

# What Are Public Data Gaps?

Naoise Coakley

Master of Arts by Research

University of York

Law

August 2023

## Abstract

The UK Government has more power than ever to gather and use data in its decision-making. Despite this power, there are still significant areas where data that should be held by public bodies is missing. Where this data is missing or otherwise suffers from poor quality, a data gap occurs. The problem of data gaps underlies an emerging legal debate about the role of the law, and public law in particular, in challenging and closing data gaps, especially when they produce harmful effects on the population. Nonetheless, to enable this debate and further legal analysis, there is a need for an improved understanding of public data gaps and a clear categorisation of different types of gaps. Existing research on this topic is disharmonised due to a lack of clarity surrounding the definition and categorisation of public data gaps. The paper establishes the need for a shared, precise, and functional language that can be utilised across academic disciplines that interact with data gaps in their research and study. Utilising an interdisciplinary scoping review of social science literature, this paper provides a new typology of gaps. This new typology presents five novel categories of public data gaps and demonstrates these through active case studies within the UK context. These case studies serve to show the prevalence of these gaps and the need for the language to be improved so that legal solutions and analysis can be explored.

## List of Contents

Abstract .....	2
List of Contents .....	3
Table of Legislation .....	4
List of Tables .....	4
Declaration.....	4
I. Introduction .....	5
II. Methodology .....	9
III. A New Typology of Data Gaps .....	15
IV. Collection gap.....	29
V. Participation gap.....	40
VI. Reporting gap .....	50
VII. Harmonisation gap.....	59
VIII. Delay gap .....	68
IX. Conclusion.....	76
Technical Appendices .....	83
Technical Appendix 1: Scoping Review Methodology .....	83
Process of the Scoping Review .....	83
Limitations of Scoping Reviews .....	91
Technical Appendix 2: Search Strings.....	93
Technical Appendix 3: Flow Chart for article selection .....	96
Technical Appendix 4: Literature Results by Research area and Country of Origin .....	97
Bibliography .....	101

## Table of Legislation

Digital Economy Act 2017

Equality Act 2010

Freedom of Information Act 2000

Police Crime and Evidence Act 1984

Road Traffic Act 1988

The Housing (Homeless Persons) Act 1977

The Housing Act 1996

## List of Tables

Figure 1: New Typology of Public Data Gaps ..... 14

Figure 2: The Giest and Samuels Typology of Data Gaps..... 19

## Declaration

I declare that this thesis is a presentation of original work, and I am the sole author. This work has not previously been presented for a degree or other qualification at this University or elsewhere. All sources are acknowledged as references.

## I. Introduction

As the UK shifts towards a “digital future”<sup>1</sup>, the ability of the Government to gather large amounts of public data has become a central point for understanding the potential of data. Across all government systems, and wherever individuals interact with public bodies, data can be collected and used to inform officials. However, despite the existing, and increasing, capacity for data collection by public bodies, the government may choose not to collect helpful data, and it often makes that choice. This missing data is what is often referred to as, and also what I will refer to as, “data gaps”.

Data gaps have the capacity to cause problems in governance and often lead to harm in a range of ways. Where data is missing, harm may be generated in decision-making that does not include relevant information. For instance, when a data gap is present, certain groups can be excluded from decision-making and policy processes due to being “invisible” in the data. A lack of data on sexual and gender minorities permitted the World Health Organisation to “do nothing” in response to reports about the health inequities sexual and gender minorities may face<sup>2</sup>. Sexual and gender minorities, therefore, experience harm from this decision as health equity is not evaluated and these minorities are locked out of policy consideration. Harm does not only affect individuals or groups; harm can be experienced by public bodies and the government itself, as missing data can frustrate the policy and decision-making processes. A lack of data also prevents the exercise of accountability over public institutions, and timely insights into the functioning of public institutions is undermined or prevented completely. Little data is taken on how certain police tactics, such as “stop and account”, affect racial minorities<sup>3</sup>. This lack of data has clear implications for the eradication

---

<sup>1</sup> Heather Wheeler, “Transforming for a Digital Future: Government’s 2022-25 Roadmap for Digital and Data” (HM Government Central Digital and Data Office, 2022) <[https://digitalanddata.campaign.gov.uk/#government\\_s\\_digital\\_and\\_data\\_strategy\\_explained](https://digitalanddata.campaign.gov.uk/#government_s_digital_and_data_strategy_explained)> accessed 10 July 2023.

<sup>2</sup> Po-Han Lee, 'Un(ac)countable no-bodies: the politics of ignorance in global health policymaking' (2023) 33 *Critical Public Health* 48.

<sup>3</sup> Anna Powell-Smith, 'Stop and Account: How is this Little-Known Police Practice Being Used?' (The Centre for Public Data, 2019) <<https://missingnumbers.org/stop-and-account/>> accessed 20 May 2023.

of discrimination in the police force, but also for preventing the police itself from understanding how policing policies are used and evaluating whether existing tactics are effective for deterring or responding to crime. Therefore, there is significant interest in understanding where and how data gaps are produced and how to manage them to minimise these harms as much as possible.

In recognition of the harms that public data gaps pose, there is a growing debate in legal scholarship and practice as to how far public law is equipped to respond to the data gaps problem<sup>4</sup>. The main questions in this debate ask whether public law can manage or close data gaps, whether there are or should be obligations from the government to collect data, and whether there should be an obligation for Government to close gaps when it is made aware of them<sup>5</sup>. Furthermore, some ask whether the existence of gaps has the potential to violate equality duties, such as those contained within the Equality Act 2010<sup>6</sup>. Furthermore, questions of equality law often tie into the debate on whether gathering data is important for ensuring duties within decision-making are fulfilled<sup>7</sup>. This legal debate also forms part of a wider policy debate about the relationship between the role of good data in the pursuit of good government and good administration. While this work will not address these questions in particular, these questions all rely on an understanding of data gaps which is, currently, extremely limited.

Underlying the important legal debate surrounding data, there is an assumption of what a “data gap” actually is. The discussion in legal scholarship thus far appears to focus on instances where the government has completely failed to collect useful data. This conception of data gaps is accurate, but this thesis will demonstrate that it is extremely limited in its scope. It is clear that the study of data gaps, especially within legal literature, is

---

<sup>4</sup> Joe Tomlinson, Jed Meers, and Cassandra Sommers-Joce, 'Judicial Review of Public Data Gaps' (2023) *Judicial Review* <<https://www.tandfonline.com/doi/full/10.1080/10854681.2023.2218256>> accessed 13 July 2023.

<sup>5</sup> Ibid.

<sup>6</sup> Ibid.

<sup>7</sup> Sarah Giest and Annemarie Samuels, "For Good Measure": Data Gaps In A Big Data World' (2020) 53 *Policy Sciences* 559.

based upon the assumption that data gaps primarily exist where data has not been collected. Whilst this assumption is correct, it is myopic in its understanding of where, why, and how gaps exist and appear. This research will establish that data gaps arise in a variety of contexts and extend beyond where data has simply not been collected. Data gaps exist anywhere where relevant information is not gathered, or where information that is gathered is obscured due to poor quality or is inaccessible. Missing or obscured information prevents the effective use of data for policy and decision-making, and therefore recognising all the ways in which information can be missing or obscured is highly relevant. Too narrow a view of data gaps will stunt efforts to find and fill gaps and avoid their harms. Given this, the purpose of this thesis is to challenge this assumption and ask a foundational question which is yet to be addressed in data gaps research: what is a data gap? More precisely, the question I will seek to address is: what are the different types of public data gaps?

In order to explore existing conceptions of data gaps, an interdisciplinary scoping review of the literature was undertaken. A review of data gaps literature across a range of disciplines and its interaction with decision-making and the behaviour of officials, has not been undertaken before. Therefore, this review will be the first of its kind to attempt to do so. Based on this review, I build my central argument: that there is a need for clear, precise, and harmonised language surrounding data gaps that can be utilised across disciplines which engage with data gaps research. To this end, a novel taxonomy of data gaps, which disaggregates different types of gaps, is developed.

This paper will be structured as follows. Firstly, section one will discuss the methodology of the scoping review. This section will establish why a scoping review approach was followed and its value to the project as a whole. This section will also lay out the process of the scoping review and how literature was gathered and sifted to select the final articles for review. Section two will present a new typology of public data gaps, which has been developed using the literature gathered in the review. This new typology identifies five categories of public data gaps. These categories are collection gaps, participation gaps, reporting gaps, harmonisation gaps, and delay gaps. Each gap will be defined, and case

study examples of each gap will be presented. Finally, this paper will make clear that the need to understand and close data gaps should be of interest to lawyers in order to ensure that harms are ended or prevented, and legal principles are upheld in the continuing legal debate into the role of public law in providing remedies.



## II. Methodology

This study adopted a scoping review method to review the existing literature. Data gaps research is a growing topic of study, and a small amount of literature is available. Scoping reviews are the most suitable method of reviewing emerging and small-scale literature bases<sup>8</sup>, which was therefore well suited to the data gaps literature base. Scoping studies have particular value in examining the “extent, range and nature” of existing research and identifying areas where more research is required on a particular topic<sup>9</sup>. Furthermore, a scoping review is considered preferable when it is needed to identify key characteristics or factors related to a concept, and to clarify key definitions in the literature<sup>10</sup>. These aims of scoping reviews directly align with the purpose of this literature review as part of this research. While scoping reviews have been more typically used in scientific literature, there has been an increase in the use of scoping studies in social sciences literature. Scoping reviews have been used to establish a policy base for old age poverty research<sup>11</sup>, map the literature on mental health, addictions, and suicide in social work education<sup>12</sup>, and to examine links between dangerous driving and other criminal behaviours<sup>13</sup>.

Research often uses systematic reviews in order to answer specific research questions. A systematic review was not deemed to be appropriate for this project. Systematic reviews aim to summarise the research available on a specific topic to answer a specific question and tend to use smaller numbers of empirical studies to do so<sup>14</sup>. For

---

<sup>8</sup> Hilary Arksey and Lisa O'Malley, 'Scoping Studies: Towards A Methodological Framework' (2005) 8 *International Journal of Social Research Methodology* 19.

<sup>9</sup> Ibid

<sup>10</sup> Zachary Munn and others, 'Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach' (2018) 18 *BMC Medical Research Methodology* < <https://bmcmmedresmethodol.biomedcentral.com/articles/10.1186/s12874-018-0611-x>> accessed 10 July 2023.

<sup>11</sup> Crystal Kwan and Christine A. Walsh, 'Old age poverty: A scoping review of the literature' (2018) 4 *Cogent Social Sciences*.

<sup>12</sup> Tuola Kourgiantakis and others, 'Social work education and training in mental health, addictions and suicide: a scoping review protocol' (2019) 9 *BMJ Open*.

<sup>13</sup> Lyndel Bates, Marina Alexander and Julianne Webster, 'The link between dangerous driving and other criminal behaviour: a scoping review' (2022) 21 *Safer Communities* 137.

<sup>14</sup> Mai T. Pham and others, 'A Scoping Review Of Scoping Reviews: Advancing The Approach And Enhancing The Consistency' (2014) 5 *Research Synthesis Methods* 371.

example, systematic reviews research highly specific questions, such as “assessing the effects of selected state and federal firearms laws on violence-related public health outcomes”<sup>15</sup>. This question is an example of the specificity required of systematic review questions to reach specific research goals based on an established literature and research base. Systematic review methodologies, furthermore, have strict inclusion and exclusion criteria in order to find the most relevant literature. These criteria are more able to be established due to the preexisting knowledge of the literature base used in systematic reviews<sup>16</sup>. In contrast, the purpose of scoping reviews is far broader; aiming to present a wider overview of the research on an area in order to better address wider questions about a research topic<sup>17</sup>. Scoping reviews tend to be an exploratory exercise in order to assess the level of research and literature that is available on a particular topic. Scoping reviews, therefore, do not have the same level of specificity as systematic reviews; in fact, it has been suggested that for emerging topics of study, scoping reviews are a necessary ‘first step’ before specific systematic reviews can take place<sup>18</sup>. The value of a scoping study in this review was to determine what literature is available on data gaps across disciplines and to determine definitions of gaps in the literature. Therefore, a scoping review was the most suited method to establish the state of existing research and provide insight into how other disciplines view data gaps. This can provide a basis for further research across disciplines into how data gaps produce harm.

It was clear at the outset of this research that the legal literature on data gaps is extremely limited, if nearly non-existent. Therefore, an interdisciplinary approach was deemed to be the appropriate avenue to find a range of literature that discusses data gaps, including literature that has studied data gaps from a non-legal perspective. By drawing upon literature from other disciplines, this literature review aims to understand and classify the

---

<sup>15</sup> Robert Hahn and others, “Firearms Laws and the Reduction of Violence: A Systematic Review” (2005) 28 *American Journal of Preventative Medicine* 40.

<sup>16</sup> Munn and others (n 10).

<sup>17</sup> Arksey and O'Malley (n 8).

<sup>18</sup> Munn and others (n 10).

manner in which data gaps are discussed, understood and demonstrated across data gaps research, in order to then apply this knowledge to the betterment of legal analysis in the future.

In order to undertake the scoping review, a search of the database “Web of Science” was completed. This database was chosen for its depth of available literature across research areas. A search string containing relevant terms for the search was applied, returning approximately three hundred pieces of literature. These articles then went through two sifting exercises in order to screen for relevance and applicability to the research question. In total, fifty-nine articles were included in the final scoping study<sup>19</sup>.

This methodology was informed by previous examples of scoping reviews used in social sciences literature, as well as the literature available on scoping review methodologies. My review did, however, diverge from these studies when it came to the search string. The final search string applied for this review was actually three separate strings, each with its own terms and research questions guiding them. This was necessary in order to access the range of articles required for a review into the connection between data gaps, reasons gaps exist, and decision-making. I would encourage further reviews which follow a similar methodology to consider if creating multiple search string “blocks” would also be a suitable way of achieving their research aims. The literature from the blocks was brought together for the final analysis and the answer to the research question for this project, making this strategy highly successful.

A review methodology of this kind could be useful for lawyers in the future, especially where the area of review is emerging. Law, in many cases, can be considered a “rendezvous subject”<sup>20</sup>, where shared interests from across disciplines meet. It is therefore beneficial for legal research to have an interest in interdisciplinary approaches, especially to

---

<sup>19</sup> For more details on the methodology of the Scoping review, and the inclusion and exclusion criteria applied to the literature, see technical appendix 1 on page 79. For copies of the search strings applied in this review, see technical appendix 2 on page 89.

<sup>20</sup> Eric Hoddy and others, 'Legal culture and climate change adaptation: An agenda for research' (2023) *WIREs Climate Change* <<https://dx.doi.org/10.1002/wcc.825>> accessed 10 July 2023.

novel problems such as data gaps. As novel problems occur, especially in the fields of data or technology, the law must continue to develop at a pace where it can provide legal solutions to issues. It may become more necessary for law and lawyers to recognise the value of co-opting research from other disciplines for the purposes of legal analysis. Scoping reviews are an attainable way to do this, as it allows for much broader questions to be asked and answered and broader research to be returned. Improving the knowledge base allows for legal analysis of emerging issues to be a thorough reflection of the problems experienced and permits novel solutions to these issues to be determined and attained.

Title	Definition	Sub-Categories
Collection Gap	Pure failure to collect data by official sources.	Type One: Data on this issue is simply not collected
		Type Two: Data is not collected on this issue with regularity/every year or within a set time frame
Participation Gap	Data is collected with limited or nonparticipation from a section or sections of the population.	Type One: This lack of participation is due to the collection method denying, erasing, or preventing the full participation of sections of the population (E.g., a Digital divide, wilful exclusion of some groups)
		Type Two: Lack of participation due to a section of the population choosing not to participate due to social stigma, lack of trust, or other structural social barriers
Reporting Gap	Gaps occur due to the method, means, or decisions made concerning the reporting of the data that has been collected.	Type One: Data is collected but is not reported or made available.
		Type Two: Data is reported in such a way that it fails to create a clear picture of the data results. This can be through a lack of disaggregation or granularity, failures to modify data discussion for layperson audiences, or

		lack of clarity in the data findings when reported.
Harmonisation Gap	Data cannot be harmonised with other data sets, preventing comparability and relevant takeaways from the data	Type One: Lack of harmonisation of definitions (of what is being collected/reported) and time scales.
		Type Two: Lack of harmonisation of access protocols/sharing data (can be due to data management/ownership, differing legislation etc)
Delay Gap	There is a significant delay in other government departments' or researchers' ability to access already collected data.	

*Figure 1: New Typology of Public Data Gaps*

Figure 1 lays out the new typology of Public Data Gaps which will be analysed in depth in Chapter III. The above figure acts as a table of reference for readers throughout the rest of this work.

### III. A New Typology of Data Gaps

Data gaps literature consistently identifies data gaps, yet there is no harmonised language used to identify and discuss gaps when they occur. In part, this is a product of the fact that data gaps literature ranges across a multitude of disciplines<sup>21</sup>, so it is expected that language and vocabulary used to describe and identify will differ<sup>22</sup>. However, in the context of developing responses to data gaps, this lack of a shared language for discussing gaps is problematic. Without a clear and precise language which is shared across disciplines which interact with data gaps research, identification and analysis of gaps in a multidisciplinary fashion is significantly more difficult. Without this, research from across disciplines cannot be fully brought together, academics and practitioners who interact with gaps cannot share their knowledge to the fullest extent, and gaps are allowed to persist even where they are identified and capable of being closed.

Furthermore, there is a developing legal debate which surrounds the ongoing analysis of the existence and impacts of public data gaps, as well as providing a basis for the exploration of potential legal solutions. The legal debate on data gaps focuses on where data has not been collected. Until now, gaps in public data which are the result of a lack of data collection have served as the primary source of gaps for legal analysis and discussion. However, the focus on where data has not been collected as the sole type of data gap in legal research terms is not a comprehensive understanding of data gaps and requires immediate revision. Other disciplines, including social sciences such as criminology, have been able to recognise that gaps exist in a multitude of forms and are not confined to simply where data has not been collected. Therefore, providing a harmonised language around all the forms of data gaps allows for any gaps analysis from other disciplines to be synthesised into legal perspectives on data gaps. Clear and precise language in relation to the existence

---

<sup>21</sup> See technical appendix four for a full breakdown of the review's research areas.

<sup>22</sup> Giest and Samuels (n 8).

and potential harms of every type of gap will therefore permit linked identification across disciplines and, in the context of the legal debate, will enable a better law-centred analysis of gaps, motivations, and harms.

The literature base upon which this project is based is a broad collection of multidisciplinary works which, at first, appear to paint data gaps in broad terms. However, most, if not all of this literature, identifies common features of data gaps and common points of analysis in their examination of the gaps that the work identifies. The recognition of these common features demonstrates that gaps in the public data set that the works interact with are systemic and common. One such common feature is that the term “data gaps” is treated as a broad term referring to missing, unclear, or poor-quality data and information. Across disciplines, the term “data gap(s)” has been generally used to describe where data is not collected, where data can’t be accessed, where data is of poor quality, as well as where data is simply inadequate in a range of scenarios. Whilst using “data gaps” as a catch-all term is beneficial for capturing the range of gaps that can exist, there is a danger that gaps analysis discusses data gaps as if they are all the same; as in, they are all caused by similar issues, which is simply not the case. The lack of a clear definition of what gaps are, as well as what different gaps are, has led to data gaps literature being disharmonious and prevented disciplines from learning from one another on how to find and manage gaps in their field of study. It has also prevented more research into the different causes, features, and harms from different types of gap. Poor quality data in public data sets will result in different harms in decision-making than gaps which result from data which is non-existent due to non-collection. It is therefore important that this work discusses these gaps as distinctly different.

In legal literature, there is a particular focus on unavailable or uncollected data. This is a valid concern, but it is myopic in the sense that it does not expand its view outwards to other ways in which information can be missing or obscured. Furthermore, with this focus, legal literature tends to pay attention to how data gaps can cause harm to individuals, such as the Centre for Public Data’s recent findings on the lack of data related to sentencing and



bail for individuals in criminal courts<sup>23</sup>. This focus on the harm felt by individuals often approaches data gaps from the perspective of equality or non-discrimination duties, or legal principles such as access to justice<sup>24</sup>. Whilst this is valid, legal literature has a blind spot for non-personal harms, such as harms that are caused to the process of policy making, and the administrative problems for the government that public data gaps cause. Recognising harm as being more than unlawful or harmful behaviour from the government to the public enables more understanding of what harms do result from data gaps. Understanding harm in different contexts of different data gaps allows better identification of impacts of data gaps and solutions to those gaps. Therefore, an improved and disaggregated definition of the types of data gaps that appear in public data is necessary.

Previous typologies of data gaps have been presented. The Geist and Samuels typology has been influential for a number of years but does not seem to have been fully adopted by all disciplines which partake in data gaps research. Giest and Samuels defined data gaps as “data for particular elements or social groups that are knowingly or unknowingly missing when policy is made on the basis of large datasets”<sup>25</sup>. This typology split data gaps into three categories: primary, secondary, and hidden<sup>26</sup>. A “primary gap” occurs when the government has an awareness of missing data but has limited opportunities to use ‘appropriate’ means to fill the gaps. Giest and Samuels identify that the use of “synthetic” or “proxy data” have been used as means to fill these gaps by the government, but that these proxy means are inappropriate. Synthetic data does not provide clear and accurate pictures of minority or social groups within datasets<sup>27</sup>, as it is often based upon already flawed or biased data.

A “secondary gap” is where there is government awareness of the existence of a data gap, but the data available may be of poor quality and not a good “fit”. The government

---

<sup>23</sup> Anna Powell-Smith and Gideon Leibowitz, 'Data and Statistical Gaps in Criminal Justice', (Center for Public Data, March 28 2023).

<sup>24</sup> Ibid.

<sup>25</sup> Giest and Samuels (n 7).

<sup>26</sup> See Figure One.

<sup>27</sup> Giest and Samuels (n 7).

may also attempt to fill secondary gaps through other sources, such as social media data. The use of secondary data sources carries significant limitations, including the potential of poor data quality, selection biases, and increased challenges in data processing. Social media data, for example, tends to overrepresent the WEIRD group: Western, Educated, Industrialised, Rich, and Democratic<sup>28</sup>. Those of higher economic status also tend to be on multiple social media platforms at once, meaning that they may generate more data points and appear more often in datasets<sup>29</sup>. This creates bias in the data as this group becomes overrepresented, and gaps are not adequately filled.

“Hidden gaps” refer to data sets that are routinely used for policymaking but contain misrepresentation, bias, or missing data, without the government being aware that it is missing. These gaps are especially problematic when data sets with hidden gaps are used by artificial intelligence or machine learning decision-making systems by government, as these methods often result in flawed decision-making or faulty inferences from the data<sup>30</sup>.

This typology has gained significant traction in the time since its publication in 2020, and has been cited in research concerning the intersection between data gaps and the potential for “big data ” and artificial intelligence in decision-making processes. For instance, the Giest and Samuels definition of data gaps has been cited as recognising data quality issues as a result of poor participation as a data gap and that these gaps can cause biases in machine learning systems<sup>31</sup>. The definition used by Geist and Samuels has been used to advocate for improvements in data collection to enable better delivery of public service, with additions of more points of collection referring to race and gender<sup>32</sup>, as well as greater efforts to include marginalised groups in the “data value chain”<sup>33</sup>. Furthermore, Giest and Samuels

---

<sup>28</sup> Ibid.

<sup>29</sup> Ibid.

<sup>30</sup> Ibid.

<sup>31</sup> Arho Suominen and Arash Hajikhani, 'Research themes in big data analytics for policymaking: Insights from a mixed-methods systematic literature review' (2021)13 *Policy and Internet* 464.

<sup>32</sup> Erna Ruijter and others, 'Social equity in the data era: A systematic literature review of data-driven public service research' (2023) 83 *Public Administration Review* 316.

<sup>33</sup> Alicja Pawluczuk, 'Digital Youth Inclusion and the Big Data Divide: Examining the Scottish Perspective' (2020) 9 *Internet Policy Review* 1.

appear to have produced the first typology of data gaps in data used for decision-making. This definition and the typology have broken ground in this area of study; drawing together public data and decision-making, the risks of inequality that poor or missing data could produce, and dividing it into three distinct forms of gap. At time of publication, this was a distinctly new way of looking at data gaps and the underlying causes, and different results that each form of gap could produce.

	<b>Data unavailable</b>	<b>Data available</b>
<b>Data gap known</b>	Primary data gap	Secondary data gap
<b>Data gap unknown</b>	Hidden data gaps	

*Figure 2: The Giest and Samuels Typology of Data Gaps*

The Giest and Samuels typology is limited in two respects. Firstly, the Secondary Gap category is too broad. Under Giest and Samuels, secondary gaps encompass any scenario where there is government awareness of a gap, and, crucially, that there is data available to fill it despite this data not being a good fit<sup>34</sup>. Under this typology, a secondary gap is a very broad category. While there is logic in having this broad category due to the numerous ways it can appear in public datasets, it does not provide a clear basis for effective analysis. Broad terminology “lumps” all data quality issues in together, despite the fact that poor-quality data can be caused by vastly different problems. The secondary gaps category under this typology includes gaps that can result from a wide variety of reasons, such as data not being collected by the government or government knowledge that official surveys have poor representation. These two scenarios are derived from different problems and have different solutions and gathering them into one category lessens the unique nature of each of these gaps. While causes of secondary gaps are recognised as including data quality issues, further disaggregation in order to understand the different impacts of quality

---

<sup>34</sup> See Figure 1.

problems is required. Furthermore, poor data quality can be the result of various flaws within data collection, processing, and reporting systems, and therefore should be disaggregated by what causes the gap. By failing to recognise the different origins of quality problems, the overly broad Secondary Gaps prevent in-depth explorations of different causes, harms, and solutions to gaps which are considered “secondary”.

Secondly, it is important to note that the Giest and Samuels typology approaches the categorisation of gaps through the lens of whether ‘big data’ can shut gaps. “Big Data” refers to data that is so large, dynamic, and varied that it requires non-conventional data processing methods, and it may have been drawn from non-conventional data sources, such as social data generated by internet users<sup>35</sup>. “Big Data” is viewed by many in the data community as having the potential to provide more, and better, insights into populations and provide more knowledge for policymakers to draw upon when making decisions<sup>36</sup>. The Giest and Samuels typology expressly identifies gaps through the lens of big data and whether it can manage gaps. By having this view, the typology does not focus on the underlying reasons why gaps may exist. As stated previously, the causes of gaps should be recentred when defining categories of gaps. Having clear categories enables easier identification of gaps and harmonised language across disciplines. This is a significant flaw in Giest and Samuels’ work, and it is shared across other definitions of data gaps that have appeared in the literature.

There have been further attempts to define data gaps, the most notable of which came as part of a research report by the Centre for Public Data into a variety of data gaps in the Courts and Justice System in England and Wales. The report defined data gaps as “[a]reas where a lack of official data means that questions of significant public interest cannot be answered”<sup>37</sup>. This definition implies two things about data gaps which must be rebutted. Firstly, this definition implies that gaps exist where there is a lack of data, and does

---

<sup>35</sup> Giest and Samuels (n 7).

<sup>36</sup> Ibid.

<sup>37</sup> Powell-Smith and Leibowitz (n 23).

not extend to where data quality causes gaps. This definition is limited and maintains its focus on only where data is unavailable. This limited view of data gaps is a common theme across data gaps literature. Only recognising data gaps as referring to missing information is myopic. The way that data is collected, how it is presented, and how it is made available, can all result in gaps occurring. These gaps can occur both alongside and simultaneously within public data sets, as gaps based on information that is not collected. A lack of data is not the only way a gap exists; gaps exist where the functioning of a public institution or public service cannot be understood due to data gaps arising from missing, obscured, or inaccessible data or information.

Secondly, the Centre for Public Data definition focuses on gaps that prevent questions from being answered, and specifically questions of “significant public interest”. This definition limits data gaps unnecessarily; gaps exist regardless of whether there is a public interest, significant or minor, in them. While it can be said that all public data gaps have significant interest due to their nature within the public sphere, limiting the definition of data gaps here prevents its applicability to data gaps which may exist outside of this sphere. As noted from the scoping literature review, data gaps research spans across the private and public spheres, and the “significant public interest” element of the definition presented by the Centre for Public Data can become difficult to apply. For example, understanding the information-gathering processes of private higher-education institutions of the numbers of degree-seeking displaced students<sup>38</sup> is a niche data gap that, while relevant to institutions and government, may not receive as much public attention or interest. Furthermore, it may be impossible to predict what will be considered within the public interest in the future, which especially impacts issues which rely on data that is collected in the present. Future levels of public interest in different questions may arise, and therefore data that is not collected or is of poor quality now will impact the ability to answer these questions later. As this thesis

---

<sup>38</sup> Lisa Unangst, Ishara Casellas Connors and Nicole Barone, 'State-Based Policy Supports for Refugee, Asylee, and TPS-Background Students in US Higher Education' (2022) 38 *Refuge: Canada's Journal on Refugees* 95.

presents a new typology of data gaps, allowing for the applicability of the typology and definition across disciplines is necessary.

Thirdly, the definition by the Centre for Public Data is flawed by having a focus on gaps that we know about, or areas where we know that data is missing. Public data, and access to public data, should allow research and questions to develop into issues where questions should be asked about, by currently aren't known. Data gaps exist regardless of whether they are known; the Geist and Samuels' typology calls these "hidden gaps". These "unknown" gaps are constantly emerging; definitions of data gaps should therefore be able to remain applicable to these gaps, which occur outside of public institution's awareness.

A revised typology of data gaps is clearly required for furthering the analysis of data gaps across disciplines and researchers. Based upon the literature reviewed for this project, for the remainder of this paper, data gaps are defined as existing where public data is either missing, of poor quality, poorly reported, or inaccessible to the government and researchers.

In order to build the new typology, I returned to the literature, especially focusing on the examples of gaps identified in this literature. While creating a bank of examples, similar themes and language relating to certain gaps began to emerge across the literature. I noted these recurring themes and their corresponding examples, and from here began to outline categories of gaps that each example could fit into<sup>39</sup>. The finalised typology identifies five categories of gaps<sup>40</sup>. Four of the five categories identify two possible scenarios of gaps occurring that fit under that category. For each category where this occurs, this has been classified as either a "type one" or a "type two" gap.

The first type of gap occurs when data has not been collected by official sources. This is a "collection gap". The literature identified two variations of collection gap, which I shall label as a "type one" collection gap, and a "type two" collection gap. A type one collection gap occurs when there is a failure to collect data on a particular issue. This could

---

<sup>39</sup> See Appendix 2.

<sup>40</sup> See Figure 2.

be due to the omission of relevant questions during data collection<sup>41</sup>. The omission may be the result of data not being deemed necessary to be collected<sup>42</sup>, despite there being relevant reasons to collect data points. A type two collection gap is where data on a particular issue is not collected with regularity, such as every year or within a set time frame, such as a census. This may be due to difficulties in consistent data collection due to the social, political or economic environment<sup>43</sup>, or failures in data collection design to collect data with set time frames to create long-term pictures of the surveyed population<sup>44</sup>. The impact of type two gaps can range from inconclusive results from the data<sup>45</sup>, to difficulties in creating cross-country studies<sup>46</sup>.

The second gap identified is a “participation gap”. Participation gaps result in the collected data being of poor quality. A participation gap occurs where data is collected with limited, or non-participation from a section of the population. Where there is limited participation, for whatever reason, any collected data may contain significant holes or provide an inaccurate outcome of the research question based on poor or biased data results<sup>47</sup>. Participation gaps can also be classed as a type one or a type two participation gap. A type one data gap arises where there is a lack of participation due to the method of data collection denying, erasing, or preventing the full participation of sections of the population. Methods of data collection can unintentionally prevent the full participation of

---

<sup>41</sup> Kerris Cooper And Polina Obolenskaya, 'Hidden Victims: The Gendered Data Gap Of Violent Crime' (2021) 61 *British Journal Of Criminology* 905. See also: Maria Sourbati and Frauke Behrendt, 'Smart Mobility, Age and Data Justice' (2021) 23 *New Media & Society* 1398.

<sup>42</sup> Jennifer Bronson and Carolyn Sufrin, 'Pregnant Women in Prison and Jail Don't Count: Data Gaps on Maternal Health and Incarceration' (2019) 134 *Public Health Reports* 57.

<sup>43</sup> Chaza Akik and others, 'Responding To Health Needs Of Women, Children And Adolescents Within Syria During Conflict: Intervention Coverage, Challenges And Adaptations' (2020) 14 *Conflict And Health* 37.

<sup>44</sup> Tetine Sentell and others, 'Data Gaps In Adolescent Fertility Surveillance In Middle-Income Countries In Latin America And South Eastern Europe: Barriers To Evidence-Based Health Promotion' (2019) 11 *South Eastern European Journal Of Public Health* 214.

<sup>45</sup> Kathleen Abu-Saad, Shlomit Avni and Ofra Kalter-Leibovici, 'Health Disparities Monitoring In The US: Lessons For Monitoring Efforts In Israel And Other Countries' (2018) 7 *Israel Journal Of Health Policy Research* 14.

<sup>46</sup> Jere R. Behrman and Mark R. Rosenzweig, 'Caveat-Emptor - Cross-Country Data On Education And The Labor-Force' (1994) 44 *Journal Of Development Economics* 147.

<sup>47</sup> Becky Pettit and Bruce Western, 'Mass Imprisonment and the Life Course: Race and Class Inequality in U.S. Incarceration' (2004) 69 *American Sociological Review* 151.

targeted groups<sup>48</sup>, and undermine the data that is successfully collected<sup>49</sup>. A type two participation gap occurs when there is a lack of participation in data collection due to sections of the population choosing not to participate. This choice could be due to social stigma, lack of trust, or other structural social barriers which prevent full participation<sup>50</sup>. When there is limited participation in data collection, the data that is gathered is not of optimal quality. When decisions are made on the basis of datasets which may have certain groups underrepresented in the data, these groups may not receive the benefits of decisions made due to their “invisibility”<sup>51</sup>.

The third gap identified in the literature is a reporting gap. Reporting gaps occur due to the method, means, or decisions made concerning the reporting of the data that is collected. The manner in which data is reported is highly significant for the effective use and translation of data for decision-making. When data is inaccurately or unclearly reported, it becomes inaccessible for decision-makers and the public to use, understand, and apply to decision-making<sup>52</sup>. Reporting gaps can also be classed into type one and type two gaps. A type one gap occurs when data has been collected but is not reported or made available. Reasons for this type of gap are often based on “reporting thresholds”. Reporting thresholds require a certain level or percentage of completed responses before there can be an official publication of the data<sup>53</sup>. Type two gaps occur when data is reported in such a way that it fails to create a clear picture of the data results. This can be through a lack of disaggregation

---

<sup>48</sup> Bina Agarwal, 'Imperatives Of Recognising The Complexities: Gendered Impacts And Responses To Covid-19 In India' (2022) 39 *Economia Politica* 31.

<sup>49</sup> Giest and Samuels (n 7).

<sup>50</sup> Geoffrey S. Holtzman, Neda A. Khoshkhoo and Elaine O. Nsoesie, 'The Racial Data Gap: Lack of Racial Data as a Barrier to Overcoming Structural Racism' (2022) 22 *The American Journal of Bioethics* 39. See also: Sentell and others (n 44)

<sup>51</sup> Henrie M. Treadwell And Others, 'Discerning Disparities: The Data Gap' (2018) 13 *American Journal Of Mens Health* <<https://doi.org/10.1177/1557988318807098>> accessed 10 July 2023.

<sup>52</sup> Stephanie. E. Galaitsi and Others, 'The Challenges Of Data Usage For The United States' Covid-19 Response' (2021) 59 *International Journal Of Information Management* <<https://www.sciencedirect.com/science/article/pii/S0268401221000451>> accessed 10 July 2023.

<sup>53</sup> “Universal Credit Statistics: Background Information and Methodology” (Department for Work and Pensions, 14 February 2023) <<https://www.gov.uk/government/publications/universal-credit-statistics-background-information-and-methodology/universal-credit-statistics-background-information-and-methodology>> accessed 10 July 2023.



or granularity in the reported results<sup>54</sup>, lack of clarity in the data findings when published<sup>55</sup>, or inability to modify data discussion and terminology for layperson audiences<sup>56</sup>.

The fourth gap identified is a harmonisation gap. A harmonisation gap occurs when collected data cannot be brought together, or harmonised, with other data sets. In reference to harmonisation, in this context, this means that data that has been collected by different sources cannot be brought together with similar data that is stored or collected separately in order to create fuller datasets. When datasets are relied upon for decision-making, it is key that all the available information on that particular issue is able to be brought together to create a clear picture for decision-makers. When data exists but cannot be brought together due to a lack of frameworks for sharing, the true picture remains fragmented<sup>57</sup>.

Harmonisation gaps can particularly affect cross-jurisdictional research and comparisons, either at state or international levels, are key ways in which researchers and public bodies can understand international and global trends across sectors such as health<sup>58</sup>, and the economy<sup>59</sup>. When there is a lack of comparable data, these efforts are hampered<sup>60</sup>. The scoping review identified that harmonisation gaps appear in two forms. The first form is where there is a lack of harmonisation between definitions of what is being collected<sup>61</sup>, a lack of harmony between what is being reported, and a lack of harmony in the time scales of data collection time scales<sup>62</sup>. This is a type one harmonisation gap. These harmonisation gaps prevent data sets from being fully integrated and the results of the data from being extracted

---

<sup>54</sup> Holtzman, Khoshkhoo and Nsoesie (n 50).

<sup>55</sup> Unangst, Casellas Connors and Barone (n 38).

<sup>56</sup> Rajiv Bhatia, Isabella Sledge and Stefan Baral, 'Missing science: A scoping study of COVID-19 epidemiological data in the United States' (2022) 17 *PLoS ONE* <<https://doi.org/10.1371/journal.pone.0248793>> accessed 10 July 2023.

<sup>57</sup> Akik and others (n 43).

<sup>58</sup> Alan Katz and others, 'Challenges Associated with Cross-Jurisdictional Analyses using Administrative Health Data and Primary Care Electronic Medical Records in Canada' (2018) 3 *International Journal of Population Data Science* 437.

<sup>59</sup> Behrman and Rosenzweig (n 46).

<sup>60</sup> Joshua R. Vest And L. Michele. Issel, 'Data Sharing Between Local Health And State Health Departments: Developing And Describing A Typology Of Data Gaps' (2013) 19 *Journal Of Public Health Management And Practice* 357.

<sup>61</sup> Sentell and others (n 44).

<sup>62</sup> Behrman and Rosenzweig (n 46).

to create a clear picture. Harmonisation gaps also occur where there is a lack of harmonisation in the access or sharing protocols across different government departments and jurisdictions<sup>63</sup>. This is a type two harmonisation gap. Differences in sharing protocols, particularly between public bodies or government departments, can prevent timely solutions<sup>64</sup>. Lack of harmonisation can also mean that useful data cannot be brought together and utilised to its fullest extent<sup>65</sup>. Lack of harmonised data access or secure sharing protocols can also result in “black spots” in data sets, where data is known to exist but cannot be included to create reliable narratives<sup>66</sup>.

The final type of gap identified in the literature is a delay gap. A delay gap occurs when there is a significant delay in government departments’ or researchers’ ability to access collected data. Delay gaps also frequently prevent Members of Parliament or policy officials from having access to data which is important for the evaluation or determination of the effectiveness of policy. The lack of timeliness in access to information can undermine the purpose of the data collection and prevent the effective use of data for policymaking, especially where there is time pressure to make policy. For example, the inability to share Covid-19 data, in near-real time, prevented the adoption of time-sensitive solutions to Covid-19 management in the United States<sup>67</sup>. Delay gaps can result in research being unable to be completed, questions being left unanswered, and an overall failure in transparency and accountability of government.

These new categories and definitions are unique when compared to the definitions of data gaps that have been utilised by recent data gaps publications. Apart from the much wider definition of data gaps that this typology presents, it also places the possibility of harm

---

<sup>63</sup> Vest and Issel (n 60). See also Akik and others (n 43).

<sup>64</sup> Gavin Freeguard and Paul Shepley, ‘Data Sharing Between National, Devolved and Local Government: A Summary of a Private Roundtable’ (*Institute for Government*, 30 January 2023) <https://www.instituteforgovernment.org.uk/publication/data-sharing-national-local-government> accessed 10 July 2023.

<sup>65</sup> Vest and Issel (n 60).

<sup>66</sup> Jennifer Kirsty Burton And Others, ‘Closing The UK Care Home Data Gap – Methodological Challenges And Solutions’ (2020) 5 *International Journal of Population Data Science* <<https://dx.doi.org/10.23889/ijpds.v5i4.1391>> accessed 10 July 2023.

<sup>67</sup> Galaiti and others (n 52).

as a central component of data gaps. Previous works fail to emphasise just how harmful gaps are and how fixable many gaps are if there is the political power, and will to do so. This argument seems obvious in the context of collection gaps; there is power in being able to determine what is and isn't worthy of collection. However, I contend that this is also applicable to the four other identifiable types of gaps within this taxonomy. As four of the five types of gaps relate to collected data of varying quality, it can be argued that the burden on public bodies to improve the quality of their data requires will and recognition of existing failures. There is a notable and asymmetric power dynamic in this relationship. When public bodies become aware of the various gaps in their data, without the will from the body to make changes, gaps will persist. That is not to say that all gaps are wilful; as will be examined in depth through the use of active UK based examples, gaps often occur due to differences in organisation practices, or differing legislation, and can even be due to rational and reasonable decisions to do with how and what data is presented. Nonetheless, when there are gaps in knowledge and information, even for understandable reasons, harm is generated.

Gaps in knowledge obscure the picture, and the real impact of an obscure picture is that there is a persistent and harmful lack of clarity, which should not be a consistent feature of a functioning government democracy. Accountability and transparency are two foundational features of good governance. The ability of the public and those outside government to review, assess, and hold decisions and policies accountable encourages better and more effective governance. Secondly, it is vital that within government and public bodies, there is an ability to assess its policies and projects, hold itself accountable, and measure successes and failures. Government assessment of policy successes or failures allows for there to be policy responsiveness, and ultimately can result in better outcomes for the public. Therefore, recognition of gaps and the different forms in which they appear helps to understand the reasons for gaps existing, the different forms that gaps can take, and the power dynamic that exists in the existence and management of gaps.

The next section of this paper will explore each gap identified in the typology in more depth. Firstly, I shall discuss the definition and structure of each gap, as well as explore reasons why the gap may exist within the context of the sociology of ignorance, the study of the powerful role that ignorance can play when harnessed by public officials<sup>68</sup>. The following section will also establish the harms the gap may cause, and the impact of each gap on public decision-making. Each gap will also be illustrated with individual case studies which identify UK data gaps, all of which remain open at the time of writing. These case studies will illustrate how data gaps appear in public life and data in the United Kingdom, and if, and how, public and academic debate seeks or is seeking to close these gaps.

---

<sup>68</sup> Linsey McGoey, 'Strategic unknowns: towards a sociology of ignorance' (2012) 41 *Economy And Society* 1

#### IV. Collection gap

A collection gap occurs where there is a failure to collect relevant data and information by official sources. There are two types of collection gap. The first type of collection gap is where data is simply not collected. By this, the Government or other public bodies fail to collect relevant data about a particular issue or from a section of the populations. This can be considered a “true” data gap, as the data that is later sought-after by researchers or decision-makers simply does not exist in official, reliable, and accessible databases. A type two collection gap occurs where data is not collected with regularity or within a set time frame, resulting in incomplete time series in data sets. This type of collection gap is focused on where data can be obtained and used for decision-making, but due to inconsistencies in data collection time frames, or lack of regularity, the utility of the data is diminished<sup>69</sup>. While certain types of data collection are designed to be within set time frames and with regularity, such as censuses, data collection design does not always follow set patterns. It is uncommon for datasets used in public governance and decision-making to be limited to a very short time period or a “one-off” data collection due to the ongoing need for comparative data when making decisions<sup>70</sup>. There are exceptions to this, such as during the Covid-19 pandemic where short-term data was required in order to make immediate decisions on health and social policy. Nevertheless, where data is not collected with regularity, large gaps in data sets occur. The result of this is that decision-makers are unable to view a snapshot of relevant information.

Collection gaps often occur due to a lack of understanding over which information is relevant and what is not at the time of data collection<sup>71</sup>. Type one gaps are frequently the result of questions not being asked, and information not being taken at points in which data collection is taking place<sup>72</sup>. Within sociological studies, there are two primary theories on why

---

<sup>69</sup> Behrman and Rosenzweig (n 46).

<sup>70</sup> Ibid.

<sup>71</sup> Bronson and Sufrin (n 42).

<sup>72</sup> Ibid.

decisions not to gather certain pieces of data could take place. These are known as “non-recognition”, and “mis-recognition”<sup>73</sup>. Non-recognition is an omissive act, where the existence or need for information is unknown or completely overlooked<sup>74</sup>. Whereas mis-recognition is commissive, where the need for information is known about but purposefully excluded. Non-recognition may be done without malicious intent; researchers may not have an interest in certain information, and its necessity is only recognised after the fact. Nevertheless, the resulting harms of a collection gap that results from mis-recognition can be severe, especially for communities with protected characteristics. Missing data on environmental features which may affect the mobility of persons with disabilities, for example, can prevent independent living and safe travels for those with reduced mobility<sup>75</sup>. This can become a failure of the government to provide accessible spaces for all and has the potential to violate duties of equality or fairness<sup>76</sup>.

Identified in the literature are suggestions that more collaboration between data collection agencies and members of marginalised groups in order to increase visibility within data collection<sup>77</sup> and data results. This would likely resolve some of the harms faced by marginalised groups’ exclusion from data capture; often, it does not become clear what is needed or missing until it is identified by those who are harmed by it<sup>78</sup>. Nevertheless, ignorance on the part of public bodies does not bear well as a justification for why collection gaps exist, especially when considering why type one collection gaps which emerge as a result of mis-recognition occurs.

---

<sup>73</sup>Both these theories are informed by the sociology of nothing, an emerging topic which explores “negatively defined phenomena”. The sociology of nothing makes a stark distinction between active commission of nothing, and passive omission when discussing the existence of “nothing”. See Susie Scott, ‘A Sociology of Nothing: Understanding the Unmarked’ (2017) 52 *Sociology* 3.

<sup>74</sup>Lee (n 2).

<sup>75</sup> Shiloh Deitz, Amy Lobben and Arielle Alferéz, ‘Squeaky Wheels: Missing Data, Disability, And Power In The Smart City’ (2021) 8 *Big Data & Society* <<https://journals.sagepub.com/doi/full/10.1177/20539517211047735>> accessed 10 July 2023.

<sup>76</sup>Tomlinson, Meers, and Summers-Joce (n 4).

<sup>77</sup> Karen Wright and others, ‘Indigenous health equity in health register ascertainment and data quality: a narrative review’ (2022) 21 *International Journal for Equity in Health* 34.

<sup>78</sup> Deitz, Lobben and Alferéz (n 75).

Mis-recognition is a commissive act, where the need for data is known but dismissed<sup>79</sup>. The ability to undertake a commissive decision to not take data is a clear example of the contexts of power within which data collection exists. Non-evidence, or the lack of evidence, says just as much as the existence of evidence does<sup>80</sup>. Non-evidence, especially that which comes about as the result of non-recognition, reflects power differences, and identifies what is considered important or prioritised in research and for decision-making<sup>81</sup>. When public bodies view non-evidence in the frame of strategic ignorance, it legitimises decisions to “do nothing”, which can further harm and silence those for which the evidence is lacking<sup>82</sup>. Strategic ignorance, in this context, refers to the political and social practices embedded in the effort to suppress or kindle new forms of ambiguity<sup>83</sup>, which can be used to the Government's benefit through the obstruction of relevant, but uncomfortable<sup>84</sup>, knowledge and information<sup>85</sup>. Data collection is a powerful tool for public bodies, and a lack of data can be used to obscure information through the dissemination of strategic ignorance. Where evidence or data is lacking, governments can intentionally harness this missing information in order to “do nothing”<sup>86</sup>, rather than view it as an opportunity to gather knowledge. Choices to do nothing, or maintain non-recognition, continue to reflect the power imbalances in decision-making. It can be idealistic to think that all decision-making is based on gathered evidence and data, and there are varying levels of how true this may be<sup>87</sup>. Research and data collection is shaped by the interests of various stakeholders<sup>88</sup>, and the application of that information to decision-making is also shaped by

---

<sup>79</sup>Lee (n 2).

<sup>80</sup>Ibid.

<sup>81</sup> Ibid.

<sup>82</sup> Ibid.

<sup>83</sup> McGoey (n 68).

<sup>84</sup> Natalie Chong, Jose-Frederic Deroubaix and Celine Bonhomme, 'Eyes wide shut: Exploring practices of negotiated ignorance in water resources modelling and management' (2018) 227 *Journal of Environmental Management* 286.

<sup>85</sup> Ibid.

<sup>86</sup> Lee (n 2).

<sup>87</sup> Erik Albaek, 'Between Knowledge And Power - Utilization Of Social-Science In Public-Policy Making' (1995) 28 *Policy Sciences* 79.

<sup>88</sup> Ibid.

various interests. Whilst research contributions provide policy and political benefits for policymakers<sup>89</sup>, active information avoidance can occur if it interferes with a decision maker's pre-held beliefs or decisions<sup>90</sup>. This information avoidance could assist in identifying why mis-recognition can take place in type one collection gaps, although it is unlikely that this is the only reason for omissions taking place.

Type two collection gaps, where data is not consistently collected resulting in gaps in datasets, can often be the result of poor data collection design. This can occur when, at the point of designing the collection of data, the wider need to collect data over a certain period or time frame is disregarded. While the need for comparable data is often a key feature of data design<sup>91</sup>, it can be overlooked by data collectors. This lack of foresight for the need for consistent data collection does not have clearly identified reasons in the literature; however, this could be recognised as another example of non-recognition, the omissive act of not gathering data<sup>92</sup>. Type 2 gaps can also be the result of social or political factors which prevent data collection, such as the outbreak of war<sup>93</sup>. These unforeseen factors disrupt the data collection process, displace surveyed populations, and can lead to data collection after-the-fact to be of poor quality<sup>94</sup>. Whilst these unforeseen barriers to consistent data collection are somewhat unavoidable, the resulting harms from this lack of collection are still notable. The inability to collect data can prevent key public functions from being effective, such as preventing health monitoring and provision<sup>95</sup>. This can further harm populations which may already face disadvantages.

---

<sup>89</sup> Karen Bogenschneider, Elizabeth Day and Emily Parrott, 'Revisiting Theory on Research Use: Turning to Policymakers for Fresh Insights' (2019) 74 *American Psychologist* 778.

<sup>90</sup> Russell Golman, David Hagmann and George Loewenstein, 'Information Avoidance' (2017) 55 *Journal of Economic Literature* 96.

<sup>91</sup> Behrman and Rosenzweig (n 46).

<sup>92</sup> Lee (n 2).

<sup>93</sup> Akik and others (n 43).

<sup>94</sup> Peter Van der Windt and Macartan Humphreys, 'Crowdsourcing in Eastern Congo: Using Cell Phones to Collect Conflict Events Data in Real Time' (2016) 60 *JOURNAL OF CONFLICT RESOLUTION* 748.

<sup>95</sup> Akik and others (n 43).



Furthermore, attempts to resolve these data issues are often flawed. Where data is collected after the fact, it often suffers from being of poor quality<sup>96</sup>. Selection and recall biases can serve to undermine the data collected<sup>97</sup> and can be unreliable narratives of events or impacts<sup>98</sup>. Efforts to fill type two collection gaps through the use of estimates or “fictionalised data”<sup>99</sup> are also widely considered problematic<sup>100</sup> by decision-makers and data analysts. Fictionalised data and estimates can often be useful for filling gaps in tables where data may be missing from certain years, but they themselves contain biases and inaccuracies<sup>101</sup> based on the quality of the data used to create these estimates. When estimates are relied upon to fill collection gaps, the data remains flawed. Harms therefore remain; where estimates themselves are based on poor quality data; the resulting data findings are not guaranteed to have accuracy for decision-makers.

#### Collection Gaps Case Study: Police Vehicle Stops

The non-collection of information about police vehicle stops provides a clear and, at the time of writing, live example of collection gaps in the action. There is very little data or information available regarding police vehicle stops, such as the time and location of the stop<sup>102</sup>. Vehicle stops are not required to be recorded, and Police body cameras are not required to be on whilst a stop is taking place.

In England and Wales, Police have the power to stop a vehicle “for any reason” under section 163 of the Road Traffic Act 1988. This allows a uniformed officer to stop any vehicle on the road, including the ability to stop bicycles and electric scooters. If a vehicle

---

<sup>96</sup> Van der Windt and Humphreys (n 94).

<sup>97</sup> Ibid.

<sup>98</sup> Ibid.

<sup>99</sup> Behrman and Rosenzweig (n 46).

<sup>100</sup> Ibid.

<sup>101</sup> Ibid.

<sup>102</sup> 'Traffic Stops (Vehicle Stops) Factsheet' (*StopWatch*, 24 October 2022) <[https://www.stop-watch.org/what-we-do/research/traffic-stops-factsheet/#:~:text=What%20is%20a%20vehicle%20stop,car%20or%20scooter\)%3B%20or](https://www.stop-watch.org/what-we-do/research/traffic-stops-factsheet/#:~:text=What%20is%20a%20vehicle%20stop,car%20or%20scooter)%3B%20or)> accessed 10 July 2023.

is flagged down by the Police and asked to stop, the driver must comply<sup>103</sup>. There are no procedural requirements<sup>104</sup> for vehicle stops under section 163 Road Traffic Act 1988, meaning that Police can stop vehicles even when there is not a suspicion that a criminal offence may have or be occurring<sup>105</sup>. Vehicle stops can lead to further use of formal police powers, such as a vehicle search if the officer has reasonable grounds to suspect they will find stolen or prohibited items in the vehicle<sup>106</sup>. Vehicle stops, therefore, is a serious police power which can be broadly unlimited in its use. However, the occurrence of a police vehicle stop is not required to be recorded by the police<sup>107</sup>.

There is very little data collected on vehicle stops by any of the 43 police forces in England and Wales. It is therefore difficult for detailed research to be conducted on the effectiveness of vehicle stops as a police tactic, and in 2015 a report by His Majesty's Inspectorate of Constabulary and Fire & Rescue Services found that none of the 43 police forces in England and Wales had conducted an audit of vehicle stops to determine if they were an effective police tactic in their area.

Academics and public interest organisations have questioned the lack of data collection on vehicle stops<sup>108</sup>. Police powers permit formal stops and searches to take place, data on which are recorded and published. Stop and Searches, for example, have been highlighted as an area in which data collection on the tactic is crucial in order to determine its effectiveness. Whilst this type of stop has also been focussed upon because of its discriminatory use against ethnic minorities<sup>109</sup>, recording its usage has revealed information about the effectiveness of these police tactics generally. Under the stop and

---

<sup>103</sup>Ibid.

<sup>104</sup> Ibid.

<sup>105</sup> Ibid.

<sup>106</sup> Section 1 Police and Criminal Evidence Act 1984.

<sup>107</sup> Mike Rowe and Geoff Pearson, 'We spent seven years observing UK police stop and search - here's what we found' (*University of Liverpool*, 2020) <://news.liverpool.ac.uk/2020/11/16/we-spent-seven-years-observing-uk-police-stop-and-search-heres-what-we-found/> accessed 18 April 2023.

<sup>108</sup> Ibid.

<sup>109</sup> 'Disproportionate Use of Police Powers: A Spotlight on Stop and Search and the Use of Force' (HM Inspectorate of Fire, Police and Rescue Services, 2021) <<https://www.justiceinspectrates.gov.uk/hmicfrs/wp-content/uploads/disproportionate-use-of-police-powers-spotlight-on-stop-search-and-use-of-force.pdf>> accessed 15 July 2023.

search statistics, the government is now able to track the change in usage through the past two years, as well as:

- demographic information including gender and ethnicity,
- number of stops and searches, and subsequent arrests carried out under different legislation.
- reason(s) for the stop and search
- day-of-week and time-of-day trends in stop and search
- stop and search hotspots<sup>110</sup>

This kind of information allows for a clearer understanding of this power. It permits research and scrutiny, ensuring that the Police, as an institution, remains accountable.

This kind of data could be recorded for vehicle stops as well<sup>111</sup>. Under sections 164 and 165 of the Road Traffic Act 1988, if a police officer asks for basic information including name and address, the driver must give the information. Knowledge of who is being stopped, when they are being stopped, where they are stopped, and how often a vehicle stop leads to further police action could be useful information. This can be identified as an area of non-recognition taking place, where the information is not seen as useful at this time.

Calls for data collection on vehicle stops, particularly on ethnicity and other demographic factors such as age and sex, were met in 2021 by a pilot run by the Metropolitan police. The Mayor of London and the Metropolitan police initially introduced the pilot after several high-profile vehicle stops by Police of people of colour, including the Labour MP Dawn Butler<sup>112</sup>. For six months in 2021, the Metropolitan Police ran a pilot

---

<sup>110</sup> 'Police Powers and Procedures: Stop and Search and Arrests, England and Wales, year ending 31 March 2022', (*HM Government Home Office*, 27 October 2022) <[<sup>111</sup> Rowe and Pearson \(n 107\).](https://www.gov.uk/government/statistics/police-powers-and-procedures-stop-and-search-and-arrests-england-and-wales-year-ending-31-march-2022/police-powers-and-procedures-stop-and-search-and-arrests-england-and-wales-year-ending-31-march-2022#:~:text=In%20the%20year%20ending%20March%202022%20there%20were%20526%2C024%20stop.and%20searches%20had%20been%20increasing.> accessed 10 July 2023.</a></p></div><div data-bbox=)

<sup>112</sup> Inzamam Rashid, 'Labour MP Dawn Butler Accuses Metropolitan Police Of Racial Profiling After Being Stopped By Officers' *Sky News* (10 August 2020) < <https://news.sky.com/story/labour-mp->

programme which required officers to ask every driver stopped under s163 of the Road Traffic Act 1988, to declare their ethnicity. Drivers were not required to answer this question, but the officer had to input either the driver's "self-defined ethnicity" or an "officer-defined ethnicity" if the driver did not provide a self-defined ethnicity<sup>113</sup>. Officers also collected information on the Driver's sex, age, the time of the stop, and the reason for the stop<sup>114</sup>. Officers were given discretion as to when to complete the provided e-form, either at the roadside or at a later time. The average time it took to record the ethnicity and other demographic features of drivers added 2 minutes to each stop<sup>115</sup>.

The pilot was well received when introduced. In Parliament, MPs widely supported the wide-ranging nature of the pilot, and campaigners for the collection of ethnicity data, such as the legal organisation Liberty, welcomed the pilot. However, MPs at the time also noted that they were doubtful that Police Officers could often see the ethnicity of drivers before they were stopped, which, they believed, diminished the likelihood of discriminatory application of the power. Police officers shared the same concern as MPs, reporting that most of the time, they cannot see the ethnicity of the driver until the vehicle has stopped. The pilot found that police were most likely to stop white drivers, with around 53% of all stopped drivers being white. This is fairly proportionate to the population of London, which is approximately 55% white<sup>116</sup>. Stops on black drivers represented 16.5% of all stops, while this ethnicity makes up 13.5% of London's overall population. These figures, according to campaigners, are evidence of racial disproportionality<sup>117</sup>. Independent analysis by Dr Krisztian Posch analysed the data from the pilot for the Guardian

---

[dawn-butler-accuses-metropolitan-police-of-racial-profiling-after-being-stopped-by-officers-12045860](https://www.theguardian.com/uk-news/2022/oct/11/met-to-stop-recording-ethnicity-of-drivers-stopped-by-its-officers)> accessed 10 July 2023.

<sup>113</sup> The Officer-defined ethnicity was only introduced for the final two months of the pilot.

<sup>114</sup> Reason for stop was introduced after the first month of the pilot, following feedback from the participating officers.

<sup>115</sup> Vikram Dodd and Mirren Gidda, 'Met to Stop Recording Ethnicity of Drivers Stopped by its Officers' *The Guardian* (11 October 2022) <<https://www.theguardian.com/uk-news/2022/oct/11/met-to-stop-recording-ethnicity-of-drivers-stopped-by-its-officers> > accessed 10 July 2023.

<sup>116</sup> This is from bringing together the data on "white British" and "white other" drivers who were stopped.

<sup>117</sup> Dodd and Gidda (n 115).

newspaper. This analysis found that, compared with their share of the population, Black drivers were 56% more likely to be stopped than White British drivers<sup>118</sup>. There has been further anecdotal evidence from high-profile individuals of colour, such as the athletes Bianca Williams<sup>119</sup> and her partner Ricardo dos Santos<sup>120</sup> whose vehicle stop caused outrage, with accusations of racial profiling and the police officers involved facing disciplinary hearings. When speaking about gathering ethnicity data on road stops, dos Santos pointed out that the pilot goes beyond just establishing whether there are disproportionate road stops for ethnic minorities; gathering the data is an act by the police to regain the trust and confidence of the public<sup>121</sup>.

Black communities in London feel less trust and confidence in the Metropolitan police<sup>122</sup>. Stop and Search statistics show that Black men are seven times more likely to be stopped and searched by police when they are on foot<sup>123</sup>, and vehicle stops under section 163 of the Road Traffic Act 1988 have also been pointed to as evidencing disproportionate policing practices<sup>124</sup>. Consistent gathering of this data helps the police to identify and seek to end racially disproportionate policing and tactics. This data can also be beneficial for understanding Police tactics and for assessing successes and areas for improvement in preventing, detecting, or stopping crime and criminal behaviour.

The Metropolitan Police decided not to continue the pilot beyond the initial six months. Approximately 86% of participating officers thought it should not be rolled out

---

<sup>118</sup> “White British” was classed as a separate ethnicity to “White Other” under the pilot.

<sup>119</sup> The pair were driving through London when their car was stopped, the pair handcuffed, and the car searched and nothing unlawful was found. The pair had also been travelling with their three-month-old baby at the time.

<sup>120</sup> ‘Action Plan – Transparency, Accountability and Trust in Policing’ (*Mayor of London and London Assembly*) < <https://www.london.gov.uk/publications/action-plan-transparency-accountability-and-trust-policing> > accessed 10 July 2023.

<sup>121</sup> Vikram Dodd and Mirren Gidda, 'Met Police To Stop Recording Ethnicity of Drivers Stopped by its Officers' (*Liberty Investigates*, 13 October 2022) <<https://libertyinvestigates.org.uk/articles/met-police-to-stop-recording-ethnicity-of-drivers-stopped-by-its-officers/>> accessed 18 May 2023.

<sup>122</sup> Action Plan - Transparency, Accountability and Trust in Policing (n 119).

<sup>123</sup> Ibid.

<sup>124</sup> Dodd and Gidda (n 115).

permanently<sup>125</sup>. 90% believed it added to their workload and 25% thought it added significantly<sup>126</sup>. Police Officers were also unsure about the purpose of the pilot, with some reporting they felt that the ethnicity of the driver had nothing to do with their role of enforcing road traffic laws<sup>127</sup>.

The Independent Office for Police Conduct included in its 2022 learning report on National Stop and Search Powers that section 163 vehicle stops should have “the grounds upon which a vehicle was stopped, the characteristics of the occupants, and any outcomes resulting from the stop”<sup>128</sup> recorded. The report found that there was a lack of basic data on the use, regularity, and outcomes of the power and that this lack of data has resulted in a lack of transparency and an inability to bring scrutiny on a matter of public interest<sup>129</sup>.

Therefore, collection gaps, both type one and type two, can cause harm. In the context of police vehicle searches, a lack of knowledge about these police practices leaves unanswered and unanswerable questions that undermine the transparency and accountability of the police as a public institution. Through the theories of mis-recognition and non-recognition, it is clear that collection gaps are the result of the context of the power structures which exist in data collection. The non-recognition of why data is necessary to be collected can be combatted. Further education for Police forces as to why the data is necessary for trust and accountability can help to close this aspect of the gap. Mis-recognition may be harder to combat, especially if the lack of data and the strategic

---

<sup>125</sup> Owen Pyle, ‘Report on the s163 RTA Ethnicity Recording Pilot’ (*The Metropolitan Police*, 17 May 2022) < [https://www.met.police.uk/SysSiteAssets/foi-media/metropolitan-police/disclosure\\_2022/november\\_2022/section163-rta-stops-recording-ethnic-background-stopped-drivers.pdf](https://www.met.police.uk/SysSiteAssets/foi-media/metropolitan-police/disclosure_2022/november_2022/section163-rta-stops-recording-ethnic-background-stopped-drivers.pdf)> accessed 10 July 2023.

<sup>126</sup> Ibid.

<sup>127</sup> Pyle (n 125).

<sup>128</sup> ‘National Stop and Search Learning Report’ (*Independent Office For Police Conduct*, April 2022) < <https://www.policeconduct.gov.uk/sites/default/files/Documents/publications/OFFICIAL%20IOPC%20National%20stop%20and%20search%20learning%20report%2020%20April%202022.pdf>> accessed 10 July 2023.

<sup>129</sup> Ibid.

ignorance it produces can be beneficial for the Police as a public institution. Nonetheless, the findings of the national learning report from the Independent Office of Police Conduct demonstrate that this mis-recognition is being combatted at the highest levels by scrutinising bodies.

## V. Participation gap

Participation gaps occur where data is collected but a section, or sections, of the population do not participate in the data collection or have limited participation. Participation gaps are noted as occurring in two scenarios, both of which are due to failures in data collection design. When data collection is poorly designed or does not consider barriers to participation, the data collected is flawed. A type one participation gap prevents the accuracy of data sets due to the method of data collection denying, erasing, or preventing full participation in data collection. A type two participation gap occurs when social forces prevent the full and willing participation of a section or sections of the population in data collection. These social forces, such as lack of trust in public institutions, social stigma, and shame, can prevent full participation in data collection. Participants may choose to avoid questions or not answer fully or truthfully due to the manner in which data collection is presented. Participation gaps are particularly problematic as they leave a lack of certainty and clarity in data sets. It can also lead to public decisions being made on the basis of inaccurate or unrepresentative data sets, which can often lead to more harm in surveyed communities as needs are left unmet.

Participation is fundamental to the effective running of Government. In the context of this research, participation in data collection can be correlated to effective participation in decision-making<sup>130</sup>. Generally, increased participation and engagement from the public results in more optimal decision-making<sup>131</sup>, as more perspectives and inputs are considered in the process of policy and decision-making. Furthermore, the more representative of the population the participation is, public officials become more likely to use it and view it as enabling higher quality policy outcomes<sup>132</sup>. Participation, therefore, has an important function for public officials in encouraging decisions which have more input legitimacy from the

---

<sup>130</sup> Albaek (n 87).

<sup>131</sup> Koen Migchelbrink and Steven Van de Walle, 'When Will Public Officials Listen? A Vignette Experiment on the Effects of Input Legitimacy on Public Officials' Willingness to Use Public Participation' (2020) 80 *Public Administration Review* 271.

<sup>132</sup> *Ibid.*



population. It is therefore problematic for decision-making when participation gaps occur, as it diminishes the likelihood of gaining legitimacy.

Participation gaps, and particularly type one participation gaps, are often the result of structural biases which, intentionally or unintentionally, result in unrepresentative data. This can be particularly harmful for vulnerable groups or those with protected characteristics. Data collection methods need to be considered carefully and with reference to their limitations and failures when determining the best course of action to take. Data collection design, therefore, must consider how to ensure the best participation levels possible through their chosen method of collection. For example, with more data being taken through online interactions, public bodies must consider whom these systems may fail to reach, primarily the less technologically literate such as the elderly<sup>133</sup>. Racial minorities historically produce less digital data than their white counterparts<sup>134</sup>, resulting in less representation of minorities in data collected in this manner. The exclusion of older adults and racial minorities from digital collection methods can be seen to reflect the levels of digital and social exclusion experienced by these groups<sup>135</sup>. Furthermore, choices about data collection, including who to include in surveys, can erase or deny participation in data collection<sup>136</sup>. For instance, Population-level data that does not include those who are incarcerated<sup>137</sup> cannot be considered truly representative of the population.

Type one participation gaps are symptomatic of wider issues with equality and require public policy to take more significant note of them in order to produce more equitable data results. A lack of visibility in data is representative of the power asymmetries between the collector of data and the subject of collection<sup>138</sup>. Data collection exists within the context in which the data is taken, as noted by Sourbati and Behrendt “data is generated in systems

---

<sup>133</sup> Sourbati and Behrendt (n 41).

<sup>134</sup> Giest and Samuels (n 7) .

<sup>135</sup> Sourbati and Behrendt (n 41).

<sup>136</sup> Treadwell and others (n 51).

<sup>137</sup> Ibid.

<sup>138</sup> Betsy Anne Williams, Catherine F. Brooks and Yotam Shmargad, 'How Algorithms Discriminate Based on Data They Lack: Challenges, Solutions, and Policy Implications' (2018) 8 *Journal of Information Policy* 78.

of power and are always captured from particular positions”. Data collection design, therefore, is a position in which there are significant power asymmetries, and existing biases have the potential to shape decisions which will later affect those who can, and cannot, participate in data. It is important to recognise the individual contexts within which participation gaps appear. The context in which the data was collected, the reasons for collection, and the later use of the data all require consideration when examining the harms that a participation gap may later cause. By recognising individual contexts, efforts to manage the production of these gaps can become more attainable.

#### Participation Gaps Case Study 1: Older populations, datafication, and public transport

Participation gaps can be identified where there has been poor or ineffective data collection design. Where public officials do not consider, the need to enable participation in data collection, the data collected is flawed. Policy decisions based on flawed data will ultimately also be flawed, which is why robust participation in data collection is important.

Poor data collection design can lead to older populations being prevented from participating in data collection about public services, such as transport, and technology. Despite digital literacy and usage for those aged 65+ increasing, with a majority of older adults in the UK using connected technologies<sup>139</sup>, perceptions of older adults not having an interest in technology persists<sup>140</sup>. These views have led to studies on technology and public services excluding older commuters from data collection exercises, effectively denying commuters above a certain age the ability to participate and provide information and data. In 2016, a Transport for London study was undertaken to assess the attitudes of commuters about the gathering of data from mobile tracking whilst using Transport for London services. The information gathering was undertaken through a series of six group interviews with different age categories from inner and outer London, with a variety of

<sup>139</sup> 'Access and Inclusion in 2016: Outcomes for Consumers in Vulnerable Circumstances' (*Ofcom*, 15 March 2017) <[https://www.ofcom.org.uk/data/assets/pdf\\_file/0030/98508/access-inclusion-report-2016.pdf](https://www.ofcom.org.uk/data/assets/pdf_file/0030/98508/access-inclusion-report-2016.pdf)> accessed 10 July 2023.

<sup>140</sup> Sourbati and Behrendt (n 41).

transport users, including car, bus, and rail<sup>141</sup>. Whilst the study was successful in reaching a cross-section of London commuters, the study did not include older commuters on any of its panels. The oldest age group included was 45+, which is a very broad category of ages when considered alongside the millions of passengers that use TFL services per year<sup>142</sup>. The study was used to inform the manner in which commuter data was gathered while using transport for London services through mobile device tracking. The study focussed on the attitudes of Transport for London customers to tracking and helped to form recommendations for how Transport for London could undertake mobile tracking with public support. The information gathered was then used to inform recommendations on the education of consumers about data and information collection, transparency about mobile tracking technology and strategy, and communication as to how data usage will benefit customers<sup>143</sup>. Transport for London rolled out its plan to track every Wi-Fi enabled device used on the whole of the underground network from July 2019<sup>144</sup>. The data collected from mobile devices is currently used to understand customer behaviour and understand where more capacity can be gained from across the Underground Rail network<sup>145</sup>. Findings from the 2016 study were directly incorporated into the overall roll-out; recommendations for using signs to inform customers that their data would be tracked were followed<sup>146</sup> and the roll-out was preceded by a public communications initiative by Transport for London<sup>147</sup>. The study, therefore, provided beneficial insights and information which was used to create policy. While older commuters were not overtly excluded from

---

<sup>141</sup> TfL Mobile Data and Privacy, (*Transport for London*, February 2016)

<<https://content.tfl.gov.uk/mobile-data-and-privacy-report.pdf>> accessed 10 July 2023.

<sup>142</sup> 'Transport for London Annual Report and Statement of Accounts 2015/16' (*Transport for London and Mayor of London*) <https://content.tfl.gov.uk/tfl-annual-report-2015-16.pdf> accessed 10 July 2023.

<sup>143</sup> TfL Mobile Data and Privacy (n 141).

<sup>144</sup> James O'Malley, 'TfL is Going to Track All London Underground Users using Wi-Fi' (*WIRED*, 22 May 2019) <<https://www.wired.co.uk/article/london-underground-wifi-tracking>> accessed 18 May 2023.

<sup>145</sup> Ibid.

<sup>146</sup> Ibid.

<sup>147</sup> Question from Gareth Bacon MP to the Mayor of London, Sadiq Khan, during Mayor's Question Time (12 September 2019) <<https://www.london.gov.uk/who-we-are/what-london-assembly-does/questions-mayor/find-an-answer/tfl-tracking-3>> accessed 10 July 2023.

data collection, insights into the feelings of older commuters about data tracking were lost. This lack of information from older commuters represents a participation gap.

The UK's public services are increasingly engaging with "datafication"; the process of taking information and transforming it into quantitative data which can then be used to transform public service delivery<sup>148</sup>. Public transport in particular, can experience benefits from the process of datafication. In recent years, the rise of Mobility as a Service providers (MaaS) has allowed for data to be gathered through "digital transport service platforms that enable users to access, pay for, and get real-time information on a range of public and private transport options"<sup>149</sup>. This can include bus and train websites and apps, which are all generally accessible to those with internet access and digital devices. This information helps to inform transport operators, and transport policy makers, about the habits, needs and wants of public transport users. As MaaS systems become more common across the UK, large data gaps have been revealed however, especially about the ability of the less-digitally literate to engage with them, and particularly the ability of the elderly<sup>150</sup>.

According to Ofcom, connected ICTs are not part of the "daily media diet" of people of an advanced age<sup>151</sup>. Of the 4.8 million adults in the UK who had never used the internet in 2017, nearly 80% were over 65 years old<sup>152</sup>. Participation, therefore, for older adults in the use of MaaS services when accessing public transport can be limited<sup>153</sup> by

---

<sup>148</sup> Jamie Bartlett and Nathaniel Tkacz 'Governance by dashboard: a policy paper' (*Demos*, March 2017) <<https://dera.ioe.ac.uk/28793/1/Demos-Governance-by-Dashboard.pdf>> accessed 10 July 2023.

<sup>149</sup> Marcus Enoch, 'Mobility as a Service (MaaS) in the UK: Change and its Implications' (*Foresight and HM Government Office for Science*, December 2018) <[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/766759/Mobilityasaservice.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/766759/Mobilityasaservice.pdf)> accessed 10 July 2023.

<sup>150</sup> Sourbati and Behrendt (n 41).

<sup>151</sup> Access and Inclusion in 2016: Outcomes for Consumers in Vulnerable Circumstances' (*Ofcom*, 15 March 2017 (n 139).

<sup>152</sup> Cecil Prescott, 'Internet users in the UK: 2017' (*Office for National Statistics*, 19 May 2017) <<https://www.ons.gov.uk/businessindustryandtrade/itandinternetindustry/bulletins/internetusers/2017#:~:text=In%202017%2C%20almost%20all%20adults,20%25%20were%20recent%20internet%20users.>> accessed 10 July 2023.

<sup>153</sup> Sourbati and Behrendt (n 41).

lack of access to the ability to use the MaaS platforms. Firstly, older adults are less likely to be able to access Mobility as a Service providers, meaning their data is harder to collect. This means that the effectiveness of Mobility as a service in fulfilling public transport needs of the community can become harder to assess<sup>154</sup>. Secondly, active exclusion of older adults due to biases also produces unrepresentative data. As the UK population ages, data on older populations becomes more relevant to understand transport needs. These two participation barriers for older people result in participation gaps, excluding older people from “services which generate datasets that can support better decision-making about transport infrastructure and operations”<sup>155</sup>.

The “digital divide” between older adults and the rest of the population, and the increasing datafication of UK public services, has led to a participation gap. Data on older, less digitally literate users of public transport are less likely to be able to be collected through MaaS services. Older populations can directly benefit from better knowledge about public transport. MaaS systems give better information about bus and train arrivals, allowing those unable to stand for long periods to better plan their journeys<sup>156</sup>. Furthermore, a better understanding of transport needs for older people helps local governments to fulfil targets for social inclusion of older adults<sup>157</sup>.

The context in which data is collected must be considered when assessing type two participation gaps. Type two gaps are the result of communities or sections of the population choosing to not participate in data collection. When communities or groups decide to not participate, the context of that decision-making must be addressed in order to gain good

---

<sup>154</sup> Ibid.

<sup>155</sup> Ed Parkes and others, ‘Using Open Data To Deliver Public Services’ (*Open Data Institute*, February 2017) <<https://theodi.org/wp-content/uploads/2018/03/Using-open-data-to-deliver-public-services.pdf>> accessed 10 July 2023.

<sup>156</sup> Sourbati and Behrendt (n 41).

<sup>157</sup> Frauke Behrendt and others, ‘Intelligent Transport Solutions for Social Inclusion (ITSSI) Project Report’, (*University of Brighton*, 2017) <<https://cris.brighton.ac.uk/ws/portalfiles/portal/456536/Behrendt+et+al+%282017%29+Intelligent+Transport+Solutions+for+Social+Inclusion+%28ITSSI%29+Project+Report.pdf>> accessed 10 July 2023.

data. The data that is being collected also has to be considered within the social context of the communities being surveyed. For example, many middle-income countries with strong religious morality and traditions as a part of the social heritage have limited data on adolescent fertility and sexual health, as communities view the questions as inappropriate for adolescents<sup>158</sup>. This view of the “appropriateness” of the data being sought prevents participation, despite it being clear that the data is necessary and important when undertaking health monitoring.

It is well noted that there is a need for trust between communities and public institutions in order to collect good data<sup>159</sup>, yet there can be significant mistrust, especially amongst minority groups<sup>160</sup>. Institutional racism and legacies of discrimination can impact community willingness to participate, which can then only exacerbate the power asymmetries that exist in data collection. As noted previously, there is significant power in deciding what data should be collected, who it should be collected from, and how to then use that data in decision-making. A lack of participation results in a lack of evidence. When the lack of evidence is the result of structural social barriers which themselves are the results of power asymmetries between populations and institutions, power remains imbalanced, and communities can be further cast into the shadows of non-evidence.

Participation Gap Case Study 2: How many Gypsies, Roma, and Travellers actually are there in the UK?

Participation gaps can also result from decisions from sections of the population to not participate in data collection. This can be seen in the lack of data surrounding the Gypsy, Roma, and Traveller populations in the United Kingdom. The Gypsy, Roma, and Traveller populations are counted in the UK through the census, which occurs every ten years. The 2021 Census contained a significant increase in the number of Gypsy, Roma,

---

<sup>158</sup> Sentell and others (n 44).

<sup>159</sup> Van der Windt and Humphreys (n 94).

<sup>160</sup> Treadwell and others (n 51).

and Travellers declaring their ethnicity, with 168,749 individuals reporting. However, this figure remains an estimate of the actual Gypsy, Roma, and Traveller populations, as it is likely that the true number is higher. The UK Government believes the true number could be close to 300,000<sup>161</sup>, while independent research from the University of Salford estimates the figure could be closer to 500,000<sup>162</sup>. Gypsy, Roma, and Traveller communities are not participating fully and freely in data collection, resulting in unclear and sub-optimal information about these communities. According to The Traveller Movement, a civil society organisation that campaigns for equality, inclusion, and anti-discrimination for Gypsy, Roma, and Traveller people, there is a “significant trust gap” between the State and Public Services and people of Gypsy, Roma, and Traveller backgrounds<sup>163</sup>. This lack of trust stems from the frequent and widespread discrimination that Gypsy, Roma, and Traveller people experience, and fears that declaring their ethnicity will be used for discriminatory purposes<sup>164</sup>. Therefore, there is a lack of knowledge about these communities and basic information on the actual number of Gypsy, Roma, and Travellers in the UK cannot be determined.

A lack of knowledge about these communities leads to harm. Gypsy, Roma, and Traveller communities are frequently at risk of “being left behind”<sup>165</sup> by public services such as healthcare and education, as their needs are unknown and unmet. Stigma and hostility towards these communities are also difficult to address successfully where information is missing; lack of trust can also result in Gypsy, Roma, and Traveller

---

<sup>161</sup> 'Gypsy, Roma and Irish Traveller Ethnicity Summary' (*Gov.uk Ethnicity Facts and Figures*, 29 March 2022) <<https://www.ethnicity-facts-figures.service.gov.uk/summaries/gypsy-roma-irish-traveller>> accessed 10 July 2023.

<sup>162</sup> Philip Brown, Philip Martin and Lisa Scullion, 'Migrant Roma In The United Kingdom And The Need To Estimate Population Size' (2014) 8 *People, Place and Policy* 19.

<sup>163</sup> '[Number of Gypsy, Roma, and Traveller People disclosing Ethnicity in Census increases by 110,000](https://travellermovement.org.uk/news/5719#_ftn1)' (*The Traveller Movement*, 29 November 2022) <[https://travellermovement.org.uk/news/5719#\\_ftn1](https://travellermovement.org.uk/news/5719#_ftn1)> accessed 10 July 2023.

<sup>164</sup> Women and Equalities Committee, *Tackling Inequalities Faced by Gypsy, Roma and Traveller Communities* (HC 2017-19 360 April 2019).

<sup>165</sup> 'England's Most Disadvantaged Groups: Gypsies, Travellers and Roma' (*Equality and Human Rights Commission*, March 2016) <<https://www.equalityhumanrights.com/sites/default/files/is-england-fairer-2016-most-disadvantaged-groups-gypsies-travellers-roma.pdf>> accessed 10 July 2023.

communities being unwilling to report harassment and discrimination. From a policy perspective, a lack of knowledge about these communities prevents effective policy decisions on how to provide support in areas where these groups are most disadvantaged, such as in education and healthcare. Furthermore, it is fundamental for the Government function to understand general ethnicity data for its population. By seeking to improve trust and encourage greater participation, the harms felt by these communities as a result of poor data can be managed and minimised where possible.

Building trust can be a challenging task for public bodies and institutions. Greater community engagement, as well as the fostering of positive, consistent relationships, have been suggested as key in order to improve upon the lack of trust<sup>166</sup>. Slowly, as demonstrated by the increasing number of Travellers, Gypsies and Roma identifying in the 2021 census, trust can be improved and data can become more representative.

Nonetheless, long-term solutions in the form of improving data collection design is required. There is a need for greater equality in data design, known as “data justice”<sup>167</sup>. Data justice provides a focus on fairness in the way people are made visible and represented as a result of the data they produce<sup>168</sup>. Data justice as a concept is generally applied to digital data<sup>169</sup>, but it should be extended to all forms of data collection. Data justice provides an ethical basis for the manner in which data is understood and collected. Greater data justice in the way that data collection is designed has the potential to prevent participation gaps before they occur. With greater collaboration with surveyed communities, as well as consideration of how collection can exclude groups from participating, the data gathered is

---

<sup>166</sup> Sarah Sweeney and Zoe Matthews, 'Friends, Families and Travellers: A Guide for Professionals Working with Gypsies, Roma and Travellers in Children's Services' (*Friends, Families and Travellers*, January 2017) < <https://www.gypsy-traveller.org/wp-content/uploads/2017/03/A-guide-for-professionals-working-with-Gypsies-and-Travellers-in-the-public-care-system.pdf>> accessed 10 July 2023.

<sup>167</sup> Sourbati and Behrendt (n 41).

<sup>168</sup> Linnet Taylor, 'What is Data Justice? The Case for Connecting Digital Rights and Freedoms Globally' (2017) *Big Data & Society* <<https://journals.sagepub.com/doi/pdf/10.1177/2053951717736335>> accessed 10 July 2023.

<sup>169</sup> Ibid.



more likely to be of better quality, more representative, and more likely to produce better decisions and policies.

Therefore, type one and type two participation gaps cause harm by preventing truly representative data from being used in decision-making. Whether gaps exist by choice, through the wilful exclusion of certain groups, or through the choice of communities to not participate due to social barriers, participation gaps result in data sets lacking information. Participation gaps can be combatted through improvements in the design of data collection, a pre-emptive step to seek to enable the best data possible. Data collectors, who are in a position of power, should therefore carefully consider how to ensure data is representative, there is access to data collection for communities and that, if necessary, there is communication and connection with communities or sections of the population in order to build trust and relationships.

## VI. Reporting gap

Reporting gaps relate to the manner in which collected data interacts with methods of reporting data findings to decision-makers, or to the public. These gaps occur due to the method, means, or decisions made concerning the reporting of the data that has been collected. Type one reporting gaps occur often in government-collected data. These gaps are due to data being collected, but there are decisions made to not report this data or make it publicly available. While these decisions are often based on “reporting thresholds”, which prevent data that may be based on too few responses from being published, the lack of accessibility to this data makes it impossible to assess the current state of the data and how to improve the data. Type two reporting gaps occur when the method of data reporting creates an unclear image of the data results, which can lead to confusion or sub-optimal decision-making. This type of reporting gap can be considered fairly broad; methods of reporting data can vary widely, and failures in reporting can also vary. Type two reporting gaps can occur where there is a lack of disaggregation and granularity<sup>170</sup> in reported data sets<sup>171</sup>, which prevents a complete understanding of the surveyed populations. Failures were also identified in the inability of data reporting to adjust its audience for laypersons compared to decision makers or experts, and lack of clarity in the results of the data, making the data results inaccessible to those without technical knowledge of the issue being surveyed.

The literature base for this project has not determined what the causes of reporting gaps are. In the case of a type one reporting gap, it can be argued that it is rational for Government or public bodies to not release data that doesn't pass reporting thresholds and is therefore known to potentially be misleading. However, in common with participation gaps, a lack of reporting results in blatant information asymmetry between public bodies and anyone with an interest in the data. Information asymmetry is representative of continued power asymmetries which can be utilised in order to benefit the power-holder. Reporting

---

<sup>170</sup> Katz and others (n 57).

<sup>171</sup> Burton and others (n 65).

gaps can be used as a tool to produce ignorance in populations, which allows those in a position of authority to utilise the tools of government and knowledge production to its own benefit<sup>172</sup>. Ignorance may be cultivated, commodified, and weaponized<sup>173</sup> to the benefit of the government and the detriment of other groups. Reporting data may not be in the best interests of public bodies, and therefore obscuring or non-reporting succeeds in maintaining the status quo<sup>174</sup>. By “holding back” information, public bodies are able to control narratives on issues, even using this ability to produce harm<sup>175</sup>. For example, lack of published information on illegal migration permits Governments to control the public characterisation of migrants in terms of their “criminality or worthiness”<sup>176</sup>.

It is important to note here that the harms which result from reporting gaps are not confined to those felt on an individual level. Reporting gaps prevent effective decision-making by individuals and throughout government as well. Where information is reported in an unclear or limited way, decisions made on the basis of the reported information cannot be done in an effective manner. Therefore, harms that are produced from reporting gaps can be experienced on an institutional and societal level, as the impact of these gaps affects wide-ranging decisions and decision-making capabilities. It is therefore important that reporting gaps are minimised wherever possible.

#### Reporting Gap Case Study 1: Who is on Universal Credit in the UK?

Decisions to not report data can result in broad reporting gaps which prevent clear information about the functioning of a public institution to take place. The Department for Work and Pensions collects ethnicity data from Universal Credit claimants at the point of application, but this information is not published. In the process of applying for Universal

---

<sup>172</sup> Sarah C. Bishop, 'An International Analysis of Governmental Media Campaigns to Deter Asylum Seekers' (2020) 14 *International Journal of Communication* 1092.

<sup>173</sup> Katharina T. Paul and Christian Haddad, 'Beyond evidence versus truthiness: toward a symmetrical approach to knowledge and ignorance in policy studies' (2019) 52 *Policy Sciences* 299.

<sup>174</sup> Ibid.

<sup>175</sup> Bishop (n 171).

<sup>176</sup> Ibid.

credit, applicants are invited to complete non-mandatory, self-declared diversity fields, which includes information about demographic factors such as ethnicity<sup>177</sup>. Due to the non-mandatory nature of these questions, there is not full participation from applicants. The Department for Work and Pensions, therefore, has instituted a “reporting threshold”, where the information gathered from these questions will only be published once completion is at, or over, 70%<sup>178</sup>. The Department of Work and Pensions, like many organisations, places this limit in order to ensure that the 70% threshold “protects from non-response bias and the drawing of false conclusions from the statistics”<sup>179</sup>. This is a valid concern for the Department. The reporting gap which has resulted from the lack of data publication has left important questions unanswered about the demographics of Universal Credit applicants. Any collected ethnicity data from when Universal Credit was rolled out in 2013 to the present has not been reported, resulting in questions which relate to trends in the demographics of those using or applying for Universal Credit cannot be answered. Secondly, it is unclear when this reporting gap will be filled. The most recent figures for completion are from December 2021, where completion rates sat around 58%<sup>180</sup>. It is therefore unknown when completion rates will reach 70% and the desired information will become available. It should also be noted that the demographic information that is unpublished also includes basic information on demographics, such as marital status and other questions about protected characteristics<sup>181</sup>.

It is justifiable that the Department of Work and Pensions does not wish to release information that could potentially be misleading. However, according to researchers, this

---

<sup>177</sup> ‘Universal Credit Statistics: Background Information and Methodology’ (HM Government Department of Work and Pensions, 14 February 2023) < [Universal Credit statistics: background information and methodology - GOV.UK \(www.gov.uk\)](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/108442/universal-credit-statistics-background-information-and-methodology.pdf)> accessed 15 July 2023.

<sup>178</sup> Freedom Of Information response from the Department of Work and Pensions to Aoife O’Reilly (27 June 2022) <<https://www.whatdotheyknow.com/request/864638/response/2068623/attach/html/3/Response%20FOI2022%2041308.pdf.html>> accessed 10 July 2023.

<sup>179</sup> Ibid.

<sup>180</sup> Ibid.

<sup>181</sup> Protected characteristics are a closed list of nine characteristics laid out under the Equality Act 2010, which prevents discrimination of any kind on the basis of the characteristics. The nine characteristics are: age, disability, gender reassignment, marriage and civil partnership, pregnancy and maternity, race, religion or belief, sex, and sexual orientation.

information is relevant and there is clear public interest<sup>182</sup>. Research by the Welfare at a Social Distance project has identified that there has been an increase in Universal Credit claims from people from ethnic minority backgrounds<sup>183</sup>. Without having the official data to confirm this, it then becomes difficult to assess the impact of the pandemic on Universal Credit, Welfare, and the experiences of Ethnic minority groups during this time. More Basic demographic information should be accessible from previous years, as without a time series of ethnicity data a large gap in understanding demographic and trends in Universal Credit applicants cannot be understood.

This reporting gap is also an example of how two types of gaps can exist at the same time and compound upon one another to make larger gaps. In this case, a participation gap exists as well as a reporting gap; in fact, the participation gap has directly led to the existence of the reporting gap. Participation gaps occur where the method of collection denies, erases, or prevents full participation in data collection. As the diversity section of the Universal Credit application is non-mandatory, full participation is therefore not likely to be achieved as there are a lack of incentives to complete the section. This directly results in the reporting threshold not being achieved as participation is too low. This is a clear example of a participation gap and a reporting gap operating simultaneously. Both of these gaps result in harm; demographic information is left unavailable, and the participation gap will likely continue.

Type two reporting gaps, where data is lacking in clarity when it is published, can occur for a series of complex reasons and decisions about the data. Opinions on the utility of the data in a certain form can impact its reporting<sup>184</sup>, but this can also produce problematic or

---

<sup>182</sup> Written Evidence Submitted by Public Law Project to Welsh Affairs Committee Inquiry into Benefits System in Wales ( 17 August 2021) <https://publiclawproject.org.uk/content/uploads/2021/11/Benefit-sanctions-evidence.pdf> accessed 16 July 2023.

<sup>183</sup> Daniel Edminston and others 'Who are the New COVID-19 Cohort of Benefit Claimants?' (Welfare – At a (Social) Distance, September 2020) <<https://hub.salford.ac.uk/welfare-at-a-social-distance/wp-content/uploads/sites/120/2020/09/WaSD-Rapid-Report-2-New-COVID-19-claimants.pdf>> accessed 16 July 2023.

<sup>184</sup> Golman, Hagmann and Loewenstein (n 89).

harmful reporting which undermines the results of the data. Inaccurate reporting can exacerbate existing issues with the representativeness of data and provide inaccurate pictures of minority communities<sup>185</sup>. Disaggregation is consistently lacking in demographic data<sup>186</sup>, making it near-impossible to gather information on how different issues are experienced by different groups across gender, ethnicity, and other characteristics. These failures in reporting accurately can be severe, resulting in continuing forms of structural racism<sup>187</sup>, and the divorce of marginalised groups from data infrastructure<sup>188</sup>.

Furthermore, reporting which is not disaggregated can result in confusion or misleading findings from the data, as occurred in the Covid-19 pandemic where national surveillance data from the USA did not include exposure history, or occupation in reporting<sup>189</sup>, which meant it was difficult to ascertain which occupations (such as medical professionals and other essential workers) were most at risk and if they had been exposed previously. Type two gaps can also be simple results of human error and are able to be resolved swiftly<sup>190</sup>. Inaccurate labelling in public data presentations occurred during the Covid-19 pandemic, resulting in confusion as to the data findings<sup>191</sup>, but were able to be resolved and relabelled correctly. Nonetheless, mistakes can cause significant harm; accidental mislabelling of ethnic minorities on death certificates has led to the inaccuracy of death figures by ethnicity<sup>192</sup>. Reporting gaps prevent a clear understanding of the data findings.

Type two reporting gaps also appear when data is reported separately or is not included in overall data reports. When data is reported separately, the resulting picture that

---

<sup>185</sup> Brittany N. Morey and others, 'No Equity without Data Equity: Data Reporting Gaps for Native Hawaiians and Pacific Islanders as Structural Racism' (2022) 47 *Journal Of Health Politics Policy And Law* 159.

<sup>186</sup> Ibid.

<sup>187</sup> Ibid.

<sup>188</sup> Wright and others (n 76).

<sup>189</sup> Bhatia, Sledge and Baral (n 55).

<sup>190</sup> Galaitsi and others (n 51).

<sup>191</sup> Ibid.

<sup>192</sup> Morey and others (n 184).

emerges from the data can be inaccurate. This can be problematic when these data sets are then used for decision-making based on inaccurate findings.

#### Reporting Gap Case Study 2: The True Nature of Violent Crime in England and Wales

The Office of National Statistics (ONS) publishes measures of crime based on two data tools: the Crime Survey of England and Wales (CSEW), and police-recorded crimes, which are recorded by police forces. The ONS publishes several different overviews of this data each year, including overviews which focus on different parts of the data, such as analysing it by household prevalence<sup>193</sup>, types of offences<sup>194</sup>, and by “violent” crime<sup>195</sup>. However, the ONS measure of “violent crime” contains a significant reporting gap; the violent crime measure does not include data on sexual or domestic violence.

The reason for this gap lies in the manner in which data on these crimes are reported. These offences are noted by Police forces when crimes are reported, but data collected through the Crime Survey of England and Wales on sexual violence is often excluded from the ONS’s measure. The Crime Survey of England and Wales is based on two parts: face-to-face interviews, and a self-completion module. However, the figures of violent crime are only based on the data collected from the face-to-face portion of the Crime Survey. The self-completion module is considered to be the most accurate form of reporting for sexual offences and domestic violence, but the data from the self-completion module is reported separately and is not part of the violent crime measure. The ONS justifies this exclusion by stating that, as the self-completion module does not contain

---

<sup>193</sup> Peter Jones, 'Crime in England and Wales: Year ending September 2022' (*Office for National Statistics: Crime and Justice*, 26 January 2023)

<<https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/bulletins/crimeinenglandandwales/yearendingseptember2022#domestic-abuse-and-sexual-offences>> accessed 10 July 2023.

<sup>194</sup> Pete Jones, 'Homicide in England and Wales: Year ending March 2022' (*Office for National Statistics: Crime and Justice*, 9 February 2023)

<<https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/articles/homicideinenglandandwales/march2022>> accessed 10 July 2023.

<sup>195</sup> Pete Jones, 'The Nature of Violent Crime in England and Wales: Year ending March 2022' (*Office For National Statistics: Crime and Justice*, 9 November 2022)

<<https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/articles/thenatureofviolentcrimeinenglandandwales/yearendingmarch2022#police-recorded-crime-trends-for-different-types-of-violent-crime>> accessed 10 July 2023.

questions on how many times violence has been experienced in the past year, the data collected is unreliable. The ONS states it is not an accurate reflection of the prevalence of this type of crime, therefore making it incompatible with the reporting required for the ONS' overall measure of violent crime<sup>196</sup>. A second reporting gap occurs due to a reporting threshold existing for certain offences in the face-to-face portion of the survey. Offences such as "rape", "attempted rape" and "indecent assault" are not included in the measure of violent crime as small numbers of these types of offences are recorded in the face-to-face portion of the Crime Survey of England and Wales. The ONS calls this small amount of data collected from the face-to-face portion too unreliable to be published as part of the ONS' violent crime figure<sup>197</sup>. Therefore, the measure of violent crime in England and Wales excludes domestic and sexual violence from its measures, reporting them separately and not as forms of violent crime, due to reporting thresholds.

Cooper and Obolonskaya point out two major harms from not including these crimes and any data on them in the ONS' measure of violent crime. Firstly, whilst concerns about having sufficient reporting rates before publication are valid, the reporting prevalence for these offences is actually higher than some offences that are included in the violent crime measure. Cooper and Obolonskaya have found that this means that "the rule is selectively applied to sexual offences"<sup>198</sup>. The second harm of reporting these offences separately is that wider trends about violence can be misleading. By excluding sexual and domestic violence, of which women are more likely to experience<sup>199</sup> violent crime becomes focussed on crimes which men are more likely to experience. This results

---

<sup>196</sup> Cooper and Obolonskaya (n 41).

<sup>197</sup> Nick Stripe, 'User Guide to Crime Statistics for England and Wales; March 2020' (*Office for National Statistics: Crime and Justice*, 4 November 2021) <<https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/methodologies/userguidetocrimestatisticsforenglandandwales>> accessed 18 May 2023.

<sup>198</sup> Cooper and Obolonskaya (n 41).

<sup>199</sup> Meghan Elkin, 'Sexual Offences Victim Characteristics, England and Wales: Year ending March 2022' (*Office for National Statistics: Crime and Justice*, 23 March 2023) <<https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/articles/sexualoffencesvictimcharacteristicsenglandandwales/yearendingmarch2022>> accessed 10 July 2023.



in indications that men are more likely to be victims of violent crime when the opposite is true when violent crime is analysed with sexual violence included<sup>200</sup>.

When figures from the Crime Survey of England and Wales have been revisited and an expanded amount of data included in the new measure, researchers have been able to show that women are the victims in “almost as many” violent offences as men<sup>201</sup>. This included the researchers considering sexual violence in its measure of violent crime<sup>202</sup>. Cooper and Obolenskaya point out that “ignoring the gendered nature of violent crime...results in less accurate estimates of the risk of violence overall”<sup>203</sup>. Both these harms result in the ONS’ figure of violent crime being misleading and not a clear image of the true nature of violent crime in England and Wales. This prevents effective policy actions aimed to prevent or reduce rates of violence. Closing this gap would require an expansion of what the ONS considered violent crime to include sexual offences, especially rape, attempted rape, and indecent assault. Including the existing data from these crimes in the measure of violent crime would give a better indication of the nature of violent crime in England and Wales in general.

Type one and type two reporting gaps cause harm. Whilst there are multiple causes for why reporting gaps exist, it is clear that the power to report data and the manner it is reported has a significant impact on the way that data and information are understood, utilised, and applied to decision-making and policy. In order to minimise reporting gaps, consistent consideration for the need to represent existing data accurately must be undertaken. While reporting thresholds are often in place for very valid reasons, their impact of them can be to cause confusion, undermine knowledge, and hold back information that has a public interest. Secondly, when data is reported separately, data is excluded, or data

---

<sup>200</sup> Cooper and Obolenskaya (n 41).

<sup>201</sup> Sylvia Walby, Jude Towers and Brian Francis, 'Mainstreaming Domestic and Gender-Based Violence into Sociology and the Criminology of Violence' (2014) 62 *The Sociological Review* 187.

<sup>202</sup> Ibid.

<sup>203</sup> Cooper and Obolenskaya (n 41).

is not disaggregated, the outcomes of the data in terms of its findings and impacts are unclear and undermined. This creates harm when applying this data to decisions or policy-making, as interpretations of the data findings may be inaccurate or limited which then does not translate to the most effective policy outcomes. In order to fill these gaps, data reporters must think about the benefits of reporting data, rather than the reasons to hold data back.

## VII. Harmonisation gap

Harmonisation gaps encompass a broad category which concerns the inability of collected data to be synchronised or brought together. Data on similar or linked issues may be collected by different public bodies and stored separately. Both entities would likely benefit from access to the other's data, which may allow for fuller data sets or more information on a particular issue to be available to both. However, where this harmonisation is not possible, a harmonisation gap occurs. Harmonisation gap can be characterised into two types. A "type one" gap exists where data is unable to be harmonised due to differences in the definitions, scale of measurements, and length of time between the repeated collection of the data<sup>204</sup>. Where these basic points of measurement are not harmonised, data is unable to be brought together, even if the data broadly measures the same issues. A "type two" harmonisation gap is related to the inability of existing datasets to be harmonised and shared between official sources. This type of gap is generally due to a lack of compatibility between the existing sharing protocols, arrangements, or agreements between public bodies, and even occurs between government departments. This type of gap receives attention, and notable frustration, from those within government and public bodies<sup>205</sup> due to knowledge of the existence of the data, but the continued inability to bring existing data sets together creates unreliable narratives from the data.

To an extent, type one harmonisation gaps are unavoidable. Cultural, societal, environmental, and governmental differences can all drive differences in definitions, measurements and time scales between jurisdictions, governments, and public bodies<sup>206</sup>. Differences in the age schooling begins in different jurisdictions can result in a lack of comparability<sup>207</sup>, as can national decisions to use alternative, but valid, definitions from

---

<sup>204</sup> This is otherwise known as "temporal resolution". In this context, temporal resolution refers to the time data collectors return to the same place to consistently measure the same information. This is a general scientific term so care must be taken when applying it to this context.

<sup>205</sup> Freeguard and Shepley (n 64).

<sup>206</sup> Galaitsi and others (n 51).

<sup>207</sup> Behrman and Rosenzweig (n 46).

official bodies in data measurements<sup>208</sup>. Nevertheless, type one harmonisation gaps can be credited with effectively preventing cross-jurisdictional and global comparisons on key issues, including on education, healthcare, and economic indicators<sup>209</sup>. When comparability is prevented, researchers and public officials are left without indicators of a jurisdiction's success or failures in relation to its neighbours or globally. This lack of indicators can prevent effective decision-making on how to improve or maintain a country's standing and can lead to longer-term harm to populations, whose needs are unrecognised and unmet. For instance, Katz and others suggest that there needs to be more oversight by official bodies, such as the United Nations Educational, Scientific and Cultural Organisation, of large, cross-jurisdictional studies in order to create more harmony in the data collection phase<sup>210</sup>. This would be effective in situations where the data is being collected for the purpose of cross-country comparisons. Furthermore, consistency and harmony in data collection would likely enable a basis of good data practice and sharing, which could enable public officials and researchers to use collected data to its fullest extent. However, this ambition to enact greater harmony could be considered wishful thinking. At present, differences in definitions may be due to different jurisdictions recognising and using valid definitions held by official bodies, and single definitions for terms such as "literacy"<sup>211</sup>, and "birth rates"<sup>212</sup> are difficult to create and enforce.

#### Harmonisation Gaps Case Study 1: Homelessness and Priority Need; Legislative

##### Divergence in the UK

Harmonisation gaps prevent the comparability of homelessness statistics across the UK. Homelessness, as a policy matter, is a devolved issue in the UK. The Housing (Homeless Persons) Act 1977 provided the first version of a legislative framework for duties for local authorities to ensure accommodation is available for households which are "eligible" for

---

<sup>208</sup> Ibid.

<sup>209</sup> Ibid.

<sup>210</sup> Katz and others (n 57).

<sup>211</sup> Behrman and Rosenzweig (n 46).

<sup>212</sup> Sentell and others (n 44).

assistance, “unintentionally homeless”, and are in “priority need”<sup>213</sup>. Since devolution, there has been significant legislative divergence, with definitions of “priority need” between England and Scotland becoming particularly notable.

Under English homelessness legislation, local housing authorities have a duty to secure accommodation for unintentionally homeless households in priority need<sup>214</sup>. Those considered to have priority need is a broad list, and can include pregnant women, those with dependent children, and vulnerable adults<sup>215</sup>. The assessment of “priority need” is a significant component in how Local Authorities assess homeless households for support and may be determinative of whether they are owed a housing duty under the Housing Act 1996. This assessment of “priority need” does not appear under the Scottish homelessness legislation. The result of this is that Scottish homeless Statistics include a far broader range of households than English statistics, making UK-wide comparisons difficult.

There is an active effort by statistical agencies in the UK to ease this harmonisation gap and make comparisons easier for the public and researchers. The Government Statistical Service created a comparison tool in 2022<sup>216</sup> which is designed to enable research into homelessness statistics across the UK. The comparison tool allows users to explore topics by concept and shows whether the statistical measures related to these concepts are comparable<sup>217</sup>. For example, when using the comparison tool to assess the ability to compare “priority need”, the tool indicates that the statistics held on priority need by Northern Ireland should not be compared with the statistics for England

---

<sup>213</sup> Suzanne Fitzpatrick and others, ‘The Impact of Devolutions: Housing and Homelessness (*The Joseph Rowntree Foundation*, January 2010)

<<https://www.jrf.org.uk/sites/default/files/jrf/migrated/files/impact-of-devolution-long-term-care-housing.pdf>> accessed 11 July 2023.

<sup>214</sup> The Housing Act 1996, pt 7.

<sup>215</sup> ‘Homelessness code of guidance for Local Authorities’ (HM Government Department for Levelling Up, Housing and Communities, 31 May 2023) <<https://www.gov.uk/guidance/homelessness-code-of-guidance-for-local-authorities/chapter-8-priority-need>> accessed 11 July 2023.

<sup>216</sup> Government Statistical Service, ‘UK Official Statistics On Homelessness: Comparisons, Definitions, And Processes (*Government Statistical Service*)’ <<https://analysisfunction.civilservice.gov.uk/dashboard/tools/homelessness-statistics/comparison.html>> accessed 18 May 2023.

<sup>217</sup> Ibid.

and Wales, and Scotland does not test for priority need<sup>218</sup>, so a comparison is not possible.

This tool is a clear step forward to attempt to find solutions to harmonisation gaps that respects the individual devolved nature of Homelessness duties. Where there may be other harmonisation gaps that similarly emerge as a result of statutory differences, whether that is due to devolution or other reasons, these comparison tools could provide a solution and allow for improved UK-wide analysis to take place.

Type two harmonisation gaps are common despite clear indications that the inability to share data can be extremely problematic<sup>219</sup>. Type two gaps are widespread across public bodies and share several common features. Firstly, a common feature is that data sharing frameworks are unsuitable or incapable to share information across departments or agencies<sup>220</sup>. Sharing frameworks govern and support the sharing of data between public institutions, allowing data to be accessible to a defined group of stakeholders with controls on the use of the data<sup>221</sup>. Sharing frameworks, especially in electronic systems, may differ between agencies, making the sharing difficult or impossible<sup>222</sup>. There is a lack of standard general protocols and quality in these data information systems, resulting in inadequate protocols for sharing and transferring information between them<sup>223</sup>. Secondly, harmonisation over data can be prevented due to the costs and expertise required to gain or manage access to data. Researchers or public officials may be required to have a high amount of knowledge or experience with different information systems in order to gain access to the data and may require more knowledge of analytic techniques in order to accurately

---

<sup>218</sup> Ibid.

<sup>219</sup> Behrman and Rosenzweig (n 46).

<sup>220</sup> Vest and Issel (n 60).

<sup>221</sup> Heather Wheeler, 'Data Sharing Governance Framework' (HM Government Central Digital & Data Office, 23 May 2022) < <https://www.gov.uk/government/publications/data-sharing-governance-framework/data-sharing-governance-framework> > accessed 11 July 2023.

<sup>222</sup> Ibid.

<sup>223</sup> Freeguard and Shepley (n 64).

harmonise the data into useable results<sup>224</sup>. Regional and organisational differences in data management systems require an understanding of multiple systems in order to harmonise them, which most public officials lack<sup>225</sup>. Thirdly, harmonisation is prevented due to differing legislation and regulations governing the protection of data in different systems<sup>226</sup>. Data protection is a sensitive and important issue when sharing or seeking to harmonise data. Data governance across organisations, agencies, and jurisdictions may differ completely, leading to data providers making “cautious and conservative”<sup>227</sup> decisions about what access is permitted under the law. Legislation compounds organisational barriers to data sharing, including a lack of awareness of powers to share data, and fears about using them<sup>228</sup>. Clear and comprehensive data-sharing legislation that facilitates the sharing of useful data is required in order to enable the harmonisation of data sets<sup>229</sup>. An inability to bring data together creates harm outside of the individual level. Harmonisation gaps frustrate research, frustrate policy, and frustrate government’s ability to implement projects and measure successes and failures<sup>230</sup>. It is therefore important that harmonisation gaps are managed and the ability to share information is harmonised where it is possible to do so.

Harmonisation Gap Case Study 2: What do we really know about the Rental Market in England and Wales?

Statistics on the rental market in England and Wales suffer from harmonisation gaps which prevent clear understanding of the market as a whole. UK-based data on the rental market is fragmented and unharmonized across measurements and agencies, suffering from both “type one” and “type two” harmonisation gaps.

---

<sup>224</sup> Katz and others (n 58).

<sup>225</sup> Galaitsi and others (n 52).

<sup>226</sup> Katz and others (n 58).

<sup>227</sup> Ibid.

<sup>228</sup> Freeguard and Shepley (n 64).

<sup>229</sup> Ibid.

<sup>230</sup> Ibid.

Data on the housing market is widely available and accessible to the public, researchers, and government. While housing market data as a whole is generally good, the data for home buyers is far more available and discernible than the equivalent data for renters. There are two reasons for the lack of data on rental properties. Firstly, data on renting a house in the UK is spread across a multitude of datasets which measure different things and are therefore generally incompatible. Secondly, official data is difficult to report as it relies on data being shared between the Office of National Statistics (ONS) and the Valuation Office Agency.

Across the UK, there are twelve sources of official statistics on the Private rental sector. Several of these datasets focus solely on individual nations within the UK. According to the Office for Statistics Regulation, the range of official statistics means that there is a disjointed, incomplete picture of the UK private rental market<sup>231</sup>. Each dataset provides a different “snapshot” of the private rental market, but these cannot be brought together. These incompatible statistics publications provide only a partial picture of the private rental market in the UK, and results in an inability to answer relevant questions. The Office for Statistics Regulation identified that there is a lack of “robust statistics about actual rent prices for all four countries that can be compared over time and across geography”<sup>232</sup>.

For information about the private rental market in England, there are two datasets which can be considered as evidence of the inability to harmonise data. Two of the official datasets on the costs to rent a house in England are the Private Rental Market Summary Statistics and the Index of Private Housing Rental Prices. The Private Rental Market Summary Statistics are published by the ONS and are based on Local Authority data collected over twelve months. However, based on the current methodology employed by

---

<sup>231</sup> 'Systematic Review of Public Value: Statistics on Housing and Planning in the UK' (Office for National Statistics, November 2017) < <https://osr.statisticsauthority.gov.uk/wp-content/uploads/2017/11/Systemic-Review-of-Statistics-on-Housing-and-Planning-in-the-UK20171110-corrected.pdf>> accessed 16 July 2023.

<sup>232</sup> *ibid.*



the ONS, a comprehensive time series is not published on this data<sup>233</sup>. The data is collected and published at set intervals, however, meaning that it is possible to look back at individual publications over time. However, without data from over time being available in a single, comprehensive manner, it is significantly more difficult to accurately assess the available data over time, and those who wish to look at data over time are faced with the onerous task of harmonising the information from across the multiple publications themselves. Furthermore, the statistics cannot be compared by geographical region<sup>234</sup>, as the Local Authority data the statistics are based on are not additionally aggregated and published based on region<sup>235</sup>. This type one harmonisation gap means that questions about changes in rent over time and across regions cannot be answered.

The Index of Private Housing Rental Prices, also published by the ONS, measures the relative change in rents by region and is UK-wide. The Index of Private Housing Rental Prices is published as average rental prices per region. By only measuring rents by region, the information cannot be harmonised and the average rental price for the UK as a whole cannot be determined. Furthermore, the Index of Private Housing Rental Prices does not include changes in rent levels over time<sup>236</sup>, again meaning that trends in rent levels cannot be determined, and comparisons over time cannot be undertaken without researchers manually comparing all points over time themselves<sup>237</sup>.

With an inability to compare the rental market across geography and time, it becomes difficult to gain a true image of the market as a whole and undertake valuable research into the benefits and successes of housing policy. Furthermore, the information contained in each dataset is not able to be harmonised with other data sources on the

---

<sup>233</sup> Aimee North, Private Rental Market Summary Statistics in England: October 2021 to September 2022 (*Office for National Statistics*, 14 December 2022) <[<sup>234</sup> 'Systematic Review of Public Value: Statistics on Housing and Planning in the UK' \(\*Office for National Statistics\*, November 2017\) \(n 231\).](https://www.ons.gov.uk/peoplepopulationandcommunity/housing/bulletins/privaterentalmarketsummarystatisticsinengland/october2021toseptember2022#:~:text=1.-,Main%20points,monthly%20rent%20at%20%C2%A3525.> accessed 11 July 2023.</a></p></div><div data-bbox=)

<sup>235</sup> Ibid.

<sup>236</sup> Ibid.

<sup>237</sup> Ibid.

private rental market, such as data held by lettings companies, due to different measurements and time frames. This means that all information on the private rental market remains individual “snapshots” of what is being measured, rather than producing images of the rental market as a whole.

On a government scale, the harmonisation gap and the resulting lack of information “hampers [our] ability to measure how much renting households actually pay and their affordability at a local level”<sup>238</sup>. Without being able to view the market, it also becomes difficult to accurately assess changes in the market over time, the responsiveness of the market to policy changes, the effectiveness of housing policy, as well as understanding wider trends about housing which affects the economy as a whole. As more than 4.5 million households in the UK live in privately rented accommodation, this information is important for local and national governments in order to produce effective housing policy which protects renters and landlords, and ensures accommodation remains affordable.

Rental statistics also suffer from a type two harmonisation gap; the inability to share, or gain access to, existing data. Rental data in England is not collected by a singular agency. The data is held by two agencies, the ONS and the Valuation Office Agency. These two agencies have historically found data difficult to share or provide access to due to data confidentiality legislation<sup>239</sup>. In 2017 the Digital Economy Act provided a “legislative gateway”<sup>240</sup> for ONS access to Valuation Office Agency microdata, but in datasets published in March of 2023 there is still recognition by the ONS that

---

<sup>238</sup> Neal Hudson, Understanding Local Housing Markets: Advice and Guidance for Local Authorities, (*Local Government Association*) <<https://www.local.gov.uk/sites/default/files/documents/LGA%20RA%20Understanding%20Local%20Housing%20Markets%20June%202019.pdf>> accessed 11 July 2023.

<sup>239</sup> Commissioners of Revenue and Customs Act 2005.

<sup>240</sup> Kat Christiansen and Rhys Lewis, UK Private Rented Sector: 2018 (*Office for National Statistics*, 18 January 2019) <[https://www.ons.gov.uk/economy/inflationandpriceindices/articles/ukprivaterentedsector/2018#:~:text=Around%20three%2Dquarters%20\(74%25\),social%20renters%20and%20owner%20occupiers](https://www.ons.gov.uk/economy/inflationandpriceindices/articles/ukprivaterentedsector/2018#:~:text=Around%20three%2Dquarters%20(74%25),social%20renters%20and%20owner%20occupiers)> accessed 11 July 2023.

information is missing due to “data access constraints”<sup>241</sup>. A key piece of missing information due to these access constraints is actual rental prices, which cannot be published in the Index of Private Housing Rental Prices<sup>242</sup>. Without this knowledge, the picture of the rental market in the UK remains unclear for all users, across Government and the public.

These case studies have demonstrated that harmonisation gaps have the capacity to cause harm through the inability to bring together knowledge and research across public institutions and government. It is clear that inability to compare and share knowledge across government creates significant knowledge gaps, which prevents active government analysis of its own policies which stretch across the UK. The assessment and creation of effective policies which are responsive to the needs and wants of the population is crucial for good government. Therefore, it is in the interest of government and public institutions to seek to close these harmonisation gaps wherever possible.

---

<sup>241</sup> Aimee North, Index of Private Housing Rental Prices, UK: April 2023 (*Office For National Statistics*, 24 May 2023)  
<<https://www.ons.gov.uk/economy/inflationandpriceindices/bulletins/indexofprivatehousingrentalprices/april2023#:~:text=Private%20rental%20prices%20paid%20by,12%20months%20to%20April%202023>> accessed 11 July 2023.

<sup>242</sup> Ibid.

## VIII. Delay gap

A delay gap exists where there are significant lags in the ability to access data that is held by a government or public body. Whilst harmonisation gaps totally prevent the ability to share or compare data, delay gaps mean that there is a lack of timely access to data. Where data is used for decision-making, data lags can become catastrophic. As seen with Covid-19 data use at the height of the pandemic<sup>243</sup>, delays in access to data can result in inaccurate decisions, or policy based on outdated information. Lack of timely access can result in unclear or incomplete pictures of data results and can prevent useful research and decision-making. Delay gaps themselves are often conflated in the literature with harmonisation gaps<sup>244</sup>. I disagree strongly with combining the two categories of gap. A delay gap is a separate entity to a harmonisation gap; in theory, a delay gap is able to be filled once there is access to the desired data. Delay gaps are, by their nature, able to be filled, whereas a harmonisation gap will result in an ongoing gap existing due to ongoing problems with the ability to share, compare, and bring data together. Harmonisation gaps, as noted above, are symptomatic of failures in sharing and comparison frameworks that go beyond a single dataset and impact the entirety of a public body's collected data.

Delay gaps can be considered a unique category of gap due to how they can interlink with other data gap categories. Often, more than a single category will appear simultaneously; reporting gaps will often come alongside a delay gap, where public bodies struggle to translate data into policy and share those data findings with researchers and the public in a timely manner<sup>245</sup>. Delay gaps and collection gaps also appear simultaneously; data that is not complete in coverage is not provided or shared within a timely manner,

---

<sup>243</sup> Galaitsi and others (n 52).

<sup>244</sup> Katz and others (n 58) identify that there are significant lags in the ability to access health data across Canadian provinces due to differences in the legislation governing the protection of the data, as well as other issues such as different IT systems, or the need for those sharing data to possess high degrees of knowledge of other systems in order to share data. When discussing these "gaps", Katz and others treats these as they are all one form of gap, rather than identifying an inability to share due to legislation, and an inability to share quickly due to the need for knowledge on the part of the person sharing the data, as two different forms of gap.

<sup>245</sup> Ibid.

meaning that even less can be understood from the data<sup>246</sup>. The lack of timeliness can undermine the purpose of the data collection and prevent the effective use of data for policymaking, especially where there is time pressure to make policy. Examples of this can be easily seen in the inability to share Covid-19 data, in near-real time, preventing the adoption of time-sensitive solutions<sup>247</sup>. There is also a risk with delay gaps that research agendas can move beyond the data that is required. Researchers can become unable to continue to wait for access to data and projects can be abandoned, resulting in research gaps and future lack of incentives to attempt to research certain areas due to known lack of access to data<sup>248</sup>. Research agendas focused on impacting policy, for example, places emphasis on the “here and now”<sup>249</sup>. By having a focus on immediate needs of research, delay gaps can remain as there is not sufficient time for research to close these gaps. When there is a focus on the use of research to impact policy and decision-making, it is necessary that access can be gained to data and information within a time frame that enables research and relevant policy decision-making.

Delay gaps, by their nature, slow down processes of decision-making and research. Therefore, the harms felt by delay gaps are often not on the individual scale; instead, delay gaps prevent effective understanding and functioning of public institutions.

Delay Gap Case Study 1: How many Unrepresented Defendants are there in English Magistrates' Courts?

Delay gaps can be difficult to identify. Many researchers are accustomed to delays and long waits before receiving the requested information. Nonetheless, information being too slow to obtain, or overly burdensome to obtain within a reasonable timeframe creates a delay gap. Delay gaps can prevent timely research, or leave relevant questions

---

<sup>246</sup> Burton and others (n 66).

<sup>247</sup> Galaitsi and others (n 52).

<sup>248</sup> Katz and others (n 58).

<sup>249</sup> Alessandro Liberati, Evridiki Chatziandreou and Olli. S. Miettinen, 'Health Care Research: What Is It About?' (1989) 1 *Quality Assurance In Health Care: The Official Journal Of The International Society For Quality Assurance In Health Care* 249.

unanswered, obscuring information for long periods of time and stopping policy responses.

Research by the Centre for Public Data published in March 2023 highlighted a number of data gaps in the UK Criminal Justice system<sup>250</sup>. One of the gaps identified related to the delay, and the denial of access, to data concerning unrepresented defendants in Magistrates Courts. The Ministry of Justice gathers data on defendants with or without representation from Courts across the UK, including from Family, Civil, Employment Tribunals, and Crown Courts. This data is held centrally, is published publicly, and is able to be requested. However, when MPs and the public have requested information from the Ministry of Justice as to the available information on representation in Magistrates Courts, the information cannot be provided<sup>251</sup>. The Ministry of Justice does not hold information on the numbers of unrepresented defendants in Magistrates Courts centrally<sup>252</sup>, but no apparent reason for the lack of centrally held information is given.

The data is available, but not in a centralised format; individual Magistrates Courts can be contacted, and the information requested. The Ministry of Justice has the capacity to do this and bring the data together themselves but refuses to do so as it would be too expensive<sup>253</sup>. Therefore, to gain specific information on Magistrates Courts, it is likely the requester would have to speak to Courts one-by-one to see if the data can be provided<sup>254</sup>. The delay gap for this data is clear; information is not held centrally which means that research and information is delayed as the information is obtained through Magistrates

<sup>250</sup> Powell-Smith and Leibowitz (n 23).

<sup>251</sup> Question from Christina Rees MP to the Ministry of Justice, June 2016, < <https://questions-statements.parliament.uk/written-questions/detail/2016-06-14/40571/>> accessed 3 July 2023.

<sup>252</sup> Question from Grahame Morris MP to Ministry of Justice (21 January 2021) <[Written questions and answers - Written questions, answers and statements - UK Parliament](#)> Accessed 12 June 202. See also: Question from Jo Stevens MP to Ministry of Justice (2 April 2019) <[Written questions and answers - Written questions, answers and statements - UK Parliament](#)> accessed 12 June 2023. See also: Question from Gloria De Piero MP to Ministry of Justice (30 November 2018) <[Written questions and answers - Written questions, answers and statements - UK Parliament](#)> accessed 12 June 2023

<sup>253</sup>Response to Robert Jones' Freedom of Information Act, 2018 [Rt Hon Michael Gove MP letterhead \(whatdotheyknow.com\)](#) .

<sup>254</sup> 'Protocol On Sharing Court Lists, Registers And Documents With The Media' (*HM Courts and Tribunals Service, 2023*) < <https://www.gov.uk/government/publications/guidance-to-staff-on-supporting-media-access-to-courts-and-tribunals/protocol-on-sharing-court-lists-registers-and-documents-with-the-media-accessible-version>> accessed 18 May 2023.

courts one by one. This is a large task; with over 150 Magistrates Courts in England and Wales which saw 1.14 million cases in 2021<sup>255</sup> gathering this data without the resources of a Government department would be difficult.

Access to this information without the above delays is important. The lack of quick access to the data prevents timely research on the scale and experience of unrepresented defendants in Magistrates Court<sup>256</sup>, and also prevents suitable policy measures from being put in place. Furthermore, this data gap directly allows for harm to continue for defendants without representation. Unrepresented defendants face harsher justice outcomes, are at a greater risk of not being able to navigate the judicial system and restrict court efficiency through a lack of knowledge of Court processes<sup>257</sup>. Without being able to understand the scale of the problem in Magistrates Courts in England and Wales, effective policy and support for unrepresented defendants is not able to be obtained.

This data gap is likely to be resolved soon. Researchers at the Centre for Public Data have been told that the legal representation of defendants is being recorded on Common Platform<sup>258</sup>, the case management system being rolled out across Magistrates Courts. Data held on Common Platform and Libra, the existing management system Common Platform is replacing, is then gathered by the Courts and Tribunals Service<sup>259</sup> and published quarterly<sup>260</sup>.

---

<sup>255</sup> Georgina Sturge, Court Statistics for England and Wales (*House of Commons Library*, 31 January 2023) <<https://commonslibrary.parliament.uk/research-briefings/cbp-8372/>> accessed 18 May 2023.

<sup>256</sup> Penelope Gibbs, 'Justice Denied? The Experience of Unrepresented Defendants in the Criminal Courts,' (*Transform Justice*, 29 April 2016) < <https://www.transformjustice.org.uk/publication/justice-denied-the-experience-of-unrepresented-defendants-in-the-criminal-courts/>> accessed 11 July 2023.

<sup>257</sup> Anna Powell-Smith and Gideon Leibowitz, Unrepresented Defendants in the Magistrates' Courts, (*The Center for Public Data*, 2023) <<https://static1.squarespace.com/static/5ee7a7d964aeed7e5c507900/t/641ae6bdbd6edd48d43e09e9/1679484606127/CFPD+Legal+Representation+Data+Gaps.pdf>> accessed 11 July 2023.

<sup>258</sup> Ibid.

<sup>259</sup> HM Court and Tribunals Service derives data from the One Performance Truth (OPT) database, which extracts its administrative data from Libra and Common Platform.

<sup>260</sup> Published in the Criminal Court Statistics Quarterly, which is published by the Ministry of Justice.

### Delay gap case Study 2: What Do We Know About Police Vetting Practices?

The available statistics on police vetting depends on the ability of each force to provide them. When statistics on vetting have been requested under Freedom of Information requests, some forces, such as the Metropolitan Police<sup>261</sup>, have been able to provide them, whilst other forces have refused the request on the basis that gathering such information would exceed the appropriate cost limit of £450 as determined in legislation<sup>262</sup>.

Police vetting operates under two regimes: national security vetting and police vetting. National security vetting is “owned” by the cabinet office and focuses on information security risks. Police vetting is “owned” by the College of Policing and is focused on an applicant’s suitability for the Police force based in attributes such as honesty and integrity. Therefore, these two vetting procedures assess different risks and have different decision-making criteria. Forces individually undertake vetting but the process itself is set by the College of Police and is governed by the Vetting Code of Practice and the Vetting Authorised Professional Practice.

Police vetting and force’s awareness of the behaviour of its officers, whether at the recruitment or during tenure as a serving officer, is an issue sharply in the public eye. High-profile reports of misconduct by Police, and especially the horrific 2020 murder of Sarah Everard by a service police officer, led to an investigation commissioned by the Home Secretary for HM Inspectorate of Constabulary and Fire & Rescue Services to assess the vetting processes of forces across England and Wales, and how well forces are able to be detected and deal with misogynistic and predatory behaviour. The report,

---

<sup>261</sup> Response to Freedom Of Information request from M Brookes to the Metropolitan Police (20 November 2019) <[https://www.whatdotheyknow.com/request/vetting\\_statistics](https://www.whatdotheyknow.com/request/vetting_statistics)> accessed 12 July 2023.

<sup>262</sup> Response to Freedom Of Information request from Louisa James to Hampshire and Isle of Wight Constabulary (18 January 2023) <[https://www.whatdotheyknow.com/request/number\\_of\\_police\\_officers\\_who\\_ha\\_19#comment-111323](https://www.whatdotheyknow.com/request/number_of_police_officers_who_ha_19#comment-111323)> accessed 12 July 2023.



published in 2022, made over 40 recommendations to forces<sup>263</sup>. Within the review, the Inspectorate reviewed hundreds of vetting files and found that the quality of vetting decision-making needed improvement, with an improved recording of the rationale for some decisions.

The Inspectorate's review faced a "worrying" gap in information<sup>264</sup> when investigating vetting units within forces. The report notes that none of the forces could produce details on cleared police officers and staff who had concerning adverse information (such as financial risks or family with criminal backgrounds) without having to manually trawl through all of their vetting cases. The Inspectorate found this troubling, expecting that every force would have this information readily available through their vetting IT systems.

This particular delay gap prevented a more robust investigation into part of the vetting process. It resulted in a manual, slow search by forces and the Inspectorate<sup>265</sup>. This, therefore, is a clear example of a delay gap in action. While data was able to be found and reviewed by the purpose of the Inspectorate's report, it was not complete, and not able to be provided in a timely manner.

This is particularly harmful when understood within the context of the Inspectorate's report. Information about the Police vetting process and the number of officers who are flagged within vetting is important information for assessing whether the process is successful at identifying individuals who could represent a risk to the public or the integrity of the Police force if they are vulnerable to corruption or to commit misconduct. The ability of the police to be accountable and assess its own processes is important.

---

<sup>263</sup> 'An Inspection of Vetting, Misconduct, and Misogyny in the Police Service' (*HM Inspectorate of Constabulary, Fire and Rescue Services*, 2 November 2022)  
<<https://www.justiceinspectores.gov.uk/hmicfrs/publication-html/an-inspection-of-vetting-misconduct-and-misogyny-in-the-police-service/#an-explanation-of-vetting-regimes-and-processes>> accessed 12 July 2023.

<sup>264</sup> Ibid.

<sup>265</sup> Ibid.

The report by HM Inspectorate recommended that by the end of April 2023, chief constables should establish processes to identify vetting clearance records within their IT systems which flag records where applicants have committed criminal offences and/or the record contains other types of concerning adverse information<sup>266</sup>. This is a clear way to close the delay gap that exists in gaining access to this information.

The process of identification of delay gaps is unique when contrasted with the process of identifying other forms of gaps. The case studies above were identified via a search of Freedom of Information requests to determine which were denied on the basis that providing the information requested would likely exceed the appropriate cost limits placed on Freedom of Information requests<sup>267</sup>. Similar research conducted by the Centre for Public Data was also able to identify delay gaps through the long process of searching through parliamentary questions to identify where members of Parliament had requested data, and the identification of the delay gap surrounding Police vetting statistics was only found through searching through Freedom of Information Requests. Often however, the language used in response to questions from Members of Parliament or Freedom of Information requests are very similar; the government response is that “the Department does hold the information requested but cannot provide it at this time”<sup>268</sup>. Often, Freedom of Information requests are refused on the basis that providing the information would be too much work and

---

<sup>266</sup> Ibid: Recommendation 2, found on page 56.

<https://www.justiceinspectors.gov.uk/hmicfrs/publication-html/an-inspection-of-vetting-misconduct-and-misogyny-in-the-police-service/#an-explanation-of-vetting-regimes-and-processes> .

<sup>267</sup> Section 12 of The Freedom of Information Act 2000 states that Public Authorities may refuse a request for information if it exceeds the “appropriate limit”. This limit relates to the cost for the department to deploy a civil servant to locate, retrieve and extract the requested information. The appropriate limit was stated to be £600 for requests to the Central government, the Armed forces, and Parliament, and £450 for all other public authorities, as determined by Regulation 4 of the Freedom of Information (Fees and Appropriate limit) Regulations 2004.

<sup>268</sup> Response to FOI request from Aoife O’Reilly to the Department of Work and Pensions (27 June 2022)

[https://www.whatdotheyknow.com/request/864638/response/2068623/attach/3/Response%20FOI2022%2041308.pdf?cookie\\_passthrough=1](https://www.whatdotheyknow.com/request/864638/response/2068623/attach/3/Response%20FOI2022%2041308.pdf?cookie_passthrough=1) > accessed 3 July 2023. See also: Response to FOI request from Robert Jones to Ministry of Justice (22 June 2018).

[https://www.whatdotheyknow.com/request/490875/response/1177296/attach/5/FOI%20180612015%20reply.pdf?cookie\\_passthrough=1](https://www.whatdotheyknow.com/request/490875/response/1177296/attach/5/FOI%20180612015%20reply.pdf?cookie_passthrough=1) > accessed 12 July 2023

would exceed the cost limit<sup>269</sup> placed on the obligation to provide information when requested. By this, it means that the data or information is able to be accessed and gathered by the Public Authority, but providing that information, whether requiring synthesising it from their records or databases, would require too much from the Public Authority.

The result of delay gaps is that information requested by Members of Parliament and the public is frequently frustrated, delaying important debates on policy and Government practice. As demonstrated by the case studies provided, delays in the receipt of the requested information, whether that delay is resolved or not, are highly problematic when understood in the context of the information requested. In order to resolve delay gaps, they must first be identified, and the data become relevant enough to be deemed necessary to either centralise the data or find a way to enable access to the data without overburdening public authorities. For the case studies provided, and especially in Police Vetting, the delay gaps only began to move towards a solution after large amounts of public interest and an investigation<sup>270</sup>. From the examples analysed in this section, it becomes apparent that the successful identification and public attention to the existence and harms of the gap provide the momentum to seek to close these gaps.

---

<sup>269</sup> Section 12 of The Freedom of Information Act 2000 states that Public Authorities may refuse a request for information if it exceeds the “appropriate limit”. The appropriate limit was stated to be £600 for central government, Parliament, and the armed forces, and £450 for all other public authorities, as determined by Regulation 4 of the Freedom of Information (Fees and Appropriate limit) Regulations 2004.

<sup>270</sup> See news stories such as: Vikram Dodd and Aina J Khan, 'Police Vetting Failures Have Allowed 'Predatory' Officers to Join Up, Watchdog Finds' The Guardian (<<https://www.theguardian.com/uk-news/2022/nov/02/prevalent-predatory-misogynistic-culture-in-police-official-report-finds>> accessed 12 July 2023 See also: An Inspection of Vetting, Misconduct, and Misogyny in the Police Service' (*HM Inspectorate of Constabulary, Fire and Rescue Services*, 2 November 2022) <<https://www.justiceinspectors.gov.uk/hmicfrs/publication-html/an-inspection-of-vetting-misconduct-and-misogyny-in-the-police-service/#an-explanation-of-vetting-regimes-and-processes>> accessed 12 July 2023.

## IX. Conclusion

The analysis of data gaps undertaken within this paper has sought to answer a fundamental question which has largely been neglected in previous work: what are data gaps? And more precisely: what are the different types of public data gaps?

Data gaps research, across disciplines, has not dealt with this question. The work by Giest and Samuels to define and categorise data gaps<sup>271</sup> has provided an inroad into answering these questions, but as made clear throughout this thesis, the language around public data gaps needs more precision and clarity than that which currently exists and is in use. The need for this precision holds significance for three reasons. Firstly, clear language is required for the identification of data gaps. This thesis has presented data gaps within five distinct categories. These are collection gaps, participation gaps, reporting gaps, harmonisation gaps, and delay gaps. Each of these gaps occurs for unique reasons and can occur in unique circumstances. Through the use of case studies, this thesis has presented active examples<sup>272</sup> of these gaps within UK public data sets and public institutions. Identification of these gaps would not have been possible without an understanding of how these different gaps may “look”. Throughout the literature reviewed for this work, all forms and types of data gaps were conflated with one another and analysed as a single data gap issue, instead of categorising the unique forms of gap that occurred in each piece of literature and how these gaps caused different issues due to their different nature. The result of the treatment of all forms of data gaps being treated as a single issue meant that the identification of these gaps was not precise. Data would be stated as “missing”<sup>273</sup> or “unavailable”<sup>274</sup> without clarifying that missing data is different from unavailable data, and that missing data results in different problems for decision makers than unavailable data, and vice versa. The typology presented in this paper is an active step towards changing this

---

<sup>271</sup> Giest and Samuels (n 7).

<sup>272</sup> All case studies were active at the time of writing.

<sup>273</sup> Powell-Smith (n 3).

<sup>274</sup> Gibbs (n 256).

misconception of data gaps as ‘all one thing’. Each form of gap, when it appears, will be different in each situation it arises in, it will have different root causes, and each data gap should therefore be analysed within its categories to truly understand and combat any negative results that the gap provides or enables. By providing a workable typology of gaps, identification and analysis of public data gaps becomes easier, and missing or unavailable data are not confused for one another as being the same thing.

Secondly, improving the language around public data gaps enables an improved analysis of the impacts of gaps and the harm they may cause. As established within the case studies presented, harm can be wide-ranging, impacting individuals<sup>275</sup>, groups within a population<sup>276</sup>, and even Government itself<sup>277</sup>. Understanding harm is important for the progression of data gaps research. When we understand that data gaps cause harm in a range of ways, the methods that can be used to ‘solve’ gaps can be identified and used to minimise harm wherever possible<sup>278</sup>. Gaps in knowledge can cause a harmful lack of clarity, which can undermine the effective functioning of Government. By recentring harm within the discussion of data gaps, the impact of these gaps upon individuals and the government can be further established. Multiple key harms have been established throughout this thesis and are illustrated through the case studies accompanying each gap. These harms can be summarised as preventing effective Government function in policy and decision-making across a broad range of issues including the justice system, welfare and social support, transportation, housing, policing and crime, and more. Data gaps also have been shown to cause harm by undermining access to justice within the courts and preventing effective policy support for vulnerable people including the elderly<sup>279</sup>, marginalised groups such as

---

<sup>275</sup> Gibbs (n 256).

<sup>276</sup> Lee (n 3).

<sup>277</sup> Freeguard and Shepley (n 64).

<sup>278</sup> Giest and Samuels (n 7). See also: Lee (n 3) and Powell-Smith and Leibowitz (n 22).

<sup>279</sup> Behrendt and others (n 157).

Gypsy communities<sup>280</sup>, those facing homelessness<sup>281</sup>, and those using the welfare system<sup>282</sup>. Furthermore, data gaps may be enabling discriminatory and harmful behaviour towards communities of colour<sup>283</sup>. These harms all result from the information that is gathered or that is available being flawed and creating data gaps. A lack of knowledge within public institutions or for those observing and scrutinising them can undermine key principles of accountability and transparency. For Government and public bodies, it is vital that policy and decision-making can be understood, assessed, and adapted based on the results that they produce, and access to good data is key for this. Without transparency and accountability, ineffective policies can continue to exist and have the potential to exacerbate existing harms. It is therefore in the interest of Government, public bodies, and the public, that the ability to categorise data gaps and understand their causes and effects is undertaken in order to produce the best governance and public services possible.

Thirdly, without this clarity surrounding the language of data gaps, and understanding of what data gaps are, research into this area remains disjointed and artificially separate. Without an accurate and applicable typology and definition of data gaps, research has reached a “dead end”, with little room for the unification of data gaps research across research areas. As evident within the breadth of literature included in the initial review<sup>284</sup>, data gaps persist across disciplines and are interacted with frequently. This thesis presents the first full attempt to draw the research from these disciplines together to create a unified language that can be used across disciplines when describing gaps. Furthermore, the adoption of a unified language will enable further identification of public data gaps across disciplines. Utilising the new typology in this thesis enabled the identification of active data

---

<sup>280</sup> ‘England’s Most Disadvantaged Groups: Gypsies, Travellers and Roma’ (*Equality and Human Rights Commission*, March 2016) <<https://www.equalityhumanrights.com/sites/default/files/is-england-fairer-2016-most-disadvantaged-groups-gypsies-travellers-roma.pdf>> accessed 10 July 2023.

<sup>281</sup> Fitzpatrick and others (n 213).

<sup>282</sup> Edminston and others (n 183).

<sup>283</sup> Dodd and Gidda (n 115).

<sup>284</sup> See technical appendix 4 for a full breakdown of the literature included in the scoping review by research area and country of origin.

gap case studies within the UK context, spanning public institutions and government. By approaching data gaps from within their unique categories, it was possible to identify these examples. For instance, the delay for the Inspectorate of Police and Fire Services to data concerning police vetting procedures<sup>285</sup> was identified by understanding the language that is commonly used for delay gaps but not other types of gaps; primarily by identifying that often Freedom of Information requests are denied on the basis of the time it would take to gather the information and the cost limits it would exceed<sup>286</sup>. Understanding gaps as belonging to unique categories with unique features allows their identification to become easier; language and features can be recognised and used to create tools for identifying data gaps in action. Work by public interest groups and researchers has demonstrated that this is possible; the Centre for Public Data, for example, searched through hundreds of questions from MPs to the Ministry of Justice in order to find if the data gathered by the Ministry was missing or unavailable<sup>287</sup>, in a project undertaken before this typology had been developed. This typology, by highlighting the specific language used for different types of gaps, was able to undertake similar search processes in a much shorter time frame and identify active case studies, making the process easier for researchers to find data gaps within public data.

The growing debate within public law has been questioning how far the law is equipped to deal with the problems that data gaps can present<sup>288</sup>. This thesis has made it clear that important legal principles, including transparency, accountability, and access to justice, are often undermined by the existence of data gaps. Throughout this work, the discussion has stated that this typology can provide a basis for improved legal analysis. This is incredibly beneficial for the future study of the interaction between law and data gaps. The legal research on data gaps is currently very limited, and any existing analysis on data gaps,

---

<sup>285</sup> An Inspection of Vetting, Misconduct, and Misogyny in the Police Service' (*HM Inspectorate of Constabulary, Fire and Rescue Services*, 2 November 2022) <<https://www.justiceinspectors.gov.uk/hmicfrs/publication-html/an-inspection-of-vetting-misconduct-and-misogyny-in-the-police-service/#an-explanation-of-vetting-regimes-and-processes>> accessed 12 July 2023.

<sup>286</sup> See Delay Gap Case Study 2 on page 71 for complete analysis of these issues.

<sup>287</sup> Powell-Smith and Leibowitz (n 23).

<sup>288</sup> Tomlinson, Meers, and Sommers-Joce (n 4).

including the gaps known by the government, is focused particularly on collection gaps<sup>289</sup>. This focus, whilst valid, currently fails to establish the range of ways in which the law, and especially public law and equalities law, can interact with data gaps in the future, as well as the avenues through which the law may be able to solve or minimise the impacts of data gaps. Through the development of this new typology and its application through the case studies identified, legal questions have arisen with regards to equality, fairness, and access to justice. These active examples of how data gaps may interact with law allows for development of the ideas contained in the public law debate, and asks whether the law, as it stands, can provide active, workable remedies to data gaps. The application of this typology will allow for the development of proposed remedies and allow them to be explored, applied, and the jurisprudence to come up to date with the existing issues with public data gaps.

This typology provides a foundation for continuing research on the nature and occurrence of data gaps. A clear and harmonised language for the identification of data gaps will allow for discussion across disciplines, interested parties, and government, and allow for novel solutions to the harms of gaps to be accessed. This improved use of language has the capacity to bring together research across disciplines into data gaps and enable improved approaches to data gaps in the future. There is potential for this typology to continue to develop and improve as the new language is applied. There is scope for the boundaries of each data gap category to be expanded as more gaps are identified and the application of the typology takes place. Furthermore, there is scope for more “types” of gaps to be discovered and analysed in recognition of their individual causes and harms in the future. Future research should therefore utilise this typology wherever possible but should also develop and apply it in a way that compliments their research and findings on public data gaps as the topic continues to evolve.

The future of data gaps research is relatively boundless. Beginning with the research and case studies presented in this thesis, there is potential for all of the data gaps explored

---

<sup>289</sup> Powell-Smith and Leibowitz (n 23).



here to be mitigated or closed. Future researchers can follow these cases and undertake similar methodologies to identify more data gaps within the UK or other public data sets. The methodologies laid out, including the use of a scoping review to identify global and UK examples of harmful data gaps can be followed and used to research further gaps which exist. Scoping reviews are growing in their use across social science research<sup>290</sup>, and the use of a scoping review within this context and its application to legal studies demonstrates that this form of literature review can be of use to researchers continuing within this field of study. Furthermore, the language that has been presented within this typology, including the categorisation of different “types” of gaps within broader gap categories can allow easier identification of data gaps, and permit a fuller charting of data gaps across public services and institutions. In common with the research undertaken into the data gaps across the Justice System in the UK by The Centre for Public Data and the Legal Education Foundation<sup>291</sup>, this research has shown that other Government departments and public bodies also face a range of data gaps that can result in harm. Future researchers can therefore use this typology and language to chart data gaps across departments and public systems such as Universal Credit and Welfare, overseen by the Department for Work and Pensions<sup>292</sup>. Charting these gaps can be incredibly useful for the Government; the gaps within the Justice system received attention from the Justice Select Committee<sup>293</sup>, which then included recommendations for closing these gaps within its report to Parliament<sup>294</sup>. Charting gaps also allows for harm to be charted, and steps can be taken to minimise and prevent the harm that can result from data gaps in the future.

In a world shaped by the constant collection of data through interactions with technology, public institutions have more data than ever before at their disposal. As

---

<sup>290</sup> Kwan and Walsh (n 11). See also Kourgiantakis and others (n 12).

<sup>291</sup> Anna Powell-Smith, “Highlighting Data Gaps in Criminal Justice”, (*Justice Lab*, 30 March 2023) <<https://justicelab.org.uk/highlighting-data-gaps-in-criminal-justice/>> accessed 10 August 2023.

<sup>292</sup> Edminston and others (n 183).

<sup>293</sup> Justice Committee, *The Role of Adult Custodial Remand in the Criminal Justice System: Government Response to the Committee’s Seventh Report* (HC 2022-23 1244 March 2023).

<sup>294</sup> *Ibid.*

evidence-based decision-making practices in the determination of policy become the norm in Government and Public Institutions, evidence and data gaps must be eliminated wherever possible. Data gaps are, fundamentally, harmful. Gaps obscure facts, frustrate policy and undermine research and decision-making. When gaps exist at a public level, the resulting effects and harms are widespread and prevent important legal principles from being achieved by the government and the public. Without data, the accountability of the Government is undermined, transparent decisions and policy-making cannot be achieved, and public discourse is prevented. Furthermore, the impact on the population can be sizeable. The duties of the government to act with equality and the prevention of discrimination in its functions can be undermined, as is access to justice, and public participation in law-making. In all, the legal principles that underpin the effective function of government are prevented and harm is perpetuated by the continued existence of data gaps. The first step in eliminating these gaps is through the effective use of accurate and clear language which can be used to discuss, analyse, and challenge public data gaps in the future.

## Technical Appendices

The following technical appendices will provide further information about the literature review that was undertaken as part of this study. The information contained in these appendices is designed to provide the reader with further explanatory detail about the process of undertaking the review, as well as providing support for the findings of the review. The contents of the appendices are as follows:

Technical Appendix 1: Scoping Review Methodology .....	79
Process of the Scoping Review .....	79
Limitations of the Scoping Review .....	87
Technical Appendix 2: Search Strings .....	89
Technical Appendix 3: Flow Chart for Article Selection .....	91
Technical Appendix 4: Analysis of Literature results by Research Area and Country of Origin .....	92

### Technical Appendix 1: Scoping Review Methodology

#### Process of the Scoping Review

As referenced on page 11 of the main thesis, this appendix contains more detail on the process of conducting the scoping review, including information on the selection process for articles that were included in the study, as well as the limitations of the review as a whole.

As stated, a scoping review aims to present a wide overview of the research on an area in order to better address wide questions about an emerging research topic<sup>295</sup>. Scoping

---

<sup>295</sup> Arksey and O'Malley (n 8).

reviews tend to be an exploratory exercise in order to assess the level of research and literature that is available on a particular topic. In this case, data gaps study is an emerging topic and therefore the broad nature of a scoping review was suitable for this literature review.

The process of the scoping review was based on the Arksey and O'Malley framework<sup>296</sup>. Arksey and O'Malley identify a 5-step process for conducting a scoping review, with an optional sixth "consultation" stage at the end. For the purposes of this review, the consultation exercise was not necessary, as there were no clear "stakeholders" in general data gaps research that could be consulted for this exercise. Therefore, the 5-stage process is as such:

1. Identifying the research question (or questions)
2. Identifying relevant studies
3. Study selection
4. Charting the data
5. Collating, summarising and reporting the results.

The first stage of the process involves determining questions which can guide the scoping review. Identifying particular research questions can help to set some parameters for review<sup>297</sup>, while still allowing for there to be "broad research within an articulated scope of inquiry"<sup>298</sup>. Determining relevant questions can also help clarify the purpose of the review and can assist in the determination of relevant studies<sup>299</sup>. The research questions which were used to steer the direction of the scoping review are contained in the main body of this project.

Broadly speaking, this project adopted the Arksey and O'Malley framework. However, when conducting the review, the actual process underwent the following 4 step process:

---

<sup>296</sup> Arksey and O'Malley (n 8).

<sup>297</sup> Ibid.

<sup>298</sup> Levac, Colquhoun and O'Brien (n 300).

<sup>299</sup> Ibid.

1. Determining the search strategy
2. Creating a search string
3. Identifying relevant studies
4. Final selection of studies for inclusion

The determination of the search strategy can be considered equivalent to the Arksey and O'Malley step of "identifying the research question". A scoping review requires guiding questions in order to focus the searches for relevant literature<sup>300</sup>. For this review, the guiding questions were:

- How do other disciplines, outside of law, consider, recognise, and define data gaps?
- Why do these data gaps happen?
- How do these data gaps cause problems and/or harm?
- Is there any discussion in the wider literature on reforming the law to try to address data gaps?

These questions remained broad in order to capture a range of disciplines in literature. Furthermore, the questions remained open for review and redrafting throughout the scoping study in case there was a need to adjust them in line with the literature returned<sup>301</sup>.

In order to inform the direction of the scoping review search string, I undertook a brief search of the existing literature, using web search engines such as google scholar. This initial search produced published academic studies as well as "grey literature". "Grey literature" refers to any writing that is not necessarily academic or was produced outside of the traditional publishing or distribution channels. Whilst I did not use "grey literature" within

---

<sup>300</sup> Danielle Levac, Heather Colquhoun and Kelly K. O'Brien, 'Scoping studies: advancing the methodology' (2010) 5 *Implementation Science* 69

<sup>301</sup> Arksey and O'Malley (n 8). See also: Helena MI Daudt, Catherine Van Mossel and Samantha J. Scott, 'Enhancing the scoping study methodology: a large, inter-professional team's experience with Arksey and O'Malley's framework' (2013) 13 *BMC Medical Research Methodology* 48.

the actual scoping review, it assisted in the identification of some key terms which were later applied to advanced search databases. The range of literature that this initial search produced was helpful as it provided an initial overview of the state of accessible literature, as well as what terms commonly appeared, and what disciplines discussed data gaps in the manner that this project wished to address.

In common with other scoping reviews in the social sciences, the database used for the literature search was the Web of Science: a database with over 350 disciplines included. This range of disciplines is reflected in the different indices within the Web of Science Core Collection, including the Arts Humanities citation index, the Social Sciences citation Index, and the Science Citation Index. For this search, the study utilised the Web of Science Core Collection database, which contains records from journals, conferences, and books from the highest impact journals worldwide.

Web of Science permits advanced searches of its database using a “search string”. A search string is a combination of search terms and exclusion terms which are then applied to a database in order to find relevant literature. The construction of a search string is a key component of the scoping review<sup>302</sup>. A successful search string must provide a suitable, relevant set of results of a manageable volume<sup>303</sup>. The process of creating the search string used in this review went through several iterations, each of which producing wholly unmanageable numbers of results (well over 100,000) articles. Therefore, a more novel approach to the search string creation process was required. In order to ensure the search string returned a sufficiently focused set of results, based on the questions and initial searching above, I separated the search string into three “blocks”, each with an individual focus and guiding question. These focuses were:

1. Data Gaps - How do other disciplines recognise, categorise, and respond to data gaps?

---

<sup>302</sup> Hoddy and others (n 20).

<sup>303</sup> Ibid.

2. Ignorance - How does data or knowledge gaps affect the behaviour of institutions and decision-makers?
3. Administrative decision-making - How do knowledge gaps affect decision-making?

After a series of trial-and-error experiments with terms within these blocks, I created three search strings and used them to search the Web of Science Core Collection database. Each block produced over 100 results from the initial search. It was therefore necessary to reduce the number of articles in each block in order to focus the review. I determined that articles which were not in English or did not have an abstract available on Web of Science should be excluded as there was not the capacity to translate non-English articles, and articles which are without an abstract were often inaccessible through Web of Science, and therefore were excluded. Abstracts and introductions of the remaining articles were then screened for relevance to the project, guided by the following questions:

“Data Gaps” block:

- Does it discuss gaps in data/evidence etc in a substantial way?
- Is the purpose of the article to reveal data gaps and their causes or effects or is it to recognise and fill these gaps?
- Does it relate to issues which have an effect or could have an effect on government policy or wide-ranging policy between institutions?

“Ignorance” block:

- Does it discuss how lack of information affects decision-making across disciplines?
- Does it discuss how individuals or organisations respond to a lack of information or attempt to fill the gaps?
- Does it discuss the harm that lack of information can cause?

“Administrative decision-making” block:

- Does it discuss the manner in which administrative decision-making/ bureaucratic decision-making is undertaken?
- Does it discuss the process of decision-making or evidence-based decision-making?
- Does it discuss participation or evidence gathering when applied to decision-making processes?

This exercise produced 94 relevant articles for the data gaps block, 57 articles for the ignorance block, and 39 articles for the decision-making block. This led to 190 articles in total, which is beyond the ordinary sample size for scoping reviews<sup>304</sup>. As a result, in line with the methodology adopted by Pham et al<sup>305</sup>, a second sifting exercise was undertaken to ensure there was a manageable number of relevant articles returned.

For the second sift, I established and applied criteria which were designed to limit the articles returned to the most relevant results possible for this review. These criteria were determined through knowledge of the content of the articles as well as in reference to the original guiding questions which had been posed for the review. Following the guidance of Levac and others<sup>306</sup>, these criteria had the capacity to be reviewed throughout the project if these criteria excluded too much of the returned literature.

Block 1: Data Gaps

Include if:

1. There is significant discussion on the underlying causes of the data or research gaps,
2. There is recognition AND discussion of the harms that these gaps have caused/are causing.

---

<sup>304</sup> Pham and others (n 14).

<sup>305</sup> Ibid.

<sup>306</sup> Levac, Colquhoun and O'Brien (n 300).



3. There is categorisation of missing information or data gaps e.g., into primary, secondary gaps.
4. There are defined issues with data that is available e.g., data is not disaggregated by gender or race.
5. There are defined issues with the collecting, reporting, or sharing of data and the reasons behind these issues are explored.

Exclude if:

1. There is recognition of data gaps, and the purpose of the article is to fill said gap through the research presented, without investigation or substantial discussion of the underlying causes of data gaps caused in the researched area.
2. The harm or problems that these data gaps have caused are not discussed in a substantial manner.
3. The article takes an approach to fill data gaps by purely focusing on data gathering - the “take more data” approach.

A total of 23 articles from this block were included in the final study.

Block 2: Ignorance:

Include if:

1. There is significant discussion as to why it can be beneficial for decision makers to not have certain information or to disregard certain information when making decisions.
2. There is significant discussion as to why it can be harmful for decision makers to not have certain information or to disregard certain information when making decisions.
3. The article identifies ignorance as harmful/beneficial within a policy making context.
4. There is significant discussion on how decision-makers or officials use ignorance to their advantage in policy contexts.

Exclude if:

1. There is substantial focus on the moral processes of decision-making or moral decision-making.
2. The article has a focus on the cognitive or psychological definitions of ignorance, rather than the anti-epistemology lens of this study.

After applying these criteria, 20 articles were selected for the final scoping review.

### Block 3: Decision-Making

Include if:

1. There is significant discussion of the manner in which decision-makers respond to information or evidence that is provided to them during the decision-making process.
2. There is significant discussion of the manner in which information is sought to be provided to decision makers, whether that is through increased public or citizen participation, or other manners.
3. There is discussion on how decision makers seek to make informed decisions, and how often decision makers seek to make informed decisions.
4. There is discussion on how the decision maker or decision-making process can be adjusted when provided with new information or missing information.

Exclude if:

1. There is insufficient attention to the process of decision-making and how decision makers respond to information during that process.
2. There is excessive focus on the process of increasing participation in governance, rather than the effect of said participation in policy decision-making processes.

After applying these criteria to the remaining results, 15 studies were selected.

Across the three blocks, a total of 59 articles fulfilled the criteria for inclusion in the final scoping study. These included three articles which were added two months after the initial search as new articles became available. These new articles went through the same two sifting exercises as were used to select articles found previously.

I would also consider the decision to divide the searches into three distinct areas to be indicative of the search strategy as a constant element that is reviewed and returned to. This approach to the search strategy expands upon the Arksey and O'Malley framework, which is actively encouraged by Levac<sup>307</sup> in their discussion of scoping review protocols.

Furthermore, the creation of the search string, especially in this case, deserves to be a step of the process itself. The creation of the search string required several hundred trial and error experiments of individual terms to see what they returned, before terms could begin to be strung together. The creation of the search string, to return manageable and relevant article results, was time-consuming and rigorous.

The identification of relevant studies was then undertaken through the use of the search strings applied in Web of Science. A copy of these search strings is attached in Appendix 2.

### Limitations of Scoping Reviews

There are limitations to scoping reviews in the literature which are important to note for the purpose of transparency. Firstly, scoping review protocols are not considered "set" in the literature, with suggestions for additions or changes to the initial protocol actively encouraged by Arksey and O'Malley<sup>308</sup>. Daudt et al suggested more guidance for the determination of the research question was necessary, and particularly called for more guidance about what questions for scoping reviews are appropriate<sup>309</sup>, and the need for researchers to consider whether scoping reviews are the best type of review to undertake. In these suggestions, Daudt et al established a further limitation of scoping reviews, in the sense that their usage can be undertaken inappropriately<sup>310</sup>. However, this is not strictly a problem within this review. A second limitation with scoping reviews, is that the necessary

---

<sup>307</sup> Ibid.

<sup>308</sup> Arksey and O'Malley (n 8).

<sup>309</sup> Daudt, Van Mossel and Scott (n 301).

<sup>310</sup> Ibid.

step of excluding literature leads to the concern that there is more relevant literature available that has been excluded or has not yet been found. This is a concern for this review, and every scoping review, and is a relevant concern to have. It is important to recognise that there could be more literature available that hasn't been found, as it fits within the role of the scoping review, which is to get a broad overview of the literature that is available. It would be impossible to read or find every relevant piece of literature which may be relevant for the purpose of this project. Therefore, it is necessary to state that some relevant literature may have been excluded. According to Gentles et al it is "unrealistic to retrieve and screen all relevant literature"<sup>311</sup> in a scoping review due to its wider focus. Furthermore, limitations are placed on this particular review due to time constraints. When compared to some other scoping reviews in the literature, the average time taken to review is 5 months<sup>312</sup>, but other reviews can last over a year<sup>313</sup>. This review did not have 12 months or more with which to undertake the review, and final selection of the articles for inclusion was completed within 3 months. One of the benefits of longer time frames is that it can be possible to have a greater breadth of studies included, as there is more time available to review these studies, as well as allowing for the potential publication of relevant articles to take place. Therefore, the time limit of 5 months on this study led to not as much breadth in the final articles being selected as there was a need to be highly focussed on the exclusion criteria.

---

<sup>311</sup> Andrew J. Gentles and others, 'The prognostic landscape of genes and infiltrating immune cells across human cancers' (2015) 21 *Nature Medicine* 938.

<sup>312</sup> Pham and others (n 14).

<sup>313</sup> Maria J. Grant and Andrew Booth, 'A typology of reviews: an analysis of 14 review types and associated methodologies' (2009) 26 *Health Information & Libraries Journal* 91.

## Technical Appendix 2: Search Strings

As referenced on page 11 of the Thesis, attached are the three search strings applied to the Web of Science database to produce the literature for this review. The three search strings are presented in their 'blocks', as referenced on page 1, and referenced again on page 87 of Technical Appendix 1.

### Data Gaps Search String

TS=("data gap\*" OR "evidence gap\*" OR "missing evidence" OR "missing knowledge" OR "missing research") OR TS=("poor quality data" OR "imperfect evidence")

NOT

TS=("fill\* the gap\*" OR "fix\*")

NOT

TS=("english teach\*" OR "foreign language education" OR "teach\*" or "information gap")

NOT

"ENVIRONMENTAL SCIENCES ECOLOGY" OR "ENGINEERING" OR "TOXICOLOGY"  
OR "COMPUTER SCIENCE" OR "MATHEMATICS" OR "PHARMACOLOGY  
PHARMACY" OR "BIODIVERSITY CONSERVATION" OR "METEOROLOGY  
ATMOSPHERIC SCIENCES" OR "ZOOLOGY" OR "AGRICULTURE" OR  
"BIOCHEMISTRY MOLECULAR BIOLOGY" OR "MATHEMATICAL COMPUTATIONAL  
BIOLOGY" OR "PHYSICAL SCIENCES OTHER TOPICS" OR "MARINE FRESHWATER  
BIOLOGY" OR "GENERAL INTERNAL MEDICINE" OR "WATER RESOURCES" OR

"CHEMISTRY" OR "PEDIATRICS" OR "GEOLOGY" OR "ENERGY FUELS" OR  
"NUTRITION DIETETICS" OR "GEOCHEMISTRY GEOPHYSICS" OR "ASTRONOMY  
ASTROPHYSICS"

REFINE BY

Language: English

REFINE BY

Article ONLY

*Table 1: Data Gaps*

Ignorance Search String

((TS=("Strategic ignorance" OR strategic unknowns OR "sociology of ignorance" OR  
"uncomfortable knowledge" OR "anti epistemology" OR antiepistemolog\* OR epistemology  
of ignorance OR "limits of knowledge" OR "structural amnesia" OR "varieties of  
uncertainty"))

REFINE BY

Language: English

REFINE BY

Article ONLY

*Table 2: Ignorance*

Administrative decision-making Search String

TS=("administrative decision-making" OR "admin\* decision\*" OR "bureaucratic decision-  
making" OR "bureauc\* decision\*"))

AND

TS=("social policy" OR "policy" OR "policy\*" OR "social science" OR "government system\*")

AND

TS=("process")

REFINE BY

Language: English

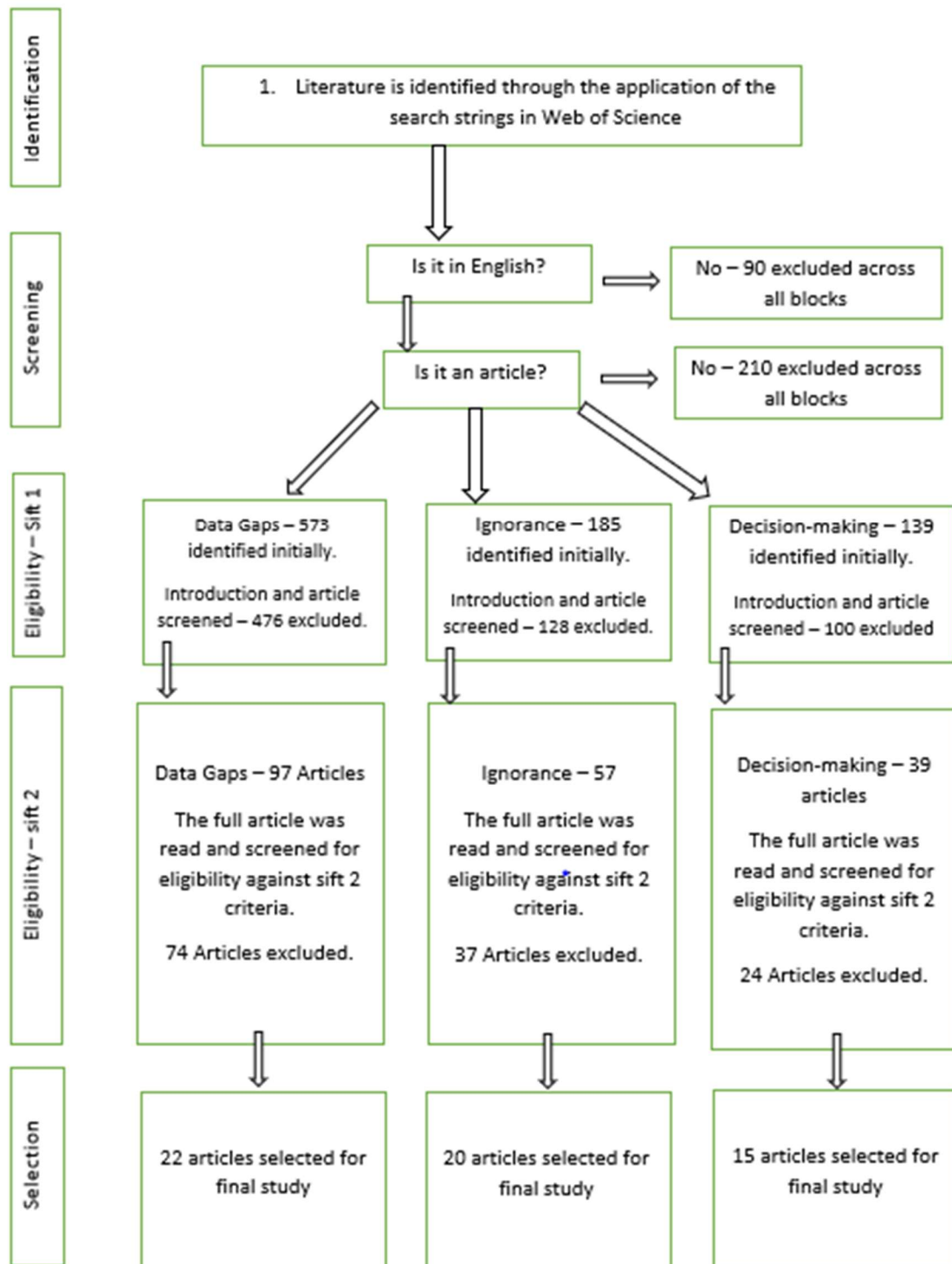
REFINE BY

Article ONLY

*Table 3: Decision-Making*

### Technical Appendix 3: Flow Chart for article selection

This flow chart lays out the process for article selection undertaken within the scoping review.





## Technical Appendix 4: Literature Results by Research Area and Country of Origin

The following tables present the research area and country of origin of the articles which were included in the final literature review. These results are broken down by 'block'. This has been included within the appendix in order to show the range of research areas included as well as the global nature of the research questions posed within this thesis.

### Data Gaps Block

#### **Field:**

<b>Research Areas</b>	<b>Record Count</b>	<b>% of 24</b>
Public Environmental Occupational Health	8	33.333%
Social Sciences Other Topics	4	16.667%
Health Care Sciences Services	3	12.500%
Biomedical Social Sciences	2	8.333%
Business Economics	2	8.333%
Government Law	2	8.333%
Social Issues	2	8.333%
Communication	1	4.167%
Criminology Penology	1	4.167%
Demography	1	4.167%
Infectious Diseases	1	4.167%
Information Science Library Science	1	4.167%
International Relations	1	4.167%
Legal Medicine	1	4.167%
Medical Ethics	1	4.167%
Public Administration	1	4.167%

<b>Field:</b>	<b>Record</b>	
<b>Countries/Regions</b>	<b>Count</b>	
USA	11	45.833%
ENGLAND	4	16.667%
AUSTRALIA	1	4.167%
IRELAND	1	4.167%
ISRAEL	1	4.167%
LEBANON	1	4.167%
NETHERLANDS	1	4.167%
NEW ZEALAND	1	4.167%
U ARAB EMIRATES	1	4.167%

#### Ignorance Block

<b>Field:</b>	<b>Record</b>	<b>% of 21</b>
<b>Research Areas</b>	<b>Count</b>	
Public Administration	5	23.810%
Business Economics	4	19.048%
Environmental Sciences Ecology	4	19.048%
Sociology	4	19.048%
Government Law	3	14.286%
History Philosophy of Science	2	9.524%
Public Environmental Occupational Health	2	9.524%
Social Sciences Other Topics	2	9.524%
Biomedical Social Sciences	1	4.762%
Communication	1	4.762%

Cultural Studies	1	4.762%
International Relations	1	4.762%
Philosophy	1	4.762%
Psychology	1	4.762%

**Field:** % of 21

Countries/Regions	Record Count	
ENGLAND	5	23.810%
FRANCE	4	19.048%
USA	3	14.286%
CANADA	2	9.524%
SPAIN	2	9.524%
SWITZERLAND	2	9.524%
AUSTRALIA	1	4.762%
AUSTRIA	1	4.762%
BRAZIL	1	4.762%
GERMANY	1	4.762%
NORWAY	1	4.762%
TAIWAN	1	4.762%

Decision-Making Block

Field: Research Areas	Record Count	% of 15
Public Administration	6	40.000%
Computer Science	3	20.000%

Government Law	3	20.000%
Psychology	3	20.000%
Health Care Sciences Services	2	13.333%
Behavioural Sciences	1	6.667%
Education Educational Research	1	6.667%
Engineering	1	6.667%
Public Environmental Occupational	1	6.667%
Health		
Social Sciences Other Topics	1	6.667%
Sociology	1	6.667%

**Field: Record % of 15**

**Countries/Regions Count**

USA	8	53.333%
BELGIUM	1	6.667%
DENMARK	1	6.667%
ENGLAND	1	6.667%
GREECE	1	6.667%
IRELAND	1	6.667%
SAUDI ARABIA	1	6.667%
SOUTH AFRICA	1	6.667%

## Bibliography

'Action Plan – Transparency, Accountability and Trust in Policing' (*Mayor of London and London Assembly*) < <https://www.london.gov.uk/publications/action-plan-transparency-accountability-and-trust-policing> > accessed 10 July 2023

'An Inspection of Vetting, Misconduct, and Misogyny in the Police Service' (*HM Inspectorate of Constabulary, Fire and Rescue Services*, 2 November 2022)  
<<https://www.justiceinspectors.gov.uk/hmicfrs/publication-html/an-inspection-of-vetting-misconduct-and-misogyny-in-the-police-service/#an-explanation-of-vetting-regimes-and-processes>> accessed 12 July 2023

'England's Most Disadvantaged Groups: Gypsies, Travellers and Roma' (*Equality and Human Rights Commission*, March 2016)  
<<https://www.equalityhumanrights.com/sites/default/files/is-england-fairer-2016-most-disadvantaged-groups-gypsies-travellers-roma.pdf>> accessed 10 July 2023

'Homelessness code of guidance for Local Authorities' (HM Government Department for Levelling Up, Housing and Communities, 31 May 2023)  
<<https://www.gov.uk/guidance/homelessness-code-of-guidance-for-local-authorities/chapter-8-priority-need>> accessed 11 July 2023

'Number of Gypsy, Roma, and Traveller People disclosing Ethnicity in Census increases by 110,000' (*The Traveller Movement*, 29 November 2022)  
<[https://travellermovement.org.uk/news/5719#\\_ftn1](https://travellermovement.org.uk/news/5719#_ftn1)> accessed 10 July 2023

'Transport for London Annual Report and Statement of Accounts 2015/16' (*Transport for London and Mayor of London*) <https://content.tfl.gov.uk/tfl-annual-report-2015-16.pdf>  
accessed 10 July 2023

"Universal Credit Statistics: Background Information and Methodology" (Department for Work and Pensions, 14 February 2023)  
<<https://www.gov.uk/government/publications/universal-credit-statistics-background->

[information-and-methodology/universal-credit-statistics-background-information-and-methodology](#)> accessed 10 July 2023

'Access and Inclusion in 2016: Outcomes for Consumers in Vulnerable Circumstances' (*Ofcom*, 15 March 2017)

<[https://www.ofcom.org.uk/data/assets/pdf\\_file/0030/98508/access-inclusion-report-2016.pdf](https://www.ofcom.org.uk/data/assets/pdf_file/0030/98508/access-inclusion-report-2016.pdf)> accessed 10 July 2023

An Inspection of Vetting, Misconduct, and Misogyny in the Police Service' (*HM Inspectorate of Constabulary, Fire and Rescue Services*, 2 November 2022)

<<https://www.justiceinspectors.gov.uk/hmicfrs/publication-html/an-inspection-of-vetting-misconduct-and-misogyny-in-the-police-service/#an-explanation-of-vetting-regimes-and-processes>> accessed 12 July 2023

Government Statistical Service, 'UK Official Statistics On Homelessness: Comparisons, Definitions, And Processes (*Government Statistical Service*)

<<https://analysisfunction.civilservice.gov.uk/dashboard/tools/homelessness-statistics/comparison.html>> accessed 18 May 2023

'Gypsy, Roma and Irish Traveller Ethnicity Summary' (*Gov.uk Ethnicity Facts and Figures*, 29 March 2022) <[https://www.ethnicity-facts-](https://www.ethnicity-facts-figures.service.gov.uk/summaries/gypsy-roma-irish-traveller)

[figures.service.gov.uk/summaries/gypsy-roma-irish-traveller](https://www.ethnicity-facts-figures.service.gov.uk/summaries/gypsy-roma-irish-traveller)> accessed 10 July 2023

Justice Committee, *The Role of Adult Custodial Remand in the Criminal Justice System: Government Response to the Committee's Seventh Report* (HC 2022-23 1244 March 2023)

'National Stop and Search Learning Report' (*Independent Office For Police Conduct*, April 2022)

<<https://www.policeconduct.gov.uk/sites/default/files/Documents/publications/OFFICIAL%20IOPC%20National%20stop%20and%20search%20learning%20report%2020%20April%202022.pdf>> accessed 10 July 2023

Police Powers and Procedures: Stop and Search and Arrests, England and Wales, year ending 31 March 2022', (*HM Government Home Office*, 27 October 2022) <

<https://www.gov.uk/government/statistics/police-powers-and-procedures-stop-and-search-and-arrests-england-and-wales-year-ending-31-march-2022/police-powers-and-procedures-stop-and-search-and-arrests-england-and-wales-year-ending-31-march-2022#:~:text=In%20the%20year%20ending%20March%202022%20there%20were%20526%2C024%20stop,and%20searches%20had%20been%20increasing.>> accessed 10 July 2023

'Protocol On Sharing Court Lists, Registers And Documents With The Media' (*HM Courts and Tribunals Service, 2023*) < <https://www.gov.uk/government/publications/guidance-to-staff-on-supporting-media-access-to-courts-and-tribunals/protocol-on-sharing-court-lists-registers-and-documents-with-the-media-accessible-version>> accessed 18 May 2023

'Systematic Review of Public Value: Statistics on Housing and Planning in the UK' (Office for National Statistics, November 2017)

'Systematic Review of Public Value: Statistics on Housing and Planning in the UK' (Office for National Statistics, November 2017)

TfL Mobile Data and Privacy, (*Transport for London, February 2016*) <<https://content.tfl.gov.uk/mobile-data-and-privacy-report.pdf>> accessed 10 July 2023

'Traffic Stops (Vehicle Stops) Factsheet' (*StopWatch, 24 October 2022*) <[https://www.stop-watch.org/what-we-do/research/traffic-stops-factsheet/#:~:text=What%20is%20a%20vehicle%20stop,car%20or%20scooter\)%3B%20or>](https://www.stop-watch.org/what-we-do/research/traffic-stops-factsheet/#:~:text=What%20is%20a%20vehicle%20stop,car%20or%20scooter)%3B%20or>) accessed 10 July 2023

Written Evidence Submitted by Public Law Project to Welsh Affairs Committee Inquiry into Benefits System in Wales ( 17 August 2021) <https://publiclawproject.org.uk/content/uploads/2021/11/Benefit-sanctions-evidence.pdf> accessed 16 July 2023

Women and Equalities Committee, *Tackling Inequalities Faced by Gypsy, Roma and Traveller Communities* (HC 2017-19 360 April 2019)

Abu-Saad K, Avni S, Kalter-Leibovici O, 'Health Disparities Monitoring In The US: Lessons For Monitoring Efforts In Israel And Other Countries' (2018) 7 Israel Journal Of Health Policy Research 14

Agarwal B, 'Imperatives Of Recognising The Complexities: Gendered Impacts And Responses To Covid-19 In India' (2022) 39 *Economia Politica* 31

Akik C, Semaan A, Shaker-Berbari L, Jamaluddine Z, Saad G, Lopes K, Constantin J, Ekzayez A, Singh N, Blanchet K, DeJong J, Ghattas H, 'Responding To Health Needs Of Women, Children And Adolescents Within Syria During Conflict: Intervention Coverage, Challenges And Adaptations' (2020) 14 *Conflict And Health* 37

Albaek E, 'Between Knowledge And Power - Utilization Of Social-Science In Public-Policy Making' (1995) 28 *Policy Sciences* 79

Arksey H and O'Malley L, 'Scoping Studies: Towards A Methodological Framework' (2005) 8 *International Journal of Social Research Methodology* 19

Bartlett J, and Tkacz N, 'Governance by dashboard: a policy paper' (*Demos*, March 2017) <<https://dera.ioe.ac.uk/28793/1/Demos-Governance-by-Dashboard.pdf>> accessed 10 July 2023

Bates L, Alexander M, Webster J, 'The link between dangerous driving and other criminal behaviour: a scoping review' (2022) 21 *Safer Communities* 137

Behrendt F, Murray L, Hancox A, Sourbati M, Huber J, 'Intelligent Transport Solutions for Social Inclusion (ITSSI) Project Report', (*University of Brighton*, 2017) <<https://cris.brighton.ac.uk/ws/portalfiles/portal/456536/Behrendt+et+al+%282017%29+Intelligent+Transport+Solutions+for+Social+Inclusion+%28ITSSI%29+Project+Report.pdf>> accessed 10 July 2023

Behrman J and Rosenzweig M, 'Caveat-Emptor - Cross-Country Data On Education And The Labor-Force' (1994) 44 *Journal Of Development Economics* 147

Bhatia R, Sledge I, Baral S, 'Missing science: A scoping study of COVID-19 epidemiological data in the United States' (2022) 17 *PLoS ONE* <<https://doi.org/10.1371/journal.pone.0248793>> accessed 10 July 2023



Bishop S.C, 'An International Analysis of Governmental Media Campaigns to Deter Asylum Seekers' (2020) 14 International Journal of Communication 1092

Bogenschneider K, Day E, Parrott E, 'Revisiting Theory on Research Use: Turning to Policymakers for Fresh Insights' (2019) 74 American Psychologist 778

Bronson J and Sufrin C, 'Pregnant Women in Prison and Jail Don't Count: Data Gaps on Maternal Health and Incarceration' (2019) 134 Public Health Reports 57

Brown P, Martin P, Scullion L, 'Migrant Roma In The United Kingdom And The Need To Estimate Population Size' (2014) 8 People, Place and Policy 19

Burton J.K, Goodman C, Guthrie B, Gordon A, Hanratty B, Quinn T, 'Closing The UK Care Home Data Gap – Methodological Challenges And Solutions' (2020) 5 International Journal of Population Data Science <<https://dx.doi.org/10.23889/ijpds.v5i4.1391>> accessed 10 July 2023

Chong N, Deroubaix J-F, Bonhomme C, 'Eyes wide shut: Exploring practices of negotiated ignorance in water resources modelling and management' (2018) 227 Journal of Environmental Management 286

Christiansen K, and Lewis R, UK Private Rented Sector: 2018 (*Office for National Statistics*, 18 January 2019)

<[https://www.ons.gov.uk/economy/inflationandpriceindices/articles/ukprivaterentedsector/2018#:~:text=Around%20three%2Dquarters%20\(74%25\),social%20renters%20and%20owner%20occupiers](https://www.ons.gov.uk/economy/inflationandpriceindices/articles/ukprivaterentedsector/2018#:~:text=Around%20three%2Dquarters%20(74%25),social%20renters%20and%20owner%20occupiers)> accessed 11 July 2023

Cooper K and Obolenskaya P, 'Hidden Victims: The Gendered Data Gap Of Violent Crime' (2021) 61 British Journal Of Criminology 905 See also: Maria Sourbati and Frauke Behrendt, 'Smart Mobility, Age and Data Justice' (2021) 23 New Media & Society 1398

Daudt H, Van Mossel C, Scott S.J, 'Enhancing the scoping study methodology: a large, inter-professional team's experience with Arksey and O'Malley's framework' (2013) 13 BMC Medical Research Methodology 48

Deitz S, Lobben A, Alferez A, 'Squeaky Wheels: Missing Data, Disability, And Power In The Smart City' (2021) 8 *Big Data & Society*

<<https://journals.sagepub.com/doi/full/10.1177/20539517211047735>> accessed 10 July 2023

Dodd V, and Gidda M, 'Met to Stop Recording Ethnicity of Drivers Stopped by its Officers' *The Guardian* (11 October 2022) <[https://www.theguardian.com/uk-](https://www.theguardian.com/uk-news/2022/oct/11/met-to-stop-recording-ethnicity-of-drivers-stopped-by-its-officers)

[news/2022/oct/11/met-to-stop-recording-ethnicity-of-drivers-stopped-by-its-officers](https://www.theguardian.com/uk-news/2022/oct/11/met-to-stop-recording-ethnicity-of-drivers-stopped-by-its-officers) > accessed 10 July 2023

— — 'Met Police To Stop Recording Ethnicity of Drivers Stopped by its Officers' (*Liberty Investigates*, 13 October 2022) <<https://libertyinvestigates.org.uk/articles/met-police-to-stop-recording-ethnicity-of-drivers-stopped-by-its-officers/>> accessed 18 May 2023

Dodd V, and Khan A.J, 'Police Vetting Failures Have Allowed 'Predatory' Officers to Join Up, Watchdog Finds' *The Guardian* (<<https://www.theguardian.com/uk-news/2022/nov/02/prevalent-predatory-misogynistic-culture-in-police-official-report-finds>> accessed 12 July 2023

Edminston D, Baumberg Geiger B, de Vries R, Scullion L, Summers K, Ingolig J, Robertshaw D, Gibbons A, and Karagiannaki E, 'Who are the New COVID-19 Cohort of Benefit Claimants?' (*Welfare – At a (Social) Distance*, September 2020)

<<https://hub.salford.ac.uk/welfare-at-a-social-distance/wp-content/uploads/sites/120/2020/09/WaSD-Rapid-Report-2-New-COVID-19-claimants.pdf>> accessed 16 July 2023

Elkin M, 'Sexual Offences Victim Characteristics, England and Wales: Year ending March 2022' (*Office for National Statistics: Crime and Justice*, 23 March 2023) <

<https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/articles/sexualoffencesvictimcharacteristicsenglandandwales/yearendingmarch2022>> accessed 10 July 2023

Enoch M, 'Mobility as a Service (MaaS) in the UK: Change and its Implications' (*Foresight and HM Government Office for Science*, December 2018) <

[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/766759/Mobilityasaservice.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/766759/Mobilityasaservice.pdf)> accessed 10 July 2023

Fitzpatrick S, Wilcox S, Stephens M, Pleace N, Wallace A, Rhodes D, 'The Impact of Devolutions: Housing and Homelessness (*The Joseph Rowntree Foundation*, January 2010) <<https://www.jrf.org.uk/sites/default/files/jrf/migrated/files/impact-of-devolution-long-term-care-housing.pdf>> accessed 11 July 2023

Freedom Of Information response from the Department of Work and Pensions to Aoife O'Reilly (27 June 2022) <<https://www.whatdotheyknow.com/request/864638/response/2068623/attach/html/3/Response%20FOI2022%2041308.pdf.html>> accessed 10 July 2023

Freeguard G and Shepley P, 'Data Sharing Between National, Devolved and Local Government: A Summary of a Private Roundtable' (*Institute for Government*, 30 January 2023) <https://www.instituteforgovernment.org.uk/publication/data-sharing-national-local-government> accessed 10 July 2023

Question from Grahame Morris MP to Ministry of Justice (21 January 2021) <Written questions and answers - Written questions, answers and statements - UK Parliament> Accessed 12 June 2023.

Question from Jo Stevens MP to Ministry of Justice (2 April 2019) <Written questions and answers - Written questions, answers and statements - UK Parliament> accessed 12 June 2023.

Question from Gloria De Piero MP to Ministry of Justice (30 November 2018) <Written questions and answers - Written questions, answers and statements - UK Parliament> accessed 12 June 2023

Galaiti S, Cegan J, Volk K, Joyner M, Trump B, Linkov I, 'The Challenges Of Data Usage For The United States' Covid-19 Response' (2021) 59 *International Journal Of Information Management* <<https://www.sciencedirect.com/science/article/pii/S0268401221000451>> accessed 10 July 2023

Gentles A, Newman A, Liu C.H, Bratman S, Feng W, Kim D, Nair V, Xu Y, Khuong A, Hoang C, Diehn M, West R, Plevritis S, Alizadeh A, 'The prognostic landscape of genes and infiltrating immune cells across human cancers' (2015) 21 Nature Medicine 938

Gibbs P, 'Justice Denied? The Experience of Unrepresented Defendants in the Criminal Courts,' (*Transform Justice*, 29 April 2016)

<<https://www.transformjustice.org.uk/publication/justice-denied-the-experience-of-unrepresented-defendants-in-the-criminal-courts/>> accessed 11 July 2023

Giest S and Samuels A, 'For Good Measure': Data Gaps In A Big Data World' (2020) 53 Policy Sciences 559

Golman R, Hagmann D, Loewenstein G, 'Information Avoidance' (2017) 55 Journal of Economic Literature 96

Grant M, and Booth A, 'A typology of reviews: an analysis of 14 review types and associated methodologies' (2009) 26 Health Information & Libraries Journal 91

Hahn R, Bilukha O, Crosby A, Fullilove M, Liberman A, Moscicki E, Snyder S, Tuma F, Briss P, "Firearms Laws and the Reduction of Violence: A Systematic Review" (2005) 28 American Journal of Preventative Medicine 40

Hoddy E, Halliday S, Ensor J, Wamsler C, Boyd E, 'Legal culture and climate change adaptation: An agenda for research' (2023) 14 WIREs Climate Change <<https://dx.doi.org/10.1002/wcc.825>> accessed 10 July 2023

Holtzman G, Khoshkhoo N, Nsoesie E, 'The Racial Data Gap: Lack of Racial Data as a Barrier to Overcoming Structural Racism' (2022) 22 The American Journal of Bioethics 39

Hudson N, Understanding Local Housing Markets: Advice and Guidance for Local Authorities, (*Local Government Association*)

<<https://www.local.gov.uk/sites/default/files/documents/LGA%20RA%20Understanding%20Local%20Housing%20Markets%20June%202019.pdf>> accessed 11 July 2023

Jones P, 'Crime in England and Wales: Year ending September 2022' (*Office for National Statistics: Crime and Justice*, 26 January 2023)

<<https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/bulletins/crimeinenglandandwales/yearendingseptember2022#domestic-abuse-and-sexual-offences>>

accessed 10 July 2023

Jones P, 'Homicide in England and Wales: Year ending March 2022' (*Office for National Statistics: Crime and Justice*, 9 February 2023)

<<https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/articles/homicideinenglandandwales/march2022>> accessed 10 July 2023

Jones P, 'The Nature of Violent Crime in England and Wales: Year ending March 2022' (*Office For National Statistics: Crime and Justice*, 9 November 2022)

<<https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/articles/thenatureofviolentcrimeinenglandandwales/yearendingmarch2022#police-recorded-crime-trends-for-different-types-of-violent-crime>> accessed 10 July 2023

Katz A, Enns J, Wong S, Williamson T, Singer A, McGrail K, Bakal J, Taylot C, Peterson S, 'Challenges Associated with Cross-Jurisdictional Analyses using Administrative Health Data and Primary Care Electronic Medical Records in Canada' (2018) 3 *International Journal of Population Data Science* 437

Kourgiantakis T, Sewell K, McNeil S, Logan J, Lee E, Adamson K, McCormick M, Kuehl D, 'Social work education and training in mental health, addictions and suicide: a scoping review protocol' (2019) 9 *BMJ Open*

Kwan C and Walsh C, 'Old age poverty: A scoping review of the literature' (2018) 4 *Cogent Social Sciences*

Lee P-H, 'Un(ac)countable no-bodies: the politics of ignorance in global health policymaking' (2023) 33 *Critical Public Health* 48

Levac D, Colquhoun H, O'Brien K.H, 'Scoping studies: advancing the methodology' (2010) 5 *Implementation Science* 69

Liberati A, Chatziandreu E, Miettinen O.S, 'Health Care Research: What Is It About?' (1989) 1 *Quality Assurance In Health Care: The Official Journal Of The International Society For Quality Assurance In Health Care* 249

McGoey L, 'Strategic unknowns: towards a sociology of ignorance' (2012) 41 *Economy And Society* 1

Migchelbrink K, and Van de Walle S, 'When Will Public Officials Listen? A Vignette Experiment on the Effects of Input Legitimacy on Public Officials' Willingness to Use Public Participation' (2020) 80 *Public Administration Review* 271

Morey B.N, Chang R.C, Thomas K.B, Tulua A, Penaia C, Tran V.D, Pierson N, Greer J.C, Bydalek M, Ponce N, 'No Equity without Data Equity: Data Reporting Gaps for Native Hawaiians and Pacific Islanders as Structural Racism' (2022) 47 *Journal Of Health Politics Policy And Law* 159

Munn Z, Peters M, Stern C, Tufanaru C, McArthur A, Aromataris E, 'Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach' (2018) 18 *BMC Medical Research Methodology* <  
<https://bmcmmedresmethodol.biomedcentral.com/articles/10.1186/s12874-018-0611-x>>  
 accessed 10 July 2023

North A, Index of Private Housing Rental Prices, UK: April 2023 (*Office For National Statistics*, 24 May 2023)  
 <<https://www.ons.gov.uk/economy/inflationandpriceindices/bulletins/indexofprivatehousingrentalprices/april2023#:~:text=Private%20rental%20prices%20paid%20by,12%20months%20to%20April%202023>> accessed 11 July 2023

North A, Private Rental Market Summary Statistics in England: October 2021 to September 2022 (*Office for National Statistics*, 14 December 2022)  
 <<https://www.ons.gov.uk/peoplepopulationandcommunity/housing/bulletins/privaterentalmarketsummarystatisticsinengland/october2021toseptember2022#:~:text=1.-,Main%20points,monthly%20rent%20at%20%C2%A3525.>> accessed 11 July 2023

O'Malley J, 'TfL is Going to Track All London Underground Users using Wi-Fi' (*WIRED*, 22 May 2019) <<https://www.wired.co.uk/article/london-underground-wifi-tracking>> accessed 18 May

Parkes E, Karger-Lerchl T, Wells P, Hardinges J, Vasilevaand R, 'Using Open Data To Deliver Public Services' (*Open Data Institute*, February 2017) <<https://theodi.org/wp-content/uploads/2018/03/Using-open-data-to-deliver-public-services.pdf>> accessed 10 July 2023

Paul K.T, and Haddad C, 'Beyond evidence versus truthiness: toward a symmetrical approach to knowledge and ignorance in policy studies' (2019) 52 *Policy Sciences* 299

Pawluczuk A, 'Digital Youth Inclusion and the Big Data Divide: Examining the Scottish Perspective' (2020) 9 *Internet Policy Review* 1

Pettit B and Western B, 'Mass Imprisonment and the Life Course: Race and Class Inequality in U.S. Incarceration' (2004) 69 *American Sociological Review* 151

Pham M, Rajić A, Greig J, Sargeant J, Papadopoulos A, McEwen S, 'A Scoping Review Of Scoping Reviews: Advancing The Approach And Enhancing The Consistency' (2014) 5 *Research Synthesis Methods* 371

Powell-Smith A and Leibowitz G, 'Data and Statistical Gaps in Criminal Justice', (Center for Public Data, March 28 2023)

— — Unrepresented Defendants in the Magistrates' Courts, (*The Center for Public Data*, 2023)

<<https://static1.squarespace.com/static/5ee7a7d964aeed7e5c507900/t/641ae6bdbd6ed448d43e09e9/1679484606127/CFPD+Legal+Representation+Data+Gaps.pdf>>

accessed 11 July 2023

Powell-Smith A, 'Stop and Account: How is this Little-Known Police Practice Being Used?' (The Centre for Public Data, 2019) <<https://missingnumbers.org/stop-and-account/>> accessed 20 May 2023

— — ‘Highlighting Data Gaps in Criminal Justice’, (Justice Lab, 30 March 2023)  
<<https://justicelab.org.uk/highlighting-data-gaps-in-criminal-justice/>> accessed 10 August 2023

Prescott C, ‘Internet users in the UK: 2017’ (*Office for National Statistics*, 19 May 2017)  
<<https://www.ons.gov.uk/businessindustryandtrade/itandinternetindustry/bulletins/internetusers/2017#:~:text=In%202017%2C%20almost%20all%20adults,20%25%20were%20re%20cent%20internet%20users.>>> accessed 10 July 2023

Pyle O, ‘Report on the s163 RTA Ethnicity Recording Pilot’ (*The Metropolitan Police*, 17 May 2022) < [https://www.met.police.uk/SysSiteAssets/foi-media/metropolitan-police/disclosure\\_2022/november\\_2022/section163-rta-stops-recording-ethnic-background-stopped-drivers.pdf](https://www.met.police.uk/SysSiteAssets/foi-media/metropolitan-police/disclosure_2022/november_2022/section163-rta-stops-recording-ethnic-background-stopped-drivers.pdf)> accessed 10 July 2023

Question from Christina Rees MP to the Ministry of Justice, June 2016, < <https://questions-statements.parliament.uk/written-questions/detail/2016-06-14/40571/>> accessed 3 July 2023.

Question from Gareth Bacon MP to the Mayor of London, Sadiq Khan, during Mayor’s Question Time (12 September 2019) < <https://www.london.gov.uk/who-we-are/what-london-assembly-does/questions-mayor/find-an-answer/tfl-tracking-3>> accessed 10 July 2023

Rashid I, ‘Labour MP Dawn Butler Accuses Metropolitan Police Of Racial Profiling After Being Stopped By Officers’ *Sky News* (10 August 2020) < <https://news.sky.com/story/labour-mp-dawn-butler-accuses-metropolitan-police-of-racial-profiling-after-being-stopped-by-officers-12045860>> accessed 10 July 2023

Response to FOI request from Aoife O’Reilly to the Department of Work and Pensions (27 June 2022)  
<[https://www.whatdotheyknow.com/request/864638/response/2068623/attach/3/Response%20FOI2022%2041308.pdf?cookie\\_passthrough=1](https://www.whatdotheyknow.com/request/864638/response/2068623/attach/3/Response%20FOI2022%2041308.pdf?cookie_passthrough=1)> accessed 3 July 2023.

Response to FOI request from Louisa James to Hampshire and Isle of Wight Constabulary (18 January 2023)



<[https://www.whatdotheyknow.com/request/number\\_of\\_police\\_officers\\_who\\_ha\\_19#comment-111323](https://www.whatdotheyknow.com/request/number_of_police_officers_who_ha_19#comment-111323)> accessed 12 July 2023

Response to FOI request from M Brookes to the Metropolitan Police (20 November 2019) <[https://www.whatdotheyknow.com/request/vetting\\_statistics](https://www.whatdotheyknow.com/request/vetting_statistics)> accessed 12 July 2023

Response to FOI request from Robert Jones to Ministry of Justice (22 June 2018) <[https://www.whatdotheyknow.com/request/490875/response/1177296/attach/5/FOI%20180612015%20reply.pdf?cookie\\_passthrough=1](https://www.whatdotheyknow.com/request/490875/response/1177296/attach/5/FOI%20180612015%20reply.pdf?cookie_passthrough=1)> accessed 12 July 2023

Response to Robert Jones' Freedom of Information Act, 2018 [Rt Hon Michael Gove MP letterhead \(whatdotheyknow.com\)](#)

Rowe M, and Pearson G, 'We spent seven years observing UK police stop and search - here's what we found' (*University of Liverpool*, 2020)

<<://news.liverpool.ac.uk/2020/11/16/we-spent-seven-years-observing-uk-police-stop-and-search-heres-what-we-found/>>> accessed 18 April 2023

Ruijter E, Porumbescu G, Porter R, Piotrowski S, 'Social equity in the data era: A systematic literature review of data-driven public service research' (2023) 83 *Public Administration Review* 316

Sentell T, da Camara S, Ylli A, Velez M, Domingues M, Bassani D, Guo M, Pirkle C, 'Data Gaps In Adolescent Fertility Surveillance In Middle-Income Countries In Latin America And South Eastern Europe: Barriers To Evidence-Based Health Promotion' (2019) 11 *South Eastern European Journal Of Public Health* 214

Stripe N, 'User Guide to Crime Statistics for England and Wales; March 2020' (Office for National Statistics: Crime and Justice, 4 November 2021)

<<https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/methodologies/userguidetocrimestatisticsforenglandandwales>> accessed 18 May 2023

Sturge G, Court Statistics for England and Wales (*House of Commons Library*, 31 January 2023) <<https://commonslibrary.parliament.uk/research-briefings/cbp-8372/>> accessed 18 May 2023

Suominen A and Hajikhani A, 'Research themes in big data analytics for policymaking: Insights from a mixed-methods systematic literature review' (2021) 13 Policy and Internet 464

Scott S, 'A Sociology of Nothing: Understanding the Unmarked' (2017) 52 Sociology 3

Sweeney S, and Matthews Z, 'Friends, Families and Travellers: A Guide for Professionals Working with Gypsies, Roma and Travellers in Children's Services' (*Friends, Families and Travellers*, January 2017) < <https://www.gypsy-traveller.org/wp-content/uploads/2017/03/A-guide-for-professionals-working-with-Gypsies-and-Travellers-in-the-public-care-system.pdf>> accessed 10 July 2023

Taylor L, 'What is Data Justice? The Case for Connecting Digital Rights and Freedoms Globally' (2017) Big Data & Society

<<https://journals.sagepub.com/doi/pdf/10.1177/2053951717736335>> accessed 10 July 2023

Tomlinson J, Meers J, Sommers-Joce C, 'Judicial Review of Public Data Gaps' (2023) Judicial Review <https://www.tandfonline.com/doi/full/10.1080/10854681.2023.2218256> accessed 13 July 2023

Treadwell H, Ro M, Sallad L, McCray E, Franklin C, 'Discerning Disparities: The Data Gap' (2018) 13 American Journal Of Mens Health <<https://doi.org/10.1177/1557988318807098>> accessed 10 July 2023

Unangst L, Casellas Connors I, Barone N, 'State-Based Policy Supports for Refugee, Asylee, and TPS-Background Students in US Higher Education' (2022) 38 Refuge: Canada's Journal on Refugees 95

Van der Windt P, and Humphreys M, 'Crowdsourcing in Eastern Congo: Using Cell Phones to Collect Conflict Events Data in Real Time' (2016) 60 JOURNAL OF CONFLICT RESOLUTION 748

Vest J and Issel L.M, 'Data Sharing Between Local Health And State Health Departments: Developing And Describing A Typology Of Data Gaps' (2013) 19 Journal Of Public Health Management And Practice 357

Walby S, Towers J, Francis B, 'Mainstreaming Domestic and Gender-Based Violence into Sociology and the Criminology of Violence' (2014) 62 The Sociological Review 187

Wheeler H, "Transforming for a Digital Future: Government's 2022-25 Roadmap for Digital and Data" (HM Government Central Digital and Data Office, 2022)

<[https://digitalanddata.campaign.gov.uk/#government\\_s\\_digital\\_and\\_data\\_strategy\\_explained](https://digitalanddata.campaign.gov.uk/#government_s_digital_and_data_strategy_explained)> accessed 10 July 2023

Wheeler H, 'Data Sharing Governance Framework' (HM Government Central Digital & Data Office, 23 May 2022) < <https://www.gov.uk/government/publications/data-sharing-governance-framework/data-sharing-governance-framework>> accessed 11 July 2023

Williams B.A, Brooks C.F, Shmargad Y, 'How Algorithms Discriminate Based on Data They Lack: Challenges, Solutions, and Policy Implications' (2018) 8 Journal of Information Policy 78

Wright K, Tapera R, Stott N.S, Sorphage A, Mackey A, Williams S.A, 'Indigenous health equity in health register ascertainment and data quality: a narrative review' (2022) 21 International Journal for Equity in Health 34