavery/archaeology/background	1
http://www.liverpoolmuseums.org.uk/wml/visit/floor-plans/index.aspx	1
http://www.liverpoolmuseums.org.uk/maritime/archive/displays/politician/stranding.aspx	1
http://www.liverpoolmuseums.org.uk/maritime/archive/listGuides.aspx	1
http://www.liverpoolmuseums.org.uk/ism/visit/floor-plan/africa/igbo-compound.aspx	1
http://www.liverpoolmuseums.org.uk/wml/events/displayevent.aspx?EventId	1
http://www.liverpoolmuseums.org.uk/maritime/visit/floor-plan/lusitania/people/peoples-stories.aspx?id	1
http://www.liverpoolmuseums.org.uk/wml/collections/antiquities/ancient-egypt/item-172129.aspx	1
http://www.liverpoolmuseums.org.uk/maritime/collections/boa/history/merchant-navy.aspx	1
http://www.liverpoolmuseums.org.uk/wml/collections/horology/results.aspx?&name=parry&town=Liverpool&trade=&page	1
http://www.liverpoolmuseums.org.uk/wml/events/planetarium-listings.aspx	1
http://www.liverpoolmuseums.org.uk/maritime/exhibitions/blitz/may.aspx	1
http://www.liverpoolmuseums.org.uk/wml/events/displayevent.aspx?EventId	1
http://www.liverpoolmuseums.org.uk/maritime/contact/maritime-enquiry.aspx	1
http://www.liverpoolmuseums.org.uk/wml/collections/index.aspx	1
http://www.liverpoolmuseums.org.uk/maritime/visit/floor-plan/titanic/liverpool.aspx	1
http://www.liverpoolmuseums.org.uk/ism/learning/index.aspx	1
http://www.liverpoolmuseums.org.uk/maritime/archive/highlights/liverpool-panoramic-photo.aspx	1
http://www.liverpoolmuseums.org.uk/maritime/archive/category-sheets.aspx?catId	1
http://www.liverpoolmuseums.org.uk/sudley/visit	1
http://www.liverpoolmuseums.org.uk/wml/schools-groups/index.aspx	1
http://www.liverpoolmuseums.org.uk/wml/events/displayevent.aspx?EventId	1

Table H.3 continued from previous page

Referrer link	Count
http://www.liverpoolmuseums.org.uk/walker/johnmoores/jm2016/prizewinners/boyd.aspx	1
http://www.liverpoolmuseums.org.uk/walker/exhibitions/victorian-treasures/item-245105.aspx	1
http://www.liverpoolmuseums.org.uk/walker/collections/sculpture/index.aspx	1
http://www.liverpoolmuseums.org.uk/ism/events/displayevent.aspx?EventId	1
http://www.liverpoolmuseums.org.uk/ism/visit/index.aspx	1
http://www.liverpoolmuseums.org.uk/wml/collections/antiquities/ivories/related-person-24343-1.aspx	1
http://www.liverpoolmuseums.org.uk/wml/events/displayevent.aspx?EventId	1
http://www.liverpoolmuseums.org.uk/wml/collections/antiquities/roman/ince-blundell/index.aspx	1
http://www.liverpoolmuseums.org.uk/wml/visit/floor-plans	1
http://www.liverpoolmuseums.org.uk/ism/events/displayevent.aspx?EventId	1
http://www.liverpoolmuseums.org.uk/ism/slavery/europe/abolition.aspx	1
http://www.liverpoolmuseums.org.uk/wml/visit/floor-plans/planningyourvisitpublic.aspx	1
http://www.liverpoolmuseums.org.uk/wml/collections/antiquities/ancient-egypt/item-295373.aspx	1
http://www.liverpoolmuseums.org.uk/ism/events/displayevent.aspx?EventId	1
http://www.liverpoolmuseums.org.uk/wml/events/displayevent.aspx?EventId	1
http://www.liverpoolmuseums.org.uk/maritime/visit/floor-plan/lusitania	1
http://www.liverpoolmuseums.org.uk/maritime/archive/collections/stewartbale/tunnelalbum/construction.aspx	1
http://www.liverpoolmuseums.org.uk/maritime/exhibitions/blitz/blitz.aspx	1
http://www.liverpoolmuseums.org.uk/ism/exhibitions/art-of-solidarity/style-and-design.aspx	1
http://www.liverpoolmuseums.org.uk/wml/visit/shop-eat.aspx	1
http://www.liverpoolmuseums.org.uk/maritime/exhibitions/worldheritagecity/three-graces.aspx	1
http://www.liverpoolmuseums.org.uk/maritime/exhibitions/meccano	1
http://www.liverpoolmuseums.org.uk/wml/visit/floor-plans/ancient-world.aspx	1
http://www.liverpoolmuseums.org.uk/wml/collections/antiquities/ancient-egypt/item-295828.aspx	1
http://www.liverpoolmuseums.org.uk/wml/events/natural-world-listings.aspx	1

Table H.3 continued from previous page

Referrer link	Count
http://www.liverpoolmuseums.org.uk/maritime/archive/collections/stewartbale/tunnelalbum/mystery_climber.aspx	1
http://www.liverpoolmuseums.org.uk/maritime/visit/floor-plan/lusitania/people/peoples-stories.aspx?id	1
http://www.liverpoolmuseums.org.uk/maritime/collections/americancivilwar/end-of-war.aspx	1
http://www.liverpoolmuseums.org.uk/ism/slavery/americas/index.aspx	1
http://www.liverpoolmuseums.org.uk/maritime/collections/americancivilwar	1
http://www.liverpoolmuseums.org.uk/walker/exhibitions/rossetti/works/beauties/thebeloved.aspx	1
Total	**513**

H.4 Group Definitions from Survey Findings

H.4.1 Academic:

The academic group comprises mainly females aged 35 to 64, university-educated, with many holding MSc/PhD degrees or currently pursuing them. They are globally dispersed, with a minority local to the physical museum, mostly employed. Their website visits are for study, work, or personal reasons, aiming to gain knowledge of collections, plan visits, or find specific items. They prefer browsing over searching, scanning pages for known words, and accessing one to 10 pages per visit, occasionally delving deeper. Access times vary due to the nature of their profession and personal interests. They predominantly use PCs, favouring Google Chrome, and exhibit high proficiency in web usage, primarily for utility purposes. This particular group explored most sections of the website, with a focus on the International Slavery Museum, the Walker Art Gallery, and the World Museum.

H.4.2 General Public:

The General Public group consists of individuals primarily aged 35 to 64, with significant representation across all age groups. Most are employed either full or part-time, with some seeking employment. They are predominantly female and have varying levels of education, from PhD down to secondary school. They visit the NML website for personal reasons, such as trip planning or exploring collections, typically viewing between 1 and 10 pages per visit. They access various sections of the website, with popular ones including the Walker Art Gallery and the World Museum. Their visits are often in the evenings, and they prefer scanning pages or viewing images. While they prefer browsing, they may search 2 or 3 times before finding desired content. Their domain knowledge is generally low to moderate, as is their knowledge of cultural heritage. They access the website from PCs or mobile devices, often using older operating systems and browsers, and exhibit proficiency in using social media and web utility sites.

H.4.3 Museum Staff:

The museum staff group primarily consists of employed individuals aged 18-34, with a high proportion being female (80%). They typically hold BSc or MSc degrees and may be engaged in further studies. Most are located in the Merseyside area and frequently visit the museum's website for work, personal, or leisure purposes. They often explore event, collection, or exhibition pages, visiting the site monthly, weekly, or even daily, accessing 2-10 pages per visit. Their visits are mainly to larger museums like the Walker Art Gallery and World Museum, occurring during the working day. They prefer scanning pages before reading them fully and equally engage in searching and browsing, although they may give up searching after a couple of attempts. Their domain knowledge varies from some to highly experienced, with a generally high level of cultural heritage knowledge. They predominantly use PCs, occasionally using mobile devices for personal reasons. Proficiency levels in web usage vary, with high proficiency in social media but medium to high proficiency in other online activities such as banking and online shopping.

H.4.4 Non-Professionals:

The non-professional group consists of individuals primarily from the local area but could also include members from around the world. They span various age groups, with slightly more 35–54-year-olds. Gender distribution is relatively even. About half are employed, with a significant portion being retired. Education levels range from PhDs to diploma or CPD level study, with ongoing education common. Their website visits are mainly for personal reasons or leisure, often in preparation for a physical museum visit. They typically view a single page or fewer than ten pages per visit, sometimes with others present, including children. Most are first-time visitors, with some visiting more frequently. They access all museums, with the Maritime Museum and Walker Art Gallery being popular choices. Website visits occur during afternoons or evenings, with a preference for browsing over searching. They tend to scan pages for content and pictures before reading fully. Domain knowledge varies, but general cultural heritage knowledge is moderate to low. They typically use larger-screened devices, with PCs being the main choice and default PC browsers. Proficiency in web usage varies, with mixed experiences in social media but regular

use of online banking and shopping.

H.4.5 Professionals:

The professional group consists of employed individuals mainly from areas outside NML, including the wider UK and worldwide. They are typically in early to mid-career stages, with an equal split between genders. Education levels range from undergraduate to master's level, with some continuing study. Their visit to the website may be their first and only, motivated by work or personal reasons. They typically seek known collections or objects, visiting collection or exhibition pages, often viewing only one page. They usually visit alone and focus on main museums and galleries like the Maritime Museum or Walker Art Gallery. Visits occur during the usual working day, with a preference for scanning pages. They prefer searching over browsing, typically searching a few times before giving up. They may have some domain knowledge but generally possess good general cultural heritage knowledge. Professionals typically use PCs, either Windows or MAC, and default browsers. Their web proficiency varies on social media but is relatively high for tasks like banking, shopping, or media consumption.

H.4.6 Students:

The student group, predominantly young adults aged 18-34, is mainly comprised of females. They are primarily full-time students, ranging from FE to PhD level, possibly also employed. Their website visits, mainly for study purposes, focus on gaining knowledge of collections and artefacts, typically accessing 1-10 pages. They usually visit alone, with occasional returns a few times a year or monthly/weekly. Accessing larger museums like the International Slavery Museum and Walker Art Gallery is common. They access the site anytime, scanning pages for content and viewing pictures first. They both search and browse, persisting in searching before giving up. Their domain knowledge ranges from novice to some experience, with a mid-level general CH knowledge. Students prefer larger screens like PCs/laptops or tablets, mostly using Mac or Linux with Safari or Chrome browsers. They are proficient in various online tasks, including social media, online banking, shopping, media consumption, and using online documents.

H.4.7 Teachers:

The teacher group, mainly aged 35-54 and predominantly female, is locally employed and typically educated to at least a B.Sc. level. They may engage in continuing study, often opting for CPD. Their website visits, for personal or work reasons, focus on gaining museum and collection knowledge, often viewing up to 10 pages per visit. They access the site alone, occasionally with children, favouring main museums like World Museum Liverpool and Walker Art Gallery. Access occurs mostly in the evening, and they prefer scanning pages for words over reading or viewing images. They browse using navigational links rather than searching, but may search a few times before giving up. Their domain knowledge ranges from mid to high levels, with mid-level general CH knowledge. Teachers are comfortable with PCs or mobile devices, often using an IOS device with the default browser. Their web page proficiency varies, excelling in Facebook use but not in other social media platforms. They are adept at online banking, shopping, audio, and video content, as well as using online documents.

Appendix I

Phase 3 Analysis

I.1 Example of Web Logs

Table I.1: Example of the the NML server log files broken down into columns

date	time	s-ip	cs-method	cs-uri-stem	cs-uri-query	s-port	cs-username	c-ip	cs(User-Agent)	sc-status	sc-substatus	sc-win32-status	time-taken
08/07/2015	00:00:11	-	GET	/whatsonnet/event_feed.aspx	-	-	-	207.46.13.134	Mozilla/5.0 (compatible; bingbot/2.0; +http://www.bing.com/bingbot.htm)	302	0	0	00:00:07
08/07/2015	00:00:18	-	GET	/onlineshop/graphics/product/thumbnails/The-steady-states-new-thumb.jpg	-	-	-	66.249.64.146	Googlebot-Image/1.0	304	0	0	00:00:02
08/07/2015	00:00:20	-	GET	/maritime/collections/artsea/models/t07760.aspx	-	-	-	157.55.39.98	Mozilla/5.0 (compatible; bingbot/2.0; +http://www.bing.com/bingbot.htm)	301	0	0	00:00:01
08/07/2015	00:00:21	-	GET	/wml/index.aspx	-	-	-	172.17.9.4	Mozilla/5.0 (Windows NT 6.1; Trident/7.0; rv:11.0) like Gecko	200	0	0	00:00:02
08/07/2015	00:00:23	-	GET	/images/sitewide/nml-logo.gif	-	-	-	66.249.93.252	Mozilla/5.0 (Windows NT 5.1; rv:11.0) Gecko Firefox/11.0 (via ggpht.com GoogleImageProxy)	200	0	0	00:00:00
08/07/2015	00:00:23	-	GET	/images/email-footer/mayas-exhibition.jpg	-	-	-	66.249.93.131	Mozilla/5.0 (Windows NT 5.1; rv:11.0) Gecko Firefox/11.0 (via ggpht.com GoogleImageProxy)	200	0	0	00:00:08
08/07/2015	00:00:31	-	GET	/walker/craftdesign/puttingstyle/costume/inspiration/hardy_blechman.aspx	-	-	-	157.55.39.98	Mozilla/5.0 (compatible; bingbot/2.0; +http://www.bing.com/bingbot.htm)	302	0	0	00:00:02
08/07/2015	00:00:33	-	GET	/about/page-not-found.aspx	aspxerrorpath=/walker/craftdesign/puttingstyle/costume/inspiration/hardy_blechman.aspx	-	-	157.55.39.98	Mozilla/5.0 (compatible; bingbot/2.0; +http://www.bing.com/bingbot.htm)	200	0	0	00:00:03
08/07/2015	00:00:36	-	GET	/onlineshop/graphics/product/thumbnails/art-of-the-samurai-new-thum.jpg	-	-	-	66.249.64.151	Googlebot-Image/1.0	304	0	0	00:00:09
08/07/2015	00:00:45	-	GET	/ism/slavery/middle_passage/	-	-	-	49.245.101.80	Mozilla/5.0 (Windows NT 6.2; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/43.0.2357.130 Safari/537.36	200	0	0	00:00:00
08/07/2015	00:00:45	-	GET	/SiteElements/redesign-implementation/css/100-main.css	version=160140?version=160140	-	-	49.245.101.80	Mozilla/5.0 (Windows NT 6.2; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/43.0.2357.130 Safari/537.36	200	0	0	00:00:01
08/07/2015	00:00:46	-	GET	/SiteElements/redesign-implementation/css/210-general.css	version=154453?version=154453	-	-	49.245.101.80	Mozilla/5.0 (Windows NT 6.2; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/43.0.2357.130 Safari/537.36	200	0	0	00:00:00

I.2 Page types by Country

Table I.2: Complete table listing page type by country with total requests and queries.

Countries	Archive	Collection	Event	General	Item	Kids	Museum	Other	Shop	Requests	Queries
United Kingdom	6005	17438	43061	111138	47842	7979	40666	21865	9096	305090	55514
United States	1315	5076	1702	15642	33574	21976	5817	10279	3422	98803	12092
Denmark	354	1278	585	5591	17180	1744	791	900	3107	31530	20368
Germany	463	1219	341	8302	3096	115	886	1120	268	15810	4081
Australia	210	309	48	1089	10902	1524	267	1336	86	15771	571
Ireland	253	771	6953	3277	1275	100	571	540	447	14187	7372
France	172	470	694	3704	3104	166	890	951	426	10577	3157
Canada	282	320	131	1293	1744	685	277	563	120	5415	834
Russian Federation	97	333	349	2308	632	222	958	269	130	5298	1436
Netherlands	59	121	124	1526	966	54	320	354	101	3625	851
Spain	62	155	163	1116	948	189	395	347	143	3518	815
India	24	96	840	1355	585	174	166	145	70	3455	1219
China	55	283	127	1150	693	43	302	181	238	3072	718
Italy	27	140	95	963	917	54	389	377	58	3020	641
Poland	21	61	69	557	695	93	216	153	48	1913	465
New Zealand	50	88	17	158	973	57	48	152	19	1562	120
Norway	24	49	51	307	283	19	142	483	10	1368	88
Brazil	15	24	81	565	174	12	126	236	13	1246	236
Korea Republic of	4	72	50	615	179	19	30	116	20	1105	190
Belgium	7	27	35	275	285	9	111	107	80	936	226
Sweden	16	33	38	324	229	16	112	121	26	915	170
Jamaica	0	457	0	66	140	4	6	157	1	831	28

Table I.2 continued from previous page

Countries	Archive	Collection	Event	General	Item	Kids	Museum	Other	Shop	Requests	Queries
Japan	13	40	48	252	246	4	113	87	17	820	158
Hong Kong	15	22	102	255	147	35	64	44	19	703	254
Portugal	14	29	17	275	136	6	78	59	41	655	92
South Africa	21	32	4	191	229	52	26	85	2	642	78
Philippines	7	14	13	342	90	71	38	61	2	638	26
Switzerland	12	48	39	218	142	5	90	66	18	638	108
Isle of Man	18	33	59	294	69	10	97	47	11	638	89
Ukraine	26	20	37	285	137	14	37	72	7	635	194
Trinidad and Tobago	0	255	2	45	204	6	4	118	0	634	23
Turkey	11	18	33	361	80	27	29	53	6	618	197
Greece	8	30	33	210	163	23	51	69	27	614	131
Czech Republic	12	21	21	213	135	3	80	32	15	532	99
Mexico	0	21	8	173	112	16	37	121	14	502	47
Israel	4	19	18	270	73	11	32	37	11	475	174
Egypt	3	7	7	232	139	33	10	37	1	469	131
Slovakia	5	27	95	168	115	7	30	11	7	465	206
Colombia	0	18	1	69	315	4	26	23	2	458	12
Romania	5	11	11	276	69	14	35	25	10	456	34
Hungary	4	11	17	162	158	2	57	26	13	450	88
Singapore	6	13	39	209	69	32	40	38	2	448	74
Argentina	5	13	7	143	127	4	46	81	5	431	38
Austria	12	27	19	161	86	22	48	47	8	430	56
Viet Nam	3	6	64	73	234	11	3	26	3	423	124
Finland	14	12	21	113	110	8	44	41	15	378	74
Indonesia	8	7	30	207	36	30	16	31	4	369	66
Malaysia	0	10	4	139	114	26	23	46	3	365	13
Nigeria	2	13	0	45	183	7	8	73	1	332	10

Table I.2 continued from previous page

Countries	Archive	Collection	Event	General	Item	Kids	Museum	Other	Shop	Requests	Queries
Pakistan	1	12	2	266	16	20	4	8	0	329	9
Serbia	1	9	2	222	51	7	17	18	1	328	17
United Arab Emirates	4	16	6	95	65	39	14	56	14	309	34
Luxembourg	5	4	18	102	86	2	15	32	39	303	156
Croatia	0	10	7	149	56	9	21	19	9	280	30
Malta	1	4	8	134	17	21	62	8	4	259	14
Saudi Arabia	7	12	2	127	60	15	8	21	3	255	35
Thailand	4	12	13	97	55	9	14	27	2	233	58
Bulgaria	2	5	6	88	50	24	34	14	8	231	35
Taiwan Province of China	2	16	7	88	65	2	14	30	6	230	37
Bangladesh	2	8	4	123	28	6	18	15	2	206	14
Estonia	0	12	19	69	38	6	36	15	10	205	50
Barbados	6	91	0	22	56	0	2	23	1	201	10
Iran Islamic Republic of	4	3	11	114	20	32	6	7	0	197	94
Algeria	2	7	3	145	16	0	3	14	0	190	3
Jordan	1	2	0	61	91	4	16	10	1	186	10
Latvia	1	4	17	98	18	5	25	2	2	172	31
Chile	11	4	3	83	29	0	8	34	0	172	11
Jersey	2	6	24	76	11	10	26	7	7	169	22
Morocco	3	1	2	111	15	11	11	15	0	169	10
Lithuania	0	7	11	80	23	2	20	5	3	151	19
Saint Kitts and Nevis	0	90	0	37	6	0	1	6	0	140	3
Qatar	0	2	1	55	37	8	10	17	3	133	12
Cyprus	1	26	8	31	25	12	10	16	2	131	22
Belize	0	43	0	8	51	0	0	14	8	124	8

Table I.2 continued from previous page

Countries	Archive	Collection	Event	General	Item	Kids	Museum	Other	Shop	Requests	Queries
Guyana	0	54	0	8	32	2	0	17	0	113	1
Slovenia	0	3	3	47	26	1	13	12	3	108	10
Kenya	3	5	0	23	39	3	2	27	0	102	6
Ghana	0	3	3	25	43	2	4	20	0	100	5
Saint Vincent and The Grenadines	1	82	0	2	10	0	0	4	0	99	1
Bahamas	2	29	1	5	36	1	1	15	0	90	1
Tunisia	1	1	3	58	6	0	6	14	1	90	5
Peru	1	1	1	35	15	4	6	25	1	89	16
Antigua and Barbuda	1	35	1	10	28	0	1	9	3	88	11
Sri Lanka	2	1	2	55	16	2	1	4	0	83	3
Bosnia and Herzegovina	1	1	0	53	15	5	3	1	0	79	0
El Salvador	0	1	0	6	8	57	2	4	0	78	6
Iceland	0	2	10	24	12	1	12	6	9	76	17
Zimbabwe	0	3	0	10	46	0	1	13	0	73	5
Grenada	0	46	0	1	23	0	0	3	0	73	0
Martinique	0	6	0	6	7	0	42	6	0	67	2
Lebanon	0	3	0	24	18	14	2	5	0	66	5
Macedonia The Former Yugoslav Republic of	0	2	0	28	20	3	2	9	0	64	2
Venezuela Bolivarian Republic of	0	3	2	42	6	1	2	7	0	63	7
Belarus	3	2	2	17	16	2	9	6	1	58	13
Saint Lucia	1	34	0	5	10	0	1	7	0	58	3
Guernsey	4	10	1	23	11	1	3	4	0	57	5

Table I.2 continued from previous page

Countries	Archive	Collection	Event	General	Item	Kids	Museum	Other	Shop	Requests	Queries
Puerto Rico	0	7	1	10	22	4	1	11	0	56	2
Costa Rica	0	1	1	19	21	0	1	8	1	52	6
Panama	0	1	2	27	5	4	8	5	0	52	7
Mauritius	0	0	0	27	16	1	1	6	0	51	0
Kuwait	0	1	3	24	8	5	5	4	0	50	3
Liberia	1	0	0	6	21	0	6	15	0	49	3
Anguilla	0	21	0	3	23	0	0	0	0	47	3
Georgia	0	2	2	20	13	1	2	7	0	47	5
Iraq	0	1	2	27	6	4	2	3	0	45	5
Dominican Republic	0	2	0	26	6	1	3	7	0	45	2
Nepal	0	0	0	31	10	0	2	1	0	44	7
Ecuador	1	2	0	21	5	0	1	10	0	40	4
Tanzania United Republic of	1	2	0	7	23	1	1	5	0	40	0
Cayman Islands	0	11	0	13	7	6	0	3	0	40	0
Gibraltar	0	0	6	22	4	0	6	1	0	39	3
Uruguay	0	3	1	17	5	2	4	4	2	38	4
Albania	1	0	0	25	7	1	2	0	0	36	1
Azerbaijan	1	0	0	21	5	2	1	6	0	36	5
Macao	0	4	2	17	9	0	1	1	1	35	10
Oman	1	0	0	20	3	7	2	2	0	35	2
Togo	0	2	2	1	1	0	29	0	0	35	2
Bermuda	0	6	0	3	7	12	0	7	0	35	5
Kazakhstan	0	9	1	6	13	1	2	3	0	35	11
Bhutan	0	0	0	1	33	0	0	0	0	34	28
Cambodia	1	2	0	8	7	10	2	3	0	33	1
Virgin Islands British	0	23	0	1	5	0	1	3	0	33	0

Table I.2 continued from previous page

Countries	Archive	Collection	Event	General	Item	Kids	Museum	Other	Shop	Requests	Queries
Bahrain	0	3	0	8	1	3	1	9	7	32	7
Dominica	0	21	0	3	4	0	0	3	0	31	0
Palestine State of	0	0	0	19	5	0	0	5	0	29	1
Haiti	0	1	2	19	1	0	4	2	0	29	2
Ethiopia	0	1	0	13	7	1	0	5	0	27	0
Cote D'ivoire	0	0	0	21	0	0	2	4	0	27	1
Reunion	0	0	0	10	11	2	1	3	0	27	3
Zambia	0	2	0	5	13	1	1	3	0	25	0
Virgin Islands U.S.	0	19	0	1	2	1	1	1	0	25	0
Moldova Republic of	0	0	1	11	4	0	4	3	1	24	3
Turks and Caicos Islands	0	6	0	2	9	1	0	5	0	23	1
Sint Maarten (Dutch Part)	0	14	0	3	3	0	0	2	0	22	6
French Guiana	0	2	6	5	2	0	3	3	1	22	7
Malawi	0	0	0	4	1	13	2	1	0	21	2
Montenegro	1	0	2	8	3	0	1	6	0	21	4
Mongolia	0	1	0	10	7	0	2	0	0	20	7
Guatemala	0	6	0	5	6	0	0	3	0	20	3
Uganda	0	0	0	3	9	1	0	6	0	19	0
Cameroon	0	0	0	10	7	0	0	1	0	18	0
Syrian Arab Republic	1	0	0	12	4	0	1	0	0	18	0
Sudan	0	0	1	4	6	0	0	7	0	18	9
Botswana	2	3	0	7	4	1	0	0	0	17	0
Bolivia Plurinational State of	0	0	0	8	4	3	0	2	0	17	1
Myanmar	0	1	0	6	0	3	2	4	0	16	1

Table I.2 continued from previous page

Countries	Archive	Collection	Event	General	Item	Kids	Museum	Other	Shop	Requests	Queries
Senegal	0	0	0	7	5	1	0	2	0	15	0
Namibia	0	2	0	4	4	1	1	2	0	14	0
Faroe Islands	3	2	3	2	1	0	0	2	0	13	4
Paraguay	1	1	0	6	3	0	0	2	0	13	0
Gabon	0	0	0	3	8	0	1	0	0	12	2
Brunei Darussalam	0	1	0	6	3	1	0	1	0	12	0
Armenia	0	1	1	4	4	0	1	0	0	11	2
Honduras	0	2	0	4	5	0	0	0	0	11	0
Falkland Islands (Malvinas)	0	1	2	5	1	0	1	1	0	11	3
Guam	1	1	0	0	5	1	1	2	0	11	0
Libya	1	0	0	4	5	0	1	0	0	11	4
Guadeloupe	0	5	0	2	1	0	0	2	0	10	0
Gambia	0	0	0	1	5	0	0	2	2	10	2
Angola	0	0	0	9	1	0	0	0	0	10	0
Mozambique	0	0	0	3	0	0	0	7	0	10	6
Uzbekistan	2	0	0	2	1	0	2	2	0	9	3
Aruba	0	3	0	3	0	2	0	1	0	9	1
Nicaragua	0	1	1	3	1	0	0	2	0	8	3
Rwanda	0	1	0	3	2	0	0	1	0	7	0
Benin	0	0	0	1	5	0	0	1	0	7	0
Maldives	0	0	0	5	0	0	0	1	0	6	0
Afghanistan	0	0	0	1	2	1	0	2	0	6	1
Monaco	3	0	0	1	0	0	0	2	0	6	2
Madagascar	0	0	0	3	1	0	0	1	0	5	0
Suriname	1	0	0	1	0	0	0	2	1	5	1
Yemen	0	1	0	2	0	0	0	2	0	5	1

Table I.2 continued from previous page

Countries	Archive	Collection	Event	General	Item	Kids	Museum	Other	Shop	Requests	Queries
Cuba	0	0	0	0	0	5	0	0	0	5	0
Mauritania	0	0	0	1	4	0	0	0	0	5	0
Somalia	0	0	0	4	1	0	0	0	0	5	0
Curacao	0	2	0	1	2	0	0	0	0	5	2
Seychelles	0	1	0	0	0	0	1	3	0	5	2
Congo	0	0	0	2	2	0	0	1	0	5	0
Cabo Verde	0	0	0	2	3	0	0	0	0	5	2
Mali	0	0	1	1	1	0	0	0	1	4	2
Northern Mariana Islands	0	0	0	2	2	0	0	0	0	4	0
Fiji	0	0	0	1	3	0	0	0	0	4	0
Saint Martin (French Part)	0	2	0	0	1	0	0	0	0	3	0
Guinea	0	1	0	0	1	0	0	1	0	3	0
Andorra	0	0	0	2	0	0	0	1	0	3	1
Aland Islands	0	0	1	0	0	0	1	0	1	3	1
Lao People's Democratic Republic	0	0	0	0	3	0	0	0	0	3	2
Liechtenstein	0	2	0	1	0	0	0	0	0	3	0
Lesotho	0	0	0	1	1	0	0	1	0	3	0
Burkina Faso	0	1	0	1	0	0	0	1	0	3	0
Sierra Leone	0	0	0	1	1	0	0	0	0	2	0
Holy See	0	0	0	1	1	0	0	0	0	2	0
Timor-Leste	0	0	0	2	0	0	0	0	0	2	0
Eswatini	0	0	0	1	0	0	0	1	0	2	0
Mayotte	0	0	0	0	1	0	0	1	0	2	0
Burundi	0	0	0	0	2	0	0	0	0	2	0

Table I.2 continued from previous page

Countries	Archive	Collection	Event	General	Item	Kids	Museum	Other	Shop	Requests	Queries
Kyrgyzstan	0	0	0	2	0	0	0	0	0	2	0
Montserrat	0	0	0	0	2	0	0	0	0	2	0
Cook Islands	0	0	0	0	1	0	0	1	0	2	0
Bonaire Sint Eustatius and Saba	0	1	0	1	0	0	0	0	0	2	0
Guinea-Bissau	0	0	0	1	0	0	0	0	0	1	0
Djibouti	0	0	0	1	0	0	0	0	0	1	0
Greenland	0	0	0	0	1	0	0	0	0	1	0
Samoa	0	0	0	0	0	0	1	0	0	1	0
American Samoa	0	0	0	0	1	0	0	0	0	1	0
Marshall Islands	0	0	0	0	1	0	0	0	0	1	0
Tajikistan	0	1	0	0	0	0	0	0	0	1	0
French Polynesia	0	0	0	1	0	0	0	0	0	1	0
Arizona	1	0	0	0	0	0	0	0	0	1	0
New Caledonia	0	0	0	0	0	0	0	1	0	1	0
Equatorial Guinea	0	0	0	1	0	0	0	0	0	1	0
San Marino	0	0	0	1	0	0	0	0	0	1	0
Niger	0	0	0	0	0	0	0	1	0	1	0
Column Total	**9847**	**30543**	**56550**	**170407**	**132058**	**36184**	**55275**	**43105**	**18354**	**552323**	**114880**

I.3 Sessions

I.3.1 All sessions

Table I.3: Complete table of sessions by country

Country	uristem	uriquery
United Kingdom	305090	55514
United States	98803	12092
Denmark	31530	20368
Germany	15810	4081
Australia	15771	571
Ireland	14187	7372
France	10577	3157
Canada	5415	834
Russian Federation	5298	1436
Netherlands	3625	851
Spain	3518	815
India	3455	1219
China	3072	718
Italy	3020	641
Poland	1913	465
New Zealand	1562	120
Norway	1368	88
Brazil	1246	236
Korea Republic of	1105	190
Belgium	936	226.0
Sweden	915	170.0
Jamaica	831	28.0
Japan	820	158.0
Hong Kong	703	254.0
Portugal	655	92.0
South Africa	642	78.0
Philippines	638	26.0
Switzerland	638	108.0
Isle of Man	638	89.0
Ukraine	635	194
Trinidad and Tobago	634	23
Turkey	618	197
Greece	614	131
Czech Republic	532	99
Mexico	502	47
Israel	475	174
Egypt	469	131
Slovakia	465	206
Colombia	458	12

Table I.3 continued from previous page

Country	uristem	uriquery
Romania	456	34
Hungary	450	88
Singapore	448	74
Argentina	431	38
Austria	430	56
Viet Nam	423	124
Finland	378	74
Indonesia	369	66
Malaysia	365	13
Nigeria	332	10
Pakistan	329	9
Serbia	328	17
United Arab Emirates	309	34
Luxembourg	303	156
Croatia	280	30
Malta	259	14
Saudi Arabia	255	35
Thailand	233	58
Bulgaria	231	35
Taiwan Province of China	230	37
Bangladesh	206	14
Estonia	205	50
Barbados	201	10
Iran Islamic Republic of	197	94
Algeria	190	3
Jordan	186	10
Latvia	172	31
Chile	172	11
Morocco	169	10
Jersey	169	22
Lithuania	151	19
Saint Kitts and Nevis	140	3
Qatar	133	12
Cyprus	131	22
Belize	124	8
Guyana	113	1
Slovenia	108	10
Kenya	102	6
Ghana	100	5
Saint Vincent and The Grenadines	99	1
Bahamas	90	1
Tunisia	90	5
Peru	89	16
Antigua and Barbuda	88	11
Sri Lanka	83	3
Bosnia and Herzegovina	79	0

Table I.3 continued from previous page

Country	uristem	uriquery
El Salvador	78	6
Iceland	76	17
Grenada	73	0
Zimbabwe	73	5
Martinique	67	2
Lebanon	66	5
Macedonia The Former Yugoslav Republic of	64	2
Venezuela Bolivarian Republic of	63	7
Belarus	58	13
Saint Lucia	58	3
Guernsey	57	5
Puerto Rico	56	2
Costa Rica	52	6
Panama	52	7
Mauritius	51	0
Kuwait	50	3
Liberia	49	3
Anguilla	47	3
Georgia	47	5
Iraq	45	5
Dominican Republic	45	2
Nepal	44	7
Ecuador	40	4
Cayman Islands	40	0
Tanzania United Republic of	40	0
Gibraltar	39	3
Uruguay	38	4
Albania	36	1
Azerbaijan	36	5
Togo	35	2
Macao	35	10
Oman	35	2
Kazakhstan	35	11
Bermuda	35	5
Bhutan	34	28
Virgin Islands British	33	0
Cambodia	33	1
Bahrain	32	7
Dominica	31	0
Palestine State of	29	1
Haiti	29	2
Reunion	27	3
Cote D'ivoire	27	1
Ethiopia	27	0
Virgin Islands U.S.	25	0
Zambia	25	0

Table I.3 continued from previous page

Country	uristem	uriquery
Moldova Republic of	24	3
Turks and Caicos Islands	23	1
Sint Maarten (Dutch Part)	22	6
French Guiana	22	7
Malawi	21	2
Montenegro	21	4
Mongolia	20	7
Guatemala	20	3
Uganda	19	0
Cameroon	18	0
Syrian Arab Republic	18	0
Sudan	18	9
Botswana	17	0
Bolivia Plurinational State of	17	1
Myanmar	16	1
Senegal	15	0
Namibia	14	0
Paraguay	13	0
Faroe Islands	13	4
Brunei Darussalam	12	0
Gabon	12	2
Guam	11	0
Armenia	11	2
Honduras	11	0
Libya	11	4
Falkland Islands (Malvinas)	11	3
Gambia	10	2
Guadeloupe	10	0
Mozambique	10	6
Angola	10	0
Aruba	9	1
Uzbekistan	9	3
Nicaragua	8	3
Benin	7	0
Rwanda	7	0
Afghanistan	6	1
Monaco	6	2
Maldives	6	0
Mauritania	5	0
Seychelles	5	2
Cabo Verde	5	2
Somalia	5	0
Yemen	5	1
Suriname	5	1
Cuba	5	0
Curacao	5	2

Table I.3 continued from previous page

Country	uristem	uriquery
Madagascar	5	0
Congo	5	0
Mali	4	2
Fiji	4	0
Northern Mariana Islands	4	0
Burkina Faso	3	0
Lesotho	3	0
Andorra	3	1
Lao People's Democratic Republic	3	2
Guinea	3	0
Aland Islands	3	1
Saint Martin (French Part)	3	0
Liechtenstein	3	0
Cook Islands	2	0
Montserrat	2	0
Mayotte	2	0
Burundi	2	0
Eswatini	2	0
Holy See	2	0
Sierra Leone	2	0
Kyrgyzstan	2	0
Bonaire Sint Eustatius and Saba	2	0
Timor-Leste	2	0
French Polynesia	1	0
New Caledonia	1	0
Greenland	1	0
American Samoa	1	0
Equatorial Guinea	1	0
Arizona	1	0
Niger	1	0
Guinea-Bissau	1	0
Djibouti	1	0
Tajikistan	1	0
Marshall Islands	1	0
San Marino	1	0
Samoa	1	0
Column Total	**552323**	**114880**

I.3.2 Single page sessions

Table I.4: Complete table of single page sessions by country

Country	uristem	uriquery
United Kingdom	129303	19859
United States	37496	4699
Germany	10370	3534
Australia	5158	255
France	3869	1097
Canada	2646	422
Ireland	2141	241
Spain	1726	287
India	1601	174
Italy	1560	228
Russian Federation	1514	766
Netherlands	1442	248
Poland	1053	221
China	908	242
Brazil	878	134
Norway	859	58
Denmark	814	38
New Zealand	751	68
Korea Republic of	672	73
Jamaica	586	17
Philippines	525	15
Sweden	521	80
Turkey	478	152
Japan	464	74
Belgium	444	70
Ukraine	404	94
Trinidad and Tobago	383	8
South Africa	374	26
Slovakia	359	150
Romania	345	17
Hong Kong	332	145
Portugal	330	29
Singapore	324	59
Egypt	322	43
Israel	320	121
Switzerland	315	54
Mexico	310	20
Pakistan	310	6
Indonesia	306	39
Czech Republic	300	35
Greece	297	40
Isle of Man	295	29

Table I.4 continued from previous page

Country	uristem	uriquery
Argentina	281	19
Serbia	275	9
Hungary	274	26
Malaysia	257	12
Viet Nam	255	87
Nigeria	254	7
United Arab Emirates	232	12
Austria	229	31
Finland	222	28
Croatia	211	9
Luxembourg	211	108
Thailand	191	41
Algeria	182	3
Saudi Arabia	173	9
Colombia	164	10
Chile	146	5
Bulgaria	141	21
Morocco	140	8
Iran Islamic Republic of	133	68
Taiwan Province of China	128	15
Barbados	126	5
Bangladesh	111	4
Estonia	110	14
Jordan	110	3
Malta	105	7
Lithuania	92	12
Guyana	91	1
Saint Kitts and Nevis	82	1
Kenya	80	6
Tunisia	80	4
Qatar	78	4
Cyprus	77	13
Ghana	75	2
Sri Lanka	74	3
Jersey	68	8
Latvia	65	10
Bosnia and Herzegovina	65	0
Saint Vincent and The Grenadines	55	1
Belize	53	0
Bahamas	53	1
Lebanon	53	2
Grenada	52	0
Peru	52	3
Antigua and Barbuda	52	11
Venezuela Bolivarian Republic of	51	2
Zimbabwe	50	2

Table I.4 continued from previous page

Country	uristem	uriquery
Slovenia	48	3
Macedonia The Former Yugoslav Republic of	46	2
Mauritius	46	0
Iceland	43	5
Panama	40	7
Iraq	40	5
Kuwait	39	2
Puerto Rico	39	1
Liberia	38	2
Dominican Republic	38	2
Tanzania United Republic of	36	0
Martinique	36	0
Albania	34	1
El Salvador	34	1
Azerbaijan	34	4
Macao	33	10
Guernsey	33	3
Georgia	33	2
Saint Lucia	31	1
Oman	30	1
Costa Rica	30	3
Palestine State of	29	1
Nepal	29	2
Ecuador	29	3
Cote D'ivoire	27	1
Dominica	27	0
Cambodia	25	1
Uruguay	25	3
Kazakhstan	23	10
Bahrain	23	3
Zambia	23	0
Togo	23	0
Ethiopia	21	0
Belarus	20	5
Virgin Islands British	20	0
Moldova Republic of	19	3
Cayman Islands	19	0
Guatemala	18	1
Bhutan	17	15
Bermuda	17	4
Uganda	16	0
Turks and Caicos Islands	16	1
Virgin Islands U.S.	16	0
Syrian Arab Republic	15	0
Gibraltar	15	0
Cameroon	15	0

Table I.4 continued from previous page

Country	uristem	uriquery
Bolivia Plurinational State of	15	1
Sint Maarten (Dutch Part)	14	1
Montenegro	14	0
Reunion	13	0
Paraguay	13	0
Sudan	13	4
Senegal	13	0
Botswana	12	0
Brunei Darussalam	12	0
Armenia	11	2
Mongolia	11	2
Namibia	11	0
Myanmar	11	1
Mozambique	10	6
Angola	10	0
Falkland Islands (Malvinas)	9	2
Honduras	9	0
French Guiana	9	1
Nicaragua	8	3
Anguilla	8	0
Haiti	8	1
Guadeloupe	8	0
Benin	7	0
Libya	7	2
Guam	7	0
Gabon	7	1
Rwanda	7	0
Afghanistan	6	1
Gambia	6	2
Malawi	6	0
Maldives	6	0
Cabo Verde	5	2
Yemen	5	1
Suriname	5	1
Madagascar	5	0
Fiji	4	0
Faroe Islands	4	1
Mali	4	2
Aruba	4	1
Liechtenstein	3	0
Curacao	3	0
Guinea	3	0
Somalia	3	0
Seychelles	3	0
Uzbekistan	3	1
Burkina Faso	3	0

Table I.4 continued from previous page

Country	uristem	uriquery
Congo	3	0
Andorra	3	1
Lesotho	3	0
Lao People's Democratic Republic	3	2
Burundi	2	0
Eswatini	2	0
Bonaire Sint Eustatius and Saba	2	0
Timor-Leste	2	0
Northern Mariana Islands	2	0
Kyrgyzstan	2	0
Sierra Leone	2	0
Montserrat	2	0
Holy See	2	0
Mauritania	2	0
Mayotte	2	0
Monaco	2	0
American Samoa	1	0
Niger	1	0
Samoa	1	0
Arizona	1	0
San Marino	1	0
New Caledonia	1	0
Greenland	1	0
French Polynesia	1	0
Marshall Islands	1	0
Djibouti	1	0
Equatorial Guinea	1	0
Guinea-Bissau	1	0
Saint Martin (French Part)	1	0
Tajikistan	1	0
Column Total	**221180**	**34764**

I.3.3 Clustering of Logs

Table I.5: Session starting pages with Items (known items) - over 50 request.

Item URL	Count of Initial session hits
/walker/exhibitions/victorian-treasures/index.aspx	4741
/ism/slavery/africa/capture_sale.aspx	2722
/walker/exhibitions/mucha/index.aspx	2657
/ism/slavery/triangle.aspx	1939
/ism/slavery/europe/liverpool.aspx	1048
/ism/slavery/africa/effect.aspx	998
/ism/slavery/middle_passage/olaudah_equiano.aspx	989
/ism/slavery/europe/abolition.aspx	907
/ism/slavery/middle_passage/index.aspx	855
/wml/collections/antiquities/ancient-egypt/item.aspx	805
/collections/research/lgbt/index.aspx	774
/ism/slavery/slave-stories/index.aspx	756
/ism/slavery/africa/index.aspx	736
/ism/slavery/middle_passage/john_newton.aspx	630
/ism/slavery/americas/abolition_americas.aspx	625
/walker/exhibitions/victorian-treasures/item.aspx	579
/ism/slavery/americas/index.aspx	564
/ism/slavery/europe/index.aspx	551
/walker/exhibitions/henry/walker.aspx	541
/ism/slavery/africa/olaudah_africa.aspx	513
/walker/collections/paintings/19c/item.aspx	492
/ism/slavery/slave-stories/index2.aspx	492
/walker/exhibitions/hockney/pool.aspx	477
/ism/slavery/slave-stories/okechukwu/okechukwu1.aspx	474
/ism/slavery/americas/plantation_life.aspx	377
/wml/collections/antiquities/ancient-egypt/search.aspx	375
/walker/collections/works-on-paper/watermark/explore.aspx	362
/mol/collections/transport/item.aspx	339
/ism/slavery/middle_passage/slave_ships.aspx	331
/wml/collections/index.aspx	320

Table I.5 continued from previous page

Item URL	Count of Initial session hits
/ism/slavery/index.aspx	309
/collections/preraphaelites/item.aspx	297
/walker/exhibitions/gay-fashion-icons/index.aspx	296
/collections/preraphaelites/index.aspx	293
/walker/exhibitions/arts-council-collection/transparency/index.aspx	283
/ism/slavery/slave-stories/oyeladun/oyeladun1.aspx	281
/ism/slavery/slave-stories/kofi/kofi1.aspx	276
/ism/slavery/slave-stories/kwame/kwame1.aspx	276
/ism/slavery/slave-stories/okechukwu/okechukwu3.aspx	263
/ism/slavery/slave-stories/okechukwu/okechukwu2.aspx	260
/collections/research/lgbt/outing-the-past.aspx	256
/collections/lgbt/health-and-wellbeing/item.aspx	255
/walker/collections/paintings/13c-16c/item.aspx	242
/ladylever/exhibitions/picasso/index.aspx	241
/ism/slavery/europe/black_people.aspx	231
/ism/slavery/slave-stories/kofi/kofi2.aspx	230
/wml/exhibitions/thebeatgoeson/thebeatgoesonline/technology/studios/1980s.aspx	225
/ism/slavery/americas/thomas_thistlewood.aspx	225
/maritime/collections/index.aspx	223
/wml/exhibitions/animal-mummies/index.aspx	220
/wml/collections/ethnology/asia/tibet/charles-bell/item.aspx	212
/ism/slavery/americas/olaudah_arrival.aspx	209
/ism/slavery/slave-stories/kwame/kwame2.aspx	194
/wml/collections/antiquities/ancient-egypt/index.aspx	192
/maritime/exhibitions/worldheritagecity/three-graces.aspx	190

Table I.5 continued from previous page

Item URL	Count of Initial session hits
/ism/slavery/slave-stories/okechukwu/okechukwu4.aspx	189
/maritime/collections/lusitania/item.aspx	189
/ladylever/collections/paintings/gallery4/salem.aspx	178
/ism/slavery/slave-stories/kwame/kwame3.aspx	176
/maritime/exhibitions/blitz/blitz.aspx	174
/walker/collections/paintings/20c/item.aspx	172
/wml/collections/antiquities/roman/ince-blundell/index.aspx	171
/ism/slavery/slave-stories/kofi/kofi3.aspx	162
/walker/exhibitions/peterchang/life_in_plastic.aspx	161
/mol/exhibitions/liverpoollives/billy_fury.aspx	158
/ism/slavery/slave-stories/kofi/kofi4.aspx	155
/ism/slavery/slave-stories/okechukwu/okechukwu5.aspx	154
/walker/collections/paintings/20c/index.aspx	153
/ism/slavery/europe/profits.aspx	147
/ism/slavery/slave-stories/oyeladun/oyeladun2.aspx	147
/walker/collections/paintings/index.aspx	145
/ism/slavery/africa/agent_letter.aspx	144
/collections/item.aspx	132
/wml/collections/ethnology/asia/tibet/item.aspx	131
/ism/slavery/slave-stories/okechukwu/okechukwu6.aspx	131
/maritime/exhibitions/cotton/material/cloth.aspx	129
/ism/slavery/slave-stories/kwame/kwame4.aspx	129
/ism/slavery/slave-stories/okechukwu/okechukwu10.aspx	127
/ism/slavery/slave-stories/oyeladun/oyeladun3.aspx	125
/ism/slavery/slave-stories/kofi/kofi7.aspx	124

Table I.5 continued from previous page

Item URL	Count of Initial session hits
/ism/slavery/slave-stories/okechukwu/okechukwu8.aspx	124
/walker/collections/works-on-paper/search.aspx	123
/walker/exhibitions/henry/whitehall.aspx	123
/collections/index.aspx	121
/mol/exhibitions/ontheroad/billyfury.aspx	120
/wml/exhibitions/terracotta-warriors/terracotta-warriors.aspx	119
/walker/collections/paintings/18c/item.aspx	119
/ism/slavery/slave-stories/okechukwu/okechukwu7.aspx	118
/ladylever/collections/wedgwood/item.aspx	116
/ism/exhibitions/broken-lives/jogini.aspx	115
/ism/slavery/slave-stories/okechukwu/okechukwu9.aspx	115
/ism/slavery/slave-stories/kofi/kofi5.aspx	115
/ism/slavery/middle_passage/resistance.aspx	115
/ism/slavery/slave-stories/kofi/kofi6.aspx	114
/wml/collections/horology/detail.aspx	113
/ism/slavery/slave-stories/oyeladun/oyeladun4.aspx	112
/ism/exhibitions/brutal-exposure/alice-seeley-harris.aspx	111
/collections/lgbt/love-and-relationships/queer-relationships/queen-anne-and-sarah-churchill/item.aspx	108
/wml/exhibitions/thebeatgoeson/thebeatgoesonline/technology/studios/multitrack.aspx	106
/mol/collections/archaeology/field/projects/roman/index.aspx	106
/wml/collections/ethnology/asia/tibet/charles-bell/index.aspx	106
/ladylever/collections/chinese/search.aspx	105
/ism/slavery/slave-stories/kofi/kofi8.aspx	104

Table I.5 continued from previous page

Item URL	Count of Initial session hits
/ism/slavery/slave-stories/kwame/kwame5.aspx	104
/ism/slavery/websites.aspx	102
/wml/collections/antiquities/roman/ince-blundell/item.aspx	101
/mol/exhibitions/ken-dodd/biography.aspx	101
/ism/slavery/slave-stories/kwame/kwame6.aspx	100
/ladylever/collections/star-objects/index.aspx	99
/wml/collections/horology/results.aspx	99
/collections/lgbt/index.aspx	99
/walker/exhibitions/arts-council-collection/coming-out/index.aspx	99
/wml/exhibitions/thebeatgoeson/thebeatgoesonline/sounds/diversecitysounds/richandpoor.aspx	98
/collections/preraphaelites/search.aspx	98
/ism/slavery/slave-stories/oyeladun/oyeladun5.aspx	95
/ism/slavery/slave-stories/oyeladun/oyeladun6.aspx	95
/ism/slavery/americas/resistance_freedom.aspx	94
/ism/slavery/slave-stories/kofi/kofi9.aspx	92
/ism/slavery/slave-stories/kofi/kofi11.aspx	92
/sudley/collections/index.aspx	91
/walker/collections/A-Z-of-artists/A-Z-of-artists.aspx	90
/walker/exhibitions/nick-mccann/index.aspx	89
/walker/collections/works-on-paper/item.aspx	89
/wml/exhibitions/snakes/snakes-quiz.aspx	88
/ism/slavery/slave-stories/kofi/kofi10.aspx	88
/maritime/collections/lusitania/search.aspx	88
/walker/collections/works-on-paper/watermark/item.aspx	85
/walker/collections/decorative-art/ceramics/ratoff/search.aspx	84
/ism/slavery/slave-stories/kwame/kwame8.aspx	84

Table I.5 continued from previous page

Item URL	Count of Initial session hits
/wml/exhibitions/thebeatgoeson/thebeatgoesonline/sounds/diversecitysounds/sailors.aspx	81
/maritime/collections/boa/history/may-blitz.aspx	80
/walker/exhibitions/rodin/jamessmith.aspx	79
/ism/slavery/slave-stories/oyeladun/oyeladun7.aspx	78
/maritime/exhibitions/cotton/traders/liverpool.aspx	77
/wml/collections/ethnology/asia/tibet/search.aspx	77
/ism/slavery/slave-stories/kwame/kwame7.aspx	77
/wml/exhibitions/thebeatgoeson/thebeatgoesonline/technology/studios/beatles.aspx	76
/ism/slavery/slave-stories/okechukwu/okechukwu_end.aspx	74
/ism/slavery/booklist.aspx	74
/ism/slavery/americas/sojourner_truth.aspx	71
/ism/slavery/slave-stories/kofi/kofi13.aspx	71
/walker/exhibitions/matisse/poem-transcript-en.aspx	71
/ism/slavery/slave-stories/kwame/kwame10.aspx	70
/wml/collections/zoology/john-gilbert/collection/index.aspx	70
/walker/collections/paintings/17c/item.aspx	70
/walker/exhibitions/arts-council-collection/zoologic/index.aspx	69
/ism/slavery/slave-stories/oyeladun/oyeladun10.aspx	69
/ism/slavery/slave-stories/kwame/kwame9.aspx	69
/wml/collections/horology/database.aspx	69
/wml/collections/antiquities/ivories/item.aspx	68
/wml/exhibitions/terracotta-warriors/index.aspx	68
/ism/slavery/slave-stories/kofi/kofi12.aspx	68

Table I.5 continued from previous page

Item URL	Count of Initial session hits
/ism/slavery/slave-stories/kwame/kwame13.aspx	67
/mol/exhibitions/april-ashley/court-case.aspx	67
/ladylever/collections/star-objects/item.aspx	65
/walker/collections/paintings/13c-16c/index.aspx	65
/ism/slavery/slave-stories/oyeladun/oyeladun11.aspx	65
/ladylever/collections/paintings/gallery2/merlin.aspx	65
/ism/slavery/slave-stories/oyeladun/oyeladun8.aspx	64
/wml/exhibitions/thebeatgoeson/thebeatgoesonline/technology/instruments/electricguitar.aspx	64
/wml/exhibitions/thebeatgoeson/thebeatgoesonline/newsounds/radio/music.aspx	63
/wml/collections/ethnology/asia/tibet/charles-bell/search.aspx	62
/ism/slavery/slave-stories/kofi/kofi14.aspx	61
/ladylever/collections/paintings/gallery4/jeunesse.aspx	60
/ism/slavery/slave-stories/kofi/kofi16.aspx	60
/ism/slavery/slave-stories/oyeladun/oyeladun13.aspx	60
/ism/slavery/slave-stories/oyeladun/oyeladun9.aspx	60
/ism/slavery/slave-stories/kwame/kwame12.aspx	59
/ladylever/collections/paintings/index.aspx	59
/maritime/exhibitions/blitz/munitions.aspx	59
/walker/collections/sculpture/item.aspx	59
/ism/slavery/slave-stories/kwame/kwame11.aspx	58
/ism/slavery/slave-stories/kofi/kofi15.aspx	58
/walker/collections/fashion/index.aspx	58
/ism/slavery/slave-stories/oyeladun/oyeladun17.aspx	57

Table I.5 continued from previous page

Item URL	Count of Initial session hits
/ism/exhibitions/art-of-solidarity/links-with-africa.aspx	57
/walker/exhibitions/romney/biography/emmawilliam.aspx	55
/walker/exhibitions/henry/holbein.aspx	54
/walker/exhibitions/rossetti/works/beauties/sibyllapalmifera.aspx	54
/ism/slavery/slave-stories/oyeladun/oyeladun12.aspx	54
/maritime/collections/boa/history/merchant-navy.aspx	53
/ladylever/collections/paintings/gallery3/girlwithdogs.aspx	52
/ism/slavery/slave-stories/oyeladun/oyeladun16.aspx	52
/walker/exhibitions/rossetti/works/beauties/venusverticordia.aspx	51
/wml/exhibitions/thebeatgoeson/thebeatgoesonline/sites/locating/coffeebars.aspx	51
/ladylever/collections/paintings/gallery3/theweddingmorning.aspx	50
...	...
Total	**58607**

Appendix J

Phase 5

J.1 Information Seeking Cluster Profiles

J.1.1 Set of Textual ISCPs for NML

J.1.1.1 Cluster 1 - First time middle age relaxers

Introduction: This description describes cluster 1. This description represents 25.28% of the total users in the sample. The following values represent what makes up 75% or over for that factor, and is limited to the highest 3 values for each factor. **Demographics:** This group is aged 35-54 (51.47%) and 55-64 (25%). The gender breakdown is female (54.41%), then male (44.12%). They have the employment status of Working (94.12%). And they are located in the World (27.94%), Country (20.59%), Regional (19.12%), Local (17.65%). These users come to the website with 3 (32.35%), 4 (30.88%), 2 (20.59%) out of 5 levels of general cultural heritage knowledge. Their domain knowledge levels are intermediate (54.41%), novice (30.88%). **Goals and Motivations:** They are visiting walker (30.88%), maritime (23.53%), wml (22.06%) sites and are visiting for Personal (44.12%), to pass time (38.24%) reason(s). In order to prepare for Other (36.76%), Museum related (25%), Pre-visit (16.18%) tasks. They are likely to visit the site first (64.71%), annual (22.06%) times a year. **Behaviours:** These users are visiting on "Desktop browser" (58.82%) and "Mobile" (41.18%) device(s). and on iOS 10.2 (27.94%), Linux (16.18%), Mac OS X 10.11 (8.82%) platforms. They prefer to find their information on the site using browse (58.82%) and search (41.18%). When searching, these users will try

a couple of times before giving up (69.12%) and a smaller portion of this group will persist till find (26.47%). They like their content in the form(s) of pictures-text (42.65%) and text-scan (33.82%). They are likely to look at just the one (57.35%) and 2-10 (36.76%) pages in a visit. **Significant Characteristics:** The characteristics that are most significant for this group are: visit-reason (p-value = 7.40940098842952e-12); visit-purpose (p-value = 3.4068837832098e-10); location (p-value = 5.76767731463641e-07); age (p-value = 0.0233076290348998); employment (p-value = 2.86447389572056e-10).

J.1.1.2 Cluster 2 - Local working pre-visit preparers

Introduction: This description describes cluster 2. This description represents 65.43% of the total users in the sample. The following values represent what makes up 75% or over for that factor, and is limited to the highest 3 values for each factor. **Demographics:** This group is aged 35-54 (36.36%), 55-64 (23.86%), 65+ (21.02%). The gender breakdown is female (69.89%), then male (28.41%). They have the employment status of Working (48.3%) and Retired (28.41%). And they are located in the Local (47.73%), Regional (26.7%), Country (13.64%). These users come to the website with 3 (39.2%), 4 (31.25%), 2 (13.64%) out of 5 levels of general cultural heritage knowledge. Their domain knowledge levels are intermediate (59.66%), novice (28.98%). **Goals and Motivations:** They are visiting walker (36.36%), wml (26.7%), maritime (17.61%) sites and are visiting for Personal (88.64%) reason(s). In order to prepare for Pre-visit (68.18%), Museum related (13.07%) tasks. They are likely to visit the site first (43.75%), annual (37.5%) times a year. **Behaviours:** These users are visiting on "Desktop browser" (50%) and "Mobile" (50%) device(s). and on iOS 10.2 (31.25%), Linux (20.45%), iOS 9.3 (9.66%) platforms. They prefer to find their information on the site using browse (65.91%) and search (34.09%). When searching, these users will try a couple of times before giving up (58.52%) and a smaller portion of this group will persist till find (33.52%). They like their content in the form(s) of text-scan (52.84%) and read-all (23.86%). They are likely to look at 2-10 (52.84%) and just the one (43.75%) pages in a visit. **Significant Characteristics:** The characteristics that are most significant for this

group are: visit-reason (p-value = 3.20567849938814e-17); visit-purpose (p-value = 8.63658696931262e-13); frequency (p-value = 0.0187918599882092); location (p-value = 1.74866144605449e-07); age (p-value = 0.00955497360790624); employment (p-value = 1.24737283818522e-08).

J.1.1.3 Cluseter 3 - Students

Introduction: This description describes cluster 3. This description represents 3.72% of the total users in the sample. The following values represent what makes up 75% or over for that factor, and is limited to the highest 3 values for each factor. **Demographics:** This group is aged 18-34 (70%) and 35-54 (30%). The gender breakdown is female (50%), then male (50%). They have the employment status of Student (70%) or Not working (20%). And they are located in the Local (40%), Country (30%), World (20%). These users come to the website with 2 (30%), 3 (30%), 5 (30%) out of 5 levels of general cultural heritage knowledge. Their domain knowledge levels are novice (40%), intermediate (40%). **Goals and Motivations:** They are visiting walker (40%), ism (30%), ladylever (10%) sites and are visiting for study (100%) reason (s). In order to prepare for Museum related (30%), Other (30%), known item (20%) tasks. They are likely to visit the site first (80%) times a year. **Behaviours:** These users are visiting on "Desktop browser" (80%) device (s). and on Linux (30%), iOS 10.2 (20%), Mac OS X 10.11 (20%) platforms. They prefer to find their information on the site using search (60%) and browse (40%). When searching, these users will try a couple of times before giving up (50%) and a smaller portion of this group will persist till find (40%). They like their content in the form (s) of text-scan (50%) and read-all (40%). They are likely to look at just the one (50%) and 2-10 (30%) pages in a visit. **Significant Characteristics:** The characteristics that are most significant for this group are: visit-reason (p-value = 3.16819421217618e-26); age (p-value = 0.000762815418791977); employment (p-value = 7.28162999515059e-15).

J.1.2 Visual ISCPs example

J.2 Qualitative Validation of Information Seeking Cluster Profiles

J.2.1 Interview script - questions

Final interview script

David Walsh

November 2023

1 Interview script

The following is the final set of interview questions after the pilot study, and amendments have been made.

PRESS RECORD ON TEAMS

2 Consent speech:

Please state your full name and then respond with "I agree" or "I disagree" after each statement is read out.

1. I confirm that I have read and understood the information sheet (previously emailed) for the study.
2. I have had the opportunity to ask questions and have received satisfactory answers.
3. I understand that my participation is voluntary and that I am free to withdraw at any time without giving a reason.
4. I agree to take part in this research.
5. I agree for my data to be used for the purposes of this research and any future related research/publications.
6. I understand that my data will be kept confidential and stored securely.
7. I consent to this interview being audio-recorded for the purposes of data collection and analysis.

Before starting, ensure participants are comfortable speaking openly and know that they won't hurt my feelings by potentially saying something negative.

3 Questions:

1. **Background and Role:**
   - What is your current role/career?
   - What does a standard day/week/design phase look like for you?
   - Please briefly describe your typical workflow for a project.
   - What type of audience do you usually work with/for?
   - When starting a new project, how well can you define your audience? (Follow up with "tell me more about X and why" ...)
2. **Experience with Personas:**
   - Do you use Personas as part of your projects? If not, why not and what alternatives do you use to understand the audience, e.g.(user stories)? Is this a personal or business-led choice?
   - What details in a Persona or information about the audience/user are most helpful to you when designing/developing? Why?
   - Can you discuss how the created Persona/user profiles are used within the UX workflow? (Gather details on any unique features.)
3. **Methods of creating Personas:**
   - How and when do you or your team usually create personas?
   - What types of data are they usually based on?
   - What are your thoughts on Data-driven Personas?
4. **Introduction to ISCP and Comparison with Personas:**

   Participants will have had time to explore and play with the demo ISCP system and see the results from: `https://www.figma.com/proto/cCQzqUfpive5CdrbqybnDP/ISCPs?type=design&node-id=1-1291&t=w3FeqMA0Sawt8NMo-0&scaling=scale-down&page-id=0%3A1&starting-point-node-id=1%3A1291`
   - I am exploring a new early survey-based data-driven type of user profiling method, a sort of pre-persona, which I am calling Information Seeking Clustering Profiles (ISCP). In simple terms, it is a way to profile users really early on in a project and UX process based on survey results, capturing users information-seeking behaviours. I provided you with a link to see how a demo of the tool could work and what the end resulting personas style ISCPS would look like and the types of information they would hold.
   - What are your thoughts on the ISCPs?
5. **Potential Value and Application of ISCP:**
   - Are ISCPs something you feel like you could use? Why/why not?

- Is there anything you would add, change, or drop from the ISCP method or end profiles to make it more relevant to your work?

6. **Feedback and Future Considerations:**

   - Beyond Personas, are there other methods or tools you currently use or are exploring to understand users/audiences better?
   - With evolving tools and methods, how do you see the future of user profiling in UX design?

Finally, Do you think I have missed anything or would like to add?

If you know anyone else you think would be good for me to interview, please email me their details, as I would not want their details to be in this data?

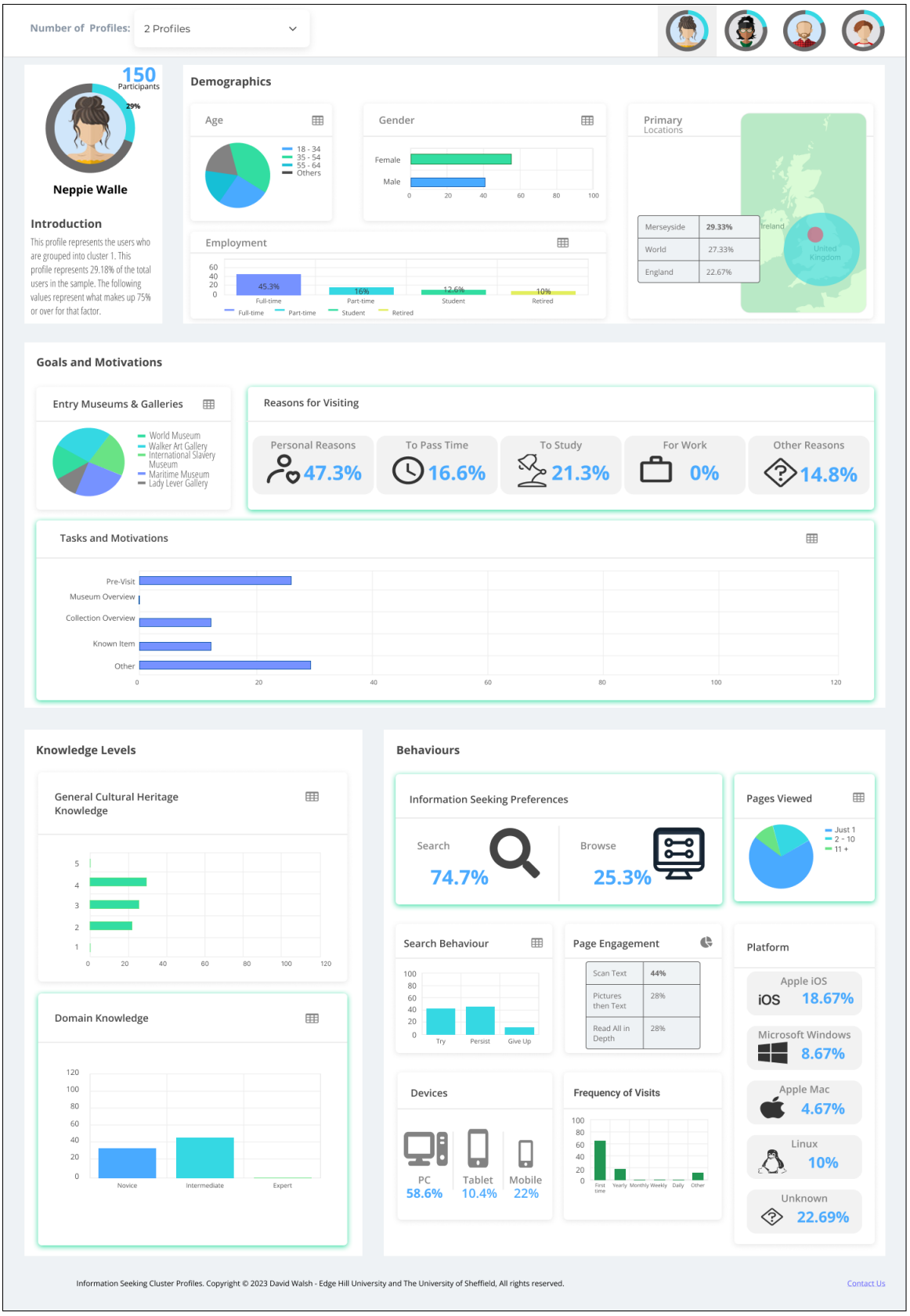


Figure J.1: An example of a persona style Information Seeking Cluster Profile for cluster 1 (Larger version)

J.2.2 Interview transcripts

J.2.2.1 Participant 1

Interviewer
The first set of questions I've got are around your background and role, so can you just tell us what you tell me what your current role is?

Participant 1
My current role is a UX designer.

That's what I've been employed as, but I.

And more of a UX researcher than a designer.

Yeah.

Interviewer
OK.

And So what does a sort of standard day week or maybe design phase look like for you then?

Participant 1
Is me sat on Figma.

Trying to make sense of designs that I'm creating for like an ancient system, a legacy system that needs to be brought into alignment with the modern.

Aesthetic of the web.

Interviewer
OK.

Participant 1
So that's a lot of my time currently.

Yeah.

Interviewer
And so how does that fit in the normal design phase that you've got that you, your team works with?

Participant 1
Well.

We've just recently had a new manager, so this is all up in the air.

So we'll normally have like, we'll try and work to our jobs.

Really trying and university concepts it's difficult.

But yeah, we're having.

Sprint planning every couple of weeks and we'll get a sign tasks to design certain aspects of a system, for example.

And.

That's it. Really. I can't really go too much in depth. It's just still in flux.

Interviewer
And when you get assigned.

Those tasks that you do 'cause you were saying that your you your role is more user researcher?

Are them are them tasks more going and naturally talking to the users? Getting some background on them and then coming back and making sense of that? Or is it something slightly different?

Participant 1
It depends on the project. So for like for the one that I'm working on currently where I'm just designing.

Creating designs for a system the user the research that I do is just it's kind of backing up my choices for like using certain elements using certain designs, but in the past I've done a lot of like.

On the field, use talking to users so.

We we were designing an app for the Museum, so I spent a lot of time in there, like observing users, talking to users, getting a feel for what they want from an app.

And the pro it's vibes. Letting them interact with high vitality, high fidelity prototypes that we created using Figma.

So yeah, it just depends on the project.

And I imagine going up, going the future, it will depend on the phase, the projects as well the intensity of certain talking to users and testing and that.

Yeah.

Interviewer
Is there a typical type of audience that you usually work with?

Participant 1
And.

What in terms of getting user feedback or?

Interviewer
Yeah. And all of that types of projects that you actually mentioned that you worked on the museum, the, the university's museum site, there was a was there a typical type of user and is that typical of all the other projects that you work on?

Participant 1
Museum one is a bit of an outlier because it doesn't follow the usual.

Audience that we'd have for for the university, for a university focused project.

So the the types that we experience in the museum.

Were quite a bit different from what we'd normally create for the university website.

So.

Yeah. So.

Interviewer
And then what? What sort of way were the different? Can you describe that?

Participant 1
Yeah, I don't wanna be offensive to the students, but they seem more engaged. They were. They were actually in the museum.

To to pass time to interact with the objects there.

For us, when you're actually talking to students.

It seems like you're fighting to get their attention because they have to be somewhere else that they're busy talking to their friends, but it also depends on the location where you're talking to them as well.

It'll be like outside the the cafeteria, however, and they're just trying to get from A to B and they don't want to interact with you. But at the museum it's a lot different. People will take the time to talk to you and give you a detailed feedback as to what their experience is.

Interviewer
OK.

Participant 1
In the museum. So yeah.

Interviewer
Cool. So it sounds like you've worked on.

A number of different projects now within your current role so.

When you're starting a new project.

Participant 1
Yeah.

Interviewer
How do you and how well can you sort of define the audience? How? How early is that when you define that audience as well? Is it really early or is it a look, you know, quite a bit of research?

Participant 1
Also, [museum], we're fortunate that we managed, we managed to get out there early before we did any designs we I did a lot of a lot of observation, a lot of talking to people there and.

For the other projects that we've done, it's not like that.

Because of time constraints, whatever.

The lack of people available to work to develop things, it's just a case of develop it, create some designs and then get it out there just to validate what you've done.

At the.

It's it's gonna change soon. What? At the moment, it's just it's it's not. It's not great. But for the [museum] stuff we did, we took a completely different approach to it because we had time.

And I think we did it really well. We've got some decent feedback.

Interviewer
Yeah. So, so you mentioned that the your current process on most of the projects isn't great. What would you mean by that? Is that?

But what did you mean by that?

Participant 1
It's it's not very user centric at the moment.

They we claim to be poor.

It. Yeah, it's.

But we're still in in the in the midst of creating creating some proto personas.

To put a bit of a context to it, within the university, there's.

Ux teams so marketing have a UX team and our with about three people with the head of UX for UX designers and then IT services, how do you? How's the UX team? So it seems like we're all doing different things. We're not really working together that well currently.

So yeah, it's just a bit of a model at the moment.

Interviewer
OK.

Participant 1
Yeah.

Yeah.

Interviewer
So you sort of mentioned a a proto persona there, which sort of leads me into my next set of questions, which is all about your experience with using personas.

Participant 1
Yeah.

Interviewer
So.

First of all, do you use personas as part of your projects?

Participant 1
Yeah.

And.

Yeah.

Interviewer
This is, you know, most of the time. Do you use them or do you aim to use them, should we say?

Participant 1
I try to use them. I try to get my team members to to use them.

With EG now we created about I think.

Five different.

Archetypes based on the people that we talk to and some of the dates that we got from like feedback from the exhibitions and talking to people on reception like in terms of numbers of visitors and stuff.

Demographic data, stuff like that.

So yeah, so we got a lot from that.

We tried to use them to like inform some of the designs that we created for for the app.

But yeah, but for the universities, for the university stuff.

We don't have.

Proper personas yet?

We are trying to get them through so we we created.

Postal facility is working with the the other UX team.

But that's just currently being put on the hold because of, like, internal politics.

Interviewer
OK.

Participant 1
But we're all in alignment with what what we created. It was all pretty good stuff, but it wasn't data-driven. It's all like anecdotal. It's all about what students have been telling us during, like on the field stuff and.

Yeah. So it's nothing like scientific or.

Compelling, really.

Interviewer
OK.

That sounds good. You know, it sounds like it's very personal driven rather than business driven though in in the favour of using personas at that point.

So when you're working with the personas, what types of details would you normally have in a persona, or what type of information about the audience or user are most helpful?

Participant 1
Yeah.

Interviewer
To you when you're designing and developing.

Participant 1
Well.

But for me, it's a basic demographic data, isn't it? So it's like age, gender, country of origin, occupation, perhaps proficiency in in IT skills perhaps?

I suppose these depend on on the actual project that you're working on. I suppose there may be some kind of like pull out quote that summarises what the user's about what the main gist of this persona is.

Because the personas have to be distinct, don't they? There's no point having.

Six different personas, and they all kind of say the same thing.

So it has to be distinct I think. And then after that I'd have things like motivations talking about what motivates them to use the system or site or whatever.

And then I'll probably have.

Goals. Obviously, yeah. What? What? They want to get from.

The system.

Behaviours perhaps?

This is based.

Yeah.

Interviewer
Is that common? Is it? Is that common in the asonas that you've worked with, that you've got all of these, or is that just something that you prefer rather than being the standard?

Participant 1
Yeah, it's this this well, we did our own little thing. Our UX team and the other UX team did their thing. It went when we brought them together. We had the same things. So it it made sense to to.

Interviewer
No.

Participant 1
Yeah, and yeah. So we have like.

Yeah, we got like a bio. We got goals, we got a background of the user as well talking about what the user is, what the persona's like. So their behaviours based on stuff that students have been telling us. And then we've got paid.

Interviewer
So what? What does that mean? What what? What do you mean by behaviours in what they're telling you?

Participant 1
So like so, if if we've been out on the field and the students have been telling us that, that they're having difficulty with with authentication like DUO sign in and stuff like that.

Then that will be reflected in the behaviour section.

And just like things that they do on campus.

Interviewer
Some more common issues rather than.

Participant 1
Yeah, yeah.

Interviewer
Preference of behaviour.

Participant 1
And.

Interviewer
OK.

Participant 1
It's it's all about the habits as well. The things that they do when they're on campus. So like they engage in academics, things that participate in campus events.

Things like that. Really. Yeah. They're just generic things about the about the persona that they would do when they're here on campus or at home. And then.

Interviewer
And generally, is all of that driven by?

Participant 1
Yeah.

Interviewer
Your observations and your interviews, you know, just general.

Positive data essentially just asking them.

Participant 1
Yeah, it's just currently it's all just based on what what we hear and see.

Interviewer
OK.

Participant 1
There's no like quantitative data.

That we that we have currently that's that will change soon because we are buying software to let us.

Look at the mobile app use in particular like scroll depth and all that kind of stuff clicks and yeah. So currently we don't have that for the mobile app and there's thousands of people using it.

So yeah, it's a massive like oversight.

So yeah, and then after that, we'll have pain points, things.

Students are finding difficult.

Interviewer
OK.

Participant 1
And so yeah, that's that's a typical sections don't have and then I'd have just a generic picture as well, just that encapsulates that persona type.

And the name and a little tagline as well. So like.

The engaged learner, the international learner for like an international student, for example.

And I think that's it.

Interviewer
And you find them useful.

Participant 1
Yeah, because you can just look at that and it's it just summarises it all. You know that that's the persona for the current student. That's the persona for the international student. So that that little strap line just captures everything, the persona and more.

Interviewer
Can you discuss how the creative persona or the user profiles that you work with are used within the UX workflow that you, you, you, you adopt or you work with?

Participant 1
Very loosely currently and I try to use the personas to inform the designs that I do.

I mean for for our latest prototype that we've done, we've we've obviously created like a student with like accessible needs. So this this one will be into will for like going forward when we're trying to.

We're also creating a design system at the moment as well, so that Persona will be fundamental in in informing the.

The design, if like developments and.

In like considering the the, the needs basically.

Interviewer
Yeah.

So.

Is it? But it sounds a bit you. You don't seem to have a.

A sort of standard workflow for gathering information on personas or a certain time frame or something within the UX process.

Participant 1
That's all, no.

Interviewer
You know you've mentioned in the museum you did you tried to do that as early as possible?

Participant 1
Yeah.

Interviewer
But on the other projects, it seems like it's a bit.

Ad hoc should we say?

Is it? Is that more reflective of of what you're what you think and what you're trying to say? Is that correct or?

Participant 1
At all. I mean it's we should be getting out there as early as possible to understand who the user is.

And the reason and we have the luxury of time we had, we didn't we didn't have a deadline for the app we we just told to get out there just speak to people and find out what they wanted and we did and we got really good feedback.

But with with.

The [university] stuff, the it's just. Everything's all about time and resources and.

And trying to get the right people to talk to as well, it's always difficult.

Market [university b] we had we always seem to use them student.

Student the whole helpers, advisors, people are paid by the university that don't give an honest.

And entirely honest.

Feedback to.

It's all like rose, tinted and very like.

So.

Interviewer
You know, experience some of the personas that are qualitative.

Driven at.

Participant 1
Both, yeah.

Yeah.

Interviewer
Bias because these these people were paid to partake essentially, or they're already they're people who they're the same type of person who would always volunteer or, you know, get engaged, whereas that's not truly representative of the full audience.

Participant 1
Exactly. Yeah. The the ones we we got for for the [museum] were a lot more.

The feedback we got was was honest and real.

But there are certain pockets in the university where you can go for honest feedback, so if you if you get, if you put designs to an audience that have domain expertise.

Or if they're within that subject, like computing students, if you put the designs forward to them, they will happily give you honest feedback.

Interviewer
OK.

Participant 1
Good. Yeah, so.

Interviewer
OK, cool.

So you you already mentioned.

That your personas when data-driven, so I'm assuming that you're you're aware of what data-driven personas are.

What are your thoughts are on data-driven personas?

Participant 1
Well, it's it's important to have them, isn't it? Because if you're just using proto personas and it's not an accurate reflection of what?

What of what the student is really, it's just a it's just a guess. So it's not a guess, it's it's, it's, it's informed, but you need to have.

Positive. The quantitative the behavioural altogether.

In order to get well-rounded persona and currently we we just have the Pro so Persona and it's not.

This is not.

Well informed.

Having the data behind it's important.

Interviewer
OK.

So.

At this point.

I want to introduce.

Part of my PhD that I've been working on.

And.

The the work that I've been doing is around studying the museum users at [Museum].

My hope is that what I'm trying to do here is gather as much quantitative information using a survey really early on in a project. I mean like right at the start of the UX process and get as much data out of the users as possible and then.

Because the teams in museums aren't necessarily.

Participant 1
Yeah.

Interviewer
Skilled in doing cluster analysis or a data analysis in some cases, then I've I've built a tool that would automatically analyse data and produce some prototypes.

Participant 1
Yes.

Interviewer
What I'm going to show you or share in a second is a screen mock up inside of figma using the actual data. I just didn't have enough time to put together the the full system, although the results are from that clustered system.

So basically what you do is you run your survey and I've got a template of the survey that people could run, gather their own data, upload it to the system, and then it will automatically do the custom in and it allows you then to explore.

The potential profiles of the users that come out of that clustering result now, these aren't true personas.

K 'cause, there's no qualitative data there behind them, but the idea is to give you a starting point within that UX process.

To get to know your users to start to think and learn about some of the habits that you some of the stuff that you've mentioned earlier to be honest.

So you're OK if I just share my screen with you?

Participant 1
Yeah. Yeah, that's right.

Interviewer
Can you see that?

Participant 1
Yep, information seeking cluster profiles, yeah.

Interviewer
That's the whole.

OK, so this is the start page of the tool. I'm just going to use the sample data for now.

And when we first go in, it's going to provide us with a little bit of information just to explain how the system actually works in terms of the drop downs and showing the different colours, the default.

Number that it provides is based on the most optimal number within the clustering that gets returned.

And So what it does is it it it?

If it catches up with itself, it's a bit slow when it's recording.

Participant 1
Yeah.

Interviewer
So in this case the default was two profiles, was was one of the optimal numbers of clusters that produce within the setting, and we start by showing the number of participate participants that are in this cluster or in this group.

And then little bit of description in the demographic data which I'm capturing, age, gender, the employment details and then the primary location.

The other types of information that I'm gathering is things on goals and motivations.

So things like what museum they entered from, in this case national museums, has a a series of museums, so it's it's worth identifying the type of museum that it's coming from.

There are reason for visiting.

In this case, we can see that the personal reasons is the highest.

Now one thing I should say at this point is that not all of these.

Areas breakdown into 100%.

Participant 1
Yeah.

Interviewer
Because what I've done in rather than take the.

The centroid or the centre object or the centre person within the cluster as a representative of that cluster, I wanted to represent the whole group in that cluster, so these are sometimes built up of what makes up 75% of that. That group in each column and each column is like so reason for visiting is 1 column.

Umm. And so these might add up to just 75% or slightly less than 100 to to whatever is 75% are over.

So we've got tasks and motivations as well. So looking at the different types of tasks that they might come with.

Participant 1
Yeah.

Interviewer
Then we've got knowledge levels which is split into two sections. So we've got the general knowledge, skills and then we've got the domain knowledge skill which is the actual the more in depth skill or sorry the more in depth knowledge on the subject that they came to visit for at that time. And then we've got the behaviours where obviously we're looking at the more obvious things like.

Device type the platforms that they're using, but we're also looking for things like their information seeking. So do they prefer to browse or search?

How many pages have they viewed on that current visit?

And then also whether they how they engage with the?

Participant 1
Yeah.

Interviewer
The content on the pages or how they perceive themselves to engage with that content. So in this case you can see that the majority prefer to scan the text on the page as opposed to looking at the pictures and then the text.

And then we've also got search behaviour, so this is if they were searching, would they?

Just try or would they try and persist? Or would they give up and what we can see is the majority here would either just try once and walk away.

Or they would try and try and try again.

Participant 1
Yeah.

Interviewer
So that's sort of 1 profile. And then obviously if we click through to the second profile, you can see that the values change on this.

Still come for personal reasons, but this is really high. It's 9590.5% here.

The location has changed slightly. We've got more full time people in this one as well. These are a lot closer to Merseyside in this group than the other group were.

And then obviously you see the the general knowledge skills are a lot higher.

These prefer to browse as opposed to the 5050 or most that we have on the other one.

Yeah, they they will try once and then give up on a search which.

It is the same backing up the browse option there and then frequency. These will come first time or yearly.

Participant 1
Yeah.

Interviewer
Umm.

And predominantly use APC. So there was difference a clear differences between these two groups.

Now obviously with clustering it's going to give you some results anyway.

I lost the, you know, the the downside to customer is that it it cannot, it will always come up with groups no matter what, and there will be some overlap. And so when we look at these, you see that I've colour coded them so that two profiles is green. That means it's the separation between the groups is quite good.

Three is OK4 is also separated and it's quite good, but five not so good. Probably not worth looking at if we just look at what happens when we go to a four clusters.

So here we open.

We've reclusted it gives us that information again, and now we've got our four different profiles.

Participant 1
Yeah.

Interviewer
The first one you can see that we're working with a lot less participants 150 now compared to 200 and something on the last one.

And again, this time predominantly full time users.

Mostly female, although there's quite a lot of males in this cluster.

Revisit is the biggest known reason that they're coming.

Pretty even on in the middling of the General General cultural harbitude knowledge and that sort of backs up that they're novice or intimate on the domain knowledge as well.

When we go to cluster 2.

Again, we're on about 126, so it's quite even between the four groups. I think looking at the dates this time, world is just slightly the most. So these are people who are coming visiting from outside of the UK.

Again, full time and student is the sort of most there, but retired is is creeping in.

Looking for pre visit or looking for museum overview. So obviously looking people come in cultural heritage knowledge is a little bit higher but the domain knowledge still is mostly in that intimate range.

Again, try which maps with the browse is the preference with that group, they will scan the text.

Nearly exclusive on the PC usage.

Most of them for time users as well so.

We're still quite cold.

And then coming into the other two groups, they're very, very similar, so.

Hey, you can see that we're looking really local at Merseyside, but we're predominantly male in this cluster.

Participant 1
Yeah.

Interviewer
And predominantly retired people, so mostly retired males and.

Everything sort of fits.

Aligns you know pre visit the local.

They're gonna look at 2 to 10 pages, which is a little bit different to the others which showed one or.

Virtually not a lot.

Yeah, that's the.

Participant 1
OK. Yeah.

Interviewer
The sort of idea behind the profiling is that we can explore these different profiles in the groups and and see the different clusters.

So.

Participant 1
Yeah.

Interviewer
That they're not quite personas, you know. They've not got that qualitative data background and they've not got that information in there. So I'm calling the, This is why I'm calling them information seeking cluster profiles because the surveys on their information seeking habits would pull in a little bit of extra demographic stuff. But then we let the cluster in, group them and decide them and form these data-driven.

Proto personas, if you like at that point because we've not got the.

Participant 1
Yeah.

Interviewer
The qualitative data to back it up, the idea is that these don't replace personas.

But these are your starting point for you to now go. OK, now I know who to go and talk to or ask about all that qualitative stuff and target you a little bit better into going and seeking them more, even if you don't use personas or you use user stories, you've now got people to go on target.

So what are your thoughts on?

On the iscps and where they fit within the UX process.

Participant 1
Think it's extremely useful? It would be really handy to see everything like this just in a in a nutshell.

And.

So I suppose it's my question is to you is how.

How many institutions will have rich data sets like this to to use and to creating these personas?

Yeah.

Interviewer

Well, I think the that's the point is that I I can provide the template of the survey. And so right at the beginning of your project, you would run that survey for two weeks.

Participant 1
Oh, OK.

Yeah.

Interviewer
Or until you've got enough results and then you just process your own survey results at that point. So you'd run it, gather that data.

However, you know you, your institution, or your your team will see fit.

Participant 1
Yeah.

Interviewer
This data was just gathered by leaving the survey alive on the museum website and asking volunteers, but obviously in your case it would be quite similar if you thought think about what you did for the Muse, the University Museum, it'd be similar sort of process. Put it on the the app or the website and and run it that way.

Participant 1
OK.

It's interesting, yeah.

OK.

OK.

Interviewer
So do you think do you think having that in there and having it at that?

Having this data really early on is it would be useful.

Participant 1
Oh, absolutely, yeah, it'll be complementary. I could use something like this to get an idea.

Of what the users are, and then I'll probably go into the museum then after that, just to observe people, to see if they match.

These Buster profiles that have been pulled out.

And.

Yeah, I think it's a good idea.

Interviewer
Yeah.

Participant 1
The.

Just just go.

Yeah.

Interviewer
Bill, is there anything that you think you would add change or drop from the ICPS that you're seeing or the method?

Participant 1
I've definitely add some kind of like quick.

Description of what the cluster profile is. I know you've got like you've got a name like Tracy Tucker there and an introduction, but to me, I want to see something quick that that tells me what this this profile is about in a nutshell. So this person is probably somebody who.

See the like the pre visit things high and the person will visit here for personal reasons. What are those personal reasons?

Are there certain things within that that we could use to create like an archetype, a title for this persona? Like is it somebody who's wanting to go to the museum just just to chill out and relax? So so this profile could be the relaxer or something like that.

Interviewer
Yeah. So the different thing I'd I forgot to point out is when the clustering happens, you can identify which of the columns in that cluster.

Are.

Participant 1
Yeah.

Yeah.

Interviewer
The most important or the most significant, shall we say, and I've highlighted them in here with the green borders. So within this profile here you can see that the agenda, the visit, reason.

The general Cultural Heritage knowledge and the device type are the most important. So do you think just pulling out that information and listing it in in the quick description might be useful?

And just pull it out and say, you know, you know this.

The the gender type or the the reason for visiting is most important. Or would you actually want that extra level of data in?

Participant 1
Then.

Interviewer
What type of personal reason? Cos that's not something that I actually asked. I just asked which what reason they came for and personal reasons was one of the options.

Yeah, and.

Participant 1
Yeah. The only reason why I went into that in such a manual kind of way is because that for our app, people were saying so people in the museum were saying that they they came to the museum just to to relax or to, like, spend a few minutes after work or during a lunch break. And and we wanted to incorporate a feature into the app so that for me, that was why I wanted to know more about the person.

More than anything.

Interviewer
Yeah. So I I do. I did offer that to pass time as an option in there. So there is a.

An option in there just to pass time as opposed to other personal reasons.

Participant 1
Yeah, yeah.

Interviewer
But yeah, I I get where you're coming from.

Participant 1
Yeah.

Interviewer
But would that not be something that you could then explore later on in your qualitative?

Participant 1
Yeah, I suppose, yeah. Drill down a bit more, yeah.

But I think it's also good that you've got all the technical stuff as well, seeing what platform people are using for all sorts important developing mobile apps need to know what.

What people are using and what features they're using dark mode, light mode, whatever.

Yeah. So that's that's good to see.

And also the search behaviour as well.

How persistent they are?

With the search.

That will inform the design in our app. How?

Ha.

So it's save somebody who's not that willing to search. Then we want to make an interface that it gives.

It's all generous interface concepts, isn't it? That you want to show everything?

That the user may find interesting without having them to like search for it manually.

So yeah, that's something that to me personally would be interesting.

Yeah, yeah.

Interviewer
OK.

Cool. I'm gonna stop sharing my screen there and then I've just got a couple more questions just to.

Finish off if you don't mind.

Beyond personas, then, are there any other methods or tools that you currently use or are exploring in order to understand your users or the audience better?

Participant 1
Well, we don't focus groups like talking to people in the university.

That's it from like a quality qualitative kind of thing. So people looking to get.

Different tools on the app to yeah, just to see what people are doing behaviourally with the app to get a bit more information out of it. Really. And.

Interviewer
OK.

Participant 1
Yeah, that's it really. It's. But like I say, it's still just.

In the next few months, that should be in place.

Yeah, yeah.

Interviewer

Right. And then tools more like glasshouse or hot jar or something that that just monitors the usage. Or is it just more?

More like Google Analytics.

Participant 1
It's both, it's analytics and.

Like hot jar? Yeah, like scroll depth and all that kind of thing. But within our apps as well, we have like feedback surveys as well that people will hopefully interact with and give us a bit more in depth.

Feedback.

Interviewer
OK. And my last question is with the evolving tools and methods that.

Are always happening, especially within the UX field. How do you see the future of user profiling UX design?

Participant 1
Mm hmm.

God, do I really want to be, say, the the obvious? Like the AI aspect of it?

We've used.

AIA lot for to help us with the development of pro personas because it kind.

It just skims like the Internet and just gives us an idea of what.

A set of personas could be.

A university. And we could and we use them to kind of.

Help us a little bit it.

It's not something we that we take word for word. It just gives us an idea. So AI is like a big it's it's a good help.

Answer.

Interviewer
And do you validate the the personas that you get from the AI?

Participant 1
We don't use the AI personas till we don't use them to that extent. It's not as if I like copy paste from from ChatGPT. It's not like that at all. We just use it to get ideas. But these proto personas that we come up with they will be.

We'll be talking to domain experts.

To further make them more robust then after that, then we'll be taken into the to the user just just to validate them and make sure that the.

The they are what they they are supposed to be.

Interviewer
OK.

Lastly, is there anything that you think I've missed? The we've not talked about that you would want to raise or have we covered everything?

Participant 1
I think it's alright. I just think that certain museums, certain institutions are.

The more receptive to doing these kind of things, the [employer] are very old fashioned and very traditional, they don't even have.

Counts of people that have come into the building.

Which I find miraculous. How are they getting funding?

So yeah, I'm just a bit suspect about how certain institutions.

Would wanna do something like this.

How interested are they in the user?

Interviewer
If they've got a UX team and you would, you would suppose that they're they are interested in in the user book.

Participant 1
Course, yeah, yeah.

Interviewer
Yeah, I think I think my tool or the idea about my tool is more about.

When you're we've got a UX team. What? Well, when you've got a a developed web team, maybe not even a full UX team who don't have the analytical or.

Cluster, you know computer science skills.

That this allows them to just fill that gap and get to something that they feel more comfortable with.

Participant 1
Yeah. OK.

That's cool. That's good idea.

What?

Interviewer
OK. Well, thank you very much for your time. I'm gonna stop the recording there, if you don't mind or stop the transcription.

J.2.2.2 Participant 2

Interviewer
So could you just describe what your current role or career or whatever your classic is?

Participant 2
I currently work as a UX designer for [employer]. I work in the funeral care section and previously I worked in various different organisations including [Company], [Museum] and [Company].

Interviewer
OK. Thanks.

And can you? What does your standard day week maybe design phase sort of look like for you?

Participant 2
Does vary a bit, but often I am either a facilitating workshops, having conversations about either a new piece of work, or an existing piece of work. I will be perhaps going into Figma doing some design I may be observing some research and taking notes, maybe talking about maybe generating ideas, maybe doing a design quit, or looking at competitive sites.

Or sometimes doing some analysis of data.

But yeah, that that it does vary a bit depending on what stage of the design process we're at with each individual piece of work.

Interviewer
OK, awesome. Cool.

Talking about that design, the you know the stages in the project, could you briefly describe the typical workflow through a project?

Participant 2
Sure. So.

It depends where the idea has the work stream has come from, whether it's been generated by someone identifying an opportunity from the data, or whether it's come from like a business demand. But usually what happens at the start of a piece of work is we say we kind of define what it is that we're trying to do, trying to frame the problem, and then we'll go through a fairly classic kind of process.

Depending on time scales as well.

And depending on the nature of the work, but we'll usually kick off with a bit of a workshop to get kind of stakeholders aligned, understand the scope of it, understand the constraints and the risks, understand what do we already know.

What do we not know? And then we might do a bit more. We might. Sometimes we'll do user research and do some analysis of the data to learn more about kind of what we can find out about that particular related to that particular work stream. Then we'll often start to come up with ideas to solve that problem.

Then we will often create a prototype or a design that we will then put in front of users and get feedback on or.

And then we'll iterate it, and then we'll usually go into engineering bits of it, and then beginning to release it into production. And then once it's in production, we will often be measuring and monitoring to see what impact that particular piece of design has had.

So that's very typical, but yeah, it does. It does depend a little bit like we have done discoveries, which is just a pure piece of discovery.

And we have done something like updating price changes on the website where it's like just give me a design that shows this new price on it. And I'm like there you go. So it does depend a bit on what The thing is, but yeah.

Interviewer
So you mentioned that you've worked at a number of different types of companies.

Oh, can you sort of describe the type of audience?

That is the usually what you work with at the moment.

Participant 2
At the moment, because I'm in funeral care, we are and we're looking working on the website a lot. It's looking at kind of people who are organising funeral or taking out a funeral plan. So there are like we do have some idea of who those people are like we have, we have a kind of demographics of like the age group, but obviously funerals affect people across the whole country. So it's not geographically restricted.

So we have some.

Yeah. So that's the kind of people that we're targeting and they're usually older people who are, but not always, who are organising funerals or sometimes it's, you know, slightly young people organising their parents funerals. So we have a bit of knowledge about ages of people, but yeah.

That's a very short description of the audience.

Interviewer
OK. And you mentioned that you used to work at [Museum].

In a previous life, how does that? What would the typical audience there be like?

Do you remember that far?

Participant 2
Well, yeah, it was sort of grouped more by different kind of purposes. So maybe more behavioural like people who are planning a visit to come to the museums, people who are interested in school teachers who are perhaps looking for resources for their kids or maybe booking a trip for their kids to come in and do a school course or visit. And then we had, like, perhaps a slightly more academic audience of people who were looking at certain collections, perhaps for their own research purposes.

And curators. So there was a few different levels of audience for that particular.

Yeah, we're working there.

Interviewer
Yeah. So quite a quite a difference between what you're looking at now and.

Participant 2
Yeah.

Interviewer
What you used to look at, you know people the audience used to work with. So when you're starting a new project.

How well can you define your audience?

Yeah, you current.

Participant 2
All or in pre visual at the museum or.

In the current how well. So usually not that well.

But we tend to focus on specific. So we did a bit of work that was focusing on people who were a terminal illness and that's quite a specific thing that characterises a person. But again, they can be people.

Of all different demographics, socioeconomic like across this span. So I don't think we're very good at categorising our audiences in funeral care. We do have some research from we have like an insights team and they are currently trying to categorise the sorts of people who buy funeral plans. So for example, they, they haven't fully shared all this with us, but we know, for example, that people who are quite wealthy, they're not necessarily going to take out a funeral plan for their future because they.

They've just got enough money in their estate to kind of deal with it and they just don't need to worry about it. And then you've got people who are at the other end scale who don't have enough money to afford one, and then you've got people in the middle who perhaps are more the likely to take out funeral plans. So we're starting to kind of categorise the different types of people, plan purchasers, and I think they've got about six or seven different profiles, but I don't know what they all are yet. And that will then help us to kind of target what we do design wise to try and reach those particular audiences. But like I say, that's very new.

Really shared it with us yet. So at the moment we just it's we're just kind of quite generic. We don't really have a great breakdown of our audiences.

Interviewer
OK. And was is that much different to what it was like when you were at the museums?

Participant 2
So I think at the museums we had.

We, our marketing department would do bits of research to understand like who was coming to our museums like what? Because often, you know, families with children. And that was like a quite a big group that we were trying to target because we knew that they were coming in. So we did do a bit of research into like who's visiting and why they're visiting.

But we didn't really have very defined groups and we were also, we had targets that we had to reach. So we were trying to reach.

More audiences of like, more ethnic minorities on the BAME skills we would look at kind of like that as well and there'd be a lot of projects trying to kind of reach people in disadvantaged groups.

So a lot of it sort of stems from like.

Trying to reach everyone.

But you know, also knowing that you've got certain people that are already coming, certain people that aren't.

Interviewer
OK, cool.

So my next section of questions is all about your experience with using personas.

First of all, do you use personas as part of your projects?

Participant 2
So at [Company] in in funeral care, we do not use personas currently.

We've talked about using personas and we've kind of decided that.

Because we haven't got very clear understanding of.

Of that, that we feel like they're not the right thing for us at the moment. I think when this new research comes out that perhaps categorises the types of people who might buy funeral plans, possibly they would sort of work as personas, but we don't currently use them at [Company], I have used personas in the past so you'd always have personas based on the type of drivers organised their parcels in quite different ways. So we tended to say right, OK, there's a there's a type of driver that does it this way and there's a type of driver that does it that way.

And we went and observed the drivers how they worked and we learned that and we kind of used them almost a shorthand for like what would this type of person do in this situation. But I don't think they worked that well. I think they helped us a little bit to remind us of the different people. But I don't we didn't use them heavily.

So I don't think I've ever really used them that successfully, because I think sometimes they are a bit made-up and not that insightful. So I think you know, obviously you're trying to bring a person to life and you're trying to base it on data and you're trying to base it on something true about your audience. But if they're not, we haven't really got that much use out of them in the places that I've

worked. So we have tended not to use them that much. Having said that, we are very interested in understanding.

Who our audience are, and knowing more about them and anything that can help us to do that is really useful. I think there's just a little bit, maybe I've observed and the attitude people around me that sometimes personas, if they're not done well, they're a little bit, I don't know, perhaps less useful than they should be sometimes.

Interviewer
OK.

So if you've not used personas per SE, if you've used anything else to sort of categorise the audience like user stories or anything like that, we used methods.

Participant 2
Yeah, we, we've used like jobs to be done and yeah, user stories.

Umm.

But we have done research with people, so we've to try and understand.

In detail, what people might be thinking, but I think we've sort of avoided a lot of categorization. I think we found it quite hard, especially with at need funerals because it can affect so many people that it's quite hard to we find it quite hard to categorise users.

Interviewer
Yeah.

Participant 2
Because quite you, you know, people dying is a universal thing, isn't it? But certainly maybe at the museums. Maybe that sort of thing might have been a bit easier, because we definitely had user groups in mind. But what techniques did we use to know we weren't very sophisticated in how we did that? It was a merely marketing led, I would say, like survey and data around that.

Yeah.

Interviewer
OK.

So do you feel that this not using personas is personal or a business led sort of approach or decision should say?

Participant 2
We don't know, actually. I don't know whether.

I think because [Company], I work very heavily with the user researcher and I think if we were I think if he was saying let's use personas, I think they'd be useful. I think I would be led by him because he's

the kind of person who we think of as being closest to our audience. So maybe that's yeah, maybe that's a cultural thing. I'm not sure how much he.

Is into the idea of personas, perhaps because we just don't have good data to base them on?

But in other places I've worked, I've said let's do personas, or if someone said let's do this, we have sort of done them and tried to use them.

But yeah, I don't know. I think it's just, maybe it's a bit.

Of a decision that's made by individual designers and teams. Maybe there may well be other bits of [Company] that are using are using personas. I just don't know.

Interviewer
OK, no worries.

So when you're thinking about an audience and maybe using a persona or using user stories or whatever, what type of information do you think is most helpful to you when you're doing your designing or you're developing?

And why?

Participant 2
I guess it's knowing about what they what they need to be able to do.

What state of mind they might be in? What are the most important things to them? What are the key things they're trying to either find out or to achieve? And how can we enable them to do that because we might want them to do one thing, but our users might want to do something different. And how do we make sure that we're designing something that takes into account what the business wants us to achieve, but also what is helpful for our users as well?

And in some certain places I've worked.

Those things have been very aligned and another place that they've been a bit different. So then you're trying to think, well, if we can help our users do what they need to do, then they're more likely to engage with things that we also need them to do.

Interviewer
OK.

Participant 2
I think I've forgotten your question, but yeah, what do they want to do? Yeah.

Interviewer
He was just looking at what was the most important information but.

Participant 2
Yeah.

Interviewer
On the other side of that then, so.

When you've got a user profile or you've got some user research on a person.

How do you use that and at what stages do you use that through the normal UX workflow or your adopted workflow?

Participant 2
So in our normal workflow, if we do some user research, we would write down our key findings.

From each user and then we would look at across them and say OK.

What? What things are standing out here that is looking like a pattern across this? What it's like is it's just one person said something like well, it's just that one person. But if we're seeing common, we start to theme it basically so.

If we look, we'll get together. Those themes of research and then we might say, OK, we need to do more research on this particular thing, but we'll start to theme the research and then try and figure out what insights we can get from that. So.

When we did research into people who were terminally ill, we did try to categorise our users based on the research we did, which was that we kind of learned we and it sounded really obvious. We kind of learned that some people want to.

Talk about stuff and some people don't, and some people want to get everything sorted and some people just aren't interested in doing that. And so when you kind of plot them together, you end up with this kind of like different sort of features that characterise those people.

So anything that can kind of give us a bit of an insight into the different mindsets that people might have is helpful.

But yeah, we've seen that. We've seen the research we pulled together the insights and then we see and then we try and make like recommend based on because we saw this we think we should do X&Y.

Because we saw this, we think doing this sort of generate hypothesis or sort of theories that we can then potentially test.

And we can design something that does that thing and then see if that works or not. So yeah, we kind of use it as the basis of where we start from often.

Interviewer
Yeah. OK.

I know you're not the doing the user research side of things, but whereabouts or whenabouts in in your project? Does that use a research normally take place? Is it early on in the middle towards the end? Whereabouts would that normally sit?

Participant 2

The first thing that we do would be, well, we might look at before we do any user search. We might look at what data we currently have like Google Analytics or look at some sessions online. So we might get whatever data we can first because that might inform what questions we have. And then from those questions, we might then take that into user research. So typically we would do it quite early on.

At the beginning and then we do more like usability type testing further down the line.

Interviewer
And in that initial.

User research what type of data is it that you usually collect and is it?

Interviews.

Observations? Or is it more qualitative quantitative survey based or?

Participant 2
We do all of those. I make it depend. It depends what we've got access to, but we typically we do all of those things or some of those things depending on we do interviews a lot, we do surveys sometimes. So we can get some calls to data, particularly if it's something that you know we might ask our call our colleagues who are working Funeral Home stuff we might do that via survey and then analyse that data.

That's come in or we? Yeah. We'll observe people through. Like, what's it called? Glass box, which is like a tool in go online. Watch people using stuff.

And.

And like I said, we've analysed Google Analytics, which is quantitative data, typically like where it's the drop off on this particular form, what can we do about that? So we'll use a mixture of all those things.

Interviewer
Alright 'cause, I'm very busy.

With your.

The analysis process that you undertake sort of, how's that analysis undertaken? Is it manual analysis? Is it automated? What how would that normally?

Participant 2
It's mainly quite a manual, so we would we do have some.

Like tools, we have a thing called. What's it called?

In is it envision or what's called? It's a bit like Google Analytics, but it's some of the kind of things that are key goals on our site are set up as like sort of like goals and Google Analytics type things. So we can see things like conversion rate and we can see funnels and where things like drop off rate. But

and we do have like dashboards where we have information that we report that we that we review but often it's more of a manual kind of go and have a look at the data, find out what you can.

So. So you'll be poking in on the data trying to understand what's happening, trying to see if there's anything interesting and doing a bit of a one off analysis on it. So yeah, we have like key metrics, if you like, that are semi kind of like automated in Google Analytics, but a lot of the time when we're looking in something because we're looking at it from a particular angle, maybe then it's a lot more manual in that process, yeah.

Interviewer
OK, I'll take you. That's quite a timely process. It takes a quite amount of time to do.

Participant 2
It can be like we do have. We have some. It's not. It's not supposed to be their job, but they do. They do it quite a lot for us because they have a bit of capacity, so they often support us on that and they have great analytical skills. So yes, they can be quite time consuming, yeah.

Interviewer
OK.

Have you heard of data-driven personas? What are your thoughts on them?

Participant 2
I mean, I don't know if I've heard of that phrase specifically other than that in my head. Ideally all personas are derived from fact rather than fiction, and often when we think about fact, it is thinking about data points in a very general way, but I haven't heard that phrase. It probably maybe that has a more specific meaning than just that. So to me, the idea of a data-driven persona would be a persona that was based on actual data rather than.

Something that was just kind of made-up out of your brain.

Interviewer
Yeah.

Participant 2
Fish.

Interviewer
So it was part of my PhD. I'm exploring this idea of very early.

Survey based data-driven type.

I don't want to call it personal because it's not. It's not got all of the qualitative.

You know, interview information or anything like that. It's missing all of that.

Oh.

You've run the survey really early on a an existing system gather.

Some information to some survey questions.

And then run them through a clustering system and that clustering system automatically.

Does the work, gathers them and separates them into groups essentially or clusters?

There is a downside to clustering in that it will always give you some groups, so not always are these clear cut if you know what I mean. They're not always perfectly separated, but I suppose in when we do any persona work, they're not perfectly separated because there's always some crossover there.

So I've been working on coming up with this idea and introducing this step that's really literally like in the first week or so of a project we run it.

Survey short period of time. Throw it at this system and then we get some instant responses, so no time lag essentially into what's there and the system will also allow you to explore different groups, different numbers of groups within that data as well. So you could recluster and come up with four groups or two groups.

And that's what I've been sort of working on as part of my PhD.

So can I just share my screen with you?

We just get to this one. So what you gonna?

See hopefully.

It allows me to share.

Can we see that?

Yeah. OK. So this is a figma mockup of the tool I have run all of the data that's in this through the actual live tool, but I couldn't get it to do the pretty pictures on the end, so I've marked it up into figma in order to save some time. But essentially, you'd run download your template for your survey questions.

Run that on your system.

Once you've got the answers, just upload them. You don't have to do any cleaning on the data, nothing like that, and then it's going to run, and it'll produce the a number of clusters for you. Ideally it will show you initially the optimum amount of clusters.

And it provides you with this little bit of information describing about how to use the system and things.

So this data might look quite.

Common to you 'cause. This is based on the NML data from when you were there in the survey that are on with you guys.

So this comes up and initially it's telling me that there's two profiles that it's worth that are separated and in the drop down you can see that the green indicates 2 profiles is good. Three there's a little bit of crossover between a couple of the groups, fours OK and fives a bit terrible.

Inside of these.

Participant 2
And how is it decided? How has it decided that there's a crossover in the three but not the four? How does that work?

Interviewer
So when the clustering algorithm works, it gets what's called.

A an overlap figure, so it looks at the data and then looks at how much overlap is within them clusters.

Participant 2
OK.

Interviewer
And so once it's clustered.

It gives out a figure and it tells us what that overlap is. It's called a silhouette figure, a silhouette score, and that's what I've used to then say whether what call that should be on, how much of that silhouette score is there and is present.

What I then do with them clusters is normally a lot of people with cluster analysis would just pick what's called the centroid. So it's a single.

Single user's response who's in the middle and everyone else is clustered into that cluster.

But the downside there is it doesn't really give you very clear clusters.

So you're basing it's like essentially goes to the persona where it's you're picking one person from a group and only using their values.

Participant 2
Mm hmm.

Interviewer
What I've done with these is I've then instead of taking that centroid, I've gone for the cluster and I've said OK. So for each of the columns of data, if you like. So if we think about age as a column of data and gender as a column of data, I've said return me the data that makes up 75% of this column.

So none of these.

Really add up to 100% all perfectly. A lot of them are a little bit under because it's whatever's made it up to that 75% to be representative of that column of data.

We get a little introduction piece that just describes how much of a percentage this is of the whole sample. So you can see it's 57% of the sample in this two clusters, which is 293 of the participants.

Participant 2
OK.

Interviewer
And then.

You get the demographic data of where they're from. We can click on the little thingy icon. Just go to the tabular versions of that data.

But we're looking also at like, what's the goals and the motivations of these people, what drives them, what museum they've come in or entered from, should we say?

Their reasons for visiting?

Their tasks that they want to try and achieve.

You'll notice that some of these are green as well, and the green.

Participant 2
OK.

Interviewer
Borders indicate that this factor is something that you know significantly different here in this group to the others. This is what makes this significantly different this group. So for this, this one persona that we've called Nappi Walla is the reason for visiting the tasks and the motivations. And then in the behaviours panel, you can see that we've got information seeking. So whether they prefer to search or browse or browse.

On the page is viewed and then how frequently they visit. That's they're the five factors that that make this profile stand out from the other profile or in the profiles in that cluster.

And when I moved to the other one, you can see that here we've got very similar things that are different.

On all five 'cause, these are obviously what make them significantly different from each other in these this simple 2 profiles.

But we can explore what the profiles would look like in four groups.

So we'll get to 4 profiles. You can see the numbers drop down quite considerably in this one.

Significant factors are the reason visiting task, motivations, information, and pages viewed still, but the domain knowledge has started to come in on this one instead of frequency for visits.

Then when we cycle through, we can see that they change quite a lot into general cultural heritage. For this one, their search behaviour, whether they're going to try or persist at getting an answer. But most of them are here just giving up after one single search.

Participant 2
And does that search mean they are using the website search or does it mean they're searching for a

specific thing and there? And is that what the browsers like? I'm just having a look around and the searches I'm looking for something specific. Just tell me a little bit more about what that means. The search and Bros.

Interviewer
Yeah. So in the searching in the browsing option, it was what is their preference? Do they prefer to search or do they prefer to browse for information?

And as you can see here, we've got nearly 90% that prefer to browse within this group.

Participant 2
Mm hmm.

Interviewer
Where is the search behaviour. Is. If you did do searching.

What? How? How would you normally go about with your search? So would you just try the once and give up if you didn't get the answer? Or would you persist?

Until you actually got the answer, if you know what I mean.

Or would you literally just give up after the first attempt? So you've got one attempt, a few attempts, and then the persist?

Participant 2
Mm hmm.

Interviewer
And then we've got the usual device stuff, so you know, most people use the PC in order to access it. And a lot of the users here are a first time.

Yeah, that's that, that group. Then in the third group, you can see the genders drastically changed onto a male group.

The ages have dropped considerably to the 18 to 34 category.

And it doesn't really pan out, but we end up with a lot in the retired category as well in there. So we do have this big portion of 65 to 70 fives in there, but they're all very local to Merseyside.

Looking pre visit to the museum, the physical museum.

But then into the knowledge domain knowledge for the exact reason as to why they've come there is actually a lot higher in the intermediate than the other groups that we've seen. And here we'll see that they persist with their searching.

So yeah, there's a lot of difference between these groups when it comes up with them.

But what's your initial thoughts on a tool like this that you could quickly just?

Gather the survey results, throw it at it, and get some initial states to then use these as a kick off point. If you'd like to go and possibly do some user research or targeted use of research to build out your user stories or your personas from that point.

Participant 2
Yeah, I mean, I'll, I'll my initial thoughts is I like the idea that it's based on.

Qualitative data. Sorry. Quantitative data that that survey's gone out. I don't know how many people have filled that in. I think I think you've got that there, haven't you, you've got the number of 500.

Interviewer
That there was 500 and something that filled it in.

Participant 2
And that's significantly more than often we're able to work with.

Data wise, so I like the fact that it's a kind of got that robustness of being solidly based on data. I think some of the questions I would have is.

What are the like? It is about the questionnaire that is given to them and the survey that is done like what questions are asked is there, you know, are is it causing a bias and the people that are willing to answer.

Questionnaire how long is the questionnaire?

Could we be missing potentially some audiences because they just can't be bothered doing that?

So I'd be really interested in the kind of like.

To understand more about the survey of it, I like the way you've presented it here, where you can see the differences between the profiles. It still wasn't immediately obvious to me at a glance exactly what was different, but I like the fact that it's highlighted and then I could say, OK, I'll go and dig into that a bit more and sort of pull out the differences almost in a way being able to have a section that that actually a dedicated section on the dashboard that says these are the differences in this profile from some of the others.

That might be helpful to even maybe enhance that even more.

But yeah, I think this is a useful tool. I think it very much depend on.

Which company and which people it was targeting, whether or not it would be of use to them, but I think I think there would be an interest in this definitely.

I think some people might say, well, what's the difference between this and some of the insights you can get out of analytics where you can sort of create segments and things like that. But I think this is a nice combination of like directly asking.

People and then categorising them based on their answers. There's a little bit of the nagging at the back of my mind that what people say they're doing doesn't always match, like they're not always super accurate, can't always rely on people to tell you what they're doing, why they're doing it. But.

That's true of any kind of stuff that you're looking at, isn't it so?

Yeah.

Interviewer
Yeah, the big, I suppose the good thing now is that 'cause you were doing this really on in the UX process or in the project workflow then whatever you get from here, you could always then go to your on a Google Analytics and back up.

Participant 2
Yeah, like, see if it. Yeah.

Interviewer
So and you can validate essentially that's to what they say they're doing against what Google Analytics is showing that the people who are logged into Google are doing or things like that.

Participant 2
Yeah, yeah, definitely. I think it's just interesting to sort of know that that's what the data's come up with. There's no, I mean, I'm sure there's bias probably in everything that's ever done isn't there, but that somehow it feels like because it's based on that data, it feels like it's potentially less biassed. I don't know if the questions lend themselves in some way to some unconscious bias and things like that, but you know, and it feels.

Interviewer
S.

Participant 2
More of a neutral way of generating these clustered kind of personas or whatever you want. Profiles which are quite like so I could see it being useful as a tool potentially for, you know if there's disagreements internally about what people are doing and and what the different people they're trying to reach are like this would be a helpful way of going oh, let's have a look and see what's actually going on here.

So yeah.

Interviewer
OK.

Participant 2
See, it might be a valuable thing to be able to use. I can also see that this would be really valuable for small museum based teams where there is no data analyst or UX team member. When I was at the [Museum] we used personas but they were less detailed than these and more wordy. I can see that these profiles based on data and in a visual format could be really useful.

Interviewer
Cool.

I've just got a couple more questions if you don't mind. It's going to stop sharing that screen.

OK. So last couple of questions on so beyond personas, are there any other methods or tools that you're you currently use or are exploring?

In order to understand users audiences better other than what we've already spoken about.

Participant 2
Mm hmm. Yeah, not nothing other than we haven't already covered. Other than that, we do at [Company]. It's big enough that we have like an insights team and they do much more like the social science kind of research, which is like more statistically kind of.

You know, they'll do large scale kind of surveys to find stuff or research so that we do have some of that going on and we do try and work with them to learn more about you know the reasons why people around plans like the cost of living. So they've done a few of these kind of big bits of research and that we then use and use that as part of our kind of insight so.

But apart from that, no, it's just the things that we've talked about.

Interviewer
OK.

And then obviously the UX field is constantly changing and evolving.

How do you see the future of user profiling within UX design?

Participant 2
Well, I do think this idea of being able to automate stuff, you know, like people talk a lot about AI, don't they? And how you can do so. The ability to kind of try and use what computers are good at to kind of help.

Go through the numbers and generate stuff based on data if that feels like where there might be value so I could see anywhere where there could be a bit of extra automation added in would be helpful I think and this this kind of thing that you've shown would be I think the future for UX in general you know is more like how do we use those tools.

I don't. I don't think we're yet at the point that all design would be automated, but.

Yeah, I think this just adding people want different tools to use in different scenarios. So I think there's a richness of tools and more will emerge as we kind of progress into the future. So yeah, something like this, I could see fitting in alongside that, but yeah.

Interviewer
OK. And finally, do you think I've missed anything or would you like to add anything?

Participant 2
Missed anything in in regards to what like just?

Interviewer
To user profile in personas or anything that you think would be relevant or not.

Participant 2
I think we'll be what might be interesting is and that this might be that I don't understand it is that could you feed this thing? You've created different survey questions and would it cope with that?

Because I think different places will have different things that are of interest to them and being able to kind of customise that like I wasn't quite sure from the way you described it, whether it's this, these are the questions and that's that or whether would you design unique questions for different businesses potentially. So like say court funeral care might be different to say the museums, so like being able to kind of feed in the things that you want to know about a bit.

At the start and how much you could do that.

We'd be good to know, but yeah.

Interviewer
OK.

Well, thank you for your time.

Participant 2
I hope it was useful. I have no idea if it was or not.

Interviewer
Very, very useful. They always are. One last thing to ask is if you do know anyone else that you think would be good to for me to do an interview as part of this study, could you e-mail me their details so that I don't have any personal details inside of my transcripts or anything like that.

Participant 2
Certainly I do, yeah.

Interviewer
If you can think of anyone, that's that'll be fine. If not, don't worry.

Participant 2
OK, I'll have a think and I'll message you separately.

Interviewer
OK. Thank you for your time. I'm gonna just stop the transcript there.

J.2.2.3 Participant 3

Interviewer
Tell me what your current role is or your what your class is your role or career at the moment.

Participant 3
So.

I am a one of the directors at [Company] and we build a product called [Product] which is a CRM system applicant tracking system for international recruitment and employment and my role at the moment is I head up the product side so I'm currently filling the role of product owner and the combined with a more technical role as well.

With that, but those are the two kind of main areas that I work in.

Interviewer
OK. And what does a standard day work or design phase sort of look like for you?

Participant 3
Like I was under day. Or how do you mean with the?

Interviewer
What did it depends on how you know when you're working on a project, what does your standard sort of day or week sort of look like?

Participant 3
So I mean in terms of my kind of standard day there's, I mean there's any number of projects running anytime. So there's a lot of project management involved. And so kind of working with.

Working with the developers working either sometimes with.

Designers or internal or external designers working with the QA side of things. So kind of.

Switching all of that together, making sure all that works, but also kind of getting involved in each one of those steps at different times as well. Sometimes it works, sometimes I'm kind of end up doing that as, as I'm sure you're aware. So. So that's kind of like a typical day for me in terms of how we approach projects.

We've got a development plan that we.

There's roughly kind of 18 months in advance, so we know kind of based on based on quarters what the kind of big rocks are that we're going to be working on in those in those quarters.

And then as we kind of approach those quarters, we'll start to look at.

Written specifications initially and then we'll get feedback from kind of key clients on those specifications to make modifications, and then we'll usually go into like a design prototype in phase where we'll kind of go through the same process again where we were we get feedback from. We've got like a group of kind of key clients that we typically bounce ideas off and things like that. So we get feedback from them on those side of things.

And then goes into development and then goes into a like a testing phase and we try to do big releases, that was that's what we're trying to work towards rather than stuff going live in bits and bobs we we're trying to work towards a quarterly kind of big release so that we can have do like here's a new version of the software and things like that even though it's not strictly that in the background but that's the way we try and try and present it.

Interviewer
And what type of audience do you usually work with or for?

Participant 3
In in terms of people or companies or?

Interviewer
All your system's built for users. What type of you? What type of users? It all starts audience there.

Participant 3
So our so we're a business to business. So our customers are the companies who use the software. So that could be a company who is running a programme, sending people to work at summer camps in America, it could be.

People selling or recruiting for gap year programmes, you know, any kind of possibly open childcare related programmes and things like that.

Uh, but we work directly with the with the company. That is kind of running those programmes essentially.

Hmm.

Interviewer
So when you start a new phase or a new project, if you like or a new version of your software.

Participant 3
Yeah.

Interviewer
How well can you define the audience the end users, the people who would actually end up using it?

Participant 3
Pretty well, to be honest. We we've got.

You know, we've got years of experience working in, in, in this industry. So, based on that.

You know, we've got a lot of information there, but it is.

When we're at a phase where we're not.

We know all of our customers quite well. You know, we do speak to them on a regular basis. So it's not that there's very few customers that we've got that we don't speak to on a on a regular basis. So we're always in in contact with them in that way, so.

Interviewer
OK. And how well do you know?

The people who.

Are actually gonna use the system, so obviously these are deal pairs. As you mentioned, all the people who want to do all pairing and things like that. How well do you know that audience?

Participant 3
Probably not as well as we know our direct customer base. So our direct customers would be the what we call the admin users. So these are the people that are employed by the by the companies that buy our software, but their customers I guess are the.

Kind of one of the real end users.

And I kind of only admit it's not. It's not an area that we're that we're super strong on knowing.

Knowing that well.

Interviewer
OK.

So the next phase of questions is around experience with personas.

So.

Participant 3
Yeah.

Interviewer
Do you or do your company use personas as part of your projects?

Participant 3
No, not really, no.

Interviewer
Right. Any reason why not?

Participant 3
Probably knowledge gap.

No, we don't have a.

A user experience expert, umm, that works for us, so there's probably a knowledge gap there. That's probably the main reason for it, I'd say.

Interviewer
OK.

So what sort of information do you?

Get about your audience or users on what, what, what do you find most useful in that information? cause. Obviously, you don't have the personas, but you say that you know the audience or you know your customers quite well, but not their customers.

From your customers, what sort of information do you find the most helpful when designing or developing?

Participant 3
Probably just.

Real, real conversations, honestly is probably the number one. So we do we, you know, we have real conversations with people. That's the one of the kind of most important ways is definitely not the most.

Modern way of doing it, but it's definitely.

You know, it definitely works for us.

Aside from that, we have got things like.

We've got things like New Relic running which tells us, you know, kind of browser starts and things like that. So we you know geographical starts.

You know, so we can we can track key transactions within the system. So we can kind of track things that way.

Another thing that we use is something called Chameleon, which is like a user engagement.

Platform. So it's chameleon dot IO is the is the URL for that. So we use that a lot for doing.

Kind of little micro surveys and right the way through to like full full-scale MPs and things like that.

But we also do kind of.

Full tours so new customers get kind of can-do full tours through the system using that, but also kind of many roll outs as well. We do like many tours and things like that with it.

Interviewer
OK.

Do you do any other form of user profiling or anything like that?

Participant 3
No, no.

Interviewer
OK, that's fine. Or we can cross that one off.

Sorry, just.

Participant 3
Done.

Interviewer
Checking where I'm at.

So if you were to do personas.

Participant 3
Yeah.

Interviewer
Just say for your clients or anything what types of data would they be based on?

Participant 3
Can you explain what you mean? I'm not quite sure.

Interviewer
If you were to go, you know, talk to you. If you were to create a template of your customer.

Participant 3
Yeah.

Interviewer
That'll help drive everyone. Then what types of information would you ask to include in that so that it became useful to you?

Participant 3
Probably.

I would say I mean this is this is gonna be very kind of industry specific. So I don't know whether that counts or not. Just tell me if I'm saying the wrong thing, but it would be.

The number of number of staff members that work for the company where they're based geographically.

From a technical aspect, I'd say.

Browser is an important thing.

But then.

Kind of very industry specific. You know what, what country they are based in, what countries they send to.

What?

I'm trying to think things that are not too, not too specific to what we what we do that would be that would be useful.

Interviewer
It's whatever would be useful to your you and your team is all I'm asking around on that so.

Participant 3
How many different programme types they offer in in terms of the in terms of what they, the kind of programmes that they offer?

How many of our other clients, do they work with? Are there any links there with other clients so that there's potential for connections, things like that?

What else?

Just general.

General technical knowledge like we, we've definitely run into issues where.

You just come across non tech people.

And that's kind of a red flag straight away. It makes it really difficult if there's not somebody who can be like a product champion at their end is definitely something that we always look out for. So.

Interviewer
OK. And obviously you half of your system is built around creating the interfaces for the customers of your customers if you like.

Participant 3
Yep.

Interviewer
So.

What sort of information do you think might be useful if you were to create a persona around that their custom? You know your customers, customers.

Participant 3
I would say.

H device.

Some way of tracking the?

The quality of the Internet signal that they can get like we we've definitely hit some issues where there's people applying from locations and countries and locations where.

It's difficult to get an Internet connection so that that's definitely something that we've that we've ran up against in the past is not as common, but it does. It does happen.

I mean.

It's hard to say because they're not our, they're not our direct customers. So we're looking for what are the absolute basics.

Of things you know. So it's not like we're going out kind of selling on behalf of our clients. It's yeah.

Yeah.

Interviewer
No worries. OK. So the next segment of my in my questioning is based around my PhD work.

And so.

I.

At the moment I I've been studying a museum audience.

And trying to think about what is the process for really understanding who these are essentially these users and I've come up with this idea behind creating.

Participant 3
Mm hmm.

Interviewer
Groups or semi personas that I'm calling Internet information seeking cluster profiles.

No, they're different to a normal persona 'cause they don't have all of the qualitative interview based.

Participant 3
Mm hmm.

Interviewer
Information in there.

They're based off a survey, so you put out the survey early doors, you gather the response really early on in the project.

Participant 3
Mm hmm.

Interviewer

And then rather than have to manually process these results, I've built a system that we upload the survey results to and it automatically does some.

Machine learning out in clustering on the data.

Participant 3
OK.

Interviewer
And it produces an output. To understand who these clusters are now as a rather than it be.

Participant 3
Mm.

OK.

Interviewer
In in normal clustering, the results are based around what's called a centroid, which is just a single user's responses that are clusters. The middle of a cluster now. Then I found that they're not necessarily that representative of a whole group, and So what I've done is instead of going with the centroid, I've gone with.

Participant 3
OK.

Interviewer
Looking at each column, if you like in the date on its own. I'm working out what makes up 75% of that column to make it representative, and then I've put them into this profile, so I'm just going to share my screen with you if you don't mind.

And then you can see what I'm talking about here.

So initially.

Participant 3
Hmm.

Yep.

Interviewer
I'm for if there was another museum studying then they could just download my template CSV, but if it was ideally someone like yourself wanting to know what your customers do then you could create your own questionnaire template.

Participant 3
OK.

Interviewer
Run that, get your data in a CSV format from all of your survey results, and then upload it. Now when we upload it all I've got here is a figma mock up, but it is based on actual data from the museum that I was working with.

Participant 3
Cool.

Interviewer
And initially when it loads it's going to load in this 2 profile format and you'll see that it's green and there's some information up here that explains how to use the interface and things like that.

Participant 3
Hmm.

Interviewer
So the green here means that once the clustering has worked or done, it's done its business, then it comes up with the optimum number of clusters ideally, and that's what it starts to show. So in this case it's showing us that.

Two clusters are the ideal.

Participant 3
Yep.

Interviewer
And with it being green, it's showing us that those clusters are separated or most separated as much as possible.

Participant 3
OK.

Interviewer
Now the downside in clustering is that it will always give you a result.

Participant 3
OK.

Hmm.

Interviewer
But that result isn't always very clean. So if I just on the drop down show you on here, 2 profiles comes out as green, so that's clean.

Participant 3
OK.

Interviewer
The yellow is. There's a little bit of overlap between some of these groups. Four is quite clean and five, there's a quite a lot of overlap.

Participant 3
OK.

Hmm.

OK.

OK.

Yep.

Interviewer
But what this allows us to do is to explore how many groups exist within our data so we can start to see who they are and what types they are. As you can see, just go through this first one. Now it gives us a little bit of an introduction of breakdown of the actual cluster and you can see that this one is representative of 293 participants, which is 57% of the total sample of the survey.

Participant 3
OK.

Interviewer
We then have groups of data, so we've got demographics, we've got goals and motivations. We have knowledge levels and then behaviours.

Participant 3
OK.

Interviewer
In the demographics.

Participant 3
Mm hmm.

Mm hmm.

Interviewer
We have a mix between well, we've got age, which is one of them columns I was talking about and what makes up 75% of this is quite a good mix, but majority of them are more 35 to 54 year olds.

Participant 3
OK.

Interviewer
You can see that the predominantly female.

Participant 3
Yep.

OK.

Interviewer
And they in full time work and live in Merseyside. So it's quite prescriptive if you like. But we still see what makes up the rest of this data, this group.

Participant 3
That's cool.

Interviewer
In the goals of motivation, you'll see that we've got these boxes, these areas or these questions where they've got a green border.

Participant 3
OK.

Interviewer
Now when the clustering works, it comes up and says these are the.

Participant 3
Hmm.

OK.

Interviewer
Factors or the columns that make or define this group, and so that's what I've presented here with the border in it being outlined in green. These are the ones that define in this group. So the reason for visiting with being over 50% for personal reasons.

Participant 3
OK.

Interviewer
Pre visit and other reasons for that what they want to do and why they're coming to the site.

The fact that in this group just over 50% prefer to search than browse.

And then most of them are only gonna come and look at one page within the website.

Participant 3
Yeah.

OK.

Interviewer
And then lastly, most of them are only coming for the this is their first visit as well. So they're you know total onboarding newbies if you like.

And then when we look at compare that to the second of the group here, you can see that the age group is significantly larger. It's a 35 to 50 fours.

Participant 3
Yep.

Interviewer
There from Merseyside, but actually they're going a bit bigger into England, covering up to England in higher areas.

Participant 3
OK.

Interviewer
They really only come for personal reasons this group.

And most of them are coming for pre visit to a the physical museum.

Participant 3
So pre visit is I'm going to visit this museum and I'm going to go to the website first to have a look to see what's there like is that.

Nice.

Interviewer

Yeah. Find out what the museum's about, what facilities? They've got that sort of thing. And then in the behaviours, you can see that majority of these prefer to browse over search.

Participant 3
That's interesting.

Interviewer
In the pages viewed.

Most people will look at 2 to 10 pages.

Participant 3
OK.

OK.

Interviewer
On a standard visit and then frequency visit is, they'll come some first time users, but a lot more yearly and weekly visitors than the other group.

And obviously we can see that most of them are using iOS.

Participant 3
Yep.

Interviewer
Devices or unknown devices in that case, in the knowledge we can see that though four and five is like the higher levels of knowledge.

Participant 3
Yep.

Interviewer
So they're quite high there. And then in the specific knowledge that they've come for, for their visit when they filled in the survey, you can see that the predominantly cluster themselves as intermediate out of a A3 level group.

Participant 3
OK.

Interviewer
So that's like the differentiation between the two groups there. If we just look at, we can explore the different profiles, sorry.

Participant 3
Yep.

Interviewer
So we go to four clusters. You can see that we get more clusters automatically done. And then as we look through these, you can see that this one's like more world based, predominantly female still working, but there's a good spread across the board.

Participant 3
Mm hmm.

Mm hmm.

Interviewer
Age group Wise, 35 to 54 again looking for pre visit but this time the very specific or the significant differences are in the cultural heritage.

Participant 3
Yeah.

Yep.

Hmm.

Interviewer
General Culture, heritage, knowledge, the information seeking preference in that most of them want to browse. They're going to look at 2 to 10 pages and they when they're doing a search, they will try a couple of times before giving up.

Participant 3
OK.

Interviewer
And most of them will use APC.

So.

It's quite, you know, significant in that one compared to the next group, which is male, predominantly very close to Merseyside.

A lot retired in the age groups are a mix between 18 to 34 and 65 to 74.

Participant 3
OK.

Interviewer
They're looking for personal reasons or to pass time quite significantly and looking to pre visit.

Participant 3
Yep.

Interviewer
You can see that the general knowledge is much more scattered.

But the specific reason it came is quite focused still on the intermediate.

Participant 3
Mm.

Interviewer
And here they will persist or a good portion of them will persist on search until they get the answer they want.

Participant 3
Yeah.

Really cool.

Interviewer
So yeah, that's just a quick whirlwind tour of the sort of profiles and then what, how they differ and everything else based on the questions and they've been clustered.

What's your initial thoughts on a tool like this?

Participant 3
That interesting.

I could ask you, I could sit and ask loads of questions, but I don't know. You've got your questions to ask me. I think it's that interesting. I think it's really cool.

In terms of how?

We could use it so a lot.

A typical way that so our customers customer, it's that we call them a participant. So they're participants, the person that's going to go on one of the programmes.

And one of the main way has always been, you know, they will, they will come to a website.

And they will click through a series of pages and try and find a programme that's relevant to them or that interests them.

All the while searching competitors websites as well and things like that. So one of the one of the real kind of.

Things that we're looking into at the moment is, is how can we?

How can we?

Properly suggest programmes to people not just based on.

You know you're this age from this country, so this programme right for you. But the domain knowledge question was particularly interesting because there's different.

Kind of programmes that that that would require you to be more.

More knowledgeable or more experienced in travel than other programmes.

So I it's I find it really interesting.

The obviously the geographical stuff as well.

Is, is helpful.

You know, I'm sorry.

Interviewer
Do you think something like this could be useful to you in?

You when you get to your design or your building phases.

Participant 3
Yes, I would say so. Yeah, there's obviously people far better than I am at design that understand these things a lot more than I do, but.

But I would say absolutely, yeah. I mean from a from a product design.

Approach there's.

Having.

Having those profiles to understand, you know, is this.

Is this like a real kind of travel savvy group of people, or is it someone that just wants to go on holiday for have an extended holiday or something like that, that having that information would really kind of affect the way we kind of build out products and things like that?

Interviewer
OK.

I'm just gonna stop sharing the screen there.

Participant 3
OK.

Interviewer
So is there anything that you would add change or drop from an ispc from the method? Because obviously this would be done within the first couple of weeks of a new project so that you've got focused profiles to then build off and do more user research or ask more questions for them. Do you think there's anything that you'd change or drop from that?

Participant 3
From that specifically, now the.

How the information is acquired is always a challenge.

Like it and.

We're kinda dealing with people who just.

Don't you just want to get something done as quickly as possible? So how we how we acquire that data might be might be a challenge, but once we've got the data there, I mean that that that information's so useful, I'd say everything that you showed me there would be could definitely could definitely influence the way we the way we do things.

N.

Interviewer
OK. So you potentially you could run your run the survey through your clients, get your clients to ask their current clients to, you know people who've already gone through the process and signed up to complete a survey on your behalf?

Participant 3
Absolutely, yeah. And I having said that, I think.

Part of the.

A good chunk of the survey may already be there in the in the application forms that people have filled out historically, and as long as the client's got permission to use the data in that way.

There's potential that you could extract that data from the application forms that people have filled in anyway, so it may well be that the data's kind of sat there already without having to. I don't, I don't know whether it explicitly has to be a survey or whether you pull the data from historical applications. I'm not sure how it's set up at your end, but.

Yeah.

Interviewer
In the moment it's designed for around surveys and the CSV, but as long as you could pull the data into the that structure, it'd work.

Participant 3
Yeah, I mean like, I mean, you know, potentially whatever the whatever the questions are, if that's

already potentially already there and you know we've got years and years of historical data with, with, with some clients that you know if as long as we can pull that data out in the right format then potentially it could be, could be run through that engine. And I guess it would give us the data. So yeah, absolutely. I mean it may well be that we don't have to convince people to fill in surveys, it might be, might be there already so.

Interviewer
Yeah, OK, cool.

So last couple of questions then.

Participant 3
OK.

Interviewer
This is around feedback and future considerations so.

Beyond personas.

Participant 3
Mm hmm.

Interviewer
Are there any other methods or tools that you currently use or potentially exploring?

Excuse me in order to understand users and the audience better.

Participant 3
It's.

Hmm.

I mean, yes, there's one of the things that we're acutely aware of at the moment is we're not, we're not, we're definitely not brilliant on gathering that kind of information. So it's definitely something that we're aware of. Like we say we, we definitely do have that that knowledge gap.

And kind of.

Building that feedback loop into the way that we work is definitely something that we that we want to work on.

Obviously we've got the kind of standard book tracking and page load times and browser stacks and all of that stuff, but.

Being able to kind of trace the way someone's using the system, and we've got a lot of this data already, we're just not able to use it effectively at the moment, so.

Like I say, we do have that that knowledge gap that we'd like to fill.

But yeah, that's I don't know whether that answers the question or not.

Interviewer
Yeah. From your perspective, it definitely does.

Participant 3
I.

Interviewer
OK so.

Participant 3
Mm hmm.

Interviewer
With the evolving tools and the methods within the UX and the development world, how do you see the future of user profiling in UX design?

Participant 3
I don't know in. In all honesty, it's not.

Interviewer
That's OK.

Participant 3
It's not.

Interviewer
No if.

Participant 3
Yeah, I'll try and say something without waffling, but it's. I don't. I don't feel like I'm. I'm probably not qualified to answer that question really.

Interviewer
That's all good. OK, so finally, do you think that I've missed anything or is there anything that you'd like to add?

Participant 3
No, I think it's. I think it's, I think it's great, I think it's.

The way you've approached it looks great. The interface makes it really easy to understand. You know the obviously there's a lot going on behind the scenes in terms of the way that that Data's crunched and worked out. There's obviously a lot of hard work going on there, but the way you've presented it makes it really straightforward to understand.

Yeah, it's a. It's a great job. I like it. Really do.

OK.

Interviewer
OK, cool. Well, that brings us the end of the interview. Got one more thing to ask is if you know anyone else that you think would be good for me to interview. I'm related to this. Then could you just e-mail me their details separately as I don't want their details to be in this.

Participant 3
OK.

Interviewer
Transcript or video?

Participant 3
From the from this company or from another company or what's there?

Interviewer
Anyone that you think would be useful if it's lower UX person or from a different company? Whatever I it's really doesn't matter. Just if you think they'd be useful.

Then if you can just send me some details and a little bit about them, then I can consider whether that'd be useful or not myself at that point.

Participant 3
Yep, definitely. Yeah, I'll definitely do that. Cool.

Interviewer
OK, cool. I'm just gonna stop the transcript.

Participant 3
OK.

J.2.2.4 Participant 4

Interviewer
The first lot of questions are around background and role. So what is your current role or career?

Participant 4
So I am a user, researcher, a user research consultant. I work for currently [Company].

Looking at people who are on what is known as the health journey, so this can be people who perhaps have a disability or a temporary illness.

Whether physical, mental or otherwise, and need financial assistance from Universal Credit.

Interviewer
OK. And what would a normal day week or maybe design phase look like for you?

Participant 4
Yeah. If I give you a design phase, that's probably an easier, easier one. So we get direction from many different places within Gov. So it might be a policy change that we need to design for or it could be ministerial change or sometimes it can be that we've had a lot of feedback and we have managed to get something in the road map that we want to want to look at. So we might.

Design phase might start with discovery about what we know already or.

What we want to learn those types of things.

Typically I'll do a lot of work in that area. There may be I'll spend time devising a research plan for that. Looking at data that we already have, that can be anything really that could be metrics. So when we're talking, you know, actual data usage, that type of thing could be past research.

Academic studies, white papers, things like that.

And then.

Typically, yeah, I work. We're quite we're working a multidisciplinary style, so I'll work with designers, interaction designers, content designers, product managers, whoever, to device, whatever we're going to work on and then I'll help them have a test that either by doing research before to inform some of their decisions or research afterwards.

You know, usability based stuff capturing.

The thoughts and opinions of people who we speak to.

And I'll try and do research with as closely as possible the demographic who are using it, sometimes in Gov. That's not always the case. If we're doing something internal or it's a very hard to reach audience.

For example.

If I want if I was doing research that was related to homelessness, I might not be able to do that research with homeless people, but I would probably do that research with someone adjacent to them who might work with them. It's like a proxy, that type of thing. Yeah. And then we'll just iterate.

That design, once we get it back, once we learn about it, yeah, that's pretty much a standard phase.

Interviewer
OK. So you've obviously talked quite a lot there about the typical workflow within that phase.

Participant 4
Hmm.

Interviewer
Is.

Pretty similar almost design phases that you've worked on.

Participant 4
Yes and no. I think what I'll give you is a classic example of sort of a design phrase taking you know from like discovery to alpha beta and taking through that. But sometimes we'll get projects that are midway through. Sometimes we'll get projects that have a very short deadline on them. So that will limit the things that we can do.

In the ideal world, we would approach everything the same way. Discovery Alphabet alive, or that way, or try and get as much research beforehand, but that's not always the case.

Does that answer your question? Sorry.

Interviewer
It does, yeah, I am. That's good. Thanks. OK.

When starting a new project, how well do you define your audience?

Participant 4
Oh well, yeah, that's probably one of like, the first things we would do, we will not. I will know exactly who the end users are.

So, and not just who the end users are, but who it might affect. So a lot of time in good we might do some sort of service mapping to understand, OK, who are all the actors in this process and where are they? So I'll know. Yeah. Who the end users are and who is involved in the stages throughout the process.

Year of the service that we're working on.

Interviewer
OK, that leads me nicely into the second block of questions, which is around the X. Your experience with using personas.

Participant 4
Mm hmm.

Interviewer
So do you use personas as part of your projects or have you in the past, and if not, why not? Is there an alternative that you use or something like that?

Participant 4
Yeah, typically, typically we don't use personas.

Now this is kind of like a Gov design thing and how it should be for everyone, so personas can sometimes be challenging to use in Gov because they end up being quite narrow. Also, there's kind of like a challenge, an additional challenge that we have in government where we might be working on new innovative digital products. But the people often leading those things are maybe like long time civil servants who perhaps haven't worked in like a digital fashion.

And personas can sometimes be the wrong tool to use in situations where we have people with a lack of digital experience because they can sometimes latch on to that a single persona. That single idea of a person.

And that could be detrimental to what we ultimately try to design or navigate through. So instead we would lean towards user needs instead.

And they come from a variety of different sources and data points and research and stuff to inform those. So I would inform them perhaps the same way I would as a persona, but I wouldn't draw an individual. Say, for example, if that makes sense.

Interviewer
Yeah.

Participant 4
I may have in the in the only kind of like caveat to that is if I'm doing something internally, so I'm doing something for colleagues or civil service or for civil servants. So maybe someone who works in a job centre for example.

Then I might have something which in some way represents a persona, but I'll probably define it like that. An example is I might refer to them as their job role and what they do in that role and what they what that role needs. That's probably like the closest thing to a persona I might have.

OK.

Interviewer
And is that typical of all the other roles that you've been in in other employments before this one?

Participant 4
I think in more commercial organisations I've used personas.

Or like market demographics?

Age groups.

Typically in commercial organisations you're you know you're designing for gens. So like this is for Gen Z. This is for this. This is for that you might do that. So yeah, perhaps in commercial organisations will have more of a sort of in Gov, not so much.

Interviewer
OK.

And so you talked about the currently you're working with user needs as instead of a persona sort of style.

What type of details or information about the audience user do you currently find helpful to help move your design or development forward?

Participant 4
When we're working with views and AIDS there, I need them to be as informed as possible. I need them to be almost bulletproof in a way like what? What is the need that a person has? And why do they need it?

That helps us then make decisions.

And consider different angles and things like that. We might not talked about if we design around the needs of someone. We can then make decisions about the progress of something.

That makes sense.

Interviewer
Yeah.

Participant 4
Yeah.

Interviewer
OK.

I.

You've already mentioned that, so we can skip that one. So in the next block of questions are talking around the methods of creating your user profiles and things like that. So how and when do you or your team usually create them user profiles? What's the process?

Participant 4
So it's typically right at the start of the project. So I'll give you 2 examples if we have, if we're doing something for colleagues, for other civil servants and things like that.

We might work with a subject matter expert, so that could be someone who works in operations, or someone who is in what we consider a frontline role and we might work with them to define the roles of the colleague that we're designing for, the type of colleagues who will work with them. That's typically a person who's got a vast amount of work experience, work experience.

You know, has been doing the job for a while, knows it in and out, maybe with one or several people like that to define. OK, who is the colleague? What do they do? And then we'll also work with the colleagues themselves that that end user to identify what their needs are. That could be something like.

Just to start off, you know we do like a group session interview to learn about the role and that type of thing.

Something as simple as that or I get on a call with someone and just chat some about their job and they tell me about it.

When it comes to working with the public again, so if we're doing a public facing service right from the outset, again we want to try and define that person as best we can.

That can be informed by past research things we already know. Again, maybe subject matter experts that work as a bit of a proxy that have dealt with a particular type of user.

But we would be it's probably it's already in the idea and that need as we go along as we learn more about the person as we learn more about the audience, we would iterate and add to the things that they need to help us like shape the idea of.

What the needs of that, that audience are and what we need to design for.

Interviewer
OK.

Have you ever come across this concept of data-driven personas?

And what are your thoughts about it?

Participant 4
I think we've spoken about it before, but.

Probably outside, outside of maybe conversations that I've had with you or maybe some things I've read online. No, not much.

Interviewer
OK.

Participant 4
OK.

Interviewer
So.

Participant 4
You.

Interviewer
My PhD is or is aiming towards coming up with this new concept of.

How to get a user profile early on in a project that then enables people to use that as like a launchpad. If you'd like to then go on and create a persona.

One of the methods that I'm looking at, in particular in, in relation to this is something called an information seeking cluster profile.

And that is where on an existing project we'd run or devise a quick questionnaire, we'd run the questionnaire really early on in the project within the first couple of days.

Gather as much information from existing customers users as we could.

Once we've got that survey results, then we throw it at a system that I've built which does clustering, so it does a cluster analysis on these survey responses.

Basically, to get rid of all of the bias when we manually start to analyse data when we start grading them so the clustering automatically comes up with these groupings or profiles if you like.

The.

Clustering we can explore with it, so we can come up with multiple different levels if you like, of how many clusters we want to work with.

The clustering, however, has a downside that it will always come up with an answer, a grouping. Now that grouping could be, should we say have overlaps. Now I know no persona is is ideal. You know what I mean when we create personas, they're never totally unique. There's always some overlap.

And so in the system I've allocated a way to sort of show which ones are cleaner clusters or cleaner profiles, less overlap between them.

And which ones are and also to identify what significant?

About them groupings.

To make them different from the others, or what? What separates the two?

So I'm going to share my screen with you if you don't mind.

If it's gonna let me.

There we go.

Can we see that now?

Participant 4
Yeah, it's come through.

Interviewer
OK, so the idea here is this data that I'm going to show you is a figma mockup. It's live data that I have run through my system, but it's just prettier in this version compared to what I'd coded in the original.

There's a template here for museums. The work I'm looking at here was a survey I run with the museums in Liverpool, and so the survey template I provide for the museums that could use it. But this mean doesn't mean that you'd have to like for your line of business. These questions probably wouldn't be relevant, but you we could design our own questionnaire and then upload that and it would do exactly the same and present them in a similar fashion for your questionnaires, if you like. So we download it, we upload it. I'm just going to use the sample data, which is the museum stuff.

Instantly, when we open, we get the instructions on what's going on in this system and how to use it.

Well, mostly this is just explaining that these profiles here don't necessarily. Each element doesn't necessarily add up to 100%.

So in clustering, normally a cluster is based around something called a centroid or a median, and what you'd normally do in clustering is take that one person in the middle's results and say this is the typical person.

But it's not actually true, because in clustering that we've got a lot of variance in each category. So for like age here.

I've taken what makes up 75% or over.

What are the answers that make up that 75% or over for that column?

Just to make it so that what we're actually getting is a template of the most typical elements for each of these groupings, if that makes sense.

Participant 4
So is that why there's kind of like a large distribution between 18 to 34 then 35 to 54 and so on? Then it's a shorter one of 55 to 64. Is that kind of what that means?

Sure. Yeah.

Yeah, I think.

Interviewer
Yeah, this these are just the results that were done. They're only what make up to 7575% are over in that column. So it's not going to be over 100%, but it might not, you know it could be anything between 75 and 100 in each of these categories just to be show you what's representative of that column, if you like, you think of it as a spreadsheet.

Participant 4
Yeah.

Interviewer
So in these profiles, obviously we get a breakdown of how many participants there are.

From the overall group, which is 57% and there's 293 in this group in.

And then the questions in the questionnaire were broken down between demographic goals and motivations. The knowledge levels of the participants and the behaviours.

The significant parts that I was talking about before and what makes these stand out are highlighted with these green borders.

So these are the columns that really define this profile. So the reason for visiting.

The task motivations the information seeking the pages viewed, and the frequency of visit are what makes this profile different and these are the significant.

Categories that make this unique.

So yeah, it gives you all these details and when we flip to the other one, you'll see that the because there's only the two profiles on this one which was clean.

The categories are the same.

On the significance, but you can see that there's dramatic difference between them.

And then to explore the clustering, we can use the drop down and the colour coding here gives us an idea around the cleanliness of the OR the separation of these profiles. So the Greens are well separated, the yellows are passable, with a little bit of overlap, but the Reds are got significant overlap and it's not advised to use them.

And if I just want to explore 4 profiles for example.

Coming into the four, you can see that we've got significantly less participants within the profile.

The data changes a little bit, so in this one we've got.

Nearly almost equal between female male with the females just taking the lead a little bit.

The predominantly from Merseyside area, although there is a bit of rough from around the world and a bit from England.

A good spread across the reasons for visiting some for personal or the majority should have, say for personal reasons, but there are people there to pass time, some to study.

No one to come to work.

And couple with other unique reasons and all the unique reasons were individuals there was, they were all one offs if you like.

The domain knowledge you can see has predominantly novice and intermediate.

And the information seeking is that they're predominantly prefer to search over browsing.

And they will usually only look at one page in the website.

Whereas if we go to the next one, you can see this prominently from around the world.

A lot more mail. Sorry. Female dominance in that group. The age group is slightly higher between 3554.

A good spread across all the different employment types.

And then when we get to the significant parts, it's really down to the high general cultural heritage knowledge.

This group prefer to browse.

And they will look at more pages between generally between 2 and 10 pages within the website per visit. And when we come to searching, they'll try quite a few times before giving up, but they won't persist until they get an answer.

And they really come on the pcpc dominant group.

And then going into third group, you can see the changes here that most demographic data that's significant.

This is a male dominant group, quite young, but also quite old. We've got a good split between the very young, the 18 to 34, and the 60 fives to 70 fours and then that is split again between the retired group is the most dominant there.

They're all local or a good junk chunk of them are local, shall we say, and that then reflects down into the reasons in that they come in for passing time and for personal reasons. I ideally planning a visit which maps we've been local as well.

General cultural heritage is quite high, although the domain knowledge which is the knowledge for the specific reason that they visited the website when the study was taken.

Is based on mostly intermediate knowledge.

When I search there will be a lot more persistent to get to. The answer, probably cause they've got more time.

But most of them are first time visitors or come yearly.

Again, in the pages viewed, it's sort of 50, well, just short of 5050 split between visiting the single page or visiting 2 to 10 pages within the site.

And then in the very last group, you can see that we're back to a very female dominated, very local to the museums, but still in mostly in work.

Significant are the.

Gender reasons for visiting with 90% wanting to come for personal reasons, you can see the pre visit dominates about 84%.

General knowledge is spread right across the board here, with most sitting in the middle.

And then the devices will see the tablets and the mobiles take up quite a huge chunk of this compared to all the other groups.

Yeah. So that's sort of a whirlwind tour of how we can get around these profiles and stuff.

What are your initial thoughts on the ISCPS and the process? The UX process of having this done really early and no manual analysis essentially?

Participant 4
Yeah, I think it's interesting. I was thinking so can I ask you a question?

Does your data comes from someone completing a survey? Is that correct?

Interviewer
Yeah, the survey was run on a museum website and it run for four weeks, so it's totally.

You know, there's no pressure, there's no bias, there's no emailing. There's nothing. It's just they were on the website. So we know that they're actually users and they volunteered. So, OK, there could be slight bias in there in the.

There's some people who will naturally want to fill in surveys and there'll be others that naturally want, so we could have missed part of that audience, but it's.

No, he's put it out there and it there's no pressure on people to fill it in, should I say.

Participant 4
Sure. And so I'm thinking.

Some of the some of that I'm thinking about how I would apply this to Universal Credit, right. I think some of that demographic data like we just have from the you know because of the nature of what we do, people apply to Universal Credit. So we take a lot of that stuff in about their age and about you know where they might be living and those things like that.

And then on the on the web itself, we have maybe.

Analytics data that can give us some of this stuff.

But it doesn't. It's a bit guesswork so, but what I think when I look at this is I think I feel like this could work well in a physical space and what I mean by that is like it could go into job centres to ask people why are they there that day.

Is it information seeking? Is it to you know, for an appointment? Is it to apply for something?

And to give a metric of the type of person that's there.

And find out a bit more about their behaviours and needs.

Essentially.

So I think it could be good in, in. Yeah, in a physical space where someone's having to be there and interact with something. I think it could be really beneficial in that space. And you could probably do that for museums too, right? You know, like people could fill this in on the way, weigh in or whatever. So yeah, I think I think that would be interesting to learn a bit more about.

The needs of people, yeah.

Interviewer
OK.

Participant 4
I would say perhaps for if this was.

Hmm, if this was for Gov.

If I was using this in Gov.

I wonder how we would approach the consent. I'm sure you've got a consent process, but I wonder what that would be like for this. It made me think about that.

Right.

Interviewer
Yeah. So we're with this survey, there was an initial consent question up front with an ethics statement that they had to agree to before they got to fill in any of the questions.

It is. It's part of a university's PhD study. So as you can imagine, the ethics is very high.

Participant 4
Yeah. Yeah, and OK. So I think then for another thing to kind of like make I would need this in Gov. The thing I would need on here to change is the gender section because at the moment there's two there and in Gov, we tried to represent others too. So I think that would be the one addition I would like to see here is that's slightly expanded or allow people to self-describe or something like that.

Interviewer
There actually was, you know, in the question there was all the options.

It's just that when this report only reports what makes up 75% of the column.

Participant 4
Oh, I understand. Ah, I see. Sure. Right. OK.

Interviewer
And so when this is reporting the female male.

That there just there was no.

Participant 4
Alternative there I guess that.

Interviewer
So the numbers were so low on the others that it doesn't go into making up that 75% if you know what I mean.

Participant 4
Sure. I do. Yeah, I understand what you're saying. OK, alright. Well, that's really good. That's really positive. So yeah. Yeah, I think I think something like this could be really useful in in a physical space to help. Help me on as a researcher on as a design team understand why people are going into these things and what their needs and wants are. And I think there's an application here. Yeah, of just learning a bit more about people.

Interviewer

OK. And do you think these would be useful from the context of taking these and then applying more use of research from these at the beginning of a project? So to give you some sort of outline to maybe plan your next bit of user research to go and create these or turn these into your user profiles or personas in you know if you use them to go and turn them into a full-blown personas with all of that extra qualitative data?

Participant 4
Yeah, I do, because one of the methods I would do is to learn like what? To inform my research. I want to know what we already know.

And I like looking at data that we have about.

Maybe the way people interact them with a service online or whatever it might be, and some of that I'll then use to inform how I write a discussion guide, for example.

I'll also do that with. I might do a survey first if I just need to know a bit more about someone.

Or about a particular user group or something, and then I'll build a more in-depth discussion guide at that point. And I think this is something that, yeah, I could use to understand those behaviours a little bit. So this might give me what people are doing and that sort of stuff. And then I could apply that then to dig a bit deeper, to understand maybe why they're doing some of those things from a research perspective. So yeah, something like this could be good for that.

OK.

Yeah, yeah.

Interviewer
OK, I'm just gonna stop sharing that screen then, and I've got a couple more questions just to finish off, if you don't mind. So.

Beyond personas and user needs, profiles that you're currently doing, are there any other methods or tools that you currently use or are possibly exploring in order to understand users or your audiences better?

Participant 4
And Google Analytics would be 1.

Just general analytics. Yeah, Google Analytics and just and just sort of general usage of certain features we might track that they'd be used with internal tools and things like that.

Interviewer
I don't want you doing that.

Participant 4
I could go into the detail about Universal Credit's constructed, but I'd probably be breaking some sort of confidentiality thing aside, then there no, but, but we can we look at sort of data of certain tasks that colleagues might get assigned and then what we do with that and that sort of stuff.

And it's not. It's for the purpose of understanding what's moving through the system, not the purpose of like tracking or holding colleagues accountable. It's not like that. But yeah, we might look at general analytics and that sort of stuff.

Do we use anything else?

I mean, yeah, generally it just comes to understanding people who I might use, like, if any academic studies have been done or things like that, that type of thing. So forget for gathering data points. Yeah, we probably just use surveys, analytics and that type of stuff. Yeah.

Interviewer
Right.

So in UX and web development in general, obviously everything's constantly evolving.

Not only the tools, but on the methods and everything else.

So how do you see the future of user profiling in UX design?

Participant 4
Interesting question.

Already in commercial organisations we know a large amount about the people.

That use.

Products, when we when we talk about commercial organisations and commercial organisations, large ones will pay money to learn even more about them, right, because it's all market driven target driven that sort of stuff.

I can't see that slowing down unless the EU make more privacy laws or.

Whatever. So that that'll probably ramp up.

In Gov.

We have more restrictions and it's a bit slower.

But.

Hmm, we build kind of like research repositories, and the helpers understand users over time and I think that that'll probably grow and get more mature and better in that sort of stuff and tools like you're describing would be good for those types of things.

So yeah, it's a tough question. It's a tough question. I feel like I could probably we could probably talk about like privacy and things like that and go into a big conversation regarding that when we talk about commercial organisations. But I don't know if that I don't know if we've got the sign.

Interviewer
No, that's OK.

Oh, thanks for that.

Finally, do you think that I've missed anything?

In our talk or is there anything else that you'd like to add to anything?

Participant 4
No, I don't think so. I I think it would be really interesting to see that applied in a physical presence. You know, whether it be at an environment, yeah, like a museum where someone's interacting with something. How does that work in that environment? How do you understand what a person's there for and what they want out of it and what their motivations and needs are? I think that could really help you then shape and design things in a physical sense. Like, it'd be interesting, you know, to play something like that in a library and then help, you know, improve the design of the library and the navigation of the library to find that type of thing. So

I think it's got some relocation in like a physical space. I think in a digital space.

There's these tools that can do some of things, but not quite as much. So yeah, it'd be interesting to see what you can do with it. Be really good.

Alright, just.

Interviewer
Yeah. So on one side of the research that I have already looked at and published was the user profiles of the physical space similar to the user profiles in the digital space because you think a museum user who goes to the physical museum?

Would be the same person you know would be the same type of person or the same profile in the digital space.

Participant 4
Hmm.

Interviewer
Actually, the research shows that whilst a lot of the people.

Come to the museum to start to plan a visit. There are also.

Certain groups of users.

Who will only ever come to the digital space and never visit the visit the physical space.

The other part of the finding was that.

When?

If you look at how the same user you know if the same person would go into places, their actual the profiles the way that they seek information and the way that the information that they want actually changes between the digital and the physical space.

So you know, if you just if you, if you surveyed them in the two spaces for the same people, it would look totally like two different groups, even though it's the same person.

And there's a there's a lot of research already done inside of museums, physically.

Participant 4
Hmm.

Interviewer
But not to this level where they get this sort of level of data.

And especially not unbiased because it's all manually curated afterwards and then they come up with these unique 5 groupings generally but.

But yeah, there was no work done in the digital space for museums, and that's where there's a lot of work in, you know, the moment is over 90% drop off rate in museum visitors.

It bounced right. So that's where the research sort of started is well, why?

Umm.

I don't know a lot of the presumptions about the users were that they were more professional, they knew what they were doing on the museum space. You know, they might have been academics, so they knew what they were going to search for and everything. And what I found is that the majority audience is just general public. They don't know what's there. So the search box doesn't serve them as much if you know what I mean.

Participant 4
Yeah, yeah.

Interviewer
So yeah, that's, that's where it all sort of stems from. And that's why in this case, I studied the digital space.

Participant 4
M.

Interviewer
But it's interesting that you say it would work in the physical space as well, cause that is useful to consider or never considered that.

Participant 4
Hmm. Yeah, because it'd be good.

Interviewer
OK.

Participant 4
No worries.

Interviewer
Well, thank you for that.

I'm gonna stop the transcribing there, if you don't mind.

J.2.2.5 Participant 5

Interviewer
OK, so let's start. The first section is all around your background and role and stuff. So what would you call your current role or career?

Participant 5
My title is front end software engineer.

But due to the size of the company, I'm not. I'm not just a best, you know, just developer developing the front end.

And will take a lot of other responsibilities, mainly UX, UI, designing, research.

Bits of project management as well testing.

So yeah, because of the size of our company.

Does only around six devs at our company. So yeah, it's almost a start kind of vibe, if you will, to the company. So we do have to work a few hats, if you will.

Interviewer
OK.

So this this question might seem a little bit strange, but can you sort of?

Describe what are the standard.

Day, week or maybe design phase looks like for you.

Participant 5
Yeah, with a reason. I would go after recent project that we've done.

Because that involved a lot of hands on with the customer right from the beginning.

With those guys, when it came to the user research at the beginning, that was a lot of interviews with people over teams or Skype or zoom, whichever platform people were using.

Usually teams and it was just interviewing people gaining use requirements.

In user stories.

Design requirements all host of things really to inform the architecture of visually and structurally of the build.

That was done by myself and one of my colleagues.

She's referred to as a solutions engineer, but again, there's quite a few hubs.

Yeah. So that phase and process was interviews and then drawn up. So as all you to research materials that then informed low level designs and then high level figure prototypes that would then roll out in further teams.

Reviews video meetings indoors that would then bring the wider team from the company in getting more feedback on in the prototypes. All the feedback and then going on to develop once a partner is agreed that.

Interviewer
And these are typical sort of workflow on most projects roughly.

Participant 5
Yeah, yeah.

There's a focus going on, user experience now more and more, which is good.

This is because when I first started a company we had few projects.

Were certain individuals in our company saw user experience and user interface design and research as.

Well, it's probably about they probably some kind of pseudo science if you will or something without merit.

The outcome of those projects and how they gravelled taught them a lesson.

Going forwards so now with every project we have, there is a UXUI research and development and design process at the beginning for a couple of weeks.

Interviewer
I'm not sure if you have a typical type of audience, but what? What type of audiences do you usually work with?

Participant 5
It's usually when we're doing a video calls.

There will always be a project manager. There'll be a team leader of a team, there'll be 2 to 3 individuals from the team who we can see as the end users, if you will.

And then from my side it will be myself one of the solutions engineers and then maybe a higher up who will be in the meetings.

That's kind of the audience we have, I would say, probably young professionals.

Administrators, you know, a breadth of people, really.

One of the projects currently is with a mileage that name of companies.

Interviewer
Yeah, I will call them out.

Participant 5
OK. Yeah.

One of the projects is with [Company].

We went up to their office in Sheffield, had a meeting with them.

That involved a series of 1 hour interviews with different people coming in and going out and people joining remotely. All the teams that ranged from, you know, people who sit on the phones taking calls and making calls. People do an administrative paperwork.

People going out and speaking to homeless people in person.

But that also then went to team leaders, to managers, all the way up to the over one of the overall directors.

So that was a complete, you know, the pyramid of and hierarchy of the workers there.

With some other companies.

Mainly with [COMPANY], who working with.

And a couple of private companies, it would just be a team leader, a solutions engineer, project manager in two to three team members who would have in our kind of meetings.

Interviewer
OK. And who are the?

Who are the systems that you build in's end users? If you like? Are they the same people, or are they the public general public sort of users?

Participant 5
Mixture really mixture? We have to obviously with the [Company]. One we can't we can't get hands on interviews with the users from you know I think we refer to them as like a person of concern.

So anyone who's homeless or faces any issue with housing or anything we can.

Interview those people. Obviously due to a whole manner of, you know, really gritty.

Issues.

But the people that they speak to at [Company], we can glean research and information and data off.

With some other projects with [COMPANY], it was actually sort of what's up project to construct conversations that will be used for refugees.

Obviously, again, we can get direct access refugees.

Because it's all kind of completely anonymous service that they use up to a point.

But we again do speak to the people who are conversing and communicating and meeting with refugees and getting to their information from those guys to build out so as to build out, design the requirements and everything.

But some other projects, the more I would say not public body but private sector private sector projects.

You know there won't be a public technically using the service, so the end users are the actual administrators or you know people that would that way.

Speaking to.

Interviewer
OK.

So.

When you're starting a new project, how well are you or how well can you define the audience within them? Them early phases of the project?

Participant 5
Can't find the length of the phase really.

And also depends upon the I would say the openness of the customer.

Or the you know, the whoever doing the project with.

You'll always find that some customers are very open. They'll be very honest. They'll state what grammes are up front and what they're expecting from a system.

And also we found that dependent upon the level of technical knowledge.

But the customer or the company?

That informs how much information you can get out of them. A lot of companies will speak to you, want a lot of hand holding. They want to be told what they need in a way, and it's very hard to get them to open up.

And give you actual insight for you to be able to provide what they need. So it's really we found no, it's about really delving and meeting people.

First, versus much we come unfortunately with some other companies, it's been our teams that's both getting faces on cameras and real estate in you know and gaining proper feedback. We're lucky that with companies like [Company] and [COMPANY], these public body companies, they've got a whole range of openness about, you know, there's no stupid questions. Everybody can be open, can communicate openly. Any thoughts can be brought to meetings. So if those customers, it's very easy.

But with the private ones. Yeah, it can be a bit harder to get them to open up.

It's almost like with the public body companies you can get your answers in words face to face or recall, but with the more private companies, almost seems like you've got to do a structural bureaucracy and send word documents, receive things back, and you've got to really put a fin structured manner to get in what you need back.

Interviewer
OK, I get that.

Participant 5
Is what we've, you know, I found personally.

Interviewer
Umm.

So you've already mentioned a little bit of this in in a couple of the other questions, but do you use personas as part of your projects?

Participant 5
Yes. Yeah we do.

Completely informed, as you say about that opening fitness in the user research. And again.

Dependent upon.

Who we're getting access to.

Is how well we can flash out and build those personas.

Interviewer
OK, So what sort of details would you have in them then?

Participant 5
We'd have, you know, if it's a basic system.

Like.

To a lot of what we do is through Twilio, so it's selling the Flex platform which is call centre. So we do a lot of car companies or a lot of educational sales companies. So that personas are quite basic you know needs and wants pinpoints desires they're quite.

Basic in what they are and the demographic details of those as well. Some of those companies tend to have a very, very narrow demographic, if you will. When it comes to age.

Yeah, because you've got, you know, say a call centre operator.

Generally, the ages from 18 to 35.

The mail, because it's a sales based role and I'm not concerned quite sexist, but it's just what we've found in our research.

Whereas the personas, when it comes to like the more complex projects like you, you know, [COMPANY], you [Company], you've got much more personas because there's much more people and touch points on the system. But there's also different requirements for the demographics are much wider.

The pinpoints.

Completely wide open and in depth.

Because obviously you're dealing with people of concern in both projects.

And those people of concern on both those projects are [Company]. [COMPANY] are facing life or death situations day-to-day.

Interviewer
Try.

Participant 5
So those personas have been very, very in depth and very numerous.

Interviewer
OK. And you mentioned earlier about doing all of the interviews and everything, so.

Are personas created?

Manually. Essentially you know, doing a lot of time and effort in getting the interviews and then manually creating them personas.

Participant 5
Yes. Yeah, yeah, yeah.

Interviewer
OK.

How are the personas then? You once you've created them, how are they then used throughout the your workflow?

Participant 5
They're always there, always referred back to every stage because essentially we're designing for users. We're not designing for ourselves.

Again, marketing it back to previous projects.

There's one or two developers in our company with the old school mentality of, you know.

Build a system that will use it and then the issues are found post.

Even deployment and going live.

And you're going back and doing the process. You should be doing at the beginning. The whole user research, the whole design, development, feedback, requirements gathering, all, all those things that you should be doing again, then get done when a system's gone previously, which was completely backwards. So now the personas are referred to at every stage because we have to be aware of, you know, who's using the system. That's who we're building it for. So, you know, that informs.

The whole research and design phase at the beginning.

You know informs or build, and then even informs testing as well because.

Where we do internal UAT and then hand it out for UAT as well to the customer. So to be able to perform the internal UAT, we need to know who the users are, how they're going to use the system, what they will be looking for, what needs and wants they have, what pinpoints they have. You know we have to ensure that all of that is met at that testing phase from the previous phases. We've been using the Sonus.

So that when we do hand it over.

The users are testing and using the system that they actually need and want.

Interviewer
OK, cool. So just delve in a little bit more into the methods that you use to create these personas. You've already mentioned that you do a lot of manual, you know, interviews and manual creation of them.

For those that you can gain contact to.

Do you use like third parties for?

The users that.

You can't, you know, the more vulnerable users that you've mentioned that you can't only get access to. So in terms of things like [Company], do you talk to the case workers maybe.

To sort of build the profile of the and vulnerable users.

Participant 5
Yeah. When it comes to companies with where it'll be involving any PII in an end system, and it personally identifiable information or people of concern we've dealt with, obviously [Company] is vulnerable homeless people.

Or displaced people. Same with [COMPANY]. We've done a big project for a company in America that was based around that advice and help. But through a public body.

So none of these projects we have access to the end user end users. So the technically the third parties were relying on our case workers.

Umm any kind of I would say that irony lists their companies, but just gleaming general information but not, you know, specific names and things like that. So to build the personas for those vulnerable end users or where details will be personally identifiable. And it would, you know, cause GDPR issues and things we do speak to the case workers, the end users, the data analysts to then build out the personas for.

Real World End users.

Outside of a company?

Interviewer
OK.

So the types of data that you'd usually base your personas on. Then obviously you've got your interviews and your third party interviews and stuff. Is there any other types of data that you would use or is that you know is that it? Is that basically where you'd farm everything from?

Participant 5
Sometimes some of the companies they are replacing one system with another.

The recent [COMPANY] one they were replacing a service they used called turned IO.

That constructed what conversations and they were wanting us to build a bespoke platform for them through Twilio.

So they gave the solutions engineer access to turn the IO to see.

A like a selection of their data that they copied across for us for expunge details from it that were very specific.

So we use that kind of information to then further inform building the personas out.

Yeah. Any kind of data we can grab from a previous system when we're replacing it is really good.

But mostly for outside of those public bodies.

It's usually the call cent stuff that we're selling, so it's not really.

Interviewer
OK.

Participant 5
A pool of data that we can pull from the not really willing to share it because it shows it shows usage amounts and things and goes into the financials of you know that could influence how much we charge them for solutions. So they're not really willing to reveal too much.

Interviewer
OK. Have you ever heard of the term data-driven personas and if so, what are your thoughts on them?

Participant 5
Yes.

We have heard of them as I just said them in the previous question.

Any of the companies where we are replacing one system with another, a lot of data is provided to us. If we can't get access to the absolute end users outside of a company.

Interviewer
OK.

Participant 5
So that data is what informs our building of personas. That's where I come to understand that as being.

Interviewer
Two. OK. So the next section is really moving into the bit of my PhD that I've come up with this concept and I'm exploring a new early survey data-driven type of user profiling method.

And it ends up creating a sort of pre persona. If you like. It's not quite a persona. It doesn't have all of the.

Qualitative data that you'd get from interviews to go into there.

And it's all obviously surveys. That's quantitative data that's informed on it.

The thing that I'm gonna show you in a minute is based on.

Museums.

Data from work that I've done previously.

And so there was a questionnaire that was put out to the audience via the website. We collected data for a few weeks and then when we closed it, what I end up with is a CSV file of interview questions. And the idea is that you could do something similar for your users on existing systems. You can pile a survey for your people, gather that information.

And then just get your CSV back and upload it to my system and the system would.

Participant 5
Yeah.

Interviewer
Automatically clean the data automatically, do a cluster analysis on that data and basically separate out all of the users and come up with groupings on its own. Now there is a caveat to clustering in that it will always come up with some groupings and so the within the interface that I'm going to show you in a minute, you'll see that.

Participant 5
Yeah.

Yeah.

Interviewer
There's a way that we can see how clean or separate them groupings are, because there's always some overlap whenever we create owners, but we don't want too much overlap if you know what I mean.

We, the other thing with clustering is normally in a cluster, we group and then we have this thing called the centroid in the cluster and that for in terms of the questionnaire that we're I would have looked at you would take that one person's answers across the board.

And say this is representative of the cluster.

But it's actually not really that true in the case of what I've seen in the clustering. So I've also come up with a heuristic, and so we do the clustering and instead of taking the centroid data and then just trusting that that is true, I go with it. If you imagine all of the answers from each cluster put into a an Excel spreadsheet, I take each column and basically work out what is up to 75%, what makes up 75% of this. And it might be numerous factors.

There might be numerous options that are joined together to make up that 75%.

Participant 5
Yeah.

Interviewer
But I try and show at least that 75% of what's making up each of the factors. So we're not just taking one person's viewpoint and saying they are the pinnacle like we would normally. When we're doing personas.

But that's the idea, and the idea is that we do this like in the first week or so of a project. So it's really early on and we're getting these profiles there. So I'm just going to share my screen with you if you're OK with that.

Participant 5
Yep.

Interviewer
Sure, that one.

There we go.

Can we see that?

Participant 5
Yeah.

Interviewer
Yeah, cool. OK, so.

Participant 5
Yeah.

Interviewer

Here I've provided a template which other museums could use for yourself in with non museum work you could create your own questionnaire and stuff like that and then upload it. I'm just going to use the sample data because it's already built in.

Well, you'll notice this is a figma interface. All the data that is in it is actually from my system, though I just haven't had the time to make the system look prettier as what I'm working with here.

Participant 5
Mm hmm.

Interviewer
So when we load the data initially we get the instructions that just inform how.

What everything is and how everything works.

And you'll see that we get 2 profiles, which is a default setting, so it doesn't always come up with two profiles. It just comes up with two because these are the first clean set of clusters.

Participant 5
Yeah.

Interviewer
Inside of these profiles, you see that we get a description. We get more of an infographic version of that, so here we're saying that this group is 293 participants, which is 57% of the total sample.

We're gathering some demographic data, some goals and motivations data.

Some knowledge level data and then the user's behaviours. So they're the four main areas that we're collecting the data on.

And the other thing that when you get the clusters?

Participant 5
Yeah.

Interviewer
You also get indicators of what is defining this cluster, so there's certain columns of data that we'll be defining this and that I've shown these with these green borders. So you see some of the internal panels have the green borders on them and these are the ones that are really making this profile different to the next one.

Participant 5
Yeah.

Yeah.

Interviewer
OK. So just quickly go through the types of data that we've got, you've you get a normal age gender, it does collect more than just the female male, it's just that these make up the 75% and all the other genders didn't really fill in there and we've got employment data that's in there and their location. So what you can see with this one is that they're predominantly in Merseyside, really close to Liverpool museums.

Mostly female, with an age of 35 to 54.

There's nothing significant in there, so it is just very similar, maybe similar to the other profile in the goals and motivations we've got entry museums and galleries. We've got a two significant areas then the reason for visiting.

And in here we can see that over half come for personal reasons, 18% come for pad just past time.

Nearly 15% come to study and just under 12% come for other reasons, but nobody comes for work reasons.

And then if we look at what tasks or what motivations that they're actually coming for?

A lot of coming for pre visit.

A lot of coming for other reasons. The other category is.

Participant 5
Yeah.

Interviewer
Basically is everyone's give me a one off and there's no catch all category that you can put them into. So there's a lot of people coming for their own unique reasons here, but pre visit is the predominant catch all category if you like.

Participant 5
Yeah.

Interviewer
In the knowledge levels, there's nothing significant here and we can see that in the general cultural heritage knowledge. It's predominantly in the middle sort of intermediate sort of levels.

And that is also backed up on the domain knowledge, which is the main reason that they've come for this visit when they filled in the survey and you see there that the intermediate is the sort of the main.

Category if you like that they would place themselves into.

Participant 5
Yeah.

Interviewer
In the behaviours we can, we've got three significant areas. So we've got the information seeking preference and we can see here that it's almost a 5050 between the users who prefer to search or the users that prefer to browse to gain their information.

The pages viewed most of them just view one single page on the website. When they come per visit and then if we look at the frequency of visits, it's really dominated by people who've been for the first time. So obviously.

Participant 5
Yeah.

Interviewer
Boarding for this profile would be really, really important, because never been before. They're only really going to look at one single page, so we need to make sure our SEO and our content is there and that we're gathering them.

And we don't know whether they're going to search or browse, so there's quite a lot of information in there.

In when they're actually searching, some will try a few times before giving up, and quite a few will just persist and try and get through it until they actually find the answer.

A lot of them are just gonna scan the content, not gonna read it in depth or look at the pictures and device wise we can see the unknown. We can't gather anything really from that, but a lot of using iOS.

Participant 5
Yeah.

Interviewer
So a lot of mobile use there coming in over anything else, even though they say that they're all, most of me using PCs, we do still have a 44% on a mobile or a tablet.

So it is quite close in there.

Participant 5
Yeah.

Interviewer
When we jump to the other profile, you can see that things change a little bit. The gender's still predominantly female, the age use is are pretty similar.

Participant 5
Yeah.

Interviewer
They split on the work and the employment is more full time people and there's more people who are local to Merseyside, so it's very, very similar to the other group in terms of how they're defining themselves.

And when we look at the significant parts, we see that the reason visiting massively changed. So it's no longer just 50% personal reasons is nearly 90%.

And nothing for passing time, nothing for study. Again, nothing for work, and almost 10% for all the reasons.

Participant 5
Yeah.

Interviewer
But the pre visit is the predominant reason that people come in here, so they're planning a visit to actually come which maps because they're local.

And they're looking for personal reasons so that they are looking to visit.

When we look at the difference in the information, so you can preferences browse 84% nearly.

So they really don't like to search, but when they do search they will try a few times and then give up.

And then they'll look at more pages this group, so they'll look between 2 and 10 pages per visit.

Participant 5
Yeah.

Interviewer
So they are definitely looking for information on what they can find, how they can find the museum, and we also see the change in the frequency of visits with the yearly being the highest.

So we still need to be on board because the first timers, but there are people who are coming back every year.

Participant 5
OK.

Interviewer
So they are like frequent visitors again in the platform 66% unknown, but again mobile 23% and that maps again into the devices it's very similar to the other group. So this is the default profiles that he provides. But in the drop down because it's clustered we can.

Play around with numbers 'cause. We don't always wanna just go with what it gives us.

So in the drop down you can see we've got two profiles that are in green. Green means that these clusters are pretty clean. They're not much crossover between them.

Participant 5
Yeah.

Interviewer
The yellow shows that there's a little bit more crossover, but the red shows that there's a lot of crossover and we really don't advise using these. So if we just explore the other green one, the four profiles.

You can see that we're obviously a number of participants is now a lot lower because we're split into four groups.

But there's big differences now between what's actually going on. So in this group, we can see employment is nearly 50% full time employed.

The male female is a lot closer. The age group still got 35 to 55 categories, the dominant one. But Merseyside is really, really high in there.

Participant 5
Right.

Interviewer
And world is the 2nd, so world is everyone that's outside lives outside of the UK and accessing so that's quite a big shift compared to the other grouping is that we've got people who are local and people who are really, really far away.

When we look into the goals and motivations, we see that the reason for visiting is still significant and the tasks and motivations is significant.

When we look at what actually the breakdown is here, we got personal reasons and then study is the 2nd. So that's drastically different to the groups in the when we look at two groups.

Pre visit is quite high in there and other which will be you know the things so they're very similar sort of waiting now compared to the other being slightly more dominant in two groups for this one.

Domain knowledge has become significant in the knowledge level in the lot of novice and intermediate levels and very little expert.

The information so you can profile, but this is predominantly search people and people are gonna look and come and visit one page.

So that's this first group. When we go to the second, we can see this changes quite a lot. So we go straight to world is the most dominant.

Grouping still got some full time, but there's a lot of students in here, a lot of retired and part time people, predominantly female.

The 35 to 54 is about 50% of the groupings.

And then we'll come in for personal reasons and to pass time again, pre visit is quite a big one. So this is people looking to visit from other countries.

And instead of the domain knowledge, now we've got the cultural general cultural heritage knowledge that's become significant and it's a lot higher. You know, we're looking from level 3-4 and five, which is the higher end. So a lot more knowledgeable generally.

Participant 5
Yeah.

Interviewer
Prefer to browse.

They will try a couple of times, then generally give up.

Use APC and they're gonna look at a few.

Just jumping into third group, we can see a big difference here. So genders massively changed. This male is dominant in this group. It's a lot younger. So we're going to look at 18 to 34 year olds and then also there's a big chunk of the older group in here. So 65 to 74.

Which maps then into the employment with mostly been retired.

And local is also now back on the cards here.

Participant 5
Yeah.

Interviewer
When we look at the other significant stuff, we've got domain knowledge which is mostly intermediate and this group are going to persist. They're going to search and keep going until they find the answer.

Or work out. There's really not going to happen.

And then the last group, we can see the gender now is pretty much all female in this group.

The reason for visiting is mostly personal reasons.

Predominantly, pre visit is the main reason they're coming.

The general had cultural heritage. Knowledge is middling to slightly high and then?

Here we see the big difference on the devices in the the mobile in this group becomes the predominant.

Way to access.

So that's.

Participant 5
Yeah.

Interviewer
Good tour if you like of the system, and obviously there's been no user interference in this or anything. It's all automatically generated and we can explore the groupings. So what are your thoughts on the iscps and the process of doing this so early on in a project through?

Qualitative data.

Participant 5
I mean, that's yeah. Like a lot better than what we do.

You know that could inform your requirements gathering stage when you're actually speaking to your customer. I'm getting because you have you have data that you're able to play with before the actual persona part. So this is almost like pre research as we would see at our company.

And yeah, I can see that would be very useful.

Interviewer
OK, so you can see this being used and.

Probably essentially being used then to profile who you're going to go and actually do the qualitative data with so could even save you some time and money.

Participant 5
Yeah, yeah, exactly, exactly.

Interviewer
OK, cool.

So do you think the iscps are something that you could potentially use in your work? Obviously there'd be a tweak to the questionnaire and how this is presented. But as long as you've got majority, it'd be some sort of thing. Do you think it's something that you might use or not within your own business or?

Participant 5
Me. Me personally. Yes, I would use it.

But yeah, I could I. Maybe there will be some resistance from certain people still. But yeah, personally I would use it. Yeah.

Interviewer
OK.

Participant 5
OK.

Interviewer
Cool. I'm just gonna stop sharing my screen there, and then we'd just got a few more questions before we sort of wrap up.

So with the iscps, do you think there's anything that you would add, change or drop?

To either the method or to the interface that would be of value at this early stage.

Make it more relevant to your work even.

Participant 5
No 'cause I think it was covering quite a wide range of demographic and platform. So that was all good.

Obviously you know the reasons thing and that would have to be tailored specifically towards the kind of industry or client or user base. That's the only thing I would change, but obviously that's a completely a part that needs to be completely customizable, but know all the other fields that fields and everything sections were you know, things that would definitely be required.

Personas and for the startup research process.

Interviewer
OK, cool.

Last two question or the last section should I say which is around feedback and future considerations. So beyond personas, are there any other methods or tools that you currently use or are potentially exploring in order to understand your users or the audience better?

Participant 5
N.

Was not do the whole prototyping process, so there's a lot of design.

Low level and then high level in that process.

We do get our hands. I mean, we do get our hands on some data, so that informs not only the, you know, the Persona building, but also how to zero in on any 10 points and issues or things for consideration when we're building a solution.

I know the speak around. I think it's probably a buzzword at the moment to really honest is AI. We're looking into a lot of like trying to use that.

More on the.

Side of things I would say rather than public facing bodies, we've got a lot of private companies that are basically just wanting to replace people with AI word again, but they don't seem to understand the drawbacks and the limitations of it.

We've stayed away from that from a research point of view or anything. We'd rather just do it human, if you will.

But no, I would say a lot.

Interviewer
But I know you mentioned user stories earlier.

How much sort of depth you go into with them?

Participant 5
I'll say quite a bit of that actually, because we're not just using them to inform the design process and the build process. We're using them to inform the internal UAT and then the external UAT at the end as well.

So yeah, you know, we use.

Use a journey mapping. Use a flow mapping.

And user stories kind of as a whole collective, if you will.

We're lucky that if some of the companies have their own.

Your teams so [Company] of their own UX team, so we can speak to those guys as well to gain and glean data and information from and share best practise. Really that really helps out when you've got a company you've got that but with some of the private companies it's not really a UX to interact with. So we're not to pick up all the legwork ourselves.

Interviewer
Yeah, I get it up. So obviously in the web world and especially in the UX world, things are always evolving and moving on extremely quickly.

How do you see the future of user profiling UX design?

Participant 5
Yeah, I could see from the system that you built. I can see it become a lot more automated and AI driven. I'm sure rather with the RCESPS system.

That's what I can see going forward. I can see it now when I'm using websites myself.

That you get in pop ups asking you to take surveys and things. It's becoming more and more used, if you will.

I can see that kind of combined with them with the whole.

Cookies are better. Hoovering that's done.

Then yeah, I can see that. Maybe the UX researcher in the future, instead of being a person, will probably be an AI that reviews data.

It's that. I'll be it. In personas, use stories, you know, heat website, heat maps and image. And you know all those kind of things. That's where I could see it probably going forward which.

Could be a good thing.

I still think there needs to be a human involved.

To delve through all that information because yeah, I use ChatGPT quite a lot of things and it's not perfect, it's just it's just a neural engine. It is an AI. Think that's where the confusion is coming for a lot of people, it's been marketed and batched as one thing, and it isn't.

Interviewer
OK. And then finally, do you think that I've missed anything or would you like to add anything to?

So this area of talking around personas and user profiling.

Participant 5
No, not really. I mean, from a personal experience, 'cause, I obviously work for a small company.

We did have a goal when I first started there, who was doing a great apprenticeship for user experience research and design.

She quit about a month 4-2 after I joined the company and that was down to a sport, you know, support issue and feeling wanted. Thankfully things have changed a little bit, but I do feel with smaller companies.

Ux is more important, but they tend to few it that it isn't. They just want to build and push out and make money.

Which you know, that's what you have to do as a start up or as a small company. But if you're pushing out products that haven't had, you know, any user input or research around them.

Then, as I've seen from previous old projects, a year or two ago, you'd gonna just encounter issues afterwards.

Interviewer
Yeah.

Participant 5
So you think you're saving money in the long term, but you're actually spending money further than a wine and more than if you've done the job properly first time?

That's personal.

Opinion.

Interviewer
That's OK. OK.

Well, thank you very much for your time.

Participant 5
No problem. OK.

Interviewer
I've got no more questions, so I'm just gonna stop the transcription. Now, if you don't mind.

Participant 5
Yeah, no worries.

J.2.2.6 Participant 6

Interviewer
What is your current role or job? How would you describe that?

Participant 6
Title is technical lead and digital media and publishing team, that's.

Manager of the development side of the digital media team. And that's primarily working on.

Main website at the museum, [Company] and then the collections website at the museum collections.[Company] and then lots of smaller websites and features.

For various parts of the website.

For different purposes attached to the activity of the museum.

Interviewer
Very busy. So what? What would a normal either a day, a week or if you work in design phases? What would a normal day, week or design phase sort of look like for you?

Participant 6
So we work in fortnightly sprints and lots of that would be working on deploying releases to either websites. The two main websites and then occasionally smaller updates for some of the other websites or features that would be either fixing bugs or adding new functionality or adding whole new features every so often.

We are not running the commercial side of the website, so we don't run the shop website, but we are connected to all of the other activities of where things go through the websites dealing with integrations and the problems come up with that is also a common part of the day or a week and then there is just lots of other.

Activity around projects, future projects.

Discussing issues.

That could come up with things people want to do in any way digitally.

Mainly where there's a website presence, but we also get involved in other aspects of digital and other departments in the museum that would respond to some of the that as well.

So yeah, so I guess it's, you know, a mixture of maintenance as with all web developments and then building new features and then discussing future projects.

Is probably a bit of yeah, there's no such thing as the average day, of course, but some a bit of that usually comes in.

Interviewer
Yeah, I appreciate that.

So on a project, do you have?

Maybe as a team rather than just yourself, but do you have like a typical workflow from?

You know, inception through to, to delivery or whatever.

Participant 6
Yeah, as much as possible. I mean trying to follow an agile to an extent.

So a project will come in for you know, various reasons it would go through definition, which is we have product managers who would try and work with whoever the projects where it's coming from to actually define the scope of the project. It would get prioritised.

For the department, based on the different criteria we have around.

What's needed, satisfying, you know, is it essential for various reasons for the museum to do this? And when does it need to be done by?

That would then go to our designers in the department who would work on if there is public facing design of it.

Umm to get the design inputs and we would collaborate with them as well. At that point on the engineering side, as we call it, to ensure that you know the we're feeding into the design to make sure.

It well, it's buildable to be honest, you know.

And then that would went, that's signed off by the product managers and the owner. If there is an owner of the project.

That would then come to development to build.

And then we'd go through build phases and until it's signed off and then it's deployed.

Interviewer
OK, thank you. I am.

So what type of audience do you normally?

Deliver for.

Participant 6
I guess.

Yeah, that, that's the big question.

We break down the two, the two main websites, I find it easier to always break it down into that.

The main website we have to think of as for the general audience primarily, so it's got to be things that are for everyone who is likely to be considering a visit to the physical museums, which we now say because we have multiple venues.

Um and the collections website, we would primarily think of people more in researcher mode or research mode, not necessarily researchers, sorry, but just people in an activity of research for whatever reason.

Excuse me.

So that's probably a bit how we would start thinking about functionality on either of those sites.

For other projects it would vary a lot depending on the project they would possibly be particularly aimed at a particular group or segment, so that would be more obvious about who that thing was being built for so.

If it's.

A digital humanities project that we might be collaborating on. Then obviously that would be targeted, say, at the people with an interest in more in that area.

So yeah, so I think.

That's.

The things built for the two main websites have two dominant different audiences. Not to say obviously either of them couldn't have a general or research audience, you know, vice versa. Researchers use the main website and the general public use the collections website. But we would be thinking more targeted for each of those.

Yeah. Or for other smaller projects there, they're more likely to have a particular target audience in mind.

Interviewer
Quite a.

Complex mixed up area in understanding these audiences.

Participant 6
Yeah.

Interviewer
When you start a project.

How well do you understand them audiences?

Participant 6
I guess that's part of the scoping of the projects and the product managers role is to define that and try and do and we have a user research part time user researcher that would be looking.

Into that.

If it's a project that's come towards from some in somewhere in the museum, that would also probably they would have some.

Area that they're already interested in, that they want to target, that's not necessarily true.

Sometimes people have ideas about the audience. They think that this is something for a general audience, and we might think it's not.

And vice versa. So yeah, so there's a bit of a mix, excuse me.

And that's part, I would say, of the definition of the projects to kind of establish what we're, what we're who we are building this for.

And to get that agreed with the before the project build starts.

Sorry.

Interviewer
And do you use personas for any of these? Do you use personas in order to document these user groups?

Participant 6
So for the collections website project we did use have used personas. I know there is like this is where slightly beyond my level of detail. I know there's different terminology for these things.

Interviewer
Yeah, yeah.

Participant 6
But yeah, we have, I don't know if segments are acceptable alternative, but yeah.

Yeah, we did develop with working with.

An agency called [Agency].

Personas or segments for the collections website to kind of identify the levels of interest in in that site.

Interviewer
Was that choice a business choice or a personal choice by the team? You know? Is it a requirement by the business essentially to go out and?

Get them created or.

Participant 6
I wouldn't say no. I would say it was more. I think that's the slightly following agile to no extent and following the philosophy, like the departmental build philosophy of like knowing who we're building something for rather than just going off on the assumption that this is a website only used by researchers in academic research, which possibly could have been the perception beforehand, because without any evidence otherwise, that is the general perception for collections websites so.

The idea was perhaps to have a bit more evidence to point to as to why we might want to build things other than just for academic researchers.

Sorry.

Yeah.

Interviewer
Oh, that's alright. I feel for you with the cold.

So other than.

These segments that you, the personas that you got highlighting the fact that there was a different audience than what might have naturally been assumed, what sort of details on them personas did you find useful?

Participant 6
Hmm.

I mean, this is going back a few years, so I'm having to re remember it now and pre pandemic even.

I think the interest, I don't know if this is so much attached to the sound as the user research side of like the audience expectations which did it was tied to personas that I can't quite remember exactly. Who said what, but the different expectations for the different personas, so things around whether the museum would have images of every object.

Very different groups of people had different perceptions based on their knowledge of how museums worked. So, you know, researchers, academic researchers had more sort of understanding that no museums don't have a picture of every object whereas.

More general audience, I think I had the kind of people with an interest in.

Culture I, and now I'm struggling to remember the exact segment name.

They were surprised and annoyed that the museum didn't have beautiful pictures of every object in the collection, because that's why they were looking at the object. So why didn't we have a picture of it that was beautiful? Because where the [MUSEUM] and why don't we have beautiful pictures of every object?

So that was quite interesting around expectations.

And what different groups of people based on their kind of knowledge of the sector, to an extent.

What they what they had as their.

Expectations for what we would do digitally.

Interviewer
And was that something that came out from?

Your extended use research. Or was that something that actually was part of the persona, do you think?

Participant 6
It was part of the development of the persona I I'm just having to try and remember this 'cause it was a few years ago, but I'm pretty sure it was from workshops they did as part of the personas.

Work.

Interviewer
OK, no worries.

And.

Participant 6
Which I think is all open. I'm not sure if.

It's all online, but I'm pretty sure we like the report, is available somewhere.

Interviewer
OK. Thank you.

So.

When you've got these personas, how are they used internally throughout the development? Are they? Are they something that you kept referring back to, or are they something that was just looked at once, or how did you? How would you describe the use of them throughout the project?

Participant 6
To this from a technical side I it's not so primarily used and I think this could be worth talking with other members of the team might be a bit more helpful because.

Sadly, a few of the people are no longer in the department, but I wouldn't say from the technical side where as much referring to them, but I think where they do come in a lot of use is.

Whitening awareness.

Of why we're building things and why it might be for different audiences and being able to point to evidence for that.

Rather than it just being assumptions.

Which again going back to, there is just a tendency to say the main website is for the public and the collections website is for academics.

And that's a shorthand that's very easy to fall into, and without having something to point to say, like the collections websites actually used by lots of different people for different reasons. And here's the.

The background to that that that gives us in discussions, it gives us some something to point to justify that view.

Yeah. So I think for me that's the kind of more abstracts not directly going back to it at low level, but just the awareness of it and being able to talk about that in discussions about why we might or might not build something in a particular way on the collections website.

That that's where. It's a really great sort of evidence base.

Interviewer

Yeah. So they were used in your discussions with the design team for maybe features building or feature design.

Or helps help steer that conversation should we say.

Participant 6
Yeah, exactly. So the you may also like feature on the object page, which is kind of to lead people on to another object that was very much built for you know the more general user, the casual you who's come from somewhere on the web, you know to submit object page often because you know people have suddenly got excited about one particular object and that was to give an onwards journey to the rest of the collection. So it's not a research.

Tall particulate that sometimes, as you know, through serendipity, of course, this thing, these things work, but. But primarily it's not aimed at research use. It's aimed at.

The more.

Casual browser, but to give an onwards journey so thinking about like that, you know, knowing who is being built for primarily is a great focus for it because otherwise especially with collections, things we you know it's very easy to just fall into this tangle of.

Is this a research tool? Is it for the general public? Are we trying to satisfy everyone you know? Should the interface have all the options? Because researchers want to be able to use it, refine their research? Or is it just a nice thing to like attracts people and give like a way in? But it's not, you know, a new way of interrogating the collection.

And you know, having that knowledge that you know what we're building this forum, why I think really helps clarify the build and design of it.

So yeah, so that's where I'd say at the abstract level, abstract isn't quite the right word, but at the higher level rather than focusing on the particular details of the segment, but at the kind of we're building it for this group in the main, that's really helpful.

OK.

Interviewer
I suppose the sort of mid level discussions within the project. So when you when you start once you've decided this is the project we're we're gonna start the design phase and we're gonna start planning out where everything was. So it sounds like it was sort of that mid level sort of conversation in the project that that they sort of added the value at this that in your projects.

Participant 6
Yeah, yeah, certainly. Where we where we come in. I I'm. I'm. I'm sure there was more granular discussion on the definition side.

That and on the user research side, but certainly for the development side, just having that shorthand of you know, you know, here's who we're building this for primarily is really useful.

And it allows us to shortcut discussions that might, you know, escalate, escalate into like.

We also add this is, you know, should we also add the filtering system into. You may also like well no because it's not meant to be a new way of searching the collection you know. So it's really good to kind of stop.

Feature group, yeah.

Interviewer
Yeah.

OK, so you've already said that your team hired that external company to create the personas and do that initial research for you.

When about did that take place in the product, you know, in the in the scope of a project was that?

Then where they hired right at the beginning to do that or part way in?

You remember that far back?

Participant 6
Yeah. It's so five, five years ago. It must be maybe now four. Yeah.

Pretty much at the start. Yes. Yeah, the yeah.

Yeah, certainly. Well before any build, so yeah.

That that work was all done.

Yeah.

Interviewer
And do you know, I know you haven't you didn't undertake the work, but you've mentioned that they did workshops and things like that. So did was there any other sort of data collection methods that they did other than just workshops that you know?

Participant 6
They did.

User interviews and screen.

Recordings of people using the sites you know to I can't remember the name of the.

People usability. Thank you. Yeah.

I think they did a few workshops of museums staff as well, so to get staff feedback.

Interviewer
But it sounds very manual.

And use a heavy.

Participant 6
Yeah, yeah.

Yeah.

Yeah.

Interviewer
Initially and time consuming, should I say more than anything?

Participant 6
Yeah.

Interviewer
OK. Have you ever heard of data-driven personas?

Participant 6
No, I don't think so.

Interviewer
OK, that's not that's fine.

OK, so this brings us on to my research part I suppose.

And I'm looking at how do we create more data-driven based personas? Now? That's not to say that most personas aren't based on real data.

And you know they are. They're ideally all personas should be data-driven.

But what I'm looking at is how do we get hold of as much information as we can from users?

At a really early phase in the project.

Very quickly. So without doing all of the manual data gathering and.

The interviews, the workshops and everything that takes a long time.

My approach was just to create a short questionnaire, post it on, in your case the museum website, let it run for a couple of weeks, download that data, and then I've devised a system that allows with to just upload that data to it and it will cluster, do a cluster analysis upon that data. So there's no biases.

Buzz, there's no misconceptions going in that you know, this is general public or this is a researcher. It is just based upon the question responses essentially.

Now obviously, for anybody who knows anything about cluster analysis, that there is one downside, which is it will always give you clusters.

You know, there is no way that it won't. It will always give you some clusters.

But they will also provide you a clarity.

How much overlap there is between these separate clusters?

So I in the within the system, I've devised it so that I'm creating something that I'm calling an information seeking cluster profile. It's not quite a persona because it doesn't have all of the.

Think the user input into that data apart from what's just captured in the survey, if you like. So the idea behind these profiles that I'll show you in a second is that.

We get these really early on in the project to start them discussions that you talked about a little bit earlier.

And they will look, they can also be useful then, for the user researcher to then possibly look into the groups that it's formed and target. Then they're further studies to go and get that extra research. So they're actually saying, OK, well, these are the types of people that we've got. So that's the concept behind what I'm trying to achieve.

Participant 6
Mm hmm.

Sure.

Interviewer
Can I just share my screen with you if you don't mind?

Can you see that Richard?

Participant 6
Yeah.

Interviewer
OK, so this is a a figma mockup because my actual clustering system.

Participant 6
Ha ha.

Interviewer
Doesn't look very pretty, so I've marked up with the actual data from.

The clustering outputs. I've imported the survey data from the NML study that I did and this is going to have the output. So initially we'd start with a page like this which gives you a template of questionnaires. If you're a museum. But if you're working on any other type of project, you could make up your own questions essentially.

And then just upload the file.

Once we upload it, does it's clustering in the background and then it's gonna present what it thinks is the optimum number of profiles initially in this group it's suggesting that two is the first optimum

number. Let's not say that there aren't others, but it's just suggesting that there is two. We get a little bit of information just to explain how the user interface is going to work.

And I'll talk through that. So we don't need to read it, but we can talk through that.

Participant 6
Mm hmm mm.

Interviewer
So ideally these are presented as much as I can in a in a Persona style format.

And within the drop down you can see that we've got the colour coding going on in here. So this is telling us whether.

Two profiles or two clusters is good because it's in green.

The three profiles is OK, it's in Amber, but clearly 5 profiles been showing in red means that there's a lot of overlap and these wouldn't necessarily be the best to go with.

But it allows you to explore the different groupings yourself. You know, as a as a researcher, you can go in and explore and see what these actually look like to you.

Participant 6
He.

Interviewer
We just look at the first one that comes up initially, you can explain the sort of structure that we've got on here. So we get a little bit of an introduction in the text.

And then we get a breakdown of.

The segments of the questions. So we've got some demographic questions.

That are going on, we've got some goals and motivations as to why they're coming to the museums.

We've then got knowledge levels on 2 levels. We've got general knowledge and then the domain knowledge for the specific reason they were coming to the website at that point that we were doing the study.

And then lastly, we've got the perceived behaviours.

So this is this includes things like that what they think their information seeking preferences, whether they prefer to search or browse.

Their search behaviour, whether they will try a few times before giving up, whether they just persist until they actually get there or whether they'll just give up full stop on after the first go and not got the answer.

Participant 6
Mm hmm.

Interviewer
We're getting device information, we're getting platform information as much as possible, and we also then get the frequency of visit.

So this is how often they come to that museum website.

And so.

In this example that we've got here, we can see. I don't know if we can see that range easier for me to zoom in a little bit.

Participant 6
Yeah.

Interviewer
We can see that they've got first visit is the predominant user there. So they come in for the first time. They're probably never been before.

The other thing that to notice on this interface is some of these boxes are outlined in green.

Participant 6
Hmm.

Interviewer
Now this indicates from the clustering.

What are the most significant elements from all of these separate elements that are made-up and that we're looking at? These are the ones that are classed as statistically significant for this profile and what separates it from the others profiles that you've got?

Participant 6
Hmm.

Interviewer
Umm, so obviously at a glance this one.

The reason for visit is significant. The task of motivations is significant and then the information seeking preferences where we're pretty much 5050 on whether they prefer to browse or search.

The number of pages viewed. They literally all want to view one page and visit once essentially so it gives us a.

A quick overview of this group and if you like, I think is the best way to explain that.

Participant 6
Mm.

Interviewer
And then if we flip to a second, you can then explore the other profiles within there and see the differences.

Participant 6
Right.

Yeah.

Interviewer
So in this one, obviously there's like two groups on this in this, so that the actual significant differences are going to be exactly the same because they're in there. But if we move to something like 4 profiles for example.

Then when we get to 4 profiles, you'll see that in this one, we've got them for visiting task motivation. The information seeking preferences, pages viewed, and the domain knowledge is what's.

Plus the significant within this.

Participant 6
Mm hmm.

Interviewer
But when we go to one of the other profiles now, it changes to like general cultural heritage. The information seeking here, they all prefer to browse pretty much.

Participant 6
Yeah.

Interviewer
They view 2 to 10 pages, which is significantly more than that first group, and then these will try a few times on search and predominantly will come and use APC in order to access that content.

If we compare that.

Group 3 This is now.

It's the mail group essentially cause all the others were female dominant.

And the AGS in here, this was split really between two age groups 18 to 34, and then the 65 to 74 year olds.

Participant 6
Alright.

Interviewer
We can see the retired in the employment is massively increased to the others as well, and they're all pretty much local to Merseyside, which is I was studying national museums Liverpool so it makes a reference to say that they've, you know they're really local there.

Participant 6
Yeah, yeah.

Interviewer
The other significant factors are the domain knowledge that they all cost themselves. Having an intermediate level pretty much.

And then search behaviour that persists these very high compared to the others in that they are going to keep going as much as possible to try and find that answer.

Which sort of lends itself into thinking. Maybe this general knowledge up here being slightly higher, having a few more fives than some of the other groups also lends it into, maybe there's a?

You know the whilst there's a lot of retired people here that's got a little bit more knowledge.

Participant 6
Yeah.

Interviewer
Maybe a little bit more motivation about what they're trying to do and when we look at the reasons for visiting that sort of backs up that idea in that we've got a high personal reasons that they're visiting, but also we've got almost 17% of people who are in this group saying that they only come to pass time.

So the general interest is also probably linked to that.

Participant 6
Yeah, yeah.

Interviewer
And.

And then obviously in the final group, this is very highly female.

The reasons for visiting nearly 90 well, 92% for personal reasons.

And very, very small amounts for the other elements that are in there.

Most people come in for a pre visit so that come into plan a visit to the actual physical museum.

Participant 6
M.

Interviewer
Knowledge levels a lot more middling on the general knowledge.

And again difference on the device. Here we can see that the tablets and mobiles are coming up closer to 50%. So it's a little bit you know smaller screen is more optimum here.

Participant 6
Yeah, yeah.

Interviewer
So that's what gives you a quick idea of the sort of things.

I was able to collect on that actual.

Survey that I undertook and how that would group this this data. I know it doesn't come up or specify.

A specific title or you know type of name or type of user, but it allows you to see what groups are naturally coming out of that data.

Participant 6
Yeah.

Interviewer
I.

So what would your thoughts be on this?

Participant 6
Interesting. I mean as alert level of data and dashboards that's immediately interesting to me.

I guess that's my you. Just that last point around the meaning of the cluster, I guess is the, that's my memory of clustering is that you know you get clusters, but then you've got to find meaning and I guess that's where how, how meaningful is it to group them in that way. I mean you've still got to review and then label in some meaningful in some meaningful way that.

It's particular that cluster is representing this meaningful group as a shorthand or.

Interviewer
Yeah. So I think the.

The significant elements maybe could be brought out into a descriptive part to sort of, you know, as textual descriptive text, I suppose.

Participant 6
Yeah, yeah.

Yeah.

Interviewer
To just say this is what really is making up this group, and why they're coming. I suppose the other thing that I forgot to mention earlier is that normally in clustering.

Your once we've got a cluster, we take a definition. We take the centroid, so the single person in the middle, and we take all of their columns if you like and say this is representative of this whole cluster.

Participant 6
Right.

Interviewer
But there's quite often that's not the case. So the other thing that I've done and which is why some of these don't add up to 100% in these elements necessarily, is because what I've done is I've got the clusters and then I've taken each of the columns, if you like, of data in that cluster.

And then said OK, well, what makes up to 75% or over?

Participant 6
Right.

Yeah.

Interviewer
Of this thing, so it might be that there's two or three things in here that are dominant as opposed to one which in natural clustering would say it's just of this one thing. Now in here we're getting, you know, so all of these we're seeing the make up of the different groupings. So we can start see that well, this one group is actually made-up of a couple of other elements within this one factor. But they're all trying to trying to make sure that we're representative the whole cluster, not just taking a centroid and going well. That's, you know, blind knowledge.

Participant 6
Yeah, yeah, yeah.

Interviewer
And I think that's where the significant parts really start to play into.

The strength of what? This is because it allows you now. Well, it would allow me to take these and say, OK, well this is a female dominant group.

You know, come in for personal reasons and maybe just putting that in text or maybe adding that into this text here might be a little bit better for what you're saying.

Participant 6
It's. I just think it's a I don't know if this is going off on a tangent, sorry, but it's kind of.

A level it gives you a level of detail thing and that obviously we could keep refining endlessly the audience for the music. Like you know we did 4 but we could you could do a, you could do 16. So this is quite nice in that you could choose the like to keep on drilling down.

In a way that you know, it would be hard to do with just user research because you got to would have to keep going back out. And so I like that this could, could let you be that level of detail you want to go into.

That's nice to my mind.

How would it work if you if you reran it again in a in a year's time, would you get? I'm just thinking about comparing like with like, so would you get entirely different clusters from the new data?

Yeah.

Yeah.

Interviewer
Yes, you would get totally new clusters based upon that data that you've got there. But you could always if you had the historic data you could run that take them.

Clusters say you decided you wanted to use four in this instance. Then you could get the four out, run your new data, put that on four, and then see what the difference would be.

Participant 6
Hmm.

Interviewer
So you might see then how the users are changing over time, or you know it might be that you want to do something like run this every beginning of every small project.

Participant 6
Hmm.

Interviewer
And there's been nothing to stop that happening, apart from maybe your users might get a bit of survey fatigue or something.

Participant 6
Yeah.

No.

Interviewer
You know that obviously running it every couple of weeks is not an option, but running it once or twice a year.

Or maybe even run it in specific sections of the website that you were targeting on your project. That might allow you to then get really specific audiences for that project.

Participant 6
Yeah.

The how 'cause it it's it a mix of analytics data and research 'cause you had kind of browser and presumably that's coming from analytics of some kind.

OK.

Interviewer
So that's just being captured by the user agent that they filled the survey in by. So when you run the survey, obviously it's naturally captured in that data.

Participant 6
Right.

Interviewer
So.

Participant 6
Yeah, yeah.

Interviewer
From the user agent, we can determine whether it's PC, tablet or mobile and also what platform they're based on or should I say most of the time we can because obviously there's an unknown category because some systems hide that. But generally we can gather that.

Participant 6
Yeah, we still get smart TV users as well, yeah.

The search and browse is that a data? Is that a question or is that based on?

OK.

He.

Interviewer
That's it. Yeah, that's a direct question to the users on what which one they prefer. Do they prefer to search when looking on this type of website or do they prefer to browse?

Participant 6
Right, yeah.

Interviewer
I've got to be honest, predominantly within the NML data set, it was 99% browsers. Is the option ideally over search?

Yes. Yeah.

Participant 6
And did you give definitions of those words? OK, just yeah, I think we've talked about this in the past as well. Yeah, I'm just curious.

I guess I was just thinking obviously to make it more.

Dynamic. Some of that like could be coming from analytics data, perhaps because you're, as you point around, you can't keep putting a survey up every month, but you could get analytics data to an extent that could.

Fresh it, perhaps a little. I don't know.

Yeah.

Yeah.

Yeah.

Interviewer
Yeah, I did. Like I say at the moment this is just 100% based on that initial survey trying to get it so that you know day one of the project starts, we know where it's going to go. Let's just issue the survey, get some data in you know within a week or two weeks or whatever you decide is an optimum number of results. Then you can pull that in and action it the cluster intakes minutes you know. And so within two weeks you've got.

These starting profiles to then go OK maybe as a user researcher now this is where we want to go. We can start a conversation about features.

To start our conversation or the conversation that you were talking about.

Are we going to?

Participant 6
Yeah.

Yeah.

Interviewer
Offer this feature we're going to not offer this feature because obviously if you've got a profile like this and they don't like to search now all of a sudden that's saying, OK, well, we really need to start

thinking about how do we allow them to browse that collection. You know, what options are we going to allow in that?

Participant 6
Yeah.

Yeah, yeah.

Interviewer
So yeah, I think the idea is that it helps steer them conversations, but really early on in the whole UX and web project format.

Participant 6
Yeah, for sure. I I mean, yeah, the, the speed, as you say of doing that versus actually organising in person, you know that's months more or less before we get any like a report.

So yeah.

That that having something back immediately is certainly.

Well, something to even start looking at umm.

Yeah.

Interviewer
OK. So do you think this is something that you might?

Possibly use in the future or on a future project at some point.

Participant 6
I would say yes, but then I'm not the person that would be making the call out. Unfortunately, Dave, Sir.

Yeah.

Interviewer
Yeah, yeah, this is your. You know, your opinion. It's not something that I'm saying that you're definitely gonna take. I'll take it on. But.

Participant 6
Yes. No, I think it's nice as a reference and it gives us sort of moving.

Moving view of the is moving the right word. A rolling view of the audience in a way that again without great resource every time, it's not possible to Commission every, every year for the website.

And it doesn't.

It doesn't answer all questions, but then none of none of them do, and it answers a different I think set of questions or gives.

Information on a different you know, area.

Which?

In a lighter light way.

That, I think is quite nice.

Interviewer
OK.

Participant 6
So yeah, I guess if I was being critical and if I perhaps wasn't so much of a data person.

And I was more of a like a design or user research in person, user research person. I guess I would maybe say the meaning the like labelling the meaning is perhaps the.

That's the tricky part. You know, it's like, well, what does it mean for this person to have these attributes? What is that? Is that a culture vulture or whatever we call, you know?

Because having that shorthand is in some way actually.

The crucial thing to be able to say, you know, giving a name to it, to that segment immediately becomes the shorthand that everyone. Then it kind of has that sense.

Interviewer
Yeah.

Participant 6
Of we're building it for this kind of person, even if now what none of us remember the exact details of that person, but the labelling of them gives an immediate.

Yeah, shorthand and on that our labelling of researcher has caused endless confusion for us because people immediately think that we're talking about academic researchers.

So you know that that just shows that the labelling is where people are finding the meaning more than drilling down into the data in some cases.

So yeah, I guess that's my only I the there is still some need for the.

Intervention of giving a label to the to the cluster that that gives some meaning.

Interviewer
Yeah, but do you think that is something that would be needed at this level or is that something that you think would be added once this once this data's been here and then you move on to generate a persona or to clarify with all of the extra user input data?

A little bit later.

Participant 6
That's a good. Yeah, that's a good question. They probably. Yeah. Hmm. I don't know. And this is possibly me talking outside my area again, really. But I.

Probably not, because it would be more. Yeah, I think it would cause more problems than it solves by trying to label them at this stage. If knowing that there was the follow on that, that would actually define the meaning to a greater extent and label it.

But this is the starting point, then yeah, probably. I would say you're right, probably not.

Because that.

And I think this is my again, going back to my memories of clustering like.

You can sort of. You can find meaning you know if you stare at if you drill. What's the phrase? You know if you shake the data hard enough, you know you get out the meaning. You want something like that. So, so you're using this as the starting point for finding the meaning, I think is an.

Yeah, a nice a nicer way of doing it to inform a guide the user research side with you know the expectations of what is being looks for.

Yeah, I mean, just going back to the number thing, I think that's, I'm just thinking back to that now. Sorry about your being able to pick the number of like 234 clusters. That's really interesting because I can't remember now exactly how we did 4, but I'm pretty sure it was just, you know, four's a nice number.

You know, I don't think there was any great.

Meaning to it than that, you know, like falls an acceptable number to deal with.

So yeah, in some ways a similar thing to data clustering. You know, at some point you are just picking a number.

Yeah.

Interviewer
Yeah, there's always gonna be exclusions or, you know, misconceptions at some point, but.

Participant 6
Yeah, and. And that's why they.

Interviewer
We do our best. We want to what we've got.

Yeah.

Participant 6
Of course. But. But that's where this has the nice advantage that, you know, you could switch to five still, which is nice, whereas we just you know for is the number we've got, you know and you're in one of those for regardless. You know we can't go back.

Interviewer
OK, cool. And So what? Just two more questions.

So beyond this idea of.

Personas or user segments that you you've been using? Are there any other methods or tools that you currently use to explore or understand your user audience? Do you use things like user stories or?

Umm.

Any alternatives or not?

Participant 6
When we're writing tickets, we tend we do it kind of as a user story, but I don't think that's quite what you're talking about.

Interviewer
If it defines your audience, then yeah, it is.

Participant 6
OK. Yeah. So I mean, when the tickets been written for the developer to do the build, you know it would be like written as the user need the user story.

Yeah.

Interviewer
OK.

Participant 6
Trying to think of any of those, I think that's probably the better question for.

Flora, the youth researcher. I think that's a bit. Yeah. I couldn't give you any of those, so.

Interviewer
OK.

And obviously within the web and the UX world, everything is constantly evolving.

You know nothing's ever standing the same.

Participant 6
Yep.

Interviewer
How do you see the future of user profiling within the industry?

Participant 6
Gosh, that's a deep question. Umm.

Interviewer
Don't worry if you don't.

Participant 6
The see I have to. I haven't given it a vast amount of thought, to be honest. It's.

Interviewer
OK, that that's fine.

Participant 6
I think that is very much more.

A question for a product manager or user research, sorry yeah.

I yeah.

Ah, ah.

Interviewer
Yep, that's OK, that's all fine. I wasn't sure whether you'd have started thinking about maybe using AI to generate the more to start coming up with them or anything in that sort of direction.

Participant 6
Gosh, that's really interesting.

I just was on one AI and cultural heritage network called yesterday and it made me think of all these different AI usages in the museum, and that's a new one. I'm going to write down. But no, I hadn't. I hadn't thought. And again, I think that goes back to that's kind of not directly my.

Interviewer
Yep.

Participant 6
Roles. I mean how? Yeah, that would be for more the product manager side, I use researcher side I'd I would be very interested if they did start using AI and I'd have lots of questions.

Interviewer
Yeah, me too. But yeah, well, I'm sure we'll get there.

Participant 6
Yes. Yeah.

Interviewer
OK so.

Very last question, do you think there's anything that I've missed or that you would like to add to the conversation?

Participant 6
I suppose if this is.

Products again, I'm just thinking about it like that. If I were, if it was being sold as a product to our user research or our product manager, I guess the.

The question is that like the integration with existing systems, you know?

How easy is it to integrate with analytics?

Yeah, I guess that's all I would like. That's probably a thing that would come up. How does it relate to the other systems we have? You know, how does it tie in with other ways we have of doing this to make it part of the sort of the steps of the project.

Yeah. So integration with existing systems and work is perhaps the only other thing, yeah.

Interviewer
OK. That's something that I need to.

Take away and think about if you can be brought in together and stuff, but that's great.

Participant 6
Oh, OK.

Interviewer
I am so I'm not sure if you know anyone else.

That would be good for me to interview you or not. But if you do, could we just not put it in the mention it here, but can you e-mail me the details?

Participant 6
Sure.

Interviewer
Separately so that then they're not in my transcript and I don't have to do the, you know, public stuff.

OK.

Participant 6
Yeah, yeah, yeah, yeah, of course. I interview over or.

Interviewer
Yep, I'll just stop the transcript there, if you don't mind.

Participant 6
Shaw.

Interviewer
Why you not?

J.2.2.7 Participant 7

Interviewer
Can you describe what your current role is?

Participant 7
Yeah.
So I'm a service designer at the moment.
My, job title changes year by year because the actual work that I do kind of is evolving all the time.
I a couple years ago I was a UX designer.
Before that, I was an information architect.
But if you tried to like bring, you know kind of.
Transparency to what I do my the purpose is to do research to gain understanding of a system, service or an application and then provide ways in which we can we can improve that for people, for human use.
UM, do the yeah.
The practice of UX is is very much around user centred design these days and then and and focusing on human beings.
So I'm actually usually on the other side of this.
I'm usually the one asking the questions and getting responses because I do do a lot of interview and I do a lot of research into people and then I test them as well.
You know, I put things in front of them and I asked them, So what do you think of this?
Do you think of that?
So in terms of my current job title is Service designer, which is a basically another.
Kind of a facet of UX where instead of looking specifically at a certain point or a certain application or a certain part of an application, I'm looking more holistically at a Service I'm trying to gain insight and understanding all the channels, all the touch points that are Service might provide their customers or users.

The trying to think of a good example.

You know, think of a company that offers an online service, but they also offer a telephone service.

They may also offer a physical letter writing service.

You know, I would look at all three of those rather than just the actual specific.

Digital application as a service designer that that, that would be my remit as a UX designer or a interaction designer of some kinds.
I would be looking directly at the digital part rather than the other parts.
Umm.
So yeah, uh, at the moment my job is to look at services and products holistically.
And I think from a wide angle from an eagle eye view, all you know, almost like rustically actually maybe that's the better way.
And am to gain insight and understanding.

And then to provide recommendations on how it could have been improved from a usability and from a design perspective, using experience and using best practices.

Interviewer
OK.
And so you've mentioned a quite a lot there and do you is the result of the standard?
Day, week or even possibly design phase that you would sort of go through?
Or is it ad hoc per project?

Participant 7
It ohh most projects are quite common in certain ways.
They have phases, you know, and the company I'm currently working for at the moment, they've because they work with a lot of local governments and a lot of government based projects.
We're using the government design system methodology, so we have discovery phases.
Then we have alpha phases and beta phases.
Umm.
And each one is kind of regulated in a certain way, so it discovery do certain things and you know, so on and so forth.
UM companies before that, though, they would have what's called design.
You know, they would still work in a kind of an agile methodology, but things would be tweaked slightly differently depending from project to project, but they'd have like kind of like either design sprints or they'd have, uh, just standard Sprint systems, you know, and and again, it basically it was dependent on you know, which project you were working on the format and which you would take.
And but for me personally, with my role, when it comes to projects, I I try and work to a kind of a they called what I called a 3D.
Kind of practice, which is uh and discovered defined design.
There's a fourth one.
It's test, but I tend to like.
I know there's the test at the end that kind of circles back to the discover, you know, so it's almost like a circle.
But yeah, discovery, discovering the issue or the problem then defining it, you know, defining a, you know, kind of a possible solution and then designing that solution and then sending out the field.
So most projects I've worked on in the past and current ones now as well kind of follow that quite loosely.
But they do follow that.
Umm.
Another kind of colour methodology, but it's like a a structure that some people also follow and I have followed as well is the double diamond.

Interviewer
Yeah.

Participant 7
Umm, you know, process, whatever.
You know when you start off, I ought to get a new branch out and you go wide and then you come back in again and you focus on the actual issue.
And then you go.
Yeah.
So on and so forth.
UM, so you you start off.
You know you don't know hardly anything and then suddenly you know something and then you don't know anything.
Yeah. Yeah.
So sorry, I'm rambling a little bit there.

Interviewer
That's OK.
That's, you know, it's good to see that you.

Participant 7
But.

Interviewer
Your standard workflow is mapping to the theoretical content that is out there as well, and that that's nice to see and.

Participant 7
Yeah, of if following those things helps me and you know, know what I need to do, at what point cause the other thing, of course about the work I do is that I can be, UM Ohm's.
Ohm's the words basically implanted into a project when it's like halfway through, you know some.
Some sometimes I can be I can be perishes, dealing a right thought.
We need you here to do this, and so it's good to know where in that flow you are at that point, because if I've, I don't need to do the work of the discovery parts already been done and I'm in a declining stage or if I'm in the designing stage, I know the definition and the discovery has been done.
It's good to know when in the part of the process, or again with the double diamond, it's good to know where you are in there so you know how to focus and what to focus on, yeah.

Interviewer
OK, fantastic.

I am.

Have you ever worked on an in cultural heritage style projects?

Participant 7
So the latest project I worked on with [Person] was [Museum].

[Museum] are an organization that is kind of a mirror to historic England.
They look after Scotland's heritage, their culture basically from the stone of Schoon dating back to century, whenever it was all the way to, you know.
More modern uh portraits and things that are currently hanging in Edinburgh Castle, which they also look after.
There archive contains millions of items and it's it's a hell of a treasure trove and we were working on a project for them.
You can imagine an organization like [Museum] Scotland, which used to be just Heritage Scotland and then they merged with another agency.
But I think it was something to do with culture and environment or something.
They merged these things together and they ended up with 10 websites, 8 databases and a whole bunch of contents and data that didn't really talk to each other.
So as a researcher or an academic person doing work in the field, you would have to visit different places and get different parts of the information to build a hot picture.
So the work we did was to combine that into one website, one search engine and one place where someone could come along and type in a keyword or a buzzword, whatever.
And it would bring up almost everything to historic improvement Scotland had on the item from, you know, Land Registry, own ownership stuff, all the way to.
Yeah.
Be a record in the Museum of a of a particular part of of of the that someone might be holding in a in a museum in Scotland's and to do with whatever the search term was.
You know, umm, there was.

Interviewer
Sounds like a really interesting project.

Participant 7
It was fascinating, absolutely fascinating.
And as a as a service designer as well as a UX designer at that point when I was working on that project as a UX designer, it was brilliant.
I was put basically in charge of the users and products part, so I had to help define how you showcase, organize and showcase all those different bits of information in one place, and that was a Herculean task considering we were dealing with 1010 websites, 8 databases and you can imagine a politics as well.

Interviewer
That's.

Participant 7
Oh my God, every single department that had its own database wanted a say in how things should be displayed and who should get priority.
But you know, you know, this should be on top then that should be on top.
Ohh my God.
Yeah, it was.
It was quite chaotic at times, but it was really insightful and rewarding.
It was a great project.
I've very much enjoyed it.

Interviewer
It sounds it and you mentioned the that you obviously have to look at the users so.
In that project and obviously and you were the projects that you work on, what are the types of audiences that you usually work with?
Can you describe the audience in any particular way?

Participant 7
Hello head, we tried to aim for audiences for our relevance to what we're talking about, what we're dealing with, so.
We don't.
It's rare that we'll just go out to general public and go vox pops, you know?
And like get people off the streets.
I have done that in the past with projects but as as time has gone on in my career has gone on because the user base that we work with has been quite focused.
Well, it's more likely that we'll either get existing customers or gets users who are known to the organization or will get volunteers.
Sometimes we'll hire third party recruitment agencies to help us gather users who are of a specific.
Demographic.
Umm, you know?
Every once in a while I've used that and then remember what it's called.
It's the letter and number that represents certain demographics of people.
And you know, I think you have a 1B2 seats, so on and so forth.
And they represent specific demographic types of general public.
I've used that before.
Umm.
And also you know if you have to do a study where you are looking for people of a specific.
Protected characteristic.
If you're doing suitability testing and stuff, then it's good to have a an agency that will.
Umm help you find the right people cause it's quite difficult to find people of that if you looking for blind people from deaf people that people with neurodivergence that's something or there's an physical disability, it's quite difficult to do that without the help of an external agency because again, you can't go on a tree and go.

Excuse me.
Are you deaf?
Wouldn't really go down very well that one.

Interviewer
No, I appreciate that.
And so when you start a new project then.

Participant 7
So yeah, hmm.
Ohh.

Interviewer
Is your process of understanding the audience pretty much the same on most projects or would you say that's different as well per project?

Participant 7
That I'm trying to remember how different projects and how we dealt with users.
I I'm not gonna say they're all the same because I think that would be too.
That would just be assuming too much.
It'd probably be easier to say that different projects project.
There may be some similarities in the way that which we recruit.
But understanding them and knowing who they are.
But I think there's going to be differences.
Project.
UM, yeah.
I'm trying to think if I can give you a decent example of where we've in the past gone and gone right, we'd right.
Yeah.
Sorry, I'm just trying to think of a a project in the past I could think of cause The thing is has was a year.
Umm have I worked with [Council]?
What do we do with their user now?
Their users were internal, so this is the other thing as well.
So a lot of the things I I I work with ohm especially for what we've done with [Company].
We're working with a lot of internal staff and lots of internal users at the local authorities or the Councils.

Interviewer
Yeah.

Participant 7

So understanding our users is relatively straightforward because we talked to them, you know on our on our regular.
So we know you know, we know that if we wanted to do interviews or to research with particular user types, we would just go to the HR department or to someone at the Council and say can you get us some names and some details for the study and then they would go out their intranets or to their.
They're not just sports, but we're looking for volunteers.
But.
Yeah, I I'm.
I'm trying to think of examples of where there might be similarities and I can't.

Interviewer
That's OK.

Participant 7
I can't think of any on top of my head.
I'm sorry.

Interviewer
That's fine.
I've got plenty more questions we can move on to.
So I don't worry.

Participant 7
So this yeah, I have to say I probably wise to say they're not.
It doesn't happen the same way every time we do it on a case by case basis, we find out what users we need and we recruit who we need to recruit.

Interviewer
OK, that's great.
I am so as part of your process.
Do you tend to use personas as part of the projects, and if not, why not?

Participant 7
Ohh.

Interviewer
And if you use any alternatives?

Participant 7
Uh, sorry, I chuckle because personas are a funny thing in my field.
Umm, I understand that they can be useful when you're dealing with large organizations and not everybody understands your user base, so giving them specific personas, assigning users to specific personas, and then revealing those personas to the wider organization can help that wider organization understand who that dealing with in terms of the user base and the key goals and the key outputs from those personas can be useful to the wider organization as a whole, especially if you're dealing with different channels.

However, some have become a kind of a check box, kind of a tick box exercise that some organizations felt we need to keep but when you dig deeper into why, they don't actually have an answer as to why they need them, they just want them.
And so, unfortunately, personas have become a little bit of a buzzword in our field.
Like sometimes they're very useful, but other times they're just not.
And there's nothing happening. Umm.

Interviewer
Sorry.
All vanished then on my back.

Participant 7
You're back?
Yeah.
Hi.
Sorry, you go.
I got some.
I got some strange feedback then for a SEC and yeah.
Did you disappear?

Interviewer
It went all funny on my end too.
Don't worry.

Participant 7
Ohh, that's OK cool.
So yeah, sorry personas.
I find them to be they can be a useful tool in certain scenarios.
They are not necessary all the time.
The practice of grouping all your users into, say, five or six key persona groups?
Umm.
Like I said, it's can be useful if you're trying to educate your wider organization.
Has the type of users that you know that are interacting with your services system on the regular.
They can be useful for that, but they're not always necessary in order for you to.
Umm, uh.
Create a new solution or to improve a piece of software.
Just knowing that you have those users and you can, you can, you can interview them and test them.
I think is enough and I think knowing their goals is enough.

Interviewer
That's alright.

Participant 7
You don't need to create a physical board that shows like the persona and the goal and stuff, so the act of grouping your users into specific persona types with key goals and things is a

good practice, but outputting that as a actual physical sheet of paper with a picture on it and goal thing, I find that practiced sometimes is not really necessary and that's why I chuckle, because then personas are something that whenever anyone brings them up, people don't know what to do with the joke about it or they argue it's not necessary or they argue that it is necessary.
It isn't.
There is no consensus on personas.

Interviewer
Yeah.
So you you're talking about someone's from?
What seems like a UX or a service designer's position which is yours?

Participant 7
Hmm.

Interviewer
I'm, but have you seen how the personas when you have been used have been used by like the development team or someone who's then not in as much contact with the users as yourself are?

Participant 7
I'll be honest with you.
No, not usually, and I can't think of any. Uh.
Examples of stuff in my head where.
It's been useful for a developer, for example, to know that this this user type requires this goal that goal because by the time the developer gets the say, the piece of work in the jar or whatever like management program they're using, umm, and the users, you know the acceptance criteria in the story has been written.

Interviewer
Yep.

Participant 7
You know the, story is specific to a key piece of functionality which has already been created based on the work done by the design team, which was based on that Persona grouping or persona need or something.
But again, like I said, sometimes personas aren't really necessary.
The physical sheet isn't necessary, so as long as these design team has an understanding of how your designing for that particular user type, we're designing for a feature, a function to help that particular person do X or do Y, and by the time the development team gets it, they're not interested.
Don't my experiences.
They're not interested in the personas, they just want to.
Know what to build and where it should go.

Interviewer
OK, I am.
So obviously, in order to be able to.
Define your audience in whatever method that you're using, whether it is creating a persona or if it is creating user stories in the Service design source of methodology, how do you?
And initially gather who them user profiles are.
How do you decide whether it is one group, two groups, or you know how?
How this the story maps together to so that you've got certain groups of users.

Participant 7
UM, yeah, that's a good question.
Again, trying to use experience from the past. Umm.
Learning about the service or the what, what the, UM.
If there's a product involved, learning what the product is supposed to do and who it's aimed at would help us then identify the users. Umm.
If there's a Service again finding out who, who they're aimed at and who they're trying to help will help us identify the users.
And their groups? UM.
Sometimes we'll be told you know.
Research will have been done already by either the clients or by previous team that will say write these.
These are our core users.
Umm we, you know, they'll show us a presentation done or you know, a piece of research done previously.
Say we identified these users that use a lot of companies will have their own analysis.
Sometimes that says, Oh yeah, we have this number of people using this platform and we have this number of, you know, so and the questions will be asked about, you know, demographics like you know, So what age range do you know, ethnicity, that kind of thing.
So from that information, we'll get a sense of who the users are as well.
UM again?
Like I said, it's rare that we'll get a program at the system or a platform that will just be aimed purely at everyone, you know the, the, the, the they will you know the I'm sure that happens, but.
There's always usually restrictions or containments of of users.
It's not just about everybody.
I mean, uh, going back to hers, the user groups, there were the widest, some of the widest I've ever seen.
So we had a range from people who were visiting a place.
So you had visitors that people, holiday makers or people who were looking for, UM, things to do activities.
Uh, and then all the other sides.
You had academic researchers, scientists and archaeologists, and you know everyone in between.

Interviewer
Yeah.

Participant 7
Uh.
And so yeah, that that was one of the widest.
But that I'd argue that doesn't include absolutely everybody.
Joe Bloggs off the public, you know?

Interviewer
Yeah.

Participant 7
UM, so yeah, we usually get a kind of a you know we either get information or we gain our own sense of who the users are, what group the user groups are and then you know we can start the process of trying to recruit them or at least gather enough of each group identified to give us a decent snapshot.
Umm.
Course with research is never 100% perfect.
You're not going to be able to get everybody you'll be able to get every single user group in the same numbers, so you know they'll always be like if, say for example you have like 7 or 8 key user groups, you might be able to get three or four of most of them and they only get one or two of another.
And so on and so forth.
This there's never perfect, but if you're looking for kind of like a benchmark number, minimum number of people, we would try and gather in terms of users.
Uhm, personally I try and get a minimum of UM.
Three to five people per user group, so the you know if one person says one thing and another person says another, there's a person in the middle.
So to offset right because you can't rely on.

Interviewer
Yeah, sounds like a good plan.

Participant 7
Yeah, three.
Yeah, exactly.
So yeah.
UM, just in case you know, and if all three of them say the same thing, that's amazing.
We understand.
But again, UM, if you've got three to five now, if you've got three to five people per user group dealing with say, 8 user groups, suddenly you've got, like my massage terabytes, 40 people.
Max, you know you've got to interview.
And that takes time.

Interviewer
Yeah.

Participant 7
The last the last project we worked on, we interviewed 52 people.

Interviewer
Wow.

Participant 7
I worked on entity interviewed 52 people.
Umm.
And it's it took two weeks.
Don't stop.
You know then.

Interviewer
Interesting.

Participant 7
Umm yeah.

Interviewer
I I Can you imagine the scheduling nightmares there?

Participant 7
Cool.

Interviewer
That sounds fun.

Participant 7
It's.

Interviewer
Umm, so you mentioned in that you in order to really understand the audience you you're gathering data through interviews.
Is there any other method that you would normally gather data from?

Participant 7
I.

Interviewer
Is there anything different or is it just interviews is the primary?

Participant 7
I know we try and gather UM through surveys, we also try and gather through either moderated tasks or unmoderated task.
You know where we could send them a link?

I mean, the beauty of the Internet today is a lot of these applications have popped up that are actually designed to help researchers do some things to do so.
That's great.
I mean, I don't know if you've heard of optimal there a like a leader in the field of of Desk based research.
So you don't have to actually.
Yeah, actually have to be in the room with the person.
You can send them a link and they'll do all the the stuff online virtually for you know.
So card source tree Jack tests.
You know the AB tests, that kind of thing.
There's all kinds of things you can do with a, with a person to gain insights into how they're feeling.
And of course, yes, the interview process, you know, asking questions and getting responses, it's tried and tested for millennia.
So of course we're going to use that, but yeah, there are plenty of ways in which we can gather information from a user and gain insight.
So no, we don't just rely on the word from interviews.
Although, as I said, that is very useful, but it's, you know, like you said, there's issues with time constraints.
People might not have the schedule that they need to do that.
So having unmoderated.
Task based.
Research is is is really quite useful and powerful for someone like me.
It also means that I can reach a larger number of people, so if I send out a survey, for example, I could reach a couple 100 people, whereas if I were to interview a couple 100 people, that would take me a long time.

Interviewer
Yeah.
OK.
And so?
Obviously, a lot of the methods that you you've spoke about are quite time consuming.
In order to be able to set up to do when abouts in the process of the project, would that normally take place and how sort of quick from the start of when you do this, what it normally be roughly?

Participant 7
So to gain insight, understanding on our users usually happens either at the beginning or the middle of the project.
Umm.
If we're problem solving and then you know, if we're looking for a problem like in the discovery that that happens the beginning.
If we're defining and we're asking them for input on how to define things like, say for example, we're doing the information architecture or we're doing some form of navigation setup or something like that, and we want their opinion, then that will happen in the middle and then at the end we test it, we put them in front of a computer with a thing and we asked them to test.

So yeah, the tasks that I mentioned earlier usually happen at the beginning or the middle, but there's also some tasks that happen at the end trying to keep the user involved or points is probably quite important thing because this technically is you sense design, right and.

Interviewer
Yeah.

Participant 7
I could make it up by as I go by myself, you know, but that wouldn't really be what we're trying to achieve here.

Interviewer
OK.

Participant 7
Umm Plus don't think I design a very good by myself, but I think I design a very good thing, yeah.

Interviewer
I am so if you ever heard of the concept of data driven personas.

Participant 7
Umm, no.
So yeah, I'm gonna have to get some clarity from you there on what you mean because all the personas we create are driven by some form of data, right?
We don't just make them up as we go along.
We have to get data from the users that we meet and talk to in order to build up personas.
So.
Yeah, I don't if that's if there's a specific meaning to what you just said or whether we're talking about the same thing.

Interviewer
So obviously what you're what you're doing there is you're creating personas that are based on the data, which is good.
Not everybody seems to do that.

Participant 7
OK.

Interviewer
After speaking some people, there's a lot of guesswork in some cases, but within the research field there is a whole area just looking at data driven.
So owners, which is creating personas from existing data or data that can be gathered very, very quickly from an existing audience.
Umm.
And that that's rather than having to go through the lengthy interviews and tasks and usability studies.

It's a case of how do we create data driven personas just to give us an insight a little bit quicker if you like.

Participant 7
So would that be more quantitative analysis then you're looking at?

Interviewer
Yeah.

Participant 7
You know good.

Interviewer
Yeah, they using so in the research of the moment, they're using social media sites linked to known users known for my systems.
So that's their drawing insights from the social media settings that people have got on their sites and things.

Participant 7
Are they?

Interviewer
Umm, there's a lot of data potentially brought him from multiple sources there.

Participant 7
Are they reading the text?
Are they reading the information that people are saying on social media?

Interviewer
I'm automatically.
Yeah.

Participant 7
Yeah, yeah.

Interviewer
So the system would automatically bring all of that in and do the analysis on that.

Participant 7
And then and then plucking keywords and plucking out keywords that may promote a mood or may promote a kind of like a feeling.
Is that right?

Interviewer
Yeah.
And the types of things that people like, you know, like in your Facebook settings, you say I like these types of films.

I I do these things and then they're also looking at potentially sentiment analysis of the posts and the things that they've liked or said within their posts as well.

Participant 7
OK.

Interviewer
Umm that that's why we're one angle of the researchers ventured into in order to be able to then generate some persona groups.

Participant 7
Hmm.

Interviewer
Umm, but there's no actual, you know, quantitative qualitative contact with the user at that point in order to generate them.

Participant 7
No.
OK.

Interviewer
Umm.

Participant 7
Uh, that's interesting.
I don't think I've come across that a practice, I mean.
It does.
I don't see why it wouldn't like bring up results.
I'd be curious to see, uh, a comparison of someone like me doing it the usual way and then getting a system to generate a persona using the method you just mentioned and seeing where the differences are, where the similarities are, that would be quite interesting.

Interviewer
Yeah, it definitely be a an Ave to look at then.
Yeah, there's, I think the validation of the personas is the being issue with that research field at the moment is that how do we validate these and where they go in, but I that's on ongoing research that some people are looking into.

Participant 7
But surely the validation comes from asking them to a user type, finding a representative of that user type and asking them specific questions around what's been generated and asking if they agree with it.

Interviewer
Yeah, I understand.
Agree.

I am.
OK.
So that sort of get brings me on to my PhD area and I've been doing some work with national museums.
Liverpool.
Uh, in particular, trying to understand who their audience is because they've got multiple museums, multiple website interfaces with different collections.
Each museum has many, many different types of connections, with many, many different types of user.
Not all the same.
You know very, very similar sort of headache to what you were suggesting that you had with the I'm Heritage Scotland stuff and and one area that their team because they're their team is very, very small.
They're not.

Participant 7
Right.

Interviewer
They're, you know, they're development team is 3 people, UM, and they don't have a user researcher.
They don't really have a UX team or anything like that, but they still wanted to have the insights in order to be able to make some jumps.
If you like and so one of the things that I I wanted to look at was how can I quickly get them some data that is usable to that team, umm.
And so I developed a way of looking and saying, OK, well, very quickly we could put out a survey asked a number of questions around who the users are, what they like, how they, you know, how they behave with the system and things like that get that data back.
And then I created a system that automatically takes that data and does a cluster analysis on.
It produces groupings.
If you like a various sizes and then allows the team to explore them clusters and look at how many different groups there are and the bonus being that there's no human bias on that data into looking in them groups, it is just the computer doing them.
They downside is that with cluster analysis there will always be groups produced and some of these groups are nicely separated and somehow have quite a lot of overlap.
So I'm just going to share my screen with you.
One second show you.
Umm, some sort of.
Maybe you'll laugh.
Maybe you will.
I don't know, I am.
And we see that.

Participant 7
Yes.

Interviewer
OK, so this is a a figma mockup of the output because I wasn't able to generate the prettier parts within my script, but basically you the use the museums or the users are the development team very early on will be able to just create their own questionnaire.
Umm.
Upload it.
On once the upload it, it will automatically do the clusters and produce not persona per SE because there's no polytype, uh, quantitative information in these, but it's all based on the quantitative information that they've been able to gather from a survey that they've run.

Participant 7
Yeah.

Interviewer
So within two weeks of the like a project starting or anything, they'll have all information to have these profiles from which they can then use as a jump off point.
So then go and do the user research and really focus in on the groups and get the gather the people that they want.
When we jump into here, there's a little bit of information about just how this works.
And in the profiles it will automatically assign which are the cleanest number of profiles.
So if for in the data was the best, it would show 4 at first, Umm, and in the actual profiles themselves we get a little bit of a description which is just describing the quantity of this user.
Umm, it generates a made up name and then assigns.
That is sort of relevant picture.
If you're like of who they are, obviously this is signed.
Two groups is the default cluster here?
Umm, I've got four major areas that I'm looking at for gathering information.
The first is the demographics.
Then we've got their goals and motivations.

Participant 7
Yeah.

Interviewer
Now we've got the knowledge levels for whether they know what they're coming for or not on 2 levels, whether it's a general cultural heritage or the domain knowledge which is, but their knowledge relevant to their specific visit to the website at that time when they filled the survey in.
And then we've got their normal behaviours section.
And when these clusters it comes up with identifying which other the factors within the grouping that make this group stand out from the other group.
And I've highlighted their them in these boxes with the green boxes.
So these green boxes are the defining factors.
If you like for this group compared to the other group that we've got, the types of information that we're producing are in the demographics, general age, gender, their

employment status, and then the primary location, because this is something that the Liverpool museums particularly wanted to understand in how they work.
Umm.
Within the goals and motivations gathering, information on what is the museum that they're actually entering from?
Obviously, with national museums Liverpool, they've got multiple different museums, which makes it a little bit different.
And then we've got their reason for visiting what we can see on this profile here is that the dominant group is people are just coming for personal reasons.

Participant 7
Umm.

Interviewer
They're not coming for academic or study or work reasons.
Umm.
And we can see that a lot of people are coming for pre visit in their task.
And what's motivating them to come to the website at that point in time?
Umm, these others are, uh, a range of basically is a count of all unique reasons.
There was no group in that could be, umm, put together for that one.
So it's a lot of people coming for their own reasoned, which are aligned with that personal reasons, if you like.

Participant 7
Yeah.

Interviewer
Umm.
In the knowledge levels, we can see that majority of people are sort of clustering themselves in the middle on a general knowledge group, and they're also saying that they've got sort of intermediate level knowledge in the domain.

Participant 7
Hmm.

Interviewer
And the behaviours we can say it's almost 5050 on whether people prefer to browse or search for this group, and most people will come and visit only one page of the website.
Umm in terms of searching, when they are doing searching a lot will try a couple of times before they give up and just a little bit less than that will persist, you know.

Participant 7
Hmm.

Interviewer
So they'll just keep searching and keep searching until we actually get the answer that they want.

Uh.
With a few just giving up after the first try if you like.
So nothing really particular there.
The other significant factor is the frequency of visits.
Umm.
And this is people who have just come in for the first time.
So the never been to the site before.
Umm.
Basically, we need to onboard them and when you were visiting one page as a quite a a big difference there.
And then obviously we got the platform differences, which is just data that I'm pulling from there.
Request, umm.
Compare that to the other group.
Still very female orientated, except this group is a little bit more different in the ages.
Umm, lot more people come in who are in full time work.
This group is very Merseyside or local based to the museum.
Umm again but slightly higher.
Personal reasons is 90.5% for this group.
Umm.
Most of them are coming from the World Museum, which is the largest, and they're looking for a pre visit.

Participant 7
Hmm.

Interviewer
Umm, but the knowledge is different with these, so these are a much higher level of general cultural heritage knowledge.
Umm.
Prefer to browse?
They will look at 2 to 10 pages on a general visit and they'll try a few times on the search and then give up.
Umm, they will come yearly a couple of more monthly, so to visits rather than just first time visits.
So there is a difference between this group.
Now the system also allows the team to explore different numbers of groups.
So in here you can see the green means that these groupings are quite separated.

Participant 7
Umm.

Interviewer
They've got quite good difference between them, but there's also like five groups where there's a lot of overlap between the groups, which is indicated with the red and saying don't really go there.

Participant 7
OK.

Interviewer
So if we just look at the four clusters, but gives those again to generate the same name.
But we can see that the this group is not the same group as before.
It's less participants in it.
There's a much more wider spread of reasonings for what they're doing and more reasons and the tasks and the motivations within that group. Umm.
Things this group is mainly predominantly search.
And they'll view just for one page again, want to compare that to some of the other groups we can see now this group's coming from outside the UK.

Participant 7
Hmm.

Interviewer
Predominantly umm.
More full time workers.
Well, this most significant is that they've got higher general cultural heritage knowledge.
They prefer to browse.
They'll look at more pages and they will try to search a few times and gently coming on a PC.

Participant 7
Hmm.

Interviewer
The third group is probably male, so in allows us to break down and explore these different groupings a little bit more from a.
Behaviour, perspective and the demographic perspective and.
So yeah, it's all from just that survey data at the moment that's gathered in within a two week period and then is clustered and the outputs are the generated from that.
What your thoughts on this?

Participant 7
OK.
Yeah.
I mean, it's very interesting it.
It's actually reminds me quite strongly of I don't have been on you, Gov.
It it reminds me a lot of the output they have when it comes to.
Revealing.
That that the results of things that you know they go out and survey quite a few people and they come back with these kinds of things which you can then purchase you know you can you can find out key information about people umm as a kind of like a demographic and the some of the information that you've revealed here like umm you know their age agenda the deployment and position you know I think that that's all you know you can get those kind of

things from you Gov as well so that's very interesting I just I I mean just a couple of things and one this will based on surveys just have curiosity are the questions required answers or are they just answer what you can and leave the rest UM.

Interviewer
For this survey, they were all required.

Participant 7
Choir.
Just so everyone has to fit everything in so that that's fine.
I was just gonna say that, you know, you might have a little situation if only some of the answers to questions are answered, then you know that that'll skew your information.
Are we?
But and that that's sometimes that can be the issue we're doing these things automatically is to you've got to trust the data 100% and if you need to put safeguards in place to allow for that?
Umm, but in in in.
I mean in terms of visual information this this is really good.
So you know, you clearly at a glance see key pieces of information.
Uh, the only thing I would ask about this level of information that you've gathered is what would be the, you know, what would be the I'm so kind of looking at the other way around when we're doing stuff like this, we're doing it because there's an issue and we need to find a solution or where we're we're being told that there might be a problem, therefore we need to find the problem and then.
And gather insight and run information.
But this seems to be.
I mean, please correct me if I'm wrong, but this seems to be the other way around.
You're gathering information about people before you've thought you know, ascertain that there's a there's an issue or a problem.
So I guess my question is what is the purpose of gathering this information together in one place?
Is it just to learn about your users or is the plan to use this information to then decide?
OK, so there's clearly a problem here because of XY and ZUM.
How do we solve that problem?
And if it is the latter, what scenarios do you see coming about?
From this you know, are you going to be looking at the reasons for visiting?
Going well, nobody's coming here to study.
Therefore, we're going to provide more.
The possibilities for those who want to study to come here and study.
Yeah, I guess what I'm trying to figure out is, you know what, what, what, what?
What's the next step from this to making things better and how?
How do you go about identifying, you know, where things could be better?
Is that another piece of analysis needs to be done on top of this.

Interviewer
Yeah.

Participant 7
You know, just thinking.

Interviewer
So they no, no, you.
These are really good things.
So in terms of where this would be run, you know, is it for just generally understanding the audience or it can it be used in a targeted way?
And I'm assuming it could be used in both in the initial research, it was because national museums Liverpool didn't really understand their audience at all.
They, the management said that everyone, everyone that came to their website was an academic and they knew exactly what they want, but they'd spent all this time and money on digitizing all their collections and they weren't getting anybody going to them to even look or understand it.
So there was a clearly a misunderstanding of who they were.

Participant 7
Umm.

Interviewer
So when the initial questionnaire went out that I've used the data from for this, it was a general discovery task at that point.
But I've also think that if we knew that there was a a problem in a certain area, then we could just target the survey up people who were using that area rather than just generally on the whole website.
Umm.

Participant 7
Yeah.

Interviewer
And so then we could start to discover who the group was and that targeted small area, but obviously we'd have to have an existing system for this in order to be able to run that.
And in terms of where I see how you get the insights and you know how we discover the major problems that we're sort of coming from, I see these as being an initial step.
These aren't like, say, these aren't a persona.
I'm calling them a cluster profile if you like, but they give us a an area now to say.
OK.
Well, let's get our user researcher or let's get some user researchers in and say, go and target these people.
See what these people you know?
Where wherever these problems that are coming in from this, uh, I could also or I've been looking into also extending this into bringing in analytical data as well.

Participant 7
Umm.

Interviewer
So linking it with something like the Sankey diagrams and just seeing if there's any obvious things like loops or drop offs and things like that with people who are in these certain areas and that could then give us some more insights, but that's still far from being implementable at this point.

Participant 7
Hmm.

Interviewer
Umm, so yeah.
Hopefully that answers a couple of your questions, but yeah, the these aren't designed to be the answer.

Participant 7
The tool.

Interviewer
They're designed to be a step off point.
Should I say?

Participant 7
OK, cool.
And why so the trying to understand the profile parts.
So originally you had two profiles and that produced 22 profiles based on the data you had, UM.
And then you show changed its four and that then turns.

Interviewer
Yeah.

Participant 7
I'm assuming the 2 into 4, right?
Umm.

Interviewer
Ain't no it what it is you've got the original data set and then you say how many clusters or how many groupings you want from that data.
And so it reprocesses that data into that number of groupings, which allows us to then say, OK, well, if I had this data and I wanted four groups, sometimes they want four, then they can look at how that data then splits up.
So the groupings, the people that were in Group one of two groups might not be in Group one or four groups.

Participant 7
Umm.

Interviewer
You know he needs the whole data set that gets reclustered and separated out at that point.

Participant 7
So and just maybe this is my misunderstanding of how data science works, but surely the more groups you have, the less likely there is to be blurring.

Interviewer
Not with clustering, no.

Participant 7
So, OK, that's me.

Interviewer
And yeah, that that was how what I thought like visually as well.

Participant 7
The not understanding the speaker could you got more options.
Surely that means that you can put people into categories better.
No, no.

Interviewer
And no, I will.
I wish you bought was the case.
No, in clustering it's it looks at, sees to and groups them as closely as it counts.
So the more groups we've got what you'd hope is that that group is more tightly cohesive.
Umm, you know, they've got more of the same characteristics with the more groups that we've got, but that isn't actually the case and especially with this data set that I got here, that wasn't really the case.

Participant 7
Umm.

Interviewer
So I had to apply a heuristic as well on top of the cluster in because normally in clustering you will get the separate groups and you will take the centroid, the one person in the middle and go that is representative of everyone else in this cluster.
And I found that that was far from the truth.

Participant 7
Umm.

Interviewer
So once I've got the clusters and what this data represents here is that I then go through each of the factors that columns for the questions and go OK, well, what actually makes this up to 75% or over within this column now within this cluster and then I show them results.

So you'll notice that some of the calculations don't add up to 100% because I'm looking for what makes up 75% or over and then that makes it more significant and more interesting in that to be more representative of that group in this there.

Participant 7
Umm, I suppose.

Interviewer
If that makes any sense whatsoever.

Participant 7
Yeah, I think so.
It's supposed the more groups you have then person number or 100 might belong in Group 4, but they also belong in Group 5 and therefore you've got the blurring.
Whereas if you have a smaller number of groups, then you have to focus that person in one of those groups and they probably won't fit into one, two or three.

Interviewer
Yeah.

Participant 7
They will fit into Group 4 and you haven't got Group 5 to blur it, so I actually I think I kind of get it, yeah.

Interviewer
Yeah.
So the whole point behind having the system is to try and eradicate that confusion of people trying to understand clustering and clustering analysis and especially when they're not the resources within the team to be able to have people who do that.
I'm but yeah, to be able to get to the groups quickly and unbiasedly is, is that the goal was suppose.

Participant 7
Hmm.
And you did you create the questionnaire that created this?

Interviewer
I did, yeah.
So the question now that I initially put out was there to discover the their users and their who they were.

Participant 7
Was it a long questioner?

Interviewer
Aims 22 questions.
It was long.

Participant 7
The I've seen longer.

Interviewer
But I wasn't allowed longer.
The museum wouldn't let me.

Participant 7
No.
I.

Interviewer
Normally they put out surveys of a maximum of five, so they raised a lot of eyebrows in me, giving 22 questions.

Participant 7
Yeah.

Interviewer
Uh, but without them, I wouldn't have been able to have the data that got back.

Participant 7
No, but that that's the other question is, is this kind of thing.
So I know we're out of time, just very quickly.

Interviewer
OK.

Participant 7
Is this kind of, uh, the you've used the Liverpool University as as example, but could this be used elsewhere and if so, what changes would need to be made to the questionnaire to meet others requirements?
So what would this be able to?
Would you be able to lift and shift basically?

Interviewer
I'm.
Allowed people to use this questionnaire as a base.
Obviously there's gonna be some questions like you're entering museum that most PCH places might only have one museum, which would be totally irrelevant.
They could eradicate that or at the same time I've also allowing people to create their own questionnaires and then they can just upload them.

Participant 7
Right.

Interviewer
So you create your own questionnaire and then it will just gather them, cluster them results and then present them results into the interface based upon whatever you presented.
So you might not do demographic.
You might not do goals, motivations or anything.
Umm, the tool would just a lot of that grouping essentially.
Foster the goal should have say.

Participant 7
Hmm yeah.

Interviewer
So yeah, OK.

Participant 7
Fascinating.

Interviewer
Fantastic, I will just stop the sharing the screen there if that's OK.

Participant 7
Yeah.

Interviewer
You know, I am so couple of final questions.
Do you think I PCP or something that you might possibly use at some point in the future on a project?
Or would they be?
OK.

Participant 7
Oh yeah, if we can create a solution that's allows an organization like ours to send out a questionnaire which could then come back with a whole host of demographic data that's automatically collated and put together in a graph setup like that, I'd use it with bells on course cuz bearing.
Bear in mind, as I said earlier, you Gov, do something similar where they go out, they ask a bunch of questions to a bunch of people, then they provide you with like a pack which has a whole host of demographic information.
Like what?
You've just shown costs a lot of money.

Interviewer
Right. OK.

Participant 7
So yeah, if a solution exists out there that provides that kind of thing and actually more targeted to a particular.
A topic or an area like culture or travel or something that if you even if you'd like almost tailor this this solution to those different areas you know UM, that that for from a research perspective that would be worth quite a bit uh.
And it would save an organization like ours time, which we're always wanting.
You know, you can imagine.

Interviewer
OK.
Is there anything that you would add change drop from what you've seen?

Participant 7
Not that I can think of, no.
I guess the only thing I would ask or put you know maybe say is I mean and again I'm only been shown brief.
And show showing of this.
But I just thought off my head.
It seems that there would need to be some parameters put in place or you know like kind of like containers or around the questions asked.
And you know, you said that you were like, anyone can create their own questionnaire and upload it.
But I assume there has to be some form of parameter like kind of like a box.
You can't ask to see anything.
Sure, sure.
I'm having maybe I'm misunderstanding, but I don't know if you can ask absolutely anything and then it'll churn out a nice like set of graphs and things because that would be quite remarkable if you can do.

Interviewer
Yeah, I think that this style of question would have to be limited as opposed in that format, yeah.

Participant 7
Yeah, exactly.
Uh, that's the only thing I would think would need to be and.

Interviewer
Yeah.

Participant 7
And in that case you would need some form of uh help or guide as to uh, what's you know, if someone wants to use this system, they would have to be taught in a way what kind of questions they could ask, how many you know and stuff like that.
So and just, yeah, off the I don't know how far you're going in this, but if you actually planning on creating a A that and at a solution or an application of some kind for doing.

Interviewer
I've got a working solution, but it's just not as pretty, which is why I've them mocked this one up and.

Participant 7
Yeah, but you may want to think about all the other bits and pieces you might need in order to put this to, say, markets or whatever.
So help information to you know, tool tips, guides, that kind of thing.
UM and I've also, uh, I don't know if you're going far into things like, you know, accessibility.
Uh.
Readability.
That kind of thing for people who cause this, this thing could be used like so such a wide scale that you probably do need to think about.
You know, people with disability criteria or characteristics.
So I I don't.
I don't know if that's something you've come across or have been part of your research, but I don't know if it's too late now or whether it's worth thinking about.
I don't know.
Yeah.

Interviewer
Well, from for my PhD is too late, but with my web developer and UX head on, it's something that I've thought about for future work and where that can go after the PhD.

Participant 7
I can.
I'm just thinking, yeah, this this is me still in service design, stroke, UX mode.
I'm just thinking, God right.

Interviewer
That's OK, it's all good.

Participant 7
But otherwise, no, I couldn't think of anything else that I would add or take away.
I mean, yeah, it looked fine.
Look fine to me.

Interviewer
OK.

Participant 7
Yeah.

Interviewer
Thanks a lot and.

Participant 7
So very quickly, what format would the questionnaires have to take?
Could you put up put up a Google form or A?
Microsoft form documents up there as CSV file.

Interviewer
As long as you got returned a CSV file.

Participant 7
I was gonna say it's the CSV file you put in there.
Yeah, OK.

Interviewer
Yeah, it's just a CSV file to upload and then everything else is fine.
Umm so as long as there's a the outputs are CSV format.
Then we're OK?

Participant 7
Yeah, yeah.

Interviewer
Umm.
OK.
So one final question, I'm with the evolving tools and methods within the UX field and how do you see the future of user profiling within the UX?

Participant 7
God, I think you're, I think what you're asking is, is AI going to affect how we do user profiles in the future?

Interviewer
I didn't put words in your mouth, but it's not what I'm saying.

Participant 7
No, no.
But that questions being thrown around quite a lot at least the premise of the questions being thrown around quite a lot.
I still think human beings talking to human beings are going to be very, very important in the future, and even more so now than ever with AI and other tools coming into the into the light.
I don't like the idea of a Computer generating a model based on what a computer thinks.
Your method still requires human input, and that is still very important.
UMI.
The don't think we're even close to a point where.
An intelligence in this I used that word very carefully cause actually technically most of the intelligence are not has been intelligence at all.
It's just clarified search programs and glorified.

Uh, like language models? UM.
That we're very far away from being able to ask, not been intelligence, generate 6 personas of all the users who currently use xyz.com and give me a you know their goals and needs cause you know you would need to have the data in place to do that and artificial intelligence.
Their models don't have that yet, and I think it'll be a long time before that's generated, but programs like yours and solutions like you, Gov and stuff.
If they did allow their data to be put into the neural net wash, that's call it of barred or ChatGPT, then you know who knows what the future might hold in terms of generating personas.
But I still would rather go from a human being to a human being and find out the answers because we.
There's insight we can gain that a computer can't from you know how.
How you know a computer can tell you that someone cares about something, but only human being can tell you how much.

Interviewer
Yeah.

Participant 7
But yeah, and that, that that's still more important than you know, just knowing that someone cares about something, it's like, yeah, but if you ask them how much they care about it, they may not care about at all.

Interviewer
Yeah.

Participant 7
They just they just mentioned it once.

Interviewer
But I also think there's the other side, which is the body language that we can read better.
Participant 7
Yeah.

Interviewer
Umm.

Participant 7
Yeah.

Interviewer
Now people will tell you one thing, but they might be saying something totally different with their face or their body, and that that's interesting from a user to user perspective as well.

Participant 7
Well, a Computer can't read sarcasm.

Interviewer
Not yet.
Well, I am.
OK, so very last, do you think there's anything that I've missed? Anything that you would like to add or that we've not covered for you, OK.

Participant 7
I'm I I think it's been great.
I've actually really enjoyed this.
Uh, it's been nice to talk to you and nice to be on the other side of this. Umm.

Interviewer
Yeah.
And please don't hold it against me.
f it was terrible.

Participant 7
No, not so.
0It's actually a very interesting topic you're working on, and the field of course has ever changing.
So yeah, good luck with the PhD and I hope it all goes well.

Interviewer
Thank you very much.

Interviewer
I really appreciate time and enjoyed the conversation?

Participant 7
That's right.
OK.

I expect to see that product on the market soon.
Interviewer
Hopefully.
OK.
1That's great.
Thanks a lot.

J.2.2.8 Participant 8

Interviewer
Umm, so how would you describe your current role?

Participant 8
Umm, so I'm currently a Service designer, at a tech design consultancy.
Yeah.
Do you want more detail than that, or that's how I would describe it?

Interviewer
No, if that's how you describe it, that's all good.
And can you describe what a standard sort of day, maybe week or even a design phase depending on how you work and would be look for you look like for you?

Participant 8
And yes, well, so it's very varied a day could look like.
And.
Creating some artifacts for a piece of research or for a project it could.
It could kind of be conducting some of the research for projects, depending on what the project is, and it could be more stakeholder engagement.
And so I think actually what might be easier or what might be more useful is to describe maybe like a phase of a project and how I tend to fit in is I what would normally come on at the start of a project or midway.
But the start of my journey kind of is looking at understanding the landscape of what the project is in.
So whatever.
Whatever the problem area is trying to understand a bit more about that landscape, both within the context of the project and sort of wider, just to make sure I understand generally what happens in similar situations and it could then be kind of conducting some research with some users to try and understand sort of now zooming in to our problem and trying to understand what it is we're trying to solve, speaking to the people who are experiencing this problem or whatever the context of the project is trying to speak to the users without projects, trying to understand a bit more about their experiences and what can be done.
And then it's usually some sort of activity or exercise to kind of.
Visually, you or represent that in some way.
So whether that is a journey map, whether that is a stakeholder map, whether that is a blueprint or a like ecosystem map it usually from that research we're trying to my a big part of my role is trying to like distil that research into something that can be understood and interacted with.
And so trying to yeah, simplify, not necessarily simplify, but trying to turn that into something that can be used to then make decisions potentially or I think making decisions is a big one.
It's kind of the artifacts will usually prompt some sort of decision making, whether that's we need to redesign the servers or we need to change this specific part of the service, or even oh, now we know where we need to focus.
And I think it's all about trying to guide us in a direction to then do something with what we've learned and a big part of again, my rule is communicating that with different people.

So if I work in a team of designers, but also maybe tech people or maybe their stakeholders who are neither design nor tech, but they're coming from a different perspective.
So policyholders and policymakers, it's kind of just making sure that we've communicated the insights in a way that they can then use to do the different things if they need to.
And then that's kind of a big part of the rule.
I think the final part is maybe then just wrapping that up in a way that can then be handed over for other people to use cause think a big part is building capability within the teams that we work with, particularly because we work with a lot of public sector or external organization does not internal.
So at some point, I'll contract will end with them and hopefully we can renew, but sometimes it ends.
So we need to hand over things that they can then do stuff with without us being there, and that's a big part as well.
I think those are the main parts I would say of my project cycle.

Interviewer
OK.
That's great.
Thanks.
And what type of audience do you usually work with?

Participant 8
And it ranges, but I think usually there is and another.
Well, my body went in two different ways.
If I'm the way that I'm thinking about audiences and one way is maybe people that we are sharing and working with.
So in those cases, I think it's usually other designers.
It's usually tech focused people, so developers or data people and then there is always like a stakeholder.
That is kind of a decision maker, so whether that's like a policy person or whoever is the service owner that we're working with and that's there was a kind of the main people that I would work with in that sense.
If you mean audiences about who I'm doing research with, then that is completely dependent on the project and can be anyone from school teacher to an.
A planning officer like that changes loads depending on the kind of project I'm working on and yeah.

Interviewer
OK, so when you're starting on a new project yourself, how well can you define the audience at that point?

Participant 8
When I'm sorry, on a new project and.
I think.
Probably not very well.
I think that I go in with a lot of assumptions and.

When I start on the project and there is no previous research that has been conducted about it and I probably I can guess and usually it's let's say about 50 or 60% accurate and probably not very accurate kind of depends on the project and how much experience I have with it and if there's been some previous research then usually I guess if I come on the project sort of midway and people have already started and then I can usually pick up quite a bit of information.
So there's usually a predefined sort of set of audiences, whether or not I then do some more work to kind of refine them.
And I think there's always some refinement that can be done with audiences.
But I think usually if you start a project where there's been previous research, you're probably pretty sure about who they are.
But unless there has been previous research, I tend to go in at the start of a project and it's very much assumptions based and I don't think it's very accurate.
And from my perspective, unless I know something very well and even then I think to probably wrapped up in all of assumptions and my experience of it.

Interviewer
That's OK.
I'm so moving into my next block of questions.
These were around use of personas.
Umm.
So first of all, do you use personas as part of your projects and if not, why not?

Participant 8
And occasionally we don't always use them.
And the reason for that is a lot of the time we can find them quite limiting and in the sense that.
Although you're hoping to use a persona for.
To kind of show a an audience type, I think sometimes because of the nature of personas being a bit more.
Or.
I don't want to say like sometimes they just don't always feel very useful because you've created this personality around them.
But it doesn't always help you actually make decisions that can that fit that work for the Service.
Sometimes they they're just so specific that it doesn't.
It doesn't allow you to do what you doesn't need to make the decision that you want to make.
Like effective decisions, and I think also sometimes we fall under like into the trap of having so many personas because you're trying to cover all your bases and you end up making loads of personas and then at the end of it, it's just not as useful to have that many personas because it then makes it unmanageable to work with and.
But the times that we have, I think what has helped sometimes is.
And when you have loads of prospect personas, you come up with loads of different audience types.
I think sometimes it's not personas that we then use it.

I think sometimes grouping personas can be quite nice and or yeah, grouping audience types by their needs and rather than like coming up with specific personas.
So I think sometimes people use like archetypes and things and instead or we do needs based personas or like yeah, I think needs to be pursued as we've done a couple of times just depending on like if they need a lot of support versus a limited amount of support rather than trying to think about the different types of people that are using this.
Because like I said, sometimes it becomes like unmanageable and another project that I'm working on at the moment we're actually trying to figure out whether or not we want to use personas, and because on the one hand it's sort of required of us going through a government assessment to have thought about users in that way and or at least it's recommended.
But on the other hand, we're at the point where we're trying to figure out if it's helpful for decision making to actually have personas, or if we're just doing it for the sake of doing it and trying to figure out if there's a different way to show the user needs and yeah.

Interviewer
OK, awesome.
Good answer.
Thank you.
So what sort of details in a persona or in your, umm, your models of the audience that you use as an alternative are most helpful to you when designing or developing and why?

Participant 8
And so I think goals is are really important just to try and figure out what it is they are, what is the one or two things that they really want to get out of this service or product I think helps to.
To make sure that we are that they can actually do what they need to do.
Really important is one of the goals is probably the main one and another one is we tried to pick out some key needs I think.
Sometimes this is where those can also get a bit tricky.
Sometimes different personas have they have a lot of needs and I think trying to figure out which ones are like the most important to actually get so they can reach their goals.
So what do they actually?
What are they trying to do and what do they need in order to do those main things and what are their challenges?
What they're currently and yeah, the current pain points or blockers like what would obstruct them from being able to do what they need to do.
And I think that those three are.
Probably in every.
Every umm type of user model that we use and then I think depending on what the service or product is, there are a couple of other ones that we use around like digital maturity I think comes up depending on what kind of service we're creating.
Something else that comes up is an.
Sort of.
And so I'm trying to remember, I did look at some to prepare for this, and now my brain has gone blank and.

Interviewer

don't worry.

Participant 8
And.
Goals and those needs challenges and digital maturity.
I think there's sometimes ones about depending again on the product or service kind of like and.
Like willingness to like time and stuff.
So how much time a person has to do?
Or like how quickly they need things.
Because I think we've had projects where it's been like a data officers that have a lot of data to process and they don't have time to go into a lot of detail.
So we need to figure out how much time they have and then whereas there are some other people on the other side who have a bit more time to get into the like nitty gritty, I think another thing that could be that we do sometimes use is how much like in depth information that they need.
And so using that example again of like a data officer, they might need access to like all the information.
Then you know all the data, whereas like the you like, policymaker may only need like outcomes or certain pieces of data.
So trying to figure out like what exactly to what like level of depth or detail do they need?
And I think those are kind of main ones that I can think of off the top of my head.
And sometimes another thing that it's not really a.
Something that we look for or like we always put in and it's not really like determines the persona, but sometimes, but it's nice to have is to pull from research like some direct quotes so that we can refer back to like this is of this person or this user type.
These are the kinds of things that they're saying when they're interacting with your service or product.
So that I think brings it back to that human element like there are people behind this thing that we need to think about and but also it's just a nice touch, I think a nice personal touch that to user groups and percentage.

Interviewer
OK, sounds good.
Excuse me and.
Can you discuss how the creative personas or user profiles are used within the normal workflow?

Participant 8
Yes, so.
They can be used to.
They're used sometimes to understand what, like the journey that a certain user is taking and how they're interacting with something.

So if we have done some research and we've created some personas based on the research and we can then identify based on those prisoners, what are the typical steps that they're taking?
And I think that it, like the journey maps in conduct conjunction with the personas, are really helpful to highlight, like really clear pain points and really which can then help with identifying some opportunities and that like that's an important part of it.
I think as soon as a helpful to understand where certain users struggling and how you can fix that problem.
I think that that that is honestly the main the main way that we've used for that, at least in work that I've done.
I've used them because the only other way that.
We have that we really do use them is if we're going through sort of assessments again to kind of show that we've done this research and we understand who our users are.
That's the only other time that person is kind of come into play is to just say these are the groups of users.
These are the clear issues so that we can have something to point towards, but the main way in terms of like the design and like yeah, the design of whatever is happening is understand where are they struggling, where the opportunities for these users and how do we make decisions based on that and.

Interviewer
OK, so whereabouts in the sort of timeline of a project from start to end the you start to then create these personas or user profiles, is it right the beginning, is it halfway in?

Participant 8
Yeah, absent they tend to.
We tend to try and start to create them after we've done some research so earlier on in the project and because we tried to do the researches like early at least the first round of research, early enough and after that first round of research, I think that's when we'll start to kind of figure out what are the key like what's repeated by different types of users and how do we fit these into personas.
But I think that.

Interviewer
Yeah.

Participant 8
Depending on how many rounds of research you do, they kind of are subject to change and so they were constantly being an iterated on throughout our project and sometimes right up until the very end because you'll always find out new things about users, whether or not you're seeking information.
And but I think relatively early on, I think in the first sort of.
Well, it depends on the length of the project and depends on how structural time you are.
But I would say in that first phase of done, the first phase of research, it's in that like following like analysis and synthesis phase, that's when you create them.
So before the first half of the project, even I would say.

Interviewer
OK.
And what types of data are they normally don't know what type of cause you mentioned you did the research early.
So what type of user research is that you undertake?

Participant 8
The main research that I've undertaken in the work that I've done is through user interviews and we have done surveys as well, but I find that.
I think it was surveys, it's a hit or miss.
How much?
Useful information you get sometimes even when you think you're being really clear and really like.
Can you give us this detail and stuff you still get people who miss certain things or do 19 and where you were expecting, which is why surveys can be a bit challenging sometimes.
But I think if you're looking particularly for and like if you're looking maybe at like ages or location based and data or yeah, like gender based data, I think surveys can be quite good for that.
Just kind of filter that information if people want to provide it, but otherwise I think a lot of the in depth stuff that feeds the personas or user types is through interviews cause that's when you can really unpick sort of their needs and like what they're struggling with more so than you can do in surveys.
And I don't think I've ever done like observations.
I think observations would be a good way to do it as well, because it's what people say and what they do are two very different things sometimes, but I just haven't been on projects where we have done observations.

Interviewer
OK, no worries.
Have you haven't heard of the term data driven personas?

Participant 8
Not until I'd seen this, so I think that I think I've heard about them in passing to be fair and but I think I've not like really looked into them in into much detail and until this and I think even to be fair, I think I did a quick scan before this call, but I still don't know.
Too much about like the research that's happening in that fear.

Interviewer
That's OK.
That's.
That's fine.

Participant 8
You.

Interviewer
That's all I needed to know. Umm.
OK.
So we're on to the section where I'm gonna introduce my so concept that I've done for my pH or coming up with for my PhD and the background to what I'm gonna show you in a second is that I initially started by doing some work with the National Museums Liverpool. There are a group of museums.
They've got 7 different museums and they really didn't understand their audience at all. They thought it was all experts.
I'm and we're wondering why nobody was going to that collections and nobody was using them.
And they'll bounce.
Rates were sky high.
Umm, so I did undertook a survey with their users initially just to see who was using the site and what their feelings were.
One of the things that I've come up with initially, after talking to quite a lot of UX researchers, was that personas are normally something that takes quite a while to get into, so they take, you know, you've got all your interviews to plan and everything else and all that sort of takes a couple of months at least.
And most projects to get into and.
One of these come up with is some way of just running the survey like within the first week of your project.

Participant 8
Umm.

Interviewer
Just gather in some quick data and then producing a profile which allows you to.
Explore what's going on and as a UX researcher, you can then use that as a I sort of guide to then move on to creating your user stories or planning on who you're gonna go, then go out and do your larger Interview sets with.
But more importantly for the museums, especially Museum then had four employees and none of them were user researchers.
None of them are UX people.

Participant 8
OK.

Interviewer
Umm, so within the website, they're mostly just developers in a lead, so they really didn't have the resources to put into understanding their audience.
Umm.
And they would still like a tool that allows them to explore what's going on and who these people are.

So the tool that I created was to get this survey, put the survey out there, you get your survey results back and you upload them and it autumn, it automatically runs a clustering algorithm.
Umm, so there's no bias from researchers if you like in in what these groups are and no assumptions are made?
Umm, so it automatically groups them and clusters them into profiles and then produces a visual.
I can't call it a persona, cause it's not.
It's more of a pre persona.

Participant 8
It's.

Interviewer
I'm calling them information seeking cluster profiles, which is rather a mouthful, but umm yeah.
Allows you to just visualize, but more importantly because it's clustering the data, it allows the museum to then play with it so they can say, Oh well, I want to look at what does two groups look like and they can see the difference between the two groups.
And then say, OK, well, what if what if I go for four groups?
What does that look like?
Cause some museums I've spoke to have set numbers that they always want to work with, but they have no idea why I'm and then others wanted just to go well and I have no idea.
So what?
What makes sense?
Umm, so do you mind if I just share my screen with you?

Participant 8
Yes, please, I'd love to see this.

Interviewer
So what I'm going to share with you is a the data has been run through my system, but I put it into a figma mockup because I couldn't make it look pretty enough.
Before I run my interviews in the actual system, I'm committee that.

Participant 8
That's yes.

Interviewer
OK, so this is just a the default load page.
I'm produced the template.
Obviously this template is set for national museums.
Liverpool.
Not probably the best template for most people, but the system ideally will allow you to create your own questions and surveys as long as they're within.

Participant 8
Yeah.

Interviewer
A type that would allow it to produce the charts from and then you can upload.
If I just run the sample data when we first go in, it automatically decides what is the best number of the clusters or groups to actually show.
Umm gives us a little bit of brief on how to utilize the system.
In here you can see it's set two profiles is the cleanest, cause the downside with clustering is it will always give you some results.

Participant 8
OK.

Interviewer
Umm.
And if I just show you the drop down I've colour coded in the drop down because when you get clusters, sometimes there's what we call overlap.

Participant 8
OK.

Interviewer
Now I'm pretty sure in every person that's ever been created that there's overlap between the personas, but the green means that these are more unique.

Participant 8
OK.

Interviewer
You know more separated whereas the red means we really don't have voice going with this because there's too much overlap.

Participant 8
OK.

Interviewer
So in this first one, just quickly show you what I've collected in here.
So we've got four groups of data.
We've got the demographic data.

Participant 8
Umm.
OK.

Interviewer

We've got the goals and motivations, the knowledge levels and then the behaviours, so this is all produced from the surveys in the demographic data I've collected, things like age, gender, location, employment, data, a pretty standard, stuff in the goals and motivations we've gathered the museum that they entered for, which in this case is useful for national museums, local, maybe not for other work you do.

Participant 8
Umm it's.

Interviewer
I'm.
But then we've also got reason for visiting, which is part like your goal that you mentioned a little bit earlier and then the task and the motivation and why they're actually coming.
You'll notice that some of these have got this green border around, and in clustering you see you're able to identify which of the columns that are contributing to make this group significantly different to the others, and that's what I've shown here.
So you seen none of the demographic stuff is making this significantly different to the to the other group that ends in the two.
But the reason for visiting and the task and motivation are and the information seeking preferences, the pages viewed and the number of pages that they sorry.

Participant 8
OK.

Interviewer
The frequency of visits is so this data basically for this person you can see it's a big mix of the age groups were predominantly female, but there is a good share of male in there, mostly full time employed, maybe a little bit retired and most of them are coming from the local Merseyside area which is right near the museums.

Participant 8
OK.

Interviewer
Reason for visiting?
They're all coming for personal reasons or majority coming for personal reasons, but there are some who would come in just a past time.

Participant 8
Yeah.

Interviewer
There's some coming to study and there's some for other random reasons.
These random reasons are everything where it's just a an individual reason that no one else has ever picked, and so it's just clustered them all together.

Participant 8
OK.

Interviewer
Umm.
And that the tasks sort of mapped to them.
You know the personal reasons.
Looking for pre visit and the other again is just all individuals grouped together which doesn't make you that useful at that point in time.

Participant 8
Yeah.

Interviewer
And we can see that there's about 5050 on the search or browse whether they prefer to search or whether they to BROWN like to browse through the system.
Most of them come in just to view one single page in the website, which is really not what the museum wants.

Participant 8
Yeah.

Interviewer
In this case, I'm and most of them are first time visitors, so obviously we can see here that we need to think about how we're on boarding, you know, and what we could do for these people to extend this pages viewed ideally we can understand the features that we're gonna need to provide.
We're gonna need to provide both types throughout that system.
Umm, most of the common pre visit but there are a few who were looking at both of the museum is an old is a an overview or just what collections are offered?
Umm, so we get quite good view of that person or should I say that profile.
I have put a name to it, but it's just a randomly assigned name.
Now the other group is 43% of the total surveyed audience.

Participant 8
Yeah.

Interviewer
Again, you see that it's mostly female, mostly full-time would retires in there.
So nothing majorly different about the demographic in this group, but this time the reason for visiting is pretty much all personal reasons and very, very dominant, predominantly pre visits.

Participant 8
Yeah.

Interviewer
Uh, the this group really prefer to browse over search?

Participant 8
The.

Interviewer
Umm, but when they do search they will try a couple of times before giving up, which is interesting compared to the other one and this one.

Participant 8
Umm.

Interviewer
We'll see that there are quite a few first time visitors, but there are quite a lot more yearly visitors, so annual visits.
I mean and they will look at 2 to 10 pages per visit, so little bit different to the other group.
Now with the all the profiles you're four group.
When we look at the fours, obviously we were using the same sample, but the these groups are based on a much smaller partition.
So this is 29% of the overall group.
Umm it's still very Merseyside based and it's still female dominant but not quite as much as what the other first group was.
It is mostly full time workers.

Participant 8
Happy.

Interviewer
Again, with seeing the significance in this group is that reason for visiting, which is quite a big spread to be honest, between personal reasons and study the read the tasks and the motivations for coming still previous it's quite high, but we're seeing known item searches now.
So we're seeing people who have a little bit more knowledge on that sort of backed up in this domain knowledge being significant with the intermediate rise in a little bit more.

Participant 8
Umm.

Interviewer
They prefer to search one page.
Yeah.
And if we compare that to like the second group and this one, you see it's more female, but it's from people outside of the UK, in the world.
Umm.

Participant 8
Yeah.

Interviewer
Dominated by the 35 to 54 year KH category.
Umm, but significant bits in here are the got a much higher level of general cultural heritage knowledge.
They prefer to browse.
They're looking 2 to 10 pages.
They'll try quite a bit.
Umm, they will mostly use a PC.
Umm, so we're seeing quite a bit on there and when they're looking at page engagement, they'll prefer to scan the text rather than looking at pictures or read the text in depth.
Umm.
And the first time visitors, most of them as well.

Participant 8
Umm.

Interviewer
If we just jump to profile three, you can see here it's clustered the mails predominantly.
Umm, these are mostly Merseyside, a younger age group.
And these three employment, because they've also got a mix of the retires in there with the 65 to 74 year olds.
So that's made them seem more significant for this group than the others.

Participant 8
Yeah.

Interviewer
And then we look at the domain knowledge intermediate and novice big shares and then the search behaviour in the some of these will persist until they find the answer they want.
Uh, no matter what I am.

Participant 8
OK.

Interviewer
But yeah, again, a lot of the data is pretty similar in there and in the last group it is predominantly females mostly in work in that 35 to 54 year group.
But again, reasons are changing in here.

Participant 8
Umm.

Interviewer

There's mostly personal reasons and previsit the domain general cultural heritage knowledge is in the middle, so they'll start on a 5 point scale in the center, and these group, if we look at what devices these are probably gonna come on some sort of mobile device or tablet.

Participant 8
Oops.

Interviewer
Umm, so yeah, this is what I've what it produces from that.
Sort of data.
What are your immediate thoughts on this?
Obviously this is at the beginning of a project.
This is 2 weeks run of survey and then you just run this and you get these results instantly.

Participant 8
OK, I was going to ask that actually is gonna ask how you I guess how you're collecting it.
Is it when people?
Was it just shared to people to fill in or was it when they logged on they had the option to do this?

Interviewer
So the survey in this case was given out when it was just when they were on the website.

Participant 8
OK.

Interviewer
So when they were on the certain parts of the website, it offered them the survey. Umm.
And it was just it was totally voluntary.

Participant 8
Yeah.
Yes.

Interviewer
But the idea behind the tool is that you could run that however you wanted really.
And then as long as you pass the CSV version of the data in, it can then produce similar sort of things.
But just for your data.

Participant 8
Yeah.
That's.
Yeah, that's cool.
And I think it's really cool.
And that was my first thought.

I thought it was like very detailed immediately and one of my first thoughts was this is a lot of information that a lot of it that wouldn't be included in a persona per SE, although I think it's helpful to help define some of these things about personas, like particularly around like age and gender, because I feel like sometimes that's just plucked out of thin air and we're just kind of like, ohh, can be an older woman and stuff.
But I think that having this where you can actually see data to show.
Oh no, it is generally older women or older men that would behave in this way, and I think it, although it's a small thing, I think it's actually quite a nice thing to be able to properly validate.
And then I thought the like the goals and motivations is really is really quite nice.
Umm, I think the reasons for visiting in this case.
It's like it's.
I just wanna clarify.
I'm assuming that these are these kind of like options that they were given to like, like with the a list like personal to pastimes to study, to work, but they just selected or do they?
Was it free text on the like algorithm has figured out which is which.
OK.
Yes.

Interviewer
It's a bit of both, so they were given a multi choice initially, but they were given the other option and once the other options been filled in, they've done interpreted where that sits or if it's unique and then it goes in the other reasons section.

Participant 8
OK, interesting.
And that's cool and similar with the custom activations is.
It's nice as well.
And like you said, it's not always I think in this form it's not like terribly useful in and of itself like you can't do much with that.
But I think it's helpful, especially if this is happening earlier on in the process to kind of.
Guide your research so I think like you said it, it's helpful to then if you are seeing some really like the task.
If you go in with them, Simpsons of what people are doing and you get a whole load of others, then you're kind of like, oh, that helps us figure out all our assumptions were wrong.
And it's quite a quick well, hopefully quick way to see immediately that you're like ohh, our assumptions were wrong about this and you don't have to wait until the interviews to get to that point and then you can use the interviews, validate and understand a bit more about what people are doing.
I think it's really helpful for a for A to guide.
Umm to guide where your research goes.
So I think that's really helpful.

Interviewer
You know, work.

Participant 8
Do you mind scrolling down a bit and just like in, see? Yeah.
I think I think the knowledge levels, I think all of it's really good and the not the knowledge levels particularly is helpful.
But I think it's one to take with the grain of salt, because I think sometimes and I think this is not like a critique of the work, I think it's just I'm thinking when I was doing this, if I'm doing research, sometimes people like a lot of self bias can come into that and can I think that it's just like a tricky 1 to I think while it's good to see kind of how people are reading themselves, I think further interviews would be really important to like validate or disprove that and just cause I know myself.
I would always put myself on the lower side.
Whether or not that's true and just for like, I don't wanna see like I know too much and but I think that's just something that came to mind for me.
And I think the behaviours is interesting and a nice because it's not something like these kind of behaviors are not something that we ever tend to include in at least in the personas that I've done we're seeing on the projects that I've been in.
But I think particularly as we kind of at least in the work I'm doing, I'm creating a lot more like digital services.
I think it's a really nice thing to be able to see and cause.
I think with working with public sector and government a lot of times people will use PC and are we work to PC and just because that's what we know people are using and I know sometimes it's a mobile first solution to that you know you can expand it if you can do it on mobile you can definitely do it on PC.
But I think a lot of the work that we do is in practice we work with like PC testing.
But I think having that.
Data which I assume is just through like Google Analytics or something.
Is that how you?

Interviewer
It's the yeah, the survey system.
You.
You collect it with your user agent and then just take it from there so.

Participant 8
Yeah, I think it would be helpful at least to guide some of that mobile.
We will first thinking if it's necessary I but and I think on the project that we're on now, it's always we're trying to find out data about how people are using like actually practically using the service.
So I think stuff like this would be really interesting for us when we think about if we were thinking about personas and how we are grouping people and the other question I wanted to ask was actually how will the how the clustering works? Umm.
Yeah, I think I don't know how to word this question.
I think I just, I'm I like want to know this like the math behind it.

Interviewer
I'm well.

Fortunately, we haven't got all week.
I'm but yeah.
What would the clustering?
What it tries to do is it identifies umm.
Was it?
Let's say you just.
It picks if we've got four groups we selected.
We want four groups.
It picks 4.

Participant 8
OK.
I E.

Interviewer
Umm, use the results at random, should we say and then it starts to compare all the others two of them and So what it tries to do is tries to find the nearest matches across all of the.
Factors and in this case the for each fact.
Each question was a factor, and that's what we see represented as a separate chart in here.

Participant 8
Yeah.

Interviewer
So it looks all of the factors and then tries to say, OK, well, what's the near, which is the nearest group in this and it keeps it just iterates thousands and thousands of times and some things move group and then within each group it then goes.
Are you the right person to be the center of this group?
If you like and so it might swap to someone else within that group, then and then it will reanalyze and look at everyone else again in that group and regroup again to try and find what it classes as the centroid or the centre.
Umm.
And then hopefully by the end when it's finished running, you've got groups where they're all.

Participant 8
Umm.

Interviewer
All similar if you like all the what we call the closest neighbours, but the in reality, and this is where I've actually changed mine a little bit.
So normally in clustering you will cluster and then you'll take that central user set of results and go.
This is the ideal person.

Participant 8
Umm.

Interviewer
So whatever this one person said, that's representative of the whole group, and I found, especially with survey stuff, that's not necessarily true.

Participant 8
Yeah.

Interviewer
And so that's why you'll notice some of these results.

Participant 8
Umm.

Interviewer
Don't add up to 100 umm, because what I've done with each.
Factor once the once the cluster is settled, I've then gone.
OK.
Well, what actually makes up 75% of this column?

Participant 8
OK.

Interviewer
So in this cluster, what made up 75% of it and then that's where I can then produce results that show in that there's a breakdown within each of these rather than saying, well, every everyone's a Tracy Tucker.
And in this case, everyone is 35 to 54 because that's where Tracy took us up.

Participant 8
OK.
And just.

Interviewer
I can actually show that we've got a mix of different groupings as well and how these are made up so you're not just looking at a?

Participant 8
Yeah.

Interviewer
I suppose the default in what you were talking about with personas before is that you have this one template and it's everyone has to fit into that where is in.
Here you've got a template, but you understand that it's made up of more than just one individual person.

Participant 8
Yeah.

Interviewer
That might be misconstrued the wrong way.

Participant 8
No, that's how that's really interesting.
Thank you.
And just a potentially stupid question and why 75%?
OK.

Interviewer
Because that thing that was a good position of what makes up the majority within that that that question, umm, it could change but it you know if I if I went with 50% it just made it very noisy.

Participant 8
OK.

Interviewer
And.

Participant 8
And then I guess if you go higher, it makes it not as useful potentially if it starts to get too similar.

Interviewer
Yeah.

Participant 8
Cool.
OK, cool.
And thank you for that.
And I think just, yeah, what another thing that was nice I think is the showing the different colours for the number of profiles and I think because sometimes we can defer to having more personas or like more profiles because it looks like necessarily not it looks like but it kind of shows more variance and users.
But I think in this case sometimes showing more personas actually has like less variance in users, and I think that's a nice thing to be able to see like actually what is the, what makes them the most different.
And I think having the colors to kind of clearly define that even though you still have the option to look at them and like suggested numbers is really helpful.
I think like I would find that really helped us to see we don't need anymore because this is what's the most difference from actual users rather than us guessing and trying to make things seem different.

So yeah, just that was, that was helpful.

Interviewer
Cool.

Participant 8
And yeah, I think it's really cool.
It's really cool to be able to get this amount of information before you've even done any research like that in depth research, cause yeah, like I said, I think it helps.
Focus.
Who you're speaking to and why you're speaking to them once, like in the context of your work and what you're actually trying to achieve sometimes.
We work on a lot of projects that are shorter in length, so we have to just start research and start speaking to people.
And as we work through the research, sometimes we then have to like it's after we've done the research we realized like, OK, this is where we need to go next or This is why we need to go next or that group of users we maybe should have spoken to this group instead.
But it's only after the fact that we've realized that and we don't have as much time to do some of that stuff.
I think this would be really useful.
To have at the start to kind of say we're time pressed, this is the context of our work.
These are who we need to focus on, and here's why.
And we have the data to prove that and then you could go and get hopefully more like valuable insights that actually are useful to the kind of problem area you're in.
And so yeah, I love data.
I think it's really cool.

Interviewer
Cool.
Do you think these are something that you could possibly use on your own projects if there's a?
If it was made available and you could ask your own questions.

Participant 8
I think so.
I think that it would be super helpful to use.
Like I would definitely advocate for that because on the project that we did with heritage work, we umm, we did do a survey to kind of get some of this information.
But like I said, people don't know results or the in the way that you're hoping and manually we had to do a lot of the classifying of what things should be or where they should go and we ended up creating kind of demographic graphs and all these charts and graphs.
But I think this would have saved so much time and would have probably been a lot more accurate, though we ended up putting together because it was manual.
There's always manual error in those things and I think this would even on the project that I'm on now, like I think it'd be super helpful to be able to use something like this.

Interviewer
Cool.

Participant 8
So yeah, 100%.

Interviewer
Just is there anything that you would add change or drop from the ISCP?

Participant 8
And so I think the thing is it would change from project to project.
So I think something like employment.
May not be as relevant on some of the projects that I work on.
Just because it yeah, it may not be.
Like even the project that I'm working on now where we're working with it's not general audiences.
It's kind of like schools and people who are stuff so that employment in that sense doesn't wouldn't necessarily work for us because we know that everyone's in full time employment in that sense.

Interviewer
Yeah.

Participant 8
So in certain cases, things like that might not be useful, and I think reasons for visiting and tasks and motivations are super and super important.
And really helpful entry I think is also quite a good one, especially when we work with government.
Projects where there's a lot of like government guidance, I think it's really helpful to know where they're starting from and how they're finding information.
I'm sorry if you Scroll down again, I forgotten what the bottoms were and in terms of knowledge levels, I think again.
From my perspective, it's and again depending on the project.
Sometimes I think knowledge levels might not be.
I think that's where maybe digital maturity could come in, in place.
And so sometimes even though again it's a self rating, but getting people to score how they feel comfortable with a certain digital tools, I think that's that could be something that we would use umm instead or as well and then.
Information seeking preferences is interesting because it's not one that we.
Always I think, yeah.
Again, it depends on the type of project, but I actually think it's a very it's an interesting one that although we don't own a lot of projects that I'm doing on, actually the only project where I have tracked search versus browse is on a heritage project.
And but I actually think that it's useful for other projects like the one that I'm working on now with teachers and schools.

And because sometimes he talk about not being able to find certain things, and it would be interesting to know how they're trying to find certain things and like if they're trying or if they're persisting or if they click one thing and they're like, I can't find it, I'm done.
It'd be interesting to know where they're kind of that getting stuck and so stuff like that I think would be quite good.
Kind of think if there's anything else that I may be would add.
I don't think of anything at the moment.

Interviewer
As OK.

Participant 8
I think the only thing off the top of my head is kind of like, yeah, someone's digital maturity and their.
Yeah, they're.
Yet how they self read that the only other one that might be nice?
I don't know where it would track.
Maybe it's in the demographics that might be helpful to know is actually.
If anyone has any sort of accessibility needs, I think would be helpful to know.
Early on because it helps with recruitment, for testing things or doing research, I think it's nice to know and because a large percentage of the population has some sort of accessibility need, it'd be good to get some of that information out so that we can make sure we're testing and doing research with people with needs so that we are and with accessibility needs so that we are designing things that are not cutting people out.
And so it would be good to kind of if we can get that information that would be another thing that would be good.

Interviewer
OK, the work.

Participant 8
And yeah, that's it.

Interviewer
So last couple of questions.
And I will.

Participant 8
They.

Interviewer
I'm just running short of time so and so beyond personas.
Are there any other methods or tools that you currently use or are exploring in order to understand your audiences better?
I know you mentioned user stories a little bit earlier.

Participant 8
Yes.
So we use user stories and we use umm.
Again, I think using sort of like journey maps is a useful way for us to know what people are doing and where they're having issues.

Interviewer
Yeah.

Participant 8
And we have in the past, but I don't use them very often, but I have in the past use like empathy maps and like, think, feel, do maps.
And but I think there are they feel more illustrative than actually useful sometimes.
And so we don't always really use them and but I think yeah, the main things are sort of user stories.
And I give you needs and then like.
Yeah.
Any sort of journey maps that are actually trying to understand like well like service blueprints like what is a user doing and what is everything that's underpinning, how they're working with this.
I think those are the main things that we tend to use.

Interviewer
OK, so obviously UX is a an ever changing world ever since it was conceived, and the tools and the techniques and the methods that are a part of that are constantly evolving.
How do you see the future of user profiling in UX?

Participant 8
Sorry, feels like a big, big question and.
I think.
I think we're, I see it moving away from personas in like the typical sense of personas, and simply because like for all the reasons I kind of stated before they they're not truly or they don't know his feel truly representative.
And there are other ways and like slightly better ways to kind of present and for it to like group people.
And I think sometimes I think moving towards a.
Like journey?
Like journey Beast, groupings of people or like needs based groupings of people and.
Yes.
So like more that way, rather than like actual personas.
But I do think and.
I do think that's because, like, I think a lot of personas are they're created by people who, yes, have done research in a number of users, but are still quite biased in how they're created.
And I think that if there's more data to kind of evidence why a persona is a certain persona, I think that then 100% they have their place and can be used and can be really helpful.

But I think that in the way that they exist of like I'm Jane and I'm 43 and I live with my cat and my son, I think that those ones are probably, I think that we see us moving away from them in a bit more like needs based and like more general archetypes or like how people behave rather than like a certain like, a face to it.

Interviewer
Yeah.

Participant 8
And yeah, I think that for me.

Interviewer
So if task and goal based rather than name and face.

Participant 8
Yeah, I think so.

Interviewer
OK cool.
I am very last question then and do you think that I've missed anything or is there anything that you'd like to add that we haven't discussed on the subject?

Participant 8
And I don't think so.
I don't think you've missed anything and I don't have any outstanding comments other than just to say thanks for reaching out.
It was really interesting and really cool to see and I hope it progresses more because I think it's a really cool tool and yeah, and I'd like to, if possible, I'd love to hear about how it progresses.
I don't know how much you can share.

Interviewer
I am.
I can well, once I've written this up, this is my final part to my PhD thesis.

Participant 8
OK.

Interviewer
Just doing the roundup and seeing what people think of where it's at the moment.
But post PhD then I'll be building on it and hopefully getting something out there.

Participant 8
Yeah, cool.
Would love to see that.

Interviewer
Yep.
Well, thank you very much for giving me your time and participating.
I've really appreciate it.
Umm and yeah, well good.

Participant 8
Thank you.
I hope it was useful.

Interviewer
It's very, very useful.
You love some of your answers.
They were really insightful and more insightful than most of the others have done so far, which is good.
So, OK, thank you.
You too.

Participant 8
What I'm glad to hear it, and thank you again and enjoy the rest of the evening and good luck with the rest of it. Bye.

Interviewer
Take care. Bye.

J.2.2.9 Participant 9

Interviewer
So could you just describe what your current role or career or whatever your career is?

Participant 9
Certainly. I'm currently the Creative Director at [Company], a digital agency focused on innovative user experiences. My role involves overseeing the UX design process, from research and conceptualisation to prototyping and user testing. It's a dynamic position that requires a balance between creative vision and practical user-cantered design principles.

Interviewer
What does a standard day or week/or design phase look like for you?

Participant 9
They are all quite different. My days are quite varied, but typically, a standard day involves team meetings to discuss our ongoing projects, reviewing design progress, and planning user research activities. Weekly, I ensure we're on track with our design sprints and deliverables. During a design phase, it's a mix of hands-on work, such as wireframing and user testing, along with strategic activities like client workshops and stakeholder presentations.

Interviewer
Please briefly describe your typical workflow for a project.

Participant 9
Our workflow at tahdar is iterative and user-centered. We start projects a client chat and then with user research to understand the needs and problems. From there, we move to ideation, creating sketches and wireframes. After some internal reviews, we develop prototypes, which we test with users. We then refine our designs based on feedback until we reach a solution that meets user needs and business goals.

Interviewer:
What type of audience do you usually work with/for?

Participant 9
The audience varies greatly depending on the project. However, I frequently work on projects

aimed at younger demographics who are tech-savvy and appreciate innovative digital experiences. We also cater to educational institutions and cultural organisations, focusing on accessibility and engagement. When working on the Museum site I remember the audience was initially really confusing. We had all, ages, all backgrounds and this did not match well with what the client was telling us as they indicated the main audience as knowledgeable researchers. But this clearly was not the case we were seeing.

Interviewer
When starting a new project, how well can you define your audience?

Participant 9
Defining our audience is a critical first step. We usually have a general idea based on client briefs, but we refine our understanding through user research. Interviews, surveys, and market analysis help us form a clear picture of who we're designing for, which then guides the entire project.

Yes, we use personas for nearly all our projects.

They're invaluable for keeping the user at the centre of our design process.

If a project requires a different approach, we'll turn to user stories or journey maps.

The method we choose really depends on what will provide the most insight for the project at hand..

When it comes to personas, the details that matter most are those that bring the user to life. For example the backgrounds.

Goals, frustrations, daily routines.

Understanding these elements helps us empathise and tailor our designs to real user needs.

It's also crucial to consider the context in which they'll use the product..

Personas are not just a tool for the design phase; they're a reference point throughout the project lifecycle.

They inform our feature prioritisation, our user testing scenarios, and even our marketing strategies.

Each persona becomes a north star, guiding us towards a product or feature that resonates with users.

Interviewer

So as part of my PhD. I'm exploring the idea of survey data driven user profiles carried out very early on in the project.

I I I don't want to call it a persona because it's not. It's not got all of the qualitative aspects.

You know, interview information or anything like that. It's missing all of that.

Ideally you run the survey really early on in the project and recruit existing system users as respondents.

Gather the responses to some survey questions.

And then run them through a clustering system and that clustering system automatically does the work, cleans them and separates them into groups essentially or clusters?

There is a downside to clustering in that it will always give you some groups, so not always are these clear cut if you know what I mean. They're not always perfectly separated, but I suppose in when we do any persona work, they're not perfectly separated because there's always some crossover there.

So I've been working on coming up with this idea and introducing this step that's really literally like in the first week or so of a project we run it.

Survey short period of time. Throw it at this system and then we get some instant responses, so no time lag essentially into what's there and the system will also allow you to explore different groups, different numbers of groups within that data as well. So you could recluster and come up with four groups or two groups.

And that's what I've been sort of working on as part of my PhD.

So can I just share my screen with you?

We just get to this one. So what you gonna?

See hopefully.

It allows me to share.

Can you see that?

Participant 9
Yes.

Interviewer
This is a mockup of the tool I have run all of the data that's in this through the actual live tool,

but I couldn't get it to do the pretty pictures on the end, so I've mocked it up into this prototype in order to save some time. But essentially you'd download the template for your survey questions.

Run that on your system.

Once all responses are gathered, upload them and it'll produce the a number of clusters for you. Ideally it will show you initially the optimum amount of clusters.

And it provides you with this little bit of information describing about how to use the system and things.

So this comes up and initially it's telling me that there's two profiles that it's worth that are separated and in the drop down you can see that the green indicates 2 profiles is good. Three there's a little bit of crossover between a couple of the groups, fours OK and fives a bit terrible.

Inside of these.

We get a little introduction piece that just describes how much of a percentage this is of the whole sample. So you can see it's 57% of the sample in this two clusters, which is 293 of the participants.

Participant 9
OK.

Interviewer
And then.

You get the demographic data of where they're from. We can click on the little icon. Just go to the tabular versions of that data.

But we're looking also at like, what's the goals and the motivations of these people, what drives them, what museum they've come in or entered from, should we say?

Their reasons for visiting?

Their tasks that they want to try and achieve.

You'll notice that some of these are green as well, and the green.

Participant 9
OK.

Interviewer
Borders indicate that this factor is something that you know significantly different here in this group to the others. This is what makes this significantly different this group. So for this, this one persona that we've called nappy, Wally is the reason for visiting the tasks and the motivations. And then in the behaviors panel, you can see that we've got information seeking. So whether they prefer to search or browse or browse. On the page is viewed and then how frequently they visit. That's they're the five factors that that make this profile stand out from the other profile or in the profiles in that cluster. And when I moved to the other one, you can see that here we've got very similar things that are different.

On all five because, this is obviously what make them significantly different from each other in these 2 profiles.

But we can explore what the profiles would look like in four groups.

So we'll get to 4 profiles. You can see the numbers drop down quite considerably in this one.

Significant factors are the reason visiting task, motivations, information, and pages viewed still, but the domain knowledge has started to come in on this one instead of frequency for visits.

Then when we cycle through, we can see that they change quite a lot into general cultural heritage. For this one, their search behaviour, whether they're going to try or persist at getting an answer. But most of them are here just giving up after one single search.

And as you can see here, we've got nearly 90% that prefer to browse within this group.

Participant 9
Mm hmm.

Interviewer
Where is the search behavior. Is. If you did do searching.

How would you normally go about with your search? So would you just try the once and give up if you didn't get the answer? Or would you persist?

Until you actually got the answer, if you know what I mean.

Or would you literally just give up after the first attempt? So you've got one attempt, a few attempts, and then the persist?

Participant 9
Mm hmm.

Interviewer
And then we've got the device data. Most people use the PC in order to access the site. And a lot of the users here are a first time.

Yeah, that's that, that group. Then in the third group, you can see the genders drastically changed onto a male group.

The ages have dropped considerably to the 18 to 34 category.

And it doesn't really pan out, but we end up with a lot in the retired category as well in there. So we do have this big portion of 65 to 70 fives in there, but they're all very local to Merseyside.

Looking pre visit to the museum, the physical museum.

But then into the knowledge domain knowledge for the exact reason as to why they've come there is actually a lot higher in the intermediate than the other groups that we've seen. And here we'll see that they persist with their searching.

So yeah, there's a lot of difference between these groups when it comes up with them.

But what's your initial thoughts on a tool like this that you could quickly just gather the survey results, throw it at it, and get some initial states to then use these as a kick off point. If you'd like to go and possibly do some user research or targeted use of research to build out your user stories or your personas from that point.

Participant 9
Initially, the ISCP tool strikes me as a highly innovative approach to persona development. It seems to harness the power of data analytics to segment users into meaningful groups of users, which could potentially streamline the persona creation process. The dashboard's ability to provide instant feedback based on survey data is particularly intriguing. It appears to offer a level of agility that could be very beneficial in the fast-paced environment of digital product development. I'd be keen to explore its capabilities further, especially in handling complex data sets, my first impression is that it's a tool with significant potential for enhancing user experience design.

Participant 9
I think this is a useful tool. I think its adoption would very much depend on, the project and needs of the project as well as which users it was targeting, but I think there would be an interest in this definitely.

There are clearly some crossover with insights you can gain analytics but I think this is a coming from a less clear and more focused sample. The speed of seeing the responses grouped is a nice thing as well as the unbiased groupings.

I can foresee a possible issue with having a biased data set though through those that completed the survey.

Interviewer
That is certainly a possibility. But not much I can control.

I've just got a couple more questions if you don't mind. It's going to stop sharing that screen.

OK. So last couple of questions on so beyond personas, are there any other methods or tools that you're you currently use or are exploring?

Participant 9
We use a lot of service designs user stories, especially on the larger projects. For the smaller projects we would fall back to personas as there are not really a lot of features on small projects to lead the user stories generation. We also have a lot less data on the end users and their wants and needs.

Interviewer
OK. And finally, do you think I've missed anything or would you like to add anything?

Participant 9
I do not think so.

Interviewer
OK.

Well, thank you for your time.

J.2.2.10 Participant 10

Interviewer
So my questions are structured into. A few groupings if you like. The first grouping is around your current background and role, so could you just explain what your current role or what you see as your career is?

Participant 10
Yeah, so it's web design management.

I'm currently a leader with a team of web designers and design and develop my websites myself as well.

Interviewer
OK. And what does your, I suppose standard day week or design phase depending on how your company works sort of look like?

Participant 10
So my standard day.

A lot of administration right now.

Emails responded to clients organising my team's workloads.

And showing that people aren't underworked or overworked in terms of how much design and development tasks they've got could be anything from tickets and content work on our current client sites to full web design and full development from start the process to the end of the process.

I also designed and developed myself so.

Yeah, I'm. I'm involved from the start of start to finish of my own projects as well, often two at a time.

Interviewer
OK. So you mentioned this start to end of the process. What does that?

Workflow I suppose usually look like for a project.

Participant 10
So we.

Have a very relaxed kind of agile setup.

We meet our initial meeting depends on the client because we take a range of clients from very small websites which can be 5 to £10,000 for a new website and they what we would consider a small website.

They would. They would. They'd skip the UX process. They they'd often skip the design process and go straight to code.

I often won't deal with something like that. Those would go to a more junior member of staff. We'll take on the higher end of the sport sites, which will include an initial onboarding meeting where it'll be the project manager myself as the designer.

A marketing a marketing lead for that project and a copyright lead for that project will have an initial discussion to get anything really important down any unique points of that website of that client wants any kind of.

And things that we wouldn't normally do.

From there, the marketing lead will go away and coordinate the marketing side of things. The copyright lead to go away and coordinate that. I'd put some initial designs down, possibly wireframe, but we don't always wire frame these days.

The so good wireframe at that point. Or if if we're doing some UX research for the for the UX meeting, then that will include things like.

Personas looking into.

Demographics of their current website, if they have one, any kind of competitors websites, what they do right or they do wrong.

I'm with them. We'll have a UX meeting which will include the whole marketing team for that project. The whole copyright team for that project and the whole web team for that project where we'll discuss these things a lot further off in about 2 hours.

From there, I'll go away and do a high fidelity design.

The copyrights will go away and do some concept copy and the marketers will do keyword research.

Once the high fidelity design's done, that goes off to the client.

They're either happy with it. Sorry. First off, it goes to a web QA, which is a a panel of designers.

I'm on 95, not designed it. If I do design it, I'm not on it. They'll come and critique it. They'll and it'll come back to the actual designer for Remez. Once that's done, it goes off to the client.

And the client will then either sign it off or come back with their own amends.

If there amends that we agree with, we'll do them. If there amends that we don't agree with for any reason, we'll feedback to the client.

At that point, they can either agree with us or not. If they don't, and they insist, then obviously we do what the client wants. They're paying for their website. We'll advise them as to why we don't think the same as them, but.

Ultimately, it's their decision once it's signed off by the client and we go into code.

At this point the copyright is are also getting sign off for their copy from the client and dealing with any amends. Leaning on the keyword research that the marketers have done.

Once the site is at a point that it's, it's got all the content in it's ready and gets sent back to the client for approval.

And the marketers start.

Their SEO and paired marketing setups, which includes things like adding Google Tag manager and all that, all that kind of stuff, stuff that I don't get paid enough to deal with.

And then once it's got final sign off, that's it. Go live process and then we do a a post live testing and checking and everything.

Interviewer
OK, cool.

What type of audience do you usually work with or for?

Participant 10
Clients range from I mean very local companies could be one independent shop, but we've also got international clients.

We have a lot of predominantly service industry. So we have a lot of landscaping clients. We do a lot of work with the.

That they're called the [Company], they're they're kind of a governing body of a landscape. Clients like a landscaping Guild. So we do a lot of work with them, a lot of work with concrete companies.

But then we have some different ones as well. I've I've just finished up on a a luxury glamping website which was really smart and I'm about to start an international motorsport team's website, which again something a little bit different.

So yeah, very wide range.

Interviewer
Cool. And do you do much analysis of their audiences so that their clients and customers when you in your process?

Participant 10
Yes, a big part. It's not often me personally, it's the project manager has a we have a set document that's it's called an on onboarding brief they'll basically sit down with the client for a good few hours and go through this entire brief. They'll get a lot of information on current clients.

Prospective clients and clients that they want to aim for.

And yeah, we got we get quite a comprehensive brief that comes to us.

At the start of the projects it comes to, it comes to myself or it comes to comes to the web lead and the marketing lead.

And we'll design pretty much the entire website and based on that, that document.

Interviewer
OK.

And.

When they're looking or gathering this data on the end, users talking to the client.

Is it pretty much that most of the sites that you're working on?

Are brand new. There is no site to replace. If you like, it's new site from the ground up.

Participant 10
I'd say it's probably.

I'd say it's probably 5050.

A lot of the a lot of the companies come to us have.

Current websites that haven't been touched in years.

They'll be service industries, concrete websites, for example. They'll and they'll have had a site from 2010 that might not even have, like a mobile view.

You know, and they've not touched the site in years and years.

All the all the content on there is way outdated and so it's essentially a new website, but it's and we it's very rare that we'll take something from that old website.

Put on the new one so it is essentially a build from the ground up, but then we we do have we do have clients that come to us with fairly new websites that just aren't happy with them and want to redesign. So it has a bit of a mix.

Interviewer
OK. And so when you've got a client that's had this existing site?

Do you ever do any user research with the actual users?

Participant 10
No.

We struggle to. I think we, I think I think we'd be a little bit worried about what legal loopholes we'd have on privacy and stuff there to contact actual clients of clients.

Is that is that what you mean by that?

Interviewer
They're essentially the people that you should be designing for, not the the person that's paying your bills, so you know.

Participant 10
Oh, no. Yeah. So we.

We wouldn't contact clients of clients. We yeah, a little bit, a little bit worried on privacy, things that it's not something that we've really explored because of a bit of a fear of kind of privacy there, but.

It it'd be a great idea. I just think it's something that would be a little bit too frying to do.

Interviewer
OK. So is it fair to assume then when you start a project?

You only have the.

Your clients, you know the bill payers view of the audience, you haven't really got a true picture of the audience.

Participant 10
Yeah, it's the.

Yeah, that view of their current clients aren't then.

Have you an opinion on what they what they want to attract?

Interviewer
Yeah.

Participant 10
Alright, I'll give you an example, right.

Right now we've got we've got another landscaping client coming right now. They currently deal with a mix of high end properties and.

Regular properties.

But they from now on, from our new website, they only want to attract the high end properties, so they have this opinion on what they're currently this this view on their current alliance and they have a way, a direction that they want to go in. They want to attract like stately homes and stuff like that.

And they want to really target that market. Bots dropping the other market so.

Interviewer
Yeah, OK, that's cool. So just moving into my second phase of questions and this is all around your experience with using personas at your current role, previous roles, whatever.

Participant 10
Mm hmm.

Interviewer
So do you use personas as part of projects and if not, why not?

And do you use any alternatives or anything?

Participant 10
We will do for the higher end clients that come in, the ones that have spent a lot more money that afford us a lot more time in the UX process.

The ones that we don't get that kind of.

We don't, we well, we don't get paid as much. We don't get the hours that are free that we so we run, we run all our projects through Salesforce which predicts how many

hours we can afford to spend on certain parts of the project. So if the UX hours are there we will use personas.

If we don't, then it's just a luxury that we can't really afford.

Interviewer
That's OK. And so obviously that's.

A business led approach. You know, on an ad hoc approach there.

Participant 10
Yep.

Interviewer
When you are using personas.

What sort of details do you have in there around the you know the what sort of information do you have around the audience or the users and which bits do you find most useful?

Participant 10
So standard would be.

It'd be like job.

Age. Location.

Things.

Goals and frustrations. We use a lot of so like what they're looking to achieve.

What the current market is frustrating them about that. Like, what's stopping them achieving what they want to achieve.

And I think the most useful thing, big one for us is age.

Knowing how old your demographic your target demographic is and knowing.

Predominantly male or female is a huge thing for us, as much as it probably shouldn't be these days. It still is.

A target in a kind of landscape client are an older generation. For example, is much easier than targeting our clientele of, say, young people who've nearly bought a house.

You have to get the kind of tone of voice right with your copy. You have to get the way the website's set up, right?

So yeah, age is a age is a big one for us.

Yeah.

Interviewer
OK, cool.

Can you describe how that persona or you know the user profiles that you create are used throughout your project development workflow? You know who uses them? How do they use them in that sort of workflow?

Participant 10
Yeah. So the onboarding document that I that I told you about will come to the marketer and the designer. The they'll often have a chat, the onboarding meeting they'll they'll have a quick chat about and they'll touch on the personas and the target audience. The web designer will create the personas. They'll often be in figma.

In a page separate to the design on the on the same design but on a different page just so they can access them at any time and bear those in mind.

They'll check with the marketers at that point, whether they've got that demographic right with those personas.

We'll often have. We'll often have clients that say they want to target everyone. That's like, OK, yeah, we'll create 50% as then shall we, you know.

Oh yeah, they will come back to them, like during the design phase.

I do anyway, at least as a designer, and I'll always push my designers to do the same where they keep referring back to who they're targeting as they work through designs.

Once we're past the design stage, that's it really. Then with personas, we don't really revisit them.

Once the design signed off, that's it. It goes to code.

So there's no real strict and formal process with our personas. They're just there as a guideline really to our designers.

Interviewer
OK, that's fine.

Umm so.

Participant 10
Yeah.

Interviewer
When whereabouts in the project are the personas usually created then? Because obviously you've talked about that they have this initial meeting with the client and then you have your wireframing and then possibly another meeting with the client and then the final mock up.

Whereabouts of the personas created in that are they created before the wireframing? After the wireframing? Where does that sort of?

Participant 10
Yeah. So the wireframing on the well, the personas would be done before the wireframing, but they're used at the same point when we. So we have the initial onboarding meeting, then all the we call them the UX documents with survey. It's like the personas and the wireframes are part of that. There's some more research as well from the marketers that they all get used in the.

The what is currently called the UX meeting, which I don't think should be called the UX meeting.

So that, yeah, so I brought up there and discussed there.

So they created should be created before the wireframe and the wireframes are done very loosely in figma so they can be moved around at the UX meaning. So it's not, it's not a strict thing that they should be done at that point, but yeah, the persona should be done before that.

Interviewer
OK.

Participant 10
And yeah, where I always do them.

Interviewer
So there's the crate before you design phase, essentially.

Participant 10
Yeah, of course.

Interviewer
Yeah. Cool so.

What types of data are the personas normally based on? Is it surveys questionnaires just the client's opinion?

Participant 10
Yeah, more often than not just the client's opinion and their experience of their company.

Interviewer
OK. And you ever find any issues with mismatches there?

Participant 10
All the time, all the time, probably every single project.

Yeah, you'll like, say, like I've previously said that clients will often say they want to target everyone they currently target everyone and it's like, well, that's just not true. Like, it can't be true.

Interviewer
OK.

Participant 10
That's there's no way you service everyone you know.

Yeah. So that's the biggest issue we get. We want to target everyone. You can't.

Interviewer
OK, So what are your thoughts on if you ever heard of the term data-driven personas?

Participant 10
Yeah, so it's.

What I understand of it is that it's personas that are created out of things like surveys where.

You've surveyed your either your current audience or your target audience.

On who they are, what like and what their goals are and things.

Interviewer
OK, what are your thoughts on them?

Participant 10
Something I'd love to utilise. I mean as much as I love design, I've always been a data-driven person like all my all my development stuff is data-driven.

I've always liked that kind of.

Hard evidence as backup as to why I've done certain things, why I've made certain choices in my designs and my development.

I'd love to implement it more. I really would. It's something that I'd love to love to do this year if possible.

Interviewer
OK, cool. So we're moving into.

The next phase of questioning, which is where I'm going to talk you through what I've been working on as part of my PhD, and I'll share my screen and demo some of this. I think my mock up of some of the workings of this, but I've been exploring a new change to the standard UX process. If you like with this idea of early survey based data-driven types of user profiling.

I wouldn't call it necessarily a persona, because there's no qualitative data there or not actually asking users who they are what their job is. That type of thing. It is very quantitative survey based.

And then what we're doing is you put the survey out there, you gather the data from natural real users. Once you've got enough data, you plug the survey into the system.

And it then creates.

A first of all does a cluster analysis on that data, which is essentially a an unbiased way of making that data into groups. There is obviously.

A caveated issue with clustering in that you throw any data at it in. It always will come up with some groups.

And so.

One thing I've done to encounter that within the process is I've got it to cluster and then normally in a cluster analysis you get given the one object in the middle. So the one survey respondents thing from the class is the centre of a cluster and that typically in cluster analysis is classed as. You know the one thing that represents the whole cluster.

What I've found is that that is very far, far from true within.

A survey based results. So one thing I've done is once the clusters have generated in order to drive the profiles that you're about to see, each of the columns I look at what makes up 75% of that column. So it's not necessarily just one object in, you know, if it's age, it's not just 35 year olds because that's the most common thing in there. It'll be a range and you might find that there's.

You know, to make up that 75% it.

50 percent 35 to 45 year olds and then some. 65 pluses or something like that. So it's whatever makes it up and that's what is represented on the screen so that you get a more true picture of who your audience is.

Participant 10
Mm hmm.

Interviewer
And obviously this is done. The idea behind this is for you guys in your situation, once you've got a client, you'll say, OK, well, have you got an existing site? Yes. OK. Can we run this survey on your site?

So actually you've done run that survey on the site for a week or two weeks at the beginning of the project, gather that data and now you've got.

All the user data against. You know you've obviously got a privacy statement on there, so all your privacy concerns will be gone.

Participant 10
Yeah.

Interviewer
And that that basically is where it sits and it comes up with these user profiles that you can then fit back in. And if you wanted to do some more qualitative data, you could then build on that to go and do your qualitative research or.

Do your keywords and stuff like that off the back of it. I think in your position the way you've described your process, it would also change. Maybe the some of the early conversations with the clients.

Participant 10
Mmm.

Interviewer
Because rather than them saying this is who I think my users are, you've now got a profile say, well, this is actually who your users are. How do we want to change this? How do we want to steer this or where do we want to go with that? Or do we want how do we want to better benefit these groups that we've got?

So it might also change that early part of the your process, but I don't want to lead you in that that's the thing. So just going to share my screen if you don't mind.

Participant 10
Mm hmm.

Interviewer
OK, let's look at sharing my screen with you.

Can we say that Participant 10?

Participant 10
All right, Kanye.

Interviewer
Yep. OK, so this is a figma mock up of the actual system that I've got running the main system doesn't look anywhere as pretty as this, so I didn't have time to tweak it up, but it runs and all the data and everything that's powering this is part of that.

So the work that you're about to see is work that I've done with a previous museum based client.

So I've used real data from them on this.

In order to drive this, there's.

Be that you could download or you could just create your own questions and post them on there. I'm just going to load the sample data.

When the ISP first starts, then you get the instructions because the clustering algorithm automatically comes up with these clusters and then clusters could be nicely separated with lots of gaps between them. Or, as is the case with cluster analysis, quite a lot, there's quite a bit of overlap, which to be honest is when we do personas in real life, there's always some overlap. We just.

And all that.

So one thing that.

I've built into this system is that it identifies using a traffic light system on which are the numbers of clusters that we could explore.

Are clean or green, essentially or slight overlap, or you know, still usable in the orange or in the red colour that they're quite overlapped and there's quite a lot of crossover, so they're not as easy to distinguish between the different groups.

Participant 10
Mm.

Interviewer
It automatically when it first starts, it will automatically profile to whatever it thinks is the best number of clusters as well, so this is defaulted to two.

And in these it comes up with a dashboard, essentially puts all this style. I suppose you'll say and we get a little bit of a an introduction and it says how many participants there were in the whole survey and this profile.

Is representative of 57% of that audience.

Participant 10
Right.

Interviewer
As you can see, I asked questions in mine around age, gender, location, employment. You know, general demographic stuff and we can see because I said before that it's about 75% of each column is what's making it up, which is why it's not just all full timers. You know there is some retired in there. There is some part time in there.

Participant 10
Yep.

Interviewer
We've got the age group where as you can see there is a dominant age group in there which is our 35 to 6 to 54 group but the others are still represented in there and whilst this is female dominated, there is still 22% of males that are in that group as well.

Location wise we would say that this is predominantly Merseyside located purely because that's the highest, but there are still people from around the world who are coming. So that's all people who are outside of the UK.

Participant 10
Mm hmm.

Mm hmm.

Interviewer
People in England, which is everybody outside of the Merseyside area but still within the UK, well within England and then in the Northwest, it's 20% make up.

The other 2-3 sections that I've looked at were behaviour, goals and motivations, so the reasons for coming why they came.

We've got an entry museum there because obviously the mute Liverpool Museum has seven different entry points, different museums to look at.

Participant 10
Mm hmm.

Interviewer
Possibly not feasible for something like you're looking up, but the reason for visiting whether they've come for personal reasons past time to study, work other reasons, and then the task and motivation. So what's the purpose of them coming to that museum website at that time? So we're looking to visit, are they looking just to see what the museum does?

Whether they've got certain collections.

Or other data that's in.

Next section is knowledge levels. So understanding what they the user already knows you know coming in. So are they. In this case we can see that there are all sorts of middling if you like, in that general cultural heritage. But in the domain knowledge.

We're also seeing that they're middling to slightly lower novice levels.

And then we've got behaviours. So this is how they prefer to.

Seek information. So do they prefer to search or browse, which is pretty much 5050 there. It's very close.

Participant 10
Mm.

Interviewer
How many pages they would normally view when they come to a site like the museum site that they were surveyed on and you can see that here. Most people would only come and look at one single page.

So we've really got to sell our information there to keep them.

Participant 10
Yeah.

Interviewer
When they are searching, would they try and persist, you know, try a few times and give up or persist until they get the answer, or just give up after the first go?

Here we can see that most people would try a few times.

Almost the same amount of people would then persist.

So some very.

Adamant people coming and using this site, knowing what they're looking for.

In the page engagement stuff, we used to look and see that people prefer to scan the text.

Before looking at the pictures.

And reading in depth is right at the bottom, so we need to make sure our keywords are really prominent and we're saying the right things in that first couple of paragraphs.

Device wise we can see that.

Mostly PC users.

And then frequency of visit, how often they'll come to the museum site. Most of them are first time users, first time visitors.

And then we've got platform do a thing to point out on here is that when we do the cluster analysis, it will identify what it thinks is the differentiating factor in that cluster. And I've identified these by changing the border colours to green. So we can see that this group, if you like, is differentiated from the other group because there's only the two groups when it did the clustering by the reasons for visit the task and motivations, the information seeking and the pages viewed.

So they're the four things that it thinks it's that differentiates this group from the others.

When we look at the other profile.

We can see it's still female dominated, but a little bit more than the other one.

Reason for visit is vastly different 90.5%.

Which is clearly why it's made this as a green boarded differentiator.

This group prefer the pre visit pretty much.

They prefer to browse and they will look at.

10 pages rather than just one, so there is a clear distinction between this group and the other grouping, and then the frequency of visits. We can see that a lot more are coming yearly. A few more are coming weekly. So there is quite a lot of difference between them differentiating factors to be able to separate them. But the good thing about clustering is that we cannot just stick with that. We can say OK, well, let's have a look at what 4 clusters looks like.

So this now redoes them and says we got 150 participants in here, which is representative of 290%.

In the next 126, then we've got 113.

From the 500.

What we can see here that there's quite a lot of differences. So in this first group, the reason for visiting the task and motivations, information seeking pages viewed on the domain knowledge of the.

Distinguishing factors in that group.

But in the second group we can see that that's changed to cultural heritage, knowledge, information seeking, pages viewed search behaviour in the device that they're using.

And then the third cluster's distinguished, mostly by demographics.

'Cause this is a male group which is.

Different to the others. And then the main domain knowledge and the search behaviour.

And then lastly, we're back into a gender group, which is mostly female reason for visiting task.

Cultural heritage, knowledge and device again.

So that's sort of the the profiles.

What was your thoughts on the iscps?

Participant 10
I got as a bit of a.

Data-driven nerd. I really like it.

The my first question would be how?

How was all the data collected? Was it? Is it all survey or is it?

Interviewer
100% survey.

And just put on the live website and left for two weeks. In this case it was left two weeks. But in your case you could just.

Put it on your client site and then leave it for awhile. Collect shut it down.

Participant 10
Yeah.

Interviewer
And then once you've got the data, you just feed it into the system and get your profiles then.

Participant 10
Yeah. Yeah, just it's obviously just as the set all the all the usual downsides of survey collection that obviously it's.

Only certain parts of the demographic are more likely to fill out a survey than others, so you don't get that you don't get that full view.

But.

It looks it looks really good, though. I mean it looks varied more varied than you'd kind of expect.

Mm hmm.

Interviewer
And the actual response to this data from the studies I've done on it are that it is very, very representative of the overall audience and it matches the Google Analytics data very well. So it is representative in this case, it's representative of everything. But yes, there are with surveys there are that potential issue of only certain people filling them in.

But I suppose the alternative.

Participant 10
Yeah.

Interviewer
In in this case, where I'm thinking about inserting this into the UX process is that you have this or you have nothing until you've been out and done your full user profiling. So some actual user data is better. I suppose in your case than your clients perceived view of what it is.

Participant 10
Yeah, absolutely.

I mean, I'd. I'd find this really useful, and I'd I'd really. I'd really like it the one. The one thing I think would.

Would stop us using this. Is that the time it would take to?

Kind of collect any data, any amount of data that we'd would then find useful.

We're kind of in a world where everyone wants their new websites ASAP, so taking the time to do this data collection right at the start of the process, which is when we when we need it, when it'd come in useful.

Would be the downside.

'Cause.

Interviewer
It was the sort of lead time into your projects normally is that.

Participant 10
We we've implemented recently at 3030 day turn around from sale to.

Having high fidelity design to the client.

Interviewer
Right. OK.

Participant 10
Alright, so yeah, it's tight now. It used to be a little bit more open than that.

Our Managing Director wanted to reduce that time. So we've now got a 30 day limit of sale to high fidelity design. So realistically, what could we afford there to do this probably a week Max, I'd like a week into 30 days. I'd be like a little bit antsy to start designing.

Interviewer
Yeah, I I get that. I I get that it's not for everyone, but it's you know.

Participant 10
I'd. I'd love it. I really would. I I honestly, if I could, I if I could do this for a month, leave it, leave it up for a month and collect that data and use it, I'd love it. I just unfortunately don't have that luxury now.

Interviewer
Yeah.

So if he did.

Participant 10
Yeah.

Interviewer
Is there anything that you would have changed or dropped from the iscps method?

Or the actual profiles to make it more relevant to your work.

Participant 10
There's a a couple of bits that we wouldn't find.

The the demographics there is fantastic. Would use everything there.

And reason for visiting, we definitely would use tasks and motivation. We would obviously the entry museums and galleries there, we wouldn't we wouldn't need that.

Knowledge levels we would for possibly for certain clients.

Some of our more.

Interviewer
Do you think you think you have to swap and change a bit project to project? Essentially make it specific?

Participant 10
Yeah, but that will, yeah, that would probably be parts that we'd we'd drop out for certain projects. We do have some quite technical clients. We have a client that's.

Just like robotics.

Where obviously you don't expect the users of that website to be knowledgeable on something like that. They go in there. They're often going there is they supply robotics to things like factories. So it'll be like factory owners and things that go in looking at robotics. They're not likely to know everything, the insurance and outs of what they're buying.

They're just looking to speed up their manufacturing process, so knowledge levels there would come in really handy because you might have guys who've researched it and they're not what they're going for and stuff, but you might have the guys who don't.

Do I need knowledge levels for like a landscaping client? Probably not.

Interviewer
No.

Participant 10
So yeah, yeah, like that customization would be great.

Interviewer
That's OK, cool.

Participant 10
From my previous projects, yeah.

Yeah.

Interviewer
OK, so I've just got a last couple of questions just to finish off and this is around feedback, future consideration and stuff. So beyond personas, are there any other methods or tools that you currently use or are exploring to understand the users or the audience better?

Participant 10
And beyond the initial onboarding document, no.

Exploring. Yes, it's it is. It's one of my goals for this year.

Every single.

Member of staff at our track from apprentices and juniors all the way up to the directors all set themselves objectives for the year. They're documented. We set tasks towards them and one of mine this year is to improve the UX process so.

Definitely exploring ways of doing that.

But right now, no.

Interviewer
OK. And with the evolving tools and the methods and things that are going on in the world, how do you see the future of user profile in and in UX design?

Participant 10
AI generated.

Interviewer
That's OK.

Participant 10
Uh, huh. I mean.

Yeah.

Hopefully based on legit, that's corrected as you've shown us.

Hopefully not just kind of predicted based on other stuff I don't know. But yeah, I think it's definitely going that way where it's going to be AI generated.

Interviewer
OK.

Umm, last question, do you think that I've missed anything that you would like or do you? Was there anything that you'd like to add on in this conversation?

Participant 10
Umm.

No, I don't think so.

Interviewer
That's OK.

Yes.

Participant 10
No, I don't think so. I think it was. Yeah. Really good. Really, really end up.

Interviewer
OK, cool. Well, thank you for that. I'm just going to stop the transcripts and the recording there.

Appendix K

Additional Material

K.1 Ethics Application and Approval

David Walsh
Registration number: 140266647
Information School
Programme: Information Studies (PhD/Info Studs (SSc) PT) INFR41

Dear David

PROJECT TITLE: Investigation of support for different users groups in digital cultural heritage
APPLICATION: Reference Number 011190

On behalf of the University ethics reviewers who reviewed your project, I am pleased to inform you that on 04/11/2016 the above-named project was **approved** on ethics grounds, on the basis that you will adhere to the following documentation that you submitted for ethics review:

- University research ethics application form 011190 (dated 07/10/2016).
- Participant information sheet 1022939 version 1 (19/08/2016).
- Participant consent form 1022941 version 1 (19/08/2016).

If during the course of the project you need to deviate significantly from the above-approved documentation please inform me since written approval will be required.

Yours sincerely

Matt Jones
Ethics Administrator
Information School

David Walsh
Registration number: 140266647
Information School [a.k.a iSchool]
Programme: Information School (PhD/Info Studs (SSc) PT) INFR41

Dear David

PROJECT TITLE: Validation Study of Infomration Seeking Cluster Profiles (ISCPs)
APPLICATION: Reference Number 055559

On behalf of the University ethics reviewers who reviewed your project, I am pleased to inform you that on 27/09/2023 the above-named project was **approved** on ethics grounds, on the basis that you will adhere to the following documentation that you submitted for ethics review:

- University research ethics application form 055559 (form submission date: 18/09/2023); (expected project end date: 28/12/2023).
- Participant information sheet 1127519 version 2 (18/09/2023).
- Participant consent form 1127520 version 2 (18/09/2023).

If during the course of the project you need to deviate significantly from the above-approved documentation please inform me since written approval will be required.

Your responsibilities in delivering this research project are set out at the end of this letter.

Yours sincerely

Peter Bath
Ethics Administrator
Information School [a.k.a iSchool]

Please note the following responsibilities of the researcher in delivering the research project:

- The project must abide by the University's Research Ethics Policy: https://www.sheffield.ac.uk/research-services/ethics-integrity/policy
- The project must abide by the University's Good Research & Innovation Practices Policy: https://www.sheffield.ac.uk/polopoly_fs/1.671066!/file/GRIPPolicy.pdf
- The researcher must inform their supervisor (in the case of a student) or Ethics Administrator (in the case of a member of staff) of any significant changes to the project or the approved documentation.
- The researcher must comply with the requirements of the law and relevant guidelines relating to security and confidentiality of personal data.
- The researcher is responsible for effectively managing the data collected both during and after the end of the project in line with best practice, and any relevant legislative, regulatory or contractual requirements.

Chapter 10

Abbreviations

AGNES ..... Agglomerative Nesting

AI ..... Artificial intelligence

ARI ..... Adjusted Rand Index

ASK ..... Anomalous State of Knowledge

CH ..... Cultural Heritage

CLARA ..... Clustering Large Applications

CPD ..... Continuing Professional Development

CSS ..... Cascading Style Sheets

DCH ..... Digital Cultural Heritage

DIANA ..... Divisive Analysis

DK ..... Domain Knowledge

DNS ..... Domain Name Server

ECA ..... Exploratory Cluster Analysis

ESS ..... Error Sum of Squares

FANNY ..... Fuzzy Analysis

GA ..... Google Analytics

GLAM ..... Galleries, Libraries, Archives and Museums

GMT ..... Greenwich Mean Time

GP ..... General Public

GUI ..... Graphical User Interface

HCI ..... Human Computer Interface

HCD ..... Human Centred Design

HCPC ..... Hierarchical Clustering on Principal Components

HTTP ..... Hypertext Transfer Protocol

ICH ..... Intangible Cultural Heritage

ICT ..... Information and Communications Technology

IEEE ..... The Institute of Electrical and Electronics Engineers

IIS ..... Internet Information Services

IMA ..... Indianapolis Museum of Art

IP ..... Internet Protocol

IR ..... Information Retrieval

ISCP ..... Information Seeking Cluster Profile

ISM ..... International Slavery Museum

ISP ..... Information Search Process

IT ..... Information technology

ISCP ..... Information Seeking Cluster Profiles

LC ..... Library of Congress

MS ..... Museum Staff

ML ..... Machine Learning

MIP ..... Museum Information Professional

MCA ..... Multiple Correspondence Analysis

MDS ..... Multidimensional scaling

MOL ..... Museum of Liverpool

NDL ..... National Digital Library

NIR ..... No Information Rate

NMI ..... Normalized Mutual Information

NML ..... National Museums Liverpool

NP ..... Non-Professional

NVI ..... Normalized Variation of Information

PC ..... Personal Computer

PCA ..... Principal Component Analysis

PAM ..... Partitioning Around Medoids

PDF ..... Portable Document Format

PRISMA ..... Preferred Reporting Items for Systematic Reviews and Meta-Analyses

QPC ..... Quantitative Persona Creation

ROCK ..... Robust Clustering using links

SLR ..... Systematic Literature Review

TK ..... Technical Knowledge

TLA ..... Transaction Log Analysis

TLD ..... Transaction Log Data

UK ..... United Kingdom

UCD ..... User Centred Design

URI ..... Uniform Resource Indicator

URL ..... Uniform Resource Locator

UX ..... User Experience

UXD ..... User Experience Design

VPN ..... Virtual Private Network

WML ..... World Museum Liverpool

WSS ..... Within-Cluster-Sum of Squared Errors

WWW ..... World Wide Web

XML ..... Extensible Markup Language

XP ..... Extreme Programming