A Model-Driven Architecture for Enterprise Document Management, Supporting Discovery and Reuse

David John Small

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds

School of Computer Studies

September 1999

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgements

To my parents and my Father. From day one, I thank you.

I would like to thank my supervisor Professor Peter Dew for his encouragement and guidance throughout the course of the work described in this thesis. I would also like to thank the many people in the School of Computer Studies who have made the past years so enjoyable. Specifically, for their friendship and support, Martin Thompson, Derek Sergeant, Suniel Curtis, Mourad Kara and John Davy, and for their guidance Richard Drew, Neil Hunter and Chris Hemingway. And for all the years and prayers, Dr Elizabeth Atherton - thank you for you.

For support and encouragement from start to end a big strictly-alphabetical thank-you to Andy Bolton, Darren Boyce, Tim Day, Shaun Hennessy, Ben Hughes, Dave Mason, Jenny Morris, Becca Mullen, Mark Reynolds, Ann Roberts, Doug Stowell, Neil Sumpter, Andy Toon and David Trice.

For assistance at the outset of this research, and again at the end, I would like to thank Paul Clark from the LSC Group (Tamworth, UK).

Finally, to Ruth. Too much laughter, altogether. Thank you for everything. 'Finished!'

Abstract

The work described in this thesis investigates the application of enterprise document management techniques to better support document discovery and reuse across a learning organisation. It examines the nature of documents and organisations, and the roles that documentation plays within modern enterprises. The thesis recognises the increasing need to reuse documents at a number of levels of abstraction, so leveraging an enterprise's valuable document resources. Enabled by Internet developments and document mark-up languages, particularly the extensible mark-up language (XML) and its associated standards, this thesis describes the development of ideas that have led to the design and initial implementation of the Model-driven Reuse Architecture (MRA).

The concepts underlying the design of the MRA have developed from an initial feasibility study undertaken with the United Kingdom Ministry of Defence (MoD) to provide management techniques for military support documentation. A range of current approaches to document management and enterprise-wide information management have added to the initial reuse approach identified by the MoD. A document-model based approach to document management permits access to and reuse of document fragments for new purposes. An extensible mechanism permits the inclusion of other document approaches to facilitate document discovery, potentially leading to reuse.

This thesis describes the approach taken in designing a new document architecture that provides a work-focussed approach to document management, encouraging social collaboration rather than automation, and supporting a range of users within the same environment. This thesis presents the overall Model-driven Reuse Architecture and a preliminary implementation that has been developed to support the specific needs of teaching and learning in higher education. The resulting web-based implementation, MRA-HE, is evaluated in terms of how it performs against a set of realistic scenarios within the domain of higher education. Upon evaluating the MRA-HE implementation, the final chapters of this thesis generalise some of the MRA approaches to the diverse demands of other types of organisation.

Contents

1 Introduction

1.1 Context and Motivation

Documents and their management have been central to the development of human history for hundreds of years. In one of its more general senses, a document is anything serving as a representation of a person's thinking by means of symbolic marks[1]. The thinking that is conveyed through a document, if accepted by the reader, becomes part of the reader's understanding. In this way the abstract use of symbols arranged on a communicative medium, whether painted onto cave walls, hand-written on paper or word-processed on-screen, conveys a fact, event or belief from the author to the reader. If the reader has sufficient belief in the communicated meaning, that meaning can serve to develop more substantial beliefs and understandings. The Latin origin of 'document' (docere: to teach) confirms the use of documents as aids to develop ideas and convey information for elaboration and reuse.

In all sectors of society documents play a number of complex roles, such as evidential, informational, communicative, contextual and argumentative. These roles are more fully expanded in chapter 2. The management of documents within an organisation is increasingly critical to its sustenance and growth, particularly for organisations engaged in the Knowledge Economy [Demarest97]. Thus the need for managing and re-using documents is of increasing importance to organisations. The following demands can be identified:

- **Increased variation of products/services** In competitive product and service sectors there is an ever-increasing demand for more customisable products and continual improvements across product ranges. Most products, except for the most innovative, are variations on existing products - design is inherently iterative[2]. Product or service designs use parts of existing products or services; it stands to reason that reuse of the associated documentation is not only viable, but also desirable.

- **Distributed working** Corporations are becoming increasingly complex and geographically separated. With regard to documents, this creates two demands: (i) appropriate management of existing documentation, and (ii) awareness of the increased amounts of documentation required for effective distributed working. A virtual enterprise (VE - DOD96) provides further

---

[1] Wordnet 1.6 - http://www.cogsci.princeton.edu/~wn/ (Last verified 8th March 2000)

[2] Learning by Design Concept Index, Georgia Tech - http://www.cc.gatech.edu/edutech/LBD/iterative-process.html (Last verified 8th March 2000)

dimension to this problem due to its largely independent business entities, each with its own business processes and culture.

- **The cost and limitations of paper** Traditionally paper documents have served to keep people informed across an organisation. According to Lexmark, paper costs for a particular business can amount to between 6% and 15% of a company's total expenditure; 98% of business information is held on paper; each paper document within an organisation is photocopied 19 times on average [Richardson96]. Where an organisation is geographically distributed the costs of printing and distribution are exacerbated. Paper documents have two associated problems: the cost of production and distribution, and the problem of currency - how best can an organisation keep a group of people up-to-date with the latest information? The static nature of paper documents often hinders access to the most recent and most useful information. Electronic document management is changing this situation, but brings its own problems such as version and access control.

- **Changes in working patterns** Modern organisations face constant change at all levels of the enterprise - local and enterprise-wide change to increase productivity, and enterprise-wide change to respond to new customer demands and new governmental legislation. It is clearly desirable to be able to adapt information systems to reflect changes within the organisation. Furthermore, frequent personnel changes call for mechanisms to retain the organisation's intellectual capital within the organisation [Branting96]. There exists a complex relationship between an organisation as a whole and its members, which is discussed further in chapter 2. Much of the knowledge that people have is said to be tacit: not explicitly recognised or codified. It is a challenging task to make explicit the tacit knowledge that people have about their role and responsibility within an organisation. Smaller organisations are equally concerned with transforming tacit knowledge into explicit knowledge to protect themselves from dependency on specific personnel. As the implicit becomes explicit documentation results, increasing demand on document management techniques.

- **Long-term storage** Many products have long lifecycles (e.g. aircraft and defence equipment). In the UK the Ministry of Defence aims to support the documentation of defence equipment with long product lifecycles (see chapter 3). Brooks describes the need within the telecommunications industry to maintain product documentation over long lifecycles, typically 20 years [Brooks94]. When a product-support document is printed on paper its life span is potentially much longer than an equivalent electronic version due to its persistence and immutability. Two significant problems exist with electronic documentation: the method by which the document is digitally encoded (the encoding model) and the method by which that digital representation is stored (the storage model). These issues are different to those of conventional paper documents

- **Increased desire to manage knowledge within the enterprise** Complex organisations are composed of many diverse, interdependent work groups, all of which have unique decision domains [Schwartz98]. The development of digital information networks is enabling the integration of work across and between these diverse groups. A learning organisation is characterised as a decentralised organisation able to continually capture 'best practice' and disseminate that practice across company boundaries. Closely associated with this is the Organisational Memory Problem which can be described as:

  > *"Information pertinent to the task at hand has passed through the organization. Was this information captured? If so, how can it be efficiently retrieved and brought to bear on the present task?"* (cited in [Schwartz98])

  There is a focus not only on collecting facts but retaining the 'lessons learned' within and between these facts. Organisation memories are motivated by the desire to preserve and share the knowledge and experiences that reside in an organisation [Sumner98]. An organisational memory is defined as an explicit, disembodied, persistent representation of knowledge and information in an organisation [vanHeijst96].

It is worth noting that the demands described above apply to documents and more structured information alike. Although the power-to-cost ratio of computing hardware has multiplied by a factor of a million in the past 30 years, the cost of organising the information within those systems has increased [Goranson92]. There is an evident need to improve the management of digital information of all types. However, according to the Gilbane Report on Open Information and Document Systems (cited in [Reinhardt94]) at least 80% of corporate electronic information is in the form of documents, as opposed to structured database records. This statistic highlights the need to focus at least as much on the management of documents as on the management of more structured information within an enterprise.

The demands on document management can be met by:

- **technological advances** in computer hardware, software and modelling languages
- the **development of user communities** that utilise those technological advances to their advantage.

The technological opportunities are similar to those that have made possible the creation and meteoric rise of the World Wide Web and corporate intranets. New technologies such as XML have the potential to raise the value of transmitted information by retaining a semantic around the content currently conveyed through HTML. The development of user communities is a more subtle but no less definite advance in recent years, underpinned by international data standards and globalisation efforts. The result is a useful tension between localisation and globalisation that maintains an organisation's outward-looking nature while striving to support current work practices. Although the development of the World Wide Web may be regarded as being driven by technology, the original WWW proposal was intended to promote interaction among researchers at CERN [Bass98] - the user community developed around the enabling technologies.

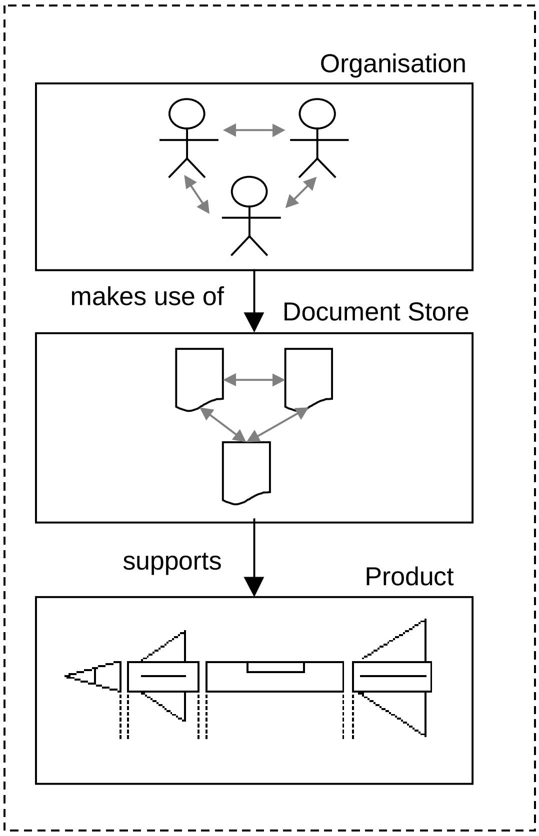


Figure 1-1 The Role of Documents in Supporting the Product Lifecycle

Figure 1-1 considers an organisation that makes use of documents to support a product through its lifecycle, illustrating the range of research and commercial communities that address certain parts of the overall document management problem. The **document management** community is concerned with reducing paper and performing storage and retrieval on electronic documents more efficiently. The **hypertext** community is concerned with the way in which electronic documents refer to each other, how documents group to form collections and how multimedia is able to extend more traditional document types.

Groupware communities are concerned with distributing electronic documentation to the right place at the right time, and **collaboration services** support shared working on those documents. The **CALS and ILS communities** (discussed in chapter 3) consider the ways in which documentation supports the lifecycle of the product or service that the organisation exists to provide. **Person-centred approaches** support the people within the organisations; developed systems support the individual and assist in collaboration between two or more users of the system. **Model-driven approaches** emphasise the changing nature of the organisation and the work that it performs; they provide an architecture within which the models change according to changes in the real world.

Each of the above approaches have developed independently in the literature; more recently projects such as MEMOIR [Pikrakis97] have combined two or more of the approaches to create more sophisticated systems. What is lacking, however, is an overall framework into which these approaches can be placed and used together. Furthermore, it seems that the furore over document management has led many to neglect one valuable potential benefit from effective document management - document reuse. This thesis addresses these issues.

1.1.1 Integrated Logistics Support

One attempt to satisfy the demands outlined in the previous section is the Integrated Logistics Support (ILS) program. In the late 1970s the American Department of Defence (DoD) realised that the cost of support of its defence equipment was often greater than the original cost of procurement. The DoD set out to optimise life cycle costs (LCC) across its range of defence products rather than concentrate on minimising procurement costs as it traditionally had done. What resulted was the ILS management approach which used international standards available at that time to implement its principles. One part of ILS describes the recommended management method for technical documentation written to support the defence equipment through its life cycle. This thesis concentrates only on the document management aspects of through-life support.

ILS aims to satisfy four of the six demands described in the previous section - increased variation of products/services, distributed working, the cost and limitations of paper and long-term storage. The principle of ILS document management is the division of large document volumes into smaller document modules (or data modules), and the subsequent unique classification of each data module. Defence products are decomposed into their constituent parts and each part is given a unique code. Every data module is classified according to which piece of equipment is describes and what operation on that equipment it describes. A set of data modules, individually stored, is retrieved according to some need and built into a publication for some purpose. One of the primary aims of ILS is to assist the reuse of single data modules across a range of publications; storing a piece of information once and using it often.

Despite a lack of initial success from the DoD, ILS principles were mandated by the UK Ministry of Defence (MoD) in 1996. The MoD approach is encapsulated in DEF STAN 00-60 (DS0060) [MoD99]. Following a successful implementation of the documentation recommendations within DS0060, described in chapter 3, trials are currently underway with the UK military armed forces. There are many strengths to the recommendations within DS0060 and a number of these recommendations form the basis for the research reporting in this thesis.

It is important to note the nature of the organisation within which DS0060 was created. The defence industry is a highly structured product-based enterprise. The MoD procures and maintains low-volume specialist equipment with high variation across its range. It is of utmost importance that information is fully audited, tracked and secure. Though powerful within the defence domain, ILS techniques are not well suited to managing the documents within lesser-structured enterprise environments. The main weakness in applying ILS to general-purpose document management is the strict classification required for each data module. A data module is often as small as a single paragraph of text; its subsequent classification has to be undertaken by a domain expert and each classification is inherently subjective. Moreover the reuse expectations from DS0060 have not yet been realised, and it is doubtful whether many commercial organisations would adopt the approach, except for those that manufacture similarly complex high-value, low-volume products.

1.1.2 Enterprise-Wide Document Management

The commercial success of current enterprise-wide document management systems is indicative of the pressing need for document management within large enterprises. In order to more effectively share documents, individual departments introduced local area networks (LANs) in the1980s, allowing users to share a common file space. Groupware products such as Lotus Notes have facilitated the flow of information more effectively around the larger enterprise, but the form of the information is more database-like than document-like. More recent document management products such as Excalibur and Autonomy have started to target lesser-structured documents, providing search across entire heterogeneous enterprise document spaces. Although semantically powerful, such information retrieval systems neglect the significance of the document to the organisation, knowing little about the document's reason for existence.

An effective enterprise-wide document management solution must consider a number of different approaches that can be used together to satisfy the demands identified earlier in this chapter. Three such approaches are workflow and collaboration ventures, the increased use of hypertext and mark-up languages, and the application of web technologies to the document management problem.

Workflow and Collaboration

Many information systems concentrate on the informational aspects of the work environment into which they are used. Much of the document management literature is concerned with the ways in which the informational content of documents can be captured and manipulated. Within the workplace many of these systems implicitly support workers in their operational and decision-making roles. Computer Supported Co-operative Work (CSCW) approaches make explicit the role that information systems accomplish for their users. [Swaby98] recognises three general motivating requirements that the CSCW discipline seeks to address:

- **Articulating co-operative work** Co-ordinating people and resources in contributing towards the performance of a common task

- **Sharing an information space** Ensuring that group members can share data, information, concepts and heuristics in a structured way

- **Adapting technology to the organisation** Creating an organisational context within which group activities may be situated, and enabling appropriate interactions within this setting.

This thesis does not replicate the review of CSCW that can be found in [Swaby98], but does acknowledge and support the need for a workflow approach to information systems. Workflow products such as Lotus Notes assist the flow of information around a distributed organisation by holding a model of the working processes of the end-users.

The term 'collaboration' traditionally suggests a level of co-operative work on a common goal. Within the context of this thesis the common goal is the construction and maintenance of a collection of document resources that satisfies a set of organisational goals. Tools that support collaboration fall into two broad categories:

- Tools to **close known distances** between distributed workers to enable joint work focussed on a common purpose.

- Tools that assist the **recognition of groups of people**, who may not otherwise know of each other's existence, in the hope that collaboration between these people will result in increased productivity or effectiveness.

Both types of collaboration tools are enabled by network-centric computing but extend beyond simple distributed access to resources. Traditionally IT has been applied to support individual activity, and co-ordination has been external to the IT used. [Greenwood95]. The first class of tools, those that close known distances between workers, is exemplified by shared workspaces and conferencing systems. One key requirement for such collaboration is the guaranteed integrity of a shared resource at all times. Products such as Lotus Notes have provided such mechanisms for some years, but in a proprietary format and within a restricted network. WebDAV [Whitehead98] is a current development that extends the WWW to provide a standard infrastructure for asynchronous collaborative authoring across the Internet.

The second category of collaborative tools emphasises the identification of potential collaborators by providing mechanisms to identify people with whom collaboration may be advantageous. The Virtual Science Park and the MEMOIR system (University of Southampton) provide explicit support for their users in discovering other users with identifiable skills or interests. The identification of suitable collaborative partners is a separate enabling process to working collaboratively, which may follow. This thesis describes an architecture that assists with both the second domain, supporting the identification of relevant resources and their owners, and the first domain, providing support tools for subsequent document reuse or co-authorship.

Hypertext and Mark-up Languages

The field of hypertext has existed since 1945 when Vannevar Bush created a machine that associated (linked) two documents, similar to the way that the mind operates by association [Bush45]. As a working definition, a hypertext is a set of nodes of information with machine-supported typed links between these nodes and a common user interface [Cunningham93]. Hypertext expresses the relationships between documents. The first electronic hypertext prototype, NLS/Augment, was demonstrated in 1968. Many of the fundamental principles characterised by today's hypermedia systems were already present in the initial prototype. The creator, Doug Engelbert, even then saw the potential for computer-support to boost the collective IQ of communities or the collective intelligence of organisations [Balasubramanian94].

The World Wide Web has popularised the notion of hypertext, with its ability to jump from one part of a document to another part of another document or to another part of the same document. The cognitive overhead of following a link is less than that of performing a search across a document space to find suitable resources. [Golovchinsky97] suggests that link-driven queries are more effective than user-specified queries in retrieving relevant information. The mechanic of hyperlinking is powerful but the quality of the link is dependent on the person who created the link. There are additional problems with hyperlinking, discussed in chapter 2, but the notion of hypertext is of great potential benefit if used appropriately.

In 1969 IBM developed the Generalised Mark-up Language (GML), intended from the outset to separate form from content in displaying documents [Berghel99]. In 1986 the Standard Generalised Mark-up

Language (SGML) became an international (ISO) standard for document mark-up. During the development of commercial word processing machines, however, designers concentrated on developing machines that worked within the limits of the hardware available at the time. The codification of the prepared document was incidental, controlled more by the amount of storage available than the clarity of the coded file.

The past two decades have seen a rapid growth within the word processing market; increasing functionality and improved user interfaces. Conventional word processors, however, use proprietary binary compressed file formats that only particular versions of a particular word processor can understand. To overcome the problem of translation between word processing packages software filters have been developed. Unfortunately these software filters often lose some of the descriptive power of the original document file. The prevalence of file formats and filters is an obstacle to groups that may want to share documents electronically but are hindered from doing so by incompatibility problems.

The current popularity and simplicity of the HyperText Mark-up Language (HTML) has introduced the general public to some of the concepts behind more powerful mark-up languages. The eXtensible Mark-up Language (XML), a simplification of SGML, is a platform-independent domain-independent method for marking up documents. Its aim is the same as for GML in the 1960s: to separate document content from document form. This concept is more fully explored in chapter 5. It is worth noting that although mark-up languages were first created to mark up document content, their use has diversified to represent simple data models as well as document models. Chapter 2 explains how the distinctions between data and documents are not clear-cut; the remainder of this thesis shows how mark-up languages can be used to mark up both documents and underlying data models.

Web Technologies

The first recognised network of computers, ARPANET, was formed in December 1969. It consisted of four nodes and was intended to decentralise military information and control. More by accident than design, the users of ARPANET began to use the network for more than the intended long-distance computing. Researchers began to use ARPANET to collaborate on projects, to trade notes and to send messages of a more social nature. There was no dependence on the type of computer used to connect to ARPANET; the only stipulation was the network protocol used to communicate. The ARPANET formed the beginnings of the Internet as it is now: a diverse collection of computers able to communicate with any other via common protocols and a widespread communication network.

Client/server computing can be thought of as 'computer-process' to 'computer-process' communication in which one process is the requestor and the other process services the request, usually across a network, using a commonly agreed upon protocol [Schwartz95]. ARPANET was conceived from the outset to implement a client-server architecture, and the modern World Wide Web is based on the same concept. The decentralisation of computing power and information enables the possibility of cross-organisation document management: local retention of 'owned' documents and global access to other documents within the organisation.

Motivated by the need to provide access to information stored on a diverse range of server technologies, network-centric computing has recently evolved, enabled by the client/server architecture and open standards. Network-centric computing is a server-centric software architecture based on open Internet standards and client independence. It provides central control over data with distributed access to that data via standards-based solutions. Network-centric computing extends the client/server architecture by providing open networks and protocols across which servers can carry data to their distributed users.

The unprecedented success of the World Wide Web (WWW) has prompted companies to introduce web technologies to their organisational infrastructure, terming the resulting networks 'intranets'. The standard network protocol and mark-up language of the WWW provide a lightweight solution to many information distribution problems, with minimum disruption across the organisation. A standard browser client is the sole requirement at the desktop, with the complexities being managed centrally at the server. However, due to the lack of semantic structure within the mark-up language currently used across the WWW (HTML), intranets have tended towards Web-based Information Systems (WIS) rather than Web-based Document Management Systems. The server's role has become to codify and disentangle information flowing between the system and the user rather than add value. It is apparent that current web technologies can not fully support enterprise-wide document management [Balasubramanian98].

Systems such as Excalibur, Autonomy and Lotus Notes are internet-enabled but not internet-driven [Bentley97]. It is evident that a convergence of document management techniques and WWW-like distribution would be advantageous in facilitating an enterprise-wide view of documents.

1.1.3 Extended Enterprises and Virtual Communities

The broad geographical reach of the Internet has advanced understanding on how disparate groups of various sizes may join to form communities and advance thought and knowledge within those communities. Usenet news groups facilitate virtual forums based around specific subject areas, allowing knowledge to flow between people who need never meet, and who need never be in direct synchronous contact with each other. Such information forums have developed informally as interests develop and diminish, influenced by wider networks and a desire for many people to share information and develop ideas. More recently, organisations have started to use the same network technologies to add structure within and between their distributed divisions and 'friendly' customer bases.

Two radically different approaches have developed in an attempt to control the terminologies used by participating members, to facilitate accurate transmission of information across traditionally disparate communities. The first is to develop a common terminology that encompasses the larger shared domain: all participants in this larger domain agree to use the new terminology, which may differ significantly from all local terminologies. The second approach is to allow all participating groups to use their own terminology internally, but to provide mappings at organisational boundaries to allow transmission of information across those boundaries. This thesis argues that the second of these is more valid. [Sumner98] suggests that supporting communities to evolve their own vocabularies and to elaborate them to create a shared domain model is a critical step towards creating a common perspective. Over time, as communities engage in negotiation and reflection about how to do their job better, their vocabularies and domain models become more elaborate and formal. This highlights the evolutionary nature of the interactions between user communities. The challenge is not representing the overall user domain, but permitting the negotiation and enrichment of the models over time. Organisational learning focuses on codifying and storing lessons learned ('tacit knowledge' - [vanHeijst96]) within the organisation.

1.2 Research Problems

The specific problems addressed in this research first arose through a project with the MoD who identified a need to store documentation pertinent to lifecycle support for defence equipment. Integrated Logistics Support (ILS) principles (section 1.1.1) support defence equipment through the initial phases of its design, in the hope to reduce lifecycle costs. DEF STAN 00-60 (DS0060) is a set of recommendations that implements a solution to the ILS requirements. The intentions and some ideas from DS0060 are sound, and have been proven in the short term by some successful implementations. However, these implementations have highlighted two weaknesses to the DS0060 recommendations. First, DS0060 does not provide direct support for product or process change within the organisation. Change is currently managed externally to the DS0060 architecture and mapped into the architecture as required. Further, the aim of reusing document parts has not yet been realised, although it is thought that reuse will become more evident when the scope of implementation broadens. In practice, the DS0060 notion of reuse is restricted to reusing a particular data module in more than one publication; the literature highlights much richer possibilities for reuse.

This thesis proposes a new document architecture - the Model-driven Reuse Architecture - that extends the ideas in DS0060 to manage change and support reuse within an organisation. Two domains are addressed: (i) MoD support for the product lifecycle and (ii) computer-supported learning at the University of Leeds. The University is a large enterprise with a need to better manage its intellectual resources. Currently course material management within the University of Leeds remains at the discretion of the individual schools that deliver taught courses. One of the demands described earlier in this chapter is of particular relevance within the University, namely the increasing variation of products/services. [Lincoln97] and [Zarhan94] describe the increasing need for more adaptable courses that better suit the learning requirements of students. This requirement suggests the reuse of existing materials to suit different purposes across the teaching domain. For successful reuse the user inevitably requires access to a broad range of resources and a range of discovery mechanisms to support the reuse process. There is further perceived value in discovering not only documents but also their owners, enabling collaboration between two users with similar interests and expertise. The Virtual Science Park provides such mechanisms for structured expertise; this thesis applies the same ideas to lesser-structured document resources.

The context for the research in the document management domain falls into two areas:

- **finding potential collaborators** through the documents associated with a particular person
- **reusing existing resources** to satisfy new and emerging demands

Both problems require solutions that are able to adapt to the continual change within the organisations they service.

1.2.1 Overview of the Model-driven Reuse Architecture

The Model-driven Reuse Architecture (MRA) provides a range of document management support for structured documents[3] that serve a valued purpose within an organisation. It supports localised groups of users and their work purposes while providing cross-enterprise document management where required. It further supports change within the enterprise, driving functionality from a set of manipulable business models. The impact of such an architecture is measured in terms of its ability to improve the management of an enterprise's document resources and its subsequent facilitation of resource re-use. Whereas DS0060 provides support for product and process, this research generalises the recommendations to further facilitate links between documents and the organisation within which they are used. The overall architecture provides an extensible model, enabling incorporation of current document management and hypertext techniques. Its focus is on supporting the reuse and repurposing of document resources across organisations within the context of a model-driven architecture. Success of the MRA is measured in terms of:

- the architecture's ability to **better manage an enterprise's document resources**, linking resources to the changing business models

- the architecture's ability to **facilitate the discovery and reuse** of document parts for a new purpose.

[3] This thesis argues in section 2.2.2 that all documents contain an intrinsic structure. To emphasise this structure the MRA is said to support 'structured documents' – those documents that have a discernible structure at any level of granularity.

1.2.2 Research Objectives

1. To study model-based architectural approaches to document management
2. To propose and analyse a new model-based architecture for enterprise document management that supports document reuse through
   - document discovery using a number of approaches
   - mechanisms to promote the reuse of documents at a number of levels
3. To build and evaluate an experimental version of the architecture

1.2.3 Contribution

This thesis describes the development of ideas founded in the UK Ministry of Defence's DEF STAN 00-60 (DS0060) document architecture. Its contribution begins with the first technical proof of concept of the theoretical and unproven DS0060 proposals. Using DS0060 as a base platform this thesis investigates current requirements from document management, placing emphasis on the need to enable and further understand document reuse within ever-changing learning organisations. The thesis proposes the design of a novel document architecture that synthesises current document management practices and introduces methods for supporting much-needed document reuse at three levels of abstraction.

1.3 Thesis Structure

The thesis is organised into the following chapters:

- Chapter 2 introduces the concepts of documents, the organisations they serve and the varied ways in which they are managed by existing document management solutions and projects.
- Chapter 3 describes the DEF STAN 00-60 document architecture created by the UK Ministry of Defence to manage support documentation for their military products. It outlines the architecture itself, a feasibility study undertaken with the LSC Group (UK) and a number of ways in which such an architecture may be extended to support other domains.
- Chapter 4 presents the design of the Model-driven Reuse Architecture (MRA) - a new document architecture based on DEF STAN 00-60 and extended to support document discovery mechanisms, hypertext and document reuse.
- Chapter 5 presents an implementation of the architecture defined in chapter 4 within the domain of supporting learning in higher education.
- Chapter 6 presents an evaluation of the MRA-HE implementation in terms of its fulfilment of scenario critical success factors, utility and usability.
- Chapter 7 reviews the work in this thesis and concludes with ways in which the MRA could be extended with future work.

2 Related Work

Chapter 1 identified an increasing need for organisations to more effectively manage and reuse their document resources across the enterprise. The purpose of this chapter is to provide an overview of the role of documents within communities and in particular organisations. It describes a number of ways in which documents are managed and valued, and a number of techniques and products that contribute towards an effective document management architecture.

2.1 Documents and Learning Organisations

Whether in digital or hard copy form, documents are more complex than data records in both their form and purpose. They are, according to Xerox, 'the vehicles within which information is created, structured, communicated and preserved'[4]. Traditionally organisations have viewed documentation as a cost item, necessary for successful internal operation and customer information. With increasingly competitive markets, however, information is becoming viewed as an enterprise asset rather than a cost item [DOD96]. Documents have been identified as a central mechanism for managing, communicating and contracting across an organisation [Amor97], and the quality of that documentation has become a key differentiating factor in the overall success of a number of commercial products [Brooks94]. Accordingly, the management of such documentation has been raised in importance within a great number of diverse organisations.

The physical embodiment of a document, whether on-screen or hard copy, provides little value unless its meaning is understood. The meaning extracted from a physical document is of little value unless there is a context within which to place it. In order to add value a document must fulfil a requirement or purpose within the organisation in which that document is located. As Mark Fox states in his microtheory of resources [Fox92], being a resource is not an innate property of an object, but is a property that is derived from the role an entity plays with respect to an activity. In short, documents support roles. When considering the qualities of effective document management, [Davies98b] argues that *'knowledge is information transformed into a capability for effective action'*. The value of documentation is only realised when it delivers a meaning that in turn creates an action or effect.

Although documents are often written to support a current organisational need, documents may also contain information that is valuable in the long-term: the area of organisational learning denotes a longer-term value of documents within organisations. Organisational learning methods advocate a disembodiment of

---

[4] The Document Cycle, Xerox - http://www.xerox-emea.com/aboutdoc/doccycle.htm (Last verified 8th March 2000)

learning experiences from the people who have those experiences, focussing on the organisation as the holder of such knowledge. Organisation memory is characterised as follows:

> *"Information pertinent to the task at hand has passed through the organization. Was this information captured? If so, how can it be efficiently retrieved and brought to bear on the present task?"*
>
> [Schwartz98]

Organisational memory is developed through the process of organisational learning. Koulopoulos's definition of a document as 'a collection of information, authored for the purpose of transferring and preserving knowledge' [Koulopoulos95] suggests that documents can clearly assist in making explicit the implicit organisational memory.

In order to help construct and maintain an organisational memory it is clear that document resources must be classified in a way that will render them useful for current and future needs. Libraries have for many years provided a classification service to their users, creating schemes for all types of documents and books. The library community has considerable experience in organising collections of documents using a coherent classification, and has much to offer to current document management problems. Digital libraries represent a convergence of computing and library services, focussing on the use of information technology to support roles within the library. The focus of much digital libraries work lies outside the scope of this thesis, but David Levy provided a critique of ongoing work in digital libraries in 1995 [Levy95]. He claimed that at the time digital libraries had a number of assumed characteristics, which he viewed as mistaken:

- Digital library collections contain **fixed permanent documents**.
- Digital libraries are based wholly on **digital technologies**.
- Digital libraries are to be **used by individuals** working alone.

Levy claimed that the initial focus on fixed, permanent materials could be traced to a preoccupation with books as the central elements and the organising principle behind earlier libraries. He further noted that much of the work on digital libraries to that point assumed an idealised model of use: the lone researcher sitting at a workstation, browsing, scanning, searching, retrieving, reading and writing. A point of criticism in [Levy95] highlights a problem common across many fields of expertise: that of over-specialisation.

> *Notably, among librarians there is sometimes a tendency to focus on collection acquisition and maintenance, and to lose sight of the library's role in supporting the community's research.*

The UK Electronic Libraries Programme (eLib), described in [Rusbridge98], addresses Levy's concerns in a number of ways:

- The Open Journal Project described in [Hitchcock98] focuses on the application of hypermedia technologies to traditionally fixed permanent library resources.

- Current projects in hybrid libraries such as Malibu [Pinfield98] are investigating the integration of electronic resources with paper and other types of resources within the services provided by real libraries.

- The eLib programme now recognises that there is a marked division in digital library activities between those centred on resources and those centred on organisational or service contexts.

The library community is beginning to understand the social nature of the documents it uses, placing an emphasis on the convergence of information and work, enabled by communications and information technology:

> *'digital library proponents must consider the role of people … if the digital library is to be truly beneficial'.* [Sloan98]

Central to using documents as components to an organisational memory is an awareness of a document's purpose. The successful use of document relies as much on interpretation of the content as on the content itself: they are socio-technical artefacts [Levy99]. Within an organisation the purpose of a document may be understood because of the learned terminology and culture of that organisation. However, documents are also used to communicate across organisations - Leigh Star calls documents 'boundary objects' [Brown96], asserting a transmission across or between organisations.

The social nature of documents, discussed in [Brown96], alludes to the reason for some of the current problems in document management.

> *Different social worlds can fight over the 'right' interpretation of a document. For there to be a 'right' way, there must be a standard and a judge external to all of the competing community-based alternatives. But there is no external fulcrum to move these social worlds that is not itself merely the internal standard of another social world.*

The above quote highlights the differing interpretations that can occur over the meaning and purpose of a document. Work on ontologies, discussed later in this thesis, and dictionaries of terms attempt to close the gap of interpretation across an organisation. The Ministry of Defence circumvents the problem within their DEF STAN 00-60 document management recommendations, described in chapter 3, by enforcing a standard interpretation over the document classification scheme. While effective at managing the information, this method is only suited to organisations which can enforce such a practice across the whole enterprise. In more commercial environments this may not be a viable solution. Alternative systems attempt to aid the communities they serve rather than define a fixed dictionary: to assist in link creation and term collection, but not to define the overall domain. Sumner argues that documents themselves are able to assist development of the relationships between communities to develop a shared understanding:

> *'a key role of documents is to support negotiation and interpretation as communities struggle to reach an understanding … representations should be regarded as starting points for discussion about the way things ought to be'* [Sumner98]

There is an increasing realisation that information, people and work elements must be synthesised to produce more useful environments. There is an increasing need to integrate users and their working roles within systems that have traditionally focussed solely on management of the information within an enterprise. It is clear that information technology is able to assist in the realisation of such an environment. Internet developments in particular provide methods of effectively distributing information into a network-centric heterogeneous computing environment.

2.2 The Nature of Documentation

It is essential from the outset to define what is meant in this thesis by the term 'document'. As with many widely-used terms there is no axiomatic definition of 'document', but the research literature points to some common themes. It also highlights that the current all-encompassing definitions of 'document' have not always been so open-ended. More traditional definitions begin and end with a document as a paper artefact 'furnishing information or evidence' [Levy99]. While it is not constructive to debate the issue at length, a brief look at the diverse attitudes to documents is useful.

The philosophical debate over the form and purpose of documents has been argued over many years, spanning many technologies. In 1951 Suzanne Briet, a pioneering information scientist, argued that the scope of 'document' extended beyond text to any material form of evidence. While useful as a marker, Briet's definition is a little too broad to be useful to this thesis[5]. Conversely, Sandy Ressler claims that *'the*

---

[5] Briet went so far as to conclude that under certain circumstances even an antelope could be a document [Levy99]

entire concept of a document is becoming increasingly obsolete' [Ressler95] as documents and databases converge. Between these opposing views are some other important and relevant definitions.

2.2.1 Documents as Black Box Objects

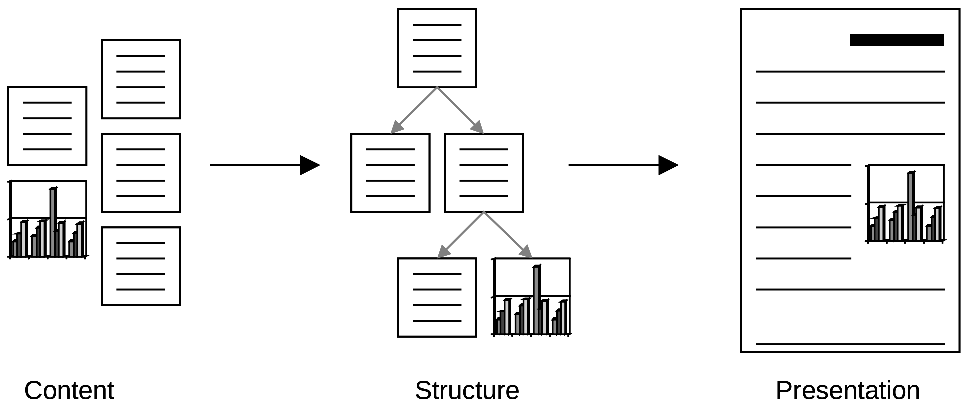
A document can be described as a unit of 'recorded information structured for human consumption' (cited in [Sprague95]). This definition implies a distinction between the storage method ('recorded information') and the means by which content is understood ('human consumption'). It constrains potential machine support to little more than a transport mechanism for black box objects. Traditional file stores and information management systems such as Lotus Notes retain this notion of a document as a binary large object (BLOB): a black box that is stored and retrieved according to some external classification. During the development of distributed networks the computing industry placed importance upon assisting the user to create, store and retrieve unitary documents, but provided no explicit support for understanding document content outside the host software application. Commercial sales of word processors over the past 28 years have flourished, yet their predominant use to more efficiently prepare page-oriented documents merely automates the typewriter and paper files of yesteryear. To regard documents as black box objects is to rely solely on the classification schemes and management methods that the user community has applied to the document space independently of the content of the stored documents.

The treatment of documents as black box objects does not scale well to the enterprise level. Not only are document classifications subjective [Nurnberg97], driven by local context and practice, but increasing amounts of documentation exacerbate the problems of using only explicitly applied classifications as a key into a document store [Apte94]. A single document classification may only partly describe the purpose or role of a document [McCarty97] and a poorly applied classification may render a document obsolete from the outset.

2.2.2 The Composition of Documents

To regard documents as black-box objects is to ignore the original intention and purpose of creating the document. To understand the relevance of documents to an organisation it is necessary to understand what a document contains. A document is a snapshot of some set of information that can incorporate many information types, exist in multiple places across a network, depend on other documents for information, change on the fly, and have an intricate structure [Sprague95]. This definition highlights the compositional nature of documents. It became apparent long before electronic document preparation tools that a document embodies a number of components. Heeman discusses two separate components: a document structure and document content [Heeman92], highlighting a distinction between the physical content of a document and its logical structure. The document mark-up community has further separated document content from its presentation, creating three broad but separate layers: content, structure and presentation (see figure 2-1).

The power of early printing presses lay in the separation of the printing plate from the printed images that were produced from it. This became a powerful idea: one plate could be used to produce a large number of identical images. Movable type was a further innovation on this idea, making it possible to create composite plates out of reusable component images. The content (type) was rearranged to form a structure



content:	the components (words, images etc) that make up a document
structure:	the organisation and inter-relationship of the components
presentation:	how a document looks and what processes are applied to it

Figure 2-1: The Compositional Nature of a Document

(page) and then presented on a suitable medium (paper). Modern methods facilitate variance in all of the three layers: customisation of the presentation, validation of the structure and reuse of the content. In more recent times, mark-up languages have continued to develop the three layers of separation and, importantly, increase awareness and acceptance of the distinctions between the layers.

The recognition of structure within documents is significant to the document versus database discussion [Ressler95]. Traditionally there has existed a distinction between structured database information and information stored and manipulated within documents. Currently that distinction is less apparent: both database data and documents store and convey information, and both have structure. The central way in which one information type differs from another is in their semantic categories: the objects that are taken as primitives and the logical relations between these objects. Conventionally the type of information that needs to be captured determines the choice of method and subsequent software support [Fillion95]. If both traditional data and documents can be described in terms of content and structure there should be no distinction in the method of their treatment. Structured data and documents are different only in granularity, not in their nature [Heeman92].

Of further value is the separation of document presentation from the underlying content and structure. Document preparation tools such as Latex have existed for many years, upholding the separation of document content from document presentation. However, the popularity of these tools has been limited, superseded by 'what you see is what you get' (WYSIWYG) document preparation tools such as Microsoft Word and Lotus WordPerfect. Advocates of mark-up languages argue for the reinstatement of the distinction between content and presentation to facilitate reuse of document content [Raymond96].

Incidental to this work but a very active research area is that of multimedia documentation. [Sprague95] alludes to the many information types that can be contained within a document without specifically discussing multimedia. This thesis regards the multimedia content of documents as secondary to the need to manage documents more effectively, but supports the use of any type of media within a document container.

2.2.3 The Document Space

The information conveyed in some documents is objective - tables of facts or statistics - but in many cases information is conveyed indirectly through the use of natural language. The meaning of a document is termed its semantic. To talk about document space is really to talk about two related spaces that may not necessarily coincide ([Neves97] and [Esposito94]):

- A **physical space** - a collection of document nodes and relationships
- A **semantic space** - the relationships among the information conveyed by documents

A data record contains the attributes of an entity such as an employee in a personnel system. A document, however, supplies information necessary to represent a concept or idea [Sprague95] - a much more complex role. A physical document space is created by the author of the document; the author's intention is to convey a meaning or argument to the reader by means of that physical space. The semantic space is the meaning that the reader attributes to that physical space, which may or may not coincide with the author's intended meaning. The problem of effectively communicating semantic meaning through a document is immense because not all words correspond directly to concepts in the world. There are no general relationships between the form of a document and the purpose intended by the document author. Karl Branting states that 'goals that documents are intended to achieve are seldom made explicit' [Branting97]. In his paper he terms such goals 'illocutionary' and attempts to attribute explicit rules to portions of document text to enable their automatic reuse for other purposes. However, the documents that Branting considers are very limited in their scope and structure.

2.3 Enterprise Demands on Documentation

The previous two sections explained the nature and value of documents to the organisations they serve. This section reviews the call for improved organisation of enterprise documentation in the ubiquitous climate of change. [Fillion95] identifies three problems that an organisation has when managing its resources across its boundaries. These are:

- **The problem of semantic diversity** Different parts of an enterprise use and target different types of information. In order to manage resources across the organisation, these information types require integration in some way. To be clear on the connections between the information types, the logical relations between semantic categories need to be precisely defined or understood.

- **The problem of representational diversity** Which modelling languages or methods are used for each part of the enterprise, and how may these be unified?

- **The problem of contextual diversity** What does a certain user group mean when they use particular phrases? The worlds of differing users are often differing, and the meaning of a certain phrase is dependent on the semantic backdrop of that particular view.

Larger, more connected enterprises with local control

While an overall organisation may have the same end-goal in terms of its product line or service, the internal structure of organisations is such that many goals are local, not global to the enterprise. There is a tension therefore between the desire to retain an enterprise-wide view of documented knowledge and the local needs of individual subdivisions of the whole. A networked environment supplies the physical connectivity required but the localisation of goals has traditionally led to an internalisation of information.

The future will see a phenomenal growth in the use of IT to address *co-ordination* between groups [Greenwood95]. Enterprise Integration (EI) is concerned with the distribution of corporate information across these groups, but as a consequence of EI efforts small innovative partners have become excluded - the effort required to integrate often exceeds the total effort required to innovate [Goranson92]. Managers are socialised people, high in relationships and low in structure [Multicosm98]. In the face of this is an organisational structure that focuses on local autonomy and insularity. Networks are often very localised [Multicosm98], and there is a requirement to enable sharing and collaborative work.

Virtual enterprises with core competencies

There is much in the research literature about the potential for development of virtual enterprises (VEs) within information-focused and product-focused organisations. The defence industry recognises that

60-80% of the value of the entire system comes from contractors [Goranson92]. A VE is a cohesive, but potentially continuously adapting business entity made up of diverse subentities. Whereas 'enterprise' commonly refers to a single organisational structure and changing goals, [DOD96] describes a virtual enterprise as a specific business goal or objective, with potentially changing structural aspects. The successful attainment of the business goals of a virtual enterprise depends on the availability to align the business processes and practices. Presley describes a VE as a temporary alliance that contains almost no employees or inventoried resources, with each member company providing its own core competencies [Presley96].

The challenges on document management within a virtual enterprise are similar to those for large distributed organisations, but more accentuated: documents are the central mechanism for managing, communicating and contracting in a virtual enterprise. Amor considers the construction industry [Amor97], which is made up of virtual companies formed for a short duration. The life of the product is longer than the life of the VE so document management is of utmost importance. Individual companies retain their own working practices, so it is even more important to provide mechanisms for communicating across company boundaries.

Variance in products and services

[Brooks94] recognises an increasing complexity and diversity of products within the telecommunications market noting that 'a company wishes to maximise reusability at minimum cost'. Within a less commercial environment the Ministry of Defence recognises that many of its low volume products are variations on a common theme. Similarly within the domain of higher education, which offers a service rather than a tangible product, there is a call to produce customised course materials on a per-student basis [Zarhan94]. All of these situations call for a high amount of variance that can only sensibly be provided by adopting a reuse policy over existing informational or intellectual resources.

Knowledge work

The research literature claims that knowledge workers are becoming increasingly common. Many see this as a new phenomenon but for some time intellectual capital has been recognised as a significant component of the business domain. Bell suggested as early as 1974 that information is the 'axial principle' of a post-industrial society in which the majority of employment is for 'information workers' rather than those engaged in manual tasks [Quintas97]. Even earlier:

> *'The intellectual and moral capital of Great Britain far exceeds all the material capital, not only in importance, but in productiveness.'* Senior 1836, cited in [Quintas97]

Knowledge management is a hard concept to define. The following are some of the contributions from the research literature. [Quintas97] and [Broadbent98] suggest that knowledge management might better be seen as a component of all forms of human and organisational activity, rather than a subject in its own right.

Knowledge management is about enhancing the use of organisational knowledge through sound practices of information management [Broadbent98], encompassing the abilities to develop new opportunities and to meet both existing and emerging needs. A more direct encapsulation is suggested in [Quintas97]: knowledge management is 'providing information to precisely the right place at precisely the right time'. This inherently implies reuse of current or historical resources for new purposes. [Broadbent98] suggests that new knowledge always begins with the personal. [Davies98] describes knowledge as information transformed into a capability for effective action, placing people at the centre of the knowledge domain. [Demarest97] suggests that the truth-value of the information being managed is incidental to its ability to generate desirable commercial performance, stressing the social nature of commercial knowledge. Groupware products currently provide an infrastructure for colleagues who are already motivated to work co-operatively on a single business process. Knowledge management techniques should build on this by facilitating co-operation among virtual communities whose members might not even discover each other's existence otherwise [Multicosm98].

Electronic document management systems play a central role in knowledge management. The goal of electronic document management systems (EDMS) is to both enhance and preserve the value of an organisation's information resources [Reagan95]. Document management techniques should further support knowledge management by including a person-aware element, facilitating innovation within and between organisations. Current document management systems can be divided into two broad approaches, minimally adapted from [Rezgui96]:

- The **integrated document management approach**, where the aim of computer support is to enable effective document retrieval using reference information supplied with the document; each document is a black box. This approach is widespread in current EDMS.

- The **document model-based approach**, where information within documents is retrieved by query and relevant new documents are produced semi-automatically. This implies access to the content and structure of the document itself, and adaptation of that content.

The following sections discuss the two approaches.

2.4 Integrated Document Management Approaches

Integrated document management approaches regard documents as binary large objects (BLOBs) with associated attributes (meta-data). The simplest of such systems is the conventional file store used by Microsoft Windows and other general-purpose operating systems. Files are arranged into a pre-determined structure and navigation to a particular document is possible by navigation of that learned file structure. This is a well-recognised and widely used method. To further enhance such structures, many systems (e.g.

Microsoft Windows, VSP resource rooms, the World Wide Web) provide the facility to attach meta-data items to each file. For a typical document, meta-data items may include document title, author, date created and a list of keywords associated with that document. Such data are either added manually or automatically extracted when the document is first entered into the file store. Meta-data are regarded by many (e.g. [Chase94], [Manola98], [Powell98], [Schuldt95], [Watson98]) as a valuable index into a document store when document content cannot be readily accessed, or as an additional controlled set of attributes that describe the document content in a systematic way. A widely accepted structure to these meta-data increases the power to find documents by people who understand that structure. Accordingly a number of user groups have defined their own collections of meta-data that they apply to documents within their domain.

The Dublin Core [Weibel97] is a well-known meta-data proposal intended to suggest a minimal set of data elements that are provided by the creators or publishers of digital objects. The Dublin Core focuses on simple resource descriptions that are usable by non-specialists, providing a core set of 15 meta-data elements: **Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage** and **Rights**[6]. The strengths of the Dublin Core are its simplicity, its semantic interoperability, its international recognition and its extensibility to incorporate more elaborate description models [Weibel97]. Another two notable meta-data projects, ARIADNE [Ariadne98] and IMS [IMS99], outline resource description schemes specifically for educational resources. They describe not only the subject area of the resource but also the education level which the resource targets. Both IMS and ARIADNE benefit from a broad recognition within the learning domain they describe.

A higher-level meta-data project currently under development is the Resource Description Framework (RDF – [Lassila96]). This is a general-purpose framework providing structured, machine-understandable meta-data for Internet resources. RDF subsumes meta-data descriptions by recommending a unified notation for their representation. It is hoped that RDF meta-data descriptions, when applied to Internet resources, will provide an integrated means of more effective resource discovery. Meta-data are currently attached to WWW documents in an ad-hoc fashion; RDF adds valuable structure to the application of meta-data. Importantly RDF provides interoperability between different meta-data schemes so that user groups and their documents do not become stranded, unable to co-operate with groups that decide to use meta-data schemes more suited to their purposes. The RDF mechanism in itself provides no recommendations for its implementation. In response to that observation, [Brophy97] calls for a National Agency for Resource Discovery in the UK to specify and oversee the working practices required to implement such a meta-data

[6] A full explanation of these terms can be found at http://purl.oclc.org/dc/about/element_set.htm (Last verified 8th March 2000)

scheme. The call for an agency external to each of the participants indicates the difficulty in developing and delivering a cross-domain solution to resource discovery.

An interesting and active area of the document management research literature involves the gathering together and clustering of documents into collections that have a similar nature. A user is able to navigate these collections to find documents that are deemed to be similar in some way. Document collections are implemented by the Hyper-G research architecture [Kappe94] from Graz University (Austria) and integrated into the HyperWave commercial product[7]. The document collections have associated searchable meta-data that may describe more about the collection than the collected meta-data associated with the component documents. A simplified version of the Hyper-G navigation model is shown in figure 2-2.

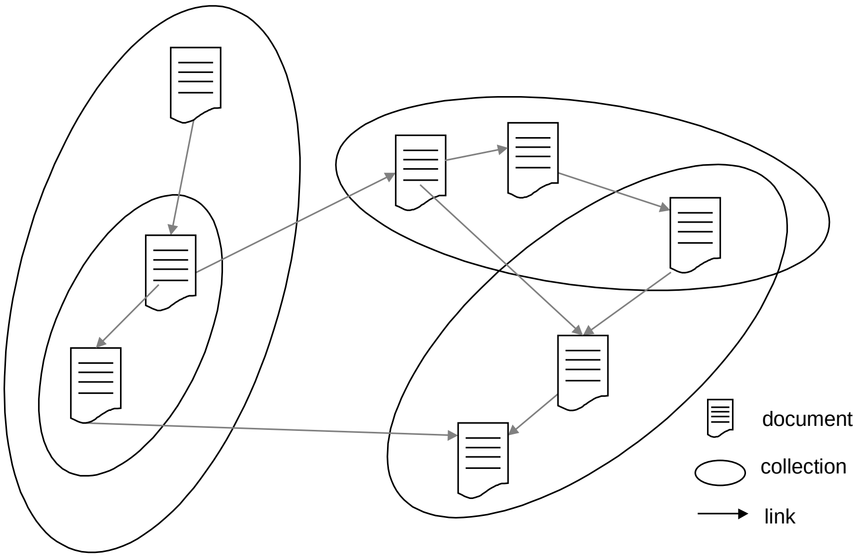


Figure 2-2: A simplified version of the Hyper-G Navigation Model

Every Hyper-G document is a member of one or more collections, which are in turn members of one or more collections. By navigating documents and the collections they belong to, Hyper-G assists the location of documents of a similar nature, without the need to perform operations over the document content itself.

Beyond the application of meta-data, document objects within an organisation can be managed through the application of co-ordination technology. Co-ordination technology is a term applied to the areas of workflow, groupware and process enactment technologies [Kawalek94]. Such technologies can be applied to a document store without accessing actual document content. Co-ordination systems have existed for a

---

[7] http://www.hyperwave.com/ (Last verified 8th March 2000)

number of years, Lotus Notes being one of the most commercially successful. The simple definition of document[8] flow within a business process permits the routing of documents from one stage of the process to the next. The power of such products is this routing of documents along recognised lines of communication, not necessarily the understanding of the document content transported.

More recently-developed systems have begun to realise the specific value of information to the organisation inside which it is stored and used. A number of systems now model the organisations inside which information is effective. The University of Leeds has developed the concept of a Virtual Working System (VWS) - an environment which combines information management with systems that enable people to visualise, navigate and access information and communicate using a variety of electronic tools [Drew97]. The Virtual Science Park (VSP) is the first implementation of a VWS.

The Virtual Science Park

The Virtual Science Park (VSP), developed at the University of Leeds, is a customisable Virtual Working Environment (VWE) intended to support industrial-academic research activities [Drew97]. The aims of the VSP project are to use networking and computing technology to create an on-line science park which meets the strategic goals of a physical science park development (eg. research collaboration). The conceptual framework of the VSP has wider applications since it provides a VWE that allows people within organisations to interact using person-to-person communication. It is virtual in the sense that the linkage between people within the environment only exists when communication is taking place [Drew97]. Each participating organisation owns a tenancy that is a holder for all information related to that organisation. One of the current focuses of the VSP is on supporting the innovation process over the expertise of the participant tenancies. Expertise is classified using recognised classification schemes from the domains of the tenants. At its outset the VSP aimed to support the innovation process; its aim was 'mediation across raw data domains to produce useful information and mechanisms for supporting communication between human networks'.

Within the VSP each tenancy has it own resource room: a store for information objects that are owned by that tenant, organised in a structured way that has significance to that tenant. An information object can be a pointer to a resource (in the form of a web address), a note, document, link to a document or a link to folder. It further allows resources and folders to be linked from other tenancies, or from other parts of the current tenancy. There is therefore a level of resource re-purposing, as the same resource can be accessed from more than one classification and potentially from more than one tenancy.

---

[8] Lotus Notes' definition of a document is a structured database record

Resources within resource rooms are typically binary files of a proprietary form and not available for full-text search. Searchable meta-data are manually supplied for each document when it is added to a resource room: title, abstract, keywords, author, date authored, document version and an external 'document ID'. Resource rooms provide simple 'store and retrieve' functionality for less structured resources (documents), providing a version control mechanism that maintains a pointer to the latest version, but also allows the user to access previous versions if required. Importantly a set of document resources within VSP resource rooms is associated with the organisation within which the resources are used. This provides an important work context, which is not provided by more general-purpose document management solutions. By associating a document with the organisations and people to whom it is relevant, the VSP enables one type of collaboration discussed in chapter 1 - assisting the recognition of groups of people in the hope that collaboration between these people will result in increased productivity or effectiveness.

The Basic Support for Co-operative Work (BSCW - [Bentley97]) project provides similar functionality to VSP resource rooms, but does not support explicit associations between documents and the enterprise inside which they are used. The Virtual Science Park contains an explicit representation of its tenant organisations and its members. A number of other systems are similarly people-aware, providing links between the information they store and the people who create, maintain, view or express an interest in that information. MEMOIR [DeRoure98] is an extension to the Microcosm hypertext system that supports researchers working with distributed information by assisting them in finding both relevant documents and researchers with related interests. When a user browses documents and follows the supplied links from one document to the other, a document trail develops for that person. MEMOIR regards these trails as first-order objects and allows another user to discover 'who else has read this document'. The proposition of such a system is that people who read the same documents may be of interest to each other, and information on other such users is supplied. Jasper [Davies98] is a recommendation system developed by British Telecom, allowing the definition of interest groups and facilitating the recommendation of web-based resources within those interest groups. Notably MEMOIR and Jasper are person-aware, not person-centred as the VSP is. Both are primarily document recommendation systems which make recommendations according to a list of documents viewed or a list of documents recommended by its users.

More broadly, effective distribution is central to the successful use and reuse of document resources. It is clear that the WWW holds many opportunities for assisting distribution: a core initial user community, the integration of existing information, the use of defacto standards, the spanning of organisational boundaries, and a software platform which is public domain, cross-platform and extensible [Dix97]. Many products now use WWW protocols and browsers to disseminate their information across networks. Most of these products, however, use the WWW only as a delivery mechanism for documents and information. Non-WWW environments currently provide more mechanisms for collaborative work, where the user is able not only to access information, but also add to it, change it and communicate with others workers. One of the aims of BCSW is to enhance the inherently stateless HTTP protocol by supporting collaborative methods such as document upload, version management, user administration and document locking mechanisms to prevent concurrency problems. The Virtual Science Park provides similar mechanisms, further providing access to Internet communication tools that enable two users to communicate and collaborate on a single document. Both the VSP and BSCW solutions circumvent the weaknesses of the current HTTP protocol, identified in [Trevor97]. A number of W3C working groups are currently developing extensions to the HTTP protocol itself (WebDAV - [Whitehead98]) to incorporate co-ordination mechanisms such as access control and version management into the native WWW environment. Such a protocol would greatly assist the development of web-based CSCW applications that enable collaborative working information over distributed networks.

2.5 Document Model-based Approaches

Model-based approaches imply an additional understanding of the content and structure of the documents they manage. While inheriting the abilities of integrated document management approaches, document model-based approaches perform further operations on the documents themselves. Through an awareness of the form of documents, document model-based approaches are able to understand and manipulate the content, structure and presentation of documents, rather than treat them solely as black box objects. The treatment of a document as a monolithic object precludes opportunities for its reuse [Matthews98].

There are two broad areas into which document model-based approaches divide – those that operate between documents and those that operate across documents. The field of hypertext defines methods for collecting and linking between sets of documents for some purpose, providing a means of navigation through the resulting collections. Search spaces and methods provide the means to discover a set of documents across the document space by applying a search criteria. A number of on-line WWW services (e.g. Yahoo!, Infoseek, About.com) combine searching and hypertext, allowing the user to search across the broad information space and then follow links that provide more localised navigation. Conversely, the user is able first to refine the area of interest by following hyperlinks and then perform a search across the document resources within that area.

2.5.1 Hypertext Document Models

The hypertext community is concerned with creating paths through documents that form threads of discourse and argument. Early hypertext systems such as NoteCards, NLS/Augment and Intermedia (reviewed in [Conklin87]) developed mechanisms for navigating through the documents that together formed a single line of thought or argument. Many of the ideas were subsumed by Tim Berners-Lee's original recommendations for the WWW [BernersLee89], but many were disregarded in favour of the much-simplified linking model currently used.

The field of hypertext provides a mechanism that may be used to more effectively organise and navigate collections of documents. A useful working definition of hypertext is 'nodes of information, machine-supported links between these nodes and a common user interface' [Cunningham93]. Hypertext allows typed links to be computed or manually added between documents that are relevant to each other in some way. Hypertext transcends the classification of document resources, supporting less structured relationships and linking between different document spaces. The WWW, for example, allows a link to be constructed between any two documents; the destination document need not know the existence of the source document. There are many identified problems with the linking scheme currently practised on the WWW, but few can dispute that its simplicity has aided the WWW's unprecedented growth.

Hypertext provides a user with the ability to pursue a line of thought or enquiry without the need to understand underlying file structures or the complexities of formulating a search query. It allows the recommendations of authors (or readers) to be highlighted to other readers so that a line of thought may be constructed and followed by other users. The effectiveness of hypertext is signalled by the current competition in the Internet portal industry. Inexperienced users are provided with links to virtual 'channels' that are created from Internet content that already exists. The strength of the Internet portal lies in filtering, classifying and linking information of a common subject. Three technical factors make the Internet portal successful: mediation of the document space, adding value through selectivity, and the provision of hyperlinks to supply that mediation to the user. Hypertext is the mechanism that communicates to the end user the set of candidate routes of further enquiry or information. The literature points to a number of issues surrounding the use of hypertext. These are:

- The problem of becoming **lost in hyperspace** is well documented and refers to the complexity of navigating the document space without limits [Conklin87]

- There is no guard against **bad paths of reasoning** created by inappropriate use of hypertext links. The situation is exacerbated on the WWW because there is no notion of link ownership or purpose beyond 'this information is related in some way'. In [Halasz88b] Frank Halasz calls for typed links ('structured relations') to resolve such issues. [Malcolm91] similarly calls for typed links, and for an association between a link and its author.

- There is a risk of **over-fragmentation of the document space** as authors become overly concerned with on-screen hypertext presentation and less concerned with the flow of document meaning and discourse. [Whalley93]

Microcosm

Microcosm [Davis92] is an exemplary hypertext product developed by the University of Southampton that allows its users to integrate disparate pieces of information into a cohesive web of documents: '*it is essentially an open framework for agents that control and relate the individual documents*' [DeRoure96]. It extends the simple unidirectional link currently prevalent on the WWW, regarding links as first-order objects, retaining their separation from document content. Microcosm identifies and implements the following types of link:

- **Static links**, where both the source and destination have been defined in advance.
- **Dynamic links**, where the source and/or destination is not defined until the link is followed. The ends of the link are decided according to some rules at the time of link activation.

Both static and dynamic links can either be created manually or computed. Computed links are dynamically created by the system according to some pre-defined rules, requiring no intervention from the author. They use information retrieval techniques such as automatic indexing to develop links that may be useful, but which may not otherwise be added manually [Li92]. The initial Microcosm model is shown in figure 2-3.

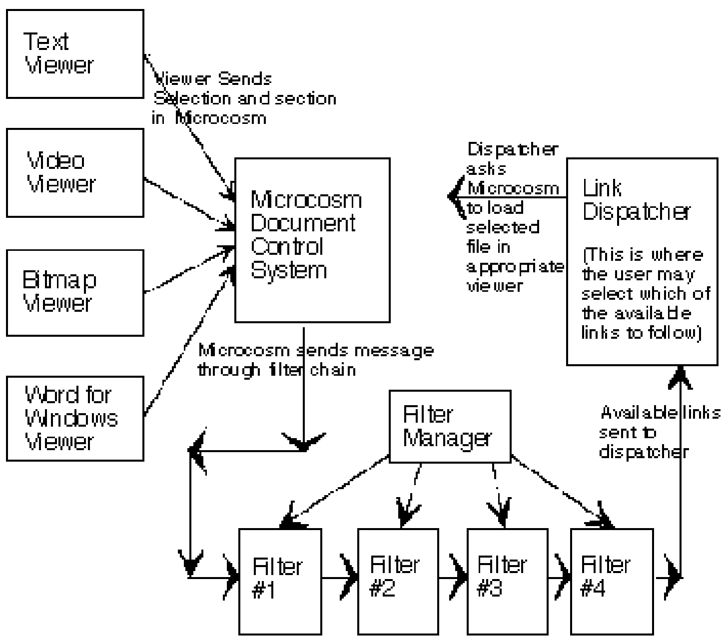


Figure 2-3: The Microcosm Model

The initial design of Microcosm facilitated the creation and navigation of links between two documents in their native file formats, allowing both manual and automatic links to be created. In recent years Microcosm has started to incorporate sessions [Goose97] that provide views on the document store.

MEMOIR [DeRoure98] further extends the hypertext provision of Microcosm to include the interests and document trails of its users.

Clustering Documents

Regardless of the hypertext approach used, increasing amounts of documentation limit the amount of manual link creation possible, and some successful automated methods have developed. Information retrieval and artificial intelligence techniques have been used effectively to uncover clusters of similar documents within limited domains. [Merkl95] describes the use of self-organising feature maps to apply statistical clustering to a collection of software documentation. The task of text classification aims to uncover the semantic similarities between various documents by constructing feature vectors for each of the candidate documents. However, even for the small domain of software documentation almost 500 distinct terms are found, forming a 500-dimensional vector for each document. [Merkl97b] acknowledges that this method is too cumbersome to scale to general-purpose document management. Similar statistical work is described in [Honkela95], [Schweighofer95] and [Apte94], based on clustering documents that have similar single word frequencies. Such keyword searches and statistical techniques have several limitations, however, noted in [Riloff94]:

- **Synonymy** Two different words may have the same meaning, but would be regarded as distinct concepts during document clustering
- **Polysemy** A single word may have multiple meanings, yet would be regarded as the same concept
- **Phrases** Some words are only meaningful in specific phrases
- **Local Context** Some words and phrases are attributed with a specific local context
- **Global Context** Some documents do not contain essential keywords because of the understood surrounding context. A football report, for example, may not contain either of the words 'football' or 'sport' because its context is understood on the sports page of a newspaper.

These problems have been approached in a number of ways: thesauruses to identify synonymy and polysemy [Wang85], recognition of word pairs or triplets to identify phrases [Croft91], specialist dictionaries [Apte94] to understand local context and analysis of document style to add a global context [Argamon98].

Caveat

Some (e.g. [Nurnberg97]) regard hyperlinks as a destructive mechanism over document spaces because they emphasise a primacy of data rather than primacy of structure within the text and its classification. Nurnberg argues that hyperlinks cut across the document structures that are required to form sound argumentative systems that reflect formal, communally defined relationships. In particular domains such as teaching and learning, [Cunningham93] argues for the case of constructivism in document delivery. Constructivism proposes that instruction is less a process in which knowledge is communicated to learners, and more a matter of nurturing the ongoing process. In this view, knowledge is an active process of construction, not the receipt of information from external sources. While not applicable to all forms of document management, the case for constructivism at least highlights the need to restrict hypertext in certain situations.

2.5.2 Document Search Space

Applications with large, unfamiliar, heterogeneously structured network are not well suited to link navigation alone [Halasz88a]. If the user is located at a point in a document space that has no defined links to the required information, no amount of link navigation will yield the required result. What is required in this situation is a higher-level search across the document space. The search may yield a document which is useful in itself, or from which the user can navigate links. Search methods across document spaces are an area of much research and commercial exploitation. Search methods can be characterised as:

- **Syntax search** A syntax search is performed across the content of the document itself and can be performed at the word or phrase level. Documents with more occurrences of the search term are considered more relevant. Such searches are the most commonly used across the WWW.

- **Meta-data search** Meta-data have been discussed earlier in this chapter, providing descriptive data about the documents to which they are attached - document title, author, etc. Searching sets of meta-data to find relevant documents enables a user to discover documents that are either not searchable or that do not contain the specific information in an identifiable format. Such an approach is adopted by the Virtual Science Park.

- **Semantic search** Forming the cornerstone of current commercial knowledge management systems such as Excalibur [Excalibur99] and Autonomy [Autonomy98], semantic searches incorporate the notions of synonyms and semantic concepts, aiming to capture and communicate intended meaning. An underlying network of semantic concepts permits a level of interpretation over the document space. The promise from such systems is a document relevant to the query that does not necessarily contain the search terms provided.

The open nature of the WWW document mark-up language HTML (HyperText Mark-up Language – [W3C98]) has led to a proliferation of on-line search engines that catalogue web resources. One of the most powerful of these, AltaVista[9], permits a syntax search across more than 40 million web pages indexed from web servers world-wide. It permits a single-word or phrase search, or any combination thereof. Results are returned to the users with a basic summary of candidate pages and appropriate links to the information sources themselves. While searches across document syntax produce an abundance of candidate documents, their effectiveness is limited [Davies98]. Such systems can be characterised by two measures - recall and precision

$$\textbf{Recall} = \frac{\text{number of relevant documents retrieved}}{\text{total number relevant}}$$

$$\textbf{Precision} = \frac{\text{number of relevant documents retrieved}}{\text{total number of documents retrieved}}$$

It has long been realised that both the recall and precision of web search engines can be quite low [Davies98]. Other search engines such as Yahoo! and AskJeeves attempt to circumvent the problems of low recall and precision by manually approving and classifying web resources. While effective to a certain extent, such manual systems are not scalable and automated alternatives are being sought. The problem is exacerbated by the lack of semantic structure within HTML. [Bosak97] explains how HTML is failing in terms of extensibility, structure and validation, and many have rejected HTML as an effective language for marking up document content.

Of increasing interest among commercial knowledge management products are semantic searches across corporate document stores. Two exemplary products of this nature are RetrievalWare by Excalibur [Excalibur99] and Autonomy's Knowledge Server [Autonomy98]. Both contain configurable semantic networks that link concepts that are not syntactically linked. The user need not therefore enter the exact words that will result in successful retrieval. Certain syntactic words activate certain semantic phrases that in turn activate other syntactic phrases. Resolving such a semantic network to a particular depth produces a list of syntactic structures that are then sought through the document store.

Another approach to semantic searches is exemplified by Carole Goble's work at the University of Manchester. Rather than provide a general purpose semantic network with limited depth, Goble provides much more depth over a more restricted domain. For a specific domain (e.g. medicine [Goble94]) a model is constructed in such a way as to permit inference of new concepts from those that are added explicitly. When a new concept is added a number of constrained inferences are made, further populating the semantic network. It is possible to associate a document with a particular concept and, through sanctioned inferences,

[9] http://www.altavista.com/ (Last verified 8th March 2000)

propagate the document through the semantic network. Such computational inference underlies much research on domain ontologies, which are discussed in a later chapter. It is proposed that attaching documents to such a domain model may assist propagation of a document to other areas of the domain where that information is of further use [Goble97].

2.5.3 Document Reuse

[Levy93] argues that documents are the localisation of communicative material for particular purposes and instances of recognisable social types. This definition points to a number of problems to overcome if documents are to be reused for more than their original purpose. These are:

- **Localisation of communicative material** identifies a wrapper of contextualisation added to the information to make it effective in its current environment. This may involve the translation from external unfamiliar language to localised terminology, so that meaning may be understood within the current context.

- **Instances of recognisable social types** highlights the importance of document type or style - the reader can assume much about the document if its document type is understood. Learning to determine the style of a document is a dual to that of determining its topic [Argamon98]. This claim is borne out by current work on community-wide document type definitions for particular document types [Weibel97] [IMS99]

In order to facilitate document reuse there is a need to clearly identify the context within which a document is useful, and to separate document content from its social type (structure and presentation). Since many documents are communicative artefacts and represent truths about the world, they may require updating as the world changes. The incremental changes in the world should be reflected as incremental changes in documents. A document that incrementally changes is one that reuses much that already exists and to support document reuse is to facilitate such incremental changes. This thesis attempts to further understanding on the reuse of documents for more than their original purpose.

David Levy identifies four stages to the reuse process which he has coined the four Cs [Levy93]. According to Levy, all documents, and versions of documents, come into existence by some combination of the following operations:

Creation:	The production of new material
Collection:	The identification and gathering together of previously existing material
Combination:	The stitching together of new and old material to form a new unity
Customisation:	The reworking of this material to fit its new setting

Karl Branting argues in [Branting96] that document reuse requires access to the original intentions underlying the document. '*Problems of document maintenance are exacerbated by the frequent personnel changes that characterise large institutions … without access to a document's authors the intent behind particular clauses may be lost*'. Branting describes the reuse process from the perspective of document purpose (or goal), proposing an automated process to find and reuse relevant documents:

Given
- A **set of goals** to be accomplished by the document to be drafted
- A library of **existing documents**

Do
- **Retrieval** Find the existing documents that best satisfy the current goals
- **Comparison** Display the differences, if any, between the goals achieved by the retrieved documents and the current goals
- **Adaptation** Remove the portions of the text whose only purpose is to satisfy goals that aren't present in the current situation (excision) and add text to satisfy any of the current goals not satisfied by the retrieved text (augmentation).

Rada emphasises that in order to successfully reuse document material, some kind of classification of existing material is vital [Rada90]. He proposes a semantic net to provide an overview of the content of document text in order to locate relevant documents for reuse. Rada, Branting and Levy all recognise that a fully automated reuse mechanism is not currently viable, and that the final phase of any reuse process must be a manual customisation of the generated document.

Examples of reuse

A changing attitude to electronic documentation is provided in [Lincoln97], which describes an interesting piece of 'future work' for the Open University (UK). The Open University is a successful distributed learning organisation that would like to deliver its courses in an increasingly dynamic way. Lincoln describes a study that was initially intended to investigate effective means of converting existing course materials from electronic text prepared for printing to the HTML format, suitable for reading on-line. However, he expresses concern that the straightforward conversion exercise may not be worthwhile without adding value. He claims that such value may be added through reuse and repackaging of existing course materials into *'smaller, inter-linked chunks'* for new learning markets. Lincoln's project report highlights an increased awareness of the potential benefits of document reuse.

A number of projects in the research literature accomplish document reuse at different levels of abstraction. Two projects described in [Gehringer98] recognise the potential advantages of reusing resources within the learning domain. The projects aim to collect together resources from disparate educational institutions and classify them according to the subject and educational level. The projects aim to provide an on-line catalogue of such resources, which may be accessed by tutors seeking course materials. A tutor with a need to produce learning resources may be able to find a set of course materials on an appropriate subject at an appropriate educational level, then be able to somehow reuse those materials.

The advantages of separating a document's content and structure from its presentation are highlighted in Karben's article 'News You Can Reuse' [Karben99]. It describes the successful re-purposing of Wall Street Journal content to a range of publishing partners, each of which requires articles in its own electronic format. Where traditional methods of content reuse involved the manual reformatting of each article for each new purpose, the system automatically produces news articles in a range of styles using the XML mark-up language and its associated standards.

Some types of documentation lend themselves to reuse more than others. [Matthews98] describes industrial project documentation that intermixes project data with more commonly-used textual descriptions. Matthews' approach embeds data elements (coded in the EXPRESS language) within structured documents, automatically generating document instances from product data. Similarly, [Vercoustre97] introduces the notion of 'virtual documents', which are part static and part query-based. Vercoustre develops a grammar that enables the retrieval of information from databases and the inclusion of that information into a new document, interleaved with fresh (static) document content. The grammar is embedded within the documents themselves, which are coded in the SGML mark-up language. While effective for its purpose, this type of document reuse is not extensible to general document production where reusable components cannot be discretised and intermixed with static document content.

2.6 Conclusions

Documents are able to assist in the construction of organisational memories that not only capture knowledge within the organisation but also provide ways of developing new opportunities. Physical document artefacts codify a semantic meaning that has value to the organisation within which they are managed. Central to a document's usefulness is its classification, its ability to be delivered to the right place at the right time and its potential for application to new problems (reuse). Notably only a small part of the research literature discusses document reuse, indicating either its irrelevance or its complexity. This thesis proposes the latter: that the issues surrounding document reuse are significantly more complex than those for structured information reuse. An understanding of a document's form - content, structure and presentation - facilitates its effective recall and presents opportunities for its continued use and reuse within an organisation. A further understanding of how a document is used within an organisation provides useful cues for potential collaboration. Through a document model-based approach, this thesis aims to provide higher-order benefits that would be possible from conventional document management.

3 ILS Case Study

3.1 Introduction

The previous chapter described the nature and value of documents, their management and further areas of research that may assist in more effective management of an enterprise's document resources. This chapter describes DEF STAN 00-60 (DS0060) - a well-developed document architecture that represents a cross-enterprise document management solution. DS0060 is based on Integrated Logistics Support (ILS) principles, which support the development of all in-service military equipment. The DS0060 document architecture focuses on the storage, retrieval and publication of technical manuals that support the lifecycle of military equipment. ILS proposals suggest that document reuse is possible through the fragmentation of documents into independent document modules and subsequent assembly of these parts into a work-specific publication. This chapter introduces the concepts of ILS and the DS0060 document architecture, describes a feasibility study undertaken, and ends with a discussion of how DS0060 principles may be extended to more general document management problems.

3.1.1 ILS and CALS

In the late 1970s the US Department of Defence (DoD) found it necessary to respond to a newly identified problem within the defence industry. The cost of in-service support for some defence equipment was proving to equal to or more than the cost of its original procurement [Kaminski95]. Clearly, to predict any necessary support early in the products design and make considerations in order to minimise that support would be very advantageous. To that end, the DoD devised a strategy called Integrated Logistics Support (ILS): a disciplined management approach aimed at optimising life cycle costs (LCC) of defence equipment. The underlying fundamental principle of ILS is "Designing for Support" as opposed to "Supporting for Design", providing a structured analysis of supportability throughout a product's life. Subsequently, in response to the DoD recommendation, ILS was mandated by the UK Ministry of Defence on 80% of new procurement for the financial year 1996/97.

ILS provides the discipline for ensuring that supportability and cost factors are identified and considered during the design stage of defence equipment so that they may influence the design, with the aim of optimising Life Cycle Costs (LCC). It encompasses Logistic Support Analysis (LSA), Integrated Supply Support Procedures (ISSP) and Electronic Documentation (ED) creation and delivery. This thesis focuses on the electronic documentation management aspects of ILS, mentioning the other elements only in-passing.

ILS focuses on the modelling of the design and development phases of a product cycle to ensure adequate provision for its through-life support. Continuous Acquisition and Lifecycle Support (CALS) broadens the scope of ILS to include procurement and maintenance phases of defence equipment. CALS is described as 'a joint government/industry strategy to re-engineer error-prone, labour-intensive, paper based business processes with an automated and integrated acquisition and logistics support processes, based on digital data and documents' [NATO96]. CALS and ILS both aim to:

- Reduce lead time, reduce cost and improve quality
- Integrate and improve design, manufacturing and logistic functions
- Bridge existing islands of automation

The aims of ILS are broad, covering a range of information and technology. The particular role of DEF STAN 00-60 is to fulfil the ILS aims for the documentation associated with the entire range of in-service military equipment.

3.1.2 DEF STAN 00-60

DEF STAN 00-60 (DS0060) defines the UK Ministry of Defence (MoD) requirements for the application of ILS principles for through life management of equipment. It is a profile of existing, internationally recognised standards which are brought together for consistent use, encompassing tri-service requirements (British Army, Royal Air Force and Royal Navy). The baseline standard for electronic documentation management under DS0060 is an Anglicised version of AECMA Specification 1000D (SPEC1000D), amended to provide support for the current range of UK military equipment and practices. The remainder of this chapter describes the recommendations of DS0060 in pursuit of through-life document support of UK-based defence equipment.

For the MoD the problem of documenting increasingly complex defence equipment is well-established. For the British Army alone, product documentation costs are estimated at £50 million per annum[10]. DS0060 aims to provide electronic methods to store and access increasing amounts of documentation for increasingly complex products with increasing life-spans. Its general aims are

- Lower document production and amendment costs
- Avoid duplication and promote reuse
- Lower distribution costs
- Increase integrity
- Widen access

[10] Courtesy of the Army technical support agency

- Allow more direct end-used feedback

Further to the support of new product documentation, there is a requirement to effectively manage legacy documentation. DS0060 includes a migration path for legacy documentation. The overall document process detailed in DS0060 is shown in figure 3-1. It shows the input of both new and legacy document data, the storage mechanism for those documents, a mediation layer that extracts documents of relevance to the user and a user interaction layer. The storage layer is of specific interest to this thesis and is discussed further in the next section.

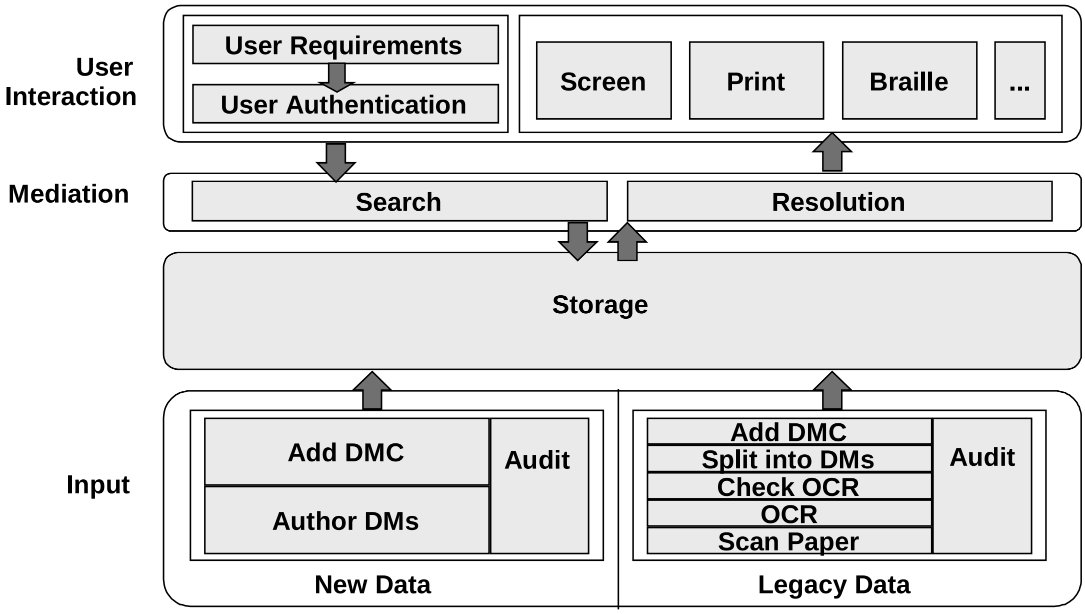


Figure 3-1: The Overall DEF STAN 0060 Document Process

3.2 The DEF STAN 00-60 Document Architecture

DS0060 establishes a uniform approach to the procurement, format, transfer medium and maintenance of electronic documentation (ED). The underlying principle of DS0060 document management is the central retrieval and generation of technical documentation from a set of data modules. A data module is defined as:

> *A self-contained unit of data for the description, operation or maintenance of an air vehicle, airborne engine, airborne equipment and support equipment. The unit of data consists of an identification and status section and contents section and is produced in such a form that it can be input into and retrieved from a database using the data module code as the identifier.*

The data module forms the lowest level of document abstraction. Each data module is marked with a unique code that identifies its role in describing some piece of defence equipment. In order to share and reuse the document information across software applications and computing platforms, The Standard Generalised Mark-up Language (SGML) is recommended as the document storage format.

DS0060 recognises two high-level document phases: the consistent creation of individual data modules (the input stage) and the construction of a selection of data modules into a final publication (the output stage). The final publication, to be viewed by the end-user, is called an Interactive Electronic Technical Publication (IETP). An IETP is a technical manual (maintenance, user, training, operations, etc.) prepared in digital form on a suitable medium, designed for electronic screen display to an end-user, and possessing the following characteristics:

- A move away from paper-based presentation of information to modular display.
- A move towards display on PCs and laptops
- A move towards interactivity in providing procedural guidance, navigation directions and supplemental information.
- A move towards multimedia
- A move towards providing a variety of paths to the right information

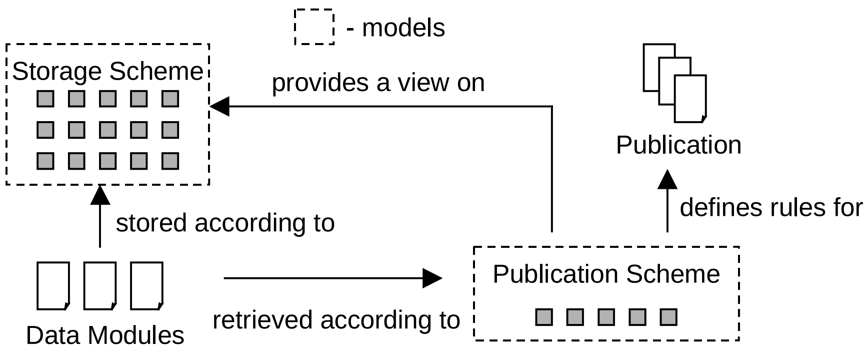


Figure 3-2: An Overview of the IETP Publication Process

The overall publication process of a publication (or manual) from a set of data modules is shown in figure 3-2. It shows a collection of data modules stored according to some classification (storage) scheme and retrieved and published for a purpose according to a publication scheme. The publication scheme provides a restricted view on the entire collection of data modules, and results in an Interactive Electronic Technical Publication (IETP) for a particular purpose. In order to build a collection of 'self-contained' data modules into a useful publication (IETP) a number of levels of abstraction are developed by DS0060; these are shown in figure 3-3 and explained below.

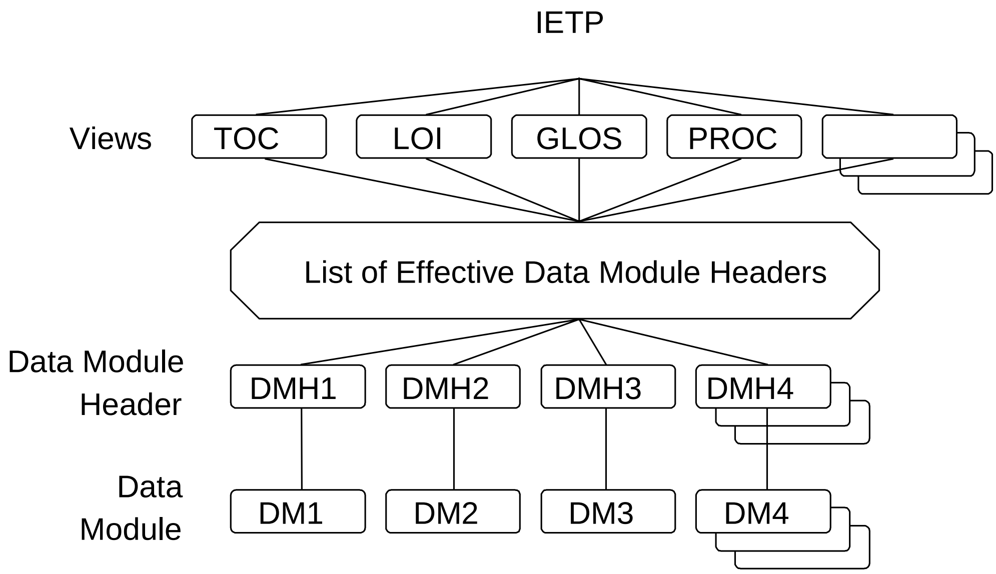


Figure 3-3: The DEF STAN 00-60 Storage Model

Data Module (DM) Layer

A Data Module (DM) is a document file that holds a 'self-contained' unit of data describing a particular operation on a particular piece of defence equipment. Each DM is an SGML file that strictly conforms to the DS0060 data module document type definition (DTD). Each DM consists of a title, a 'status and identification' section and document content. The status represents the security classification and issue number of the data module, and can be disregarded for the purposes of this thesis. The identification section is, however, of significant importance, codifying the document content with respect to the MoD equipment range. The identification code is formally known as the Document Module Code (DMC) and is discussed later in this chapter.

Data module header (DMH) layer

A Data Module Header (DMH) is defined for each data module in the database. The DMH is the only point of access to its corresponding data-module**,** provides both a wrapper of abstraction and entry points into the DM's contents. The DMH may contain any of the following:

- A generated table of contents (TOC) for each data-module
- A glossary for all tagged terms used within the DM
- Access links to different points within the DM content.

A DMH can be dynamically generated from a data module when it is first entered into the database and can also be manually extended. Entry to DMs is only granted at certain points to ensure that the reader does not miss essential document content such as warnings or cautions. This is an obvious concern for a document system that delivers safety-critical information that must not be ignored by the user.

List of Effective Data Module Headers (LOEDMH) layer

The LOEDMH serves two purposes: it provides a catalogue of all the DMs in the database and a switching mechanism for linking public identifiers to system identifiers. The separation of the public identifier from the system identifier is a valuable concept, allowing a data module to be referenced from any of the higher layers without the need to refer to the physical file location.

'Views' layer

The views layer, also known as the output specification, contains an abstraction of the DMs required to form a particular publication. It is at the views layer that all relevant DMs are extracted from the database and arranged for publication. The views layer provides the end-user with a variety of paths to the right information via a method that best suits the current purpose. Figure 3-3 shows a number of typical views:

- The **TOC** (Table of Contents) view does not specify a subset of DMs to extract and publish, but does specify mechanisms for ordering and displaying the DMs contained within the database. Only at the TOC level is there an awareness of a required ordering. Early recommendations for the TOC view go no further than suggesting an ordering identical to that of the corresponding paper-based publications.

- The **LOI** (List of Illustrations) view extracts all illustration DMs from the database and orders them in some way. As with the TOC view, early recommendations merely recreate the structure of equivalent paper-based publications.

- The **GLOS** (Glossary) view recommends the extraction of all glossary entries from the database and their subsequent ordering. Technical documentation inherently contains a number of specialist terms which are conventionally defined in a separate glossary. The GLOS view provides the recommendations for building that glossary.

- The **PROC** (procedural) view is a complex view that retrieves DMs relevant to a specific task. There exists a PROC description for each task that an engineer may need to perform, which gathers together documentary and illustrative materials into a single publication. The PROC view hints at the possibilities for document reuse: a single DM may be referenced in more than one procedural view. A paper-based equivalent contains high levels of duplication to ensure that all printed procedures contain all the required information.

Interactive Electronic Technical Publication (IETP) layer
An IETP is a collection of DMs relating to a pre-determined topic, aimed at being viewed by the end-user. The lightweight IETP layer contains publication-specific information such as a generated table of contents and the current publication status. Access to a publication is recommended strictly through the IETP layer, which provides access to all the DMs within the specified publication.

3.2.1 The DS0060 Document Module Code

The document module code (DMC) is the central mechanism with which to identify the purpose of each data module. The structure of this 32 character enterprise-specific alphanumeric code is shown in figure 3-4. In order to identify the applicability of a DM to a specific piece of defence equipment (type and model), a model identification code (MIC) (13 characters) is provided. A new MIC is used whenever a new type, model or variant is thought to necessitate establishment of a separate database. To qualify the model identification code a single character material item category code specifies which of the tri-service categories the MIC applies to - one of 'ship', 'electronic/automated software', 'aircraft/missile', 'surface vehicle' or 'ordnance'.

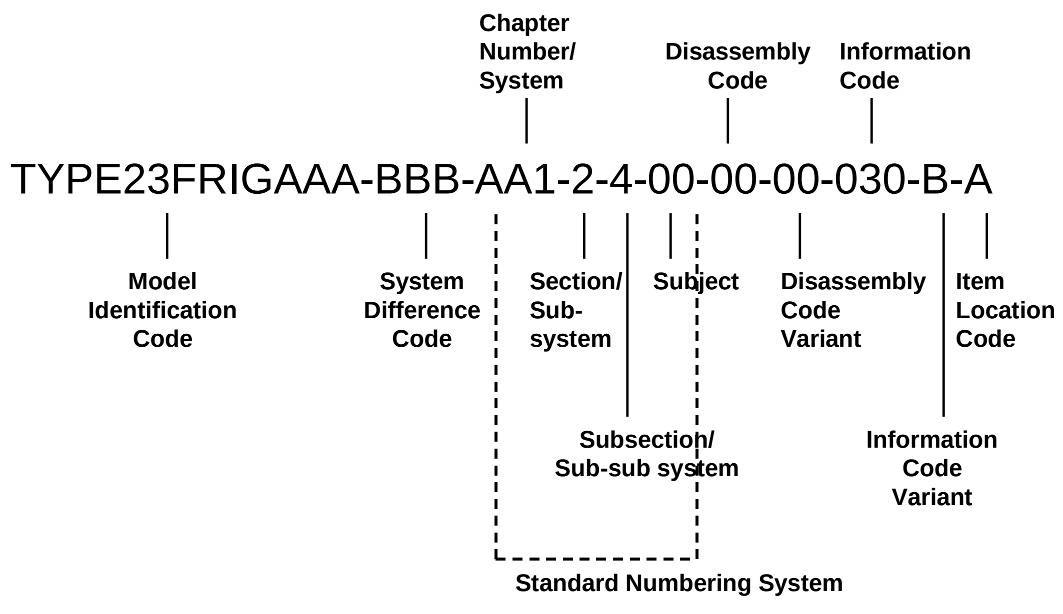


Figure 3-4: The DEF STAN 00-60 Document Module Code (DMC)

The Standard Numbering System (SNS) provides a system/sub-system/sub-sub-system breakdown of the equipment identified by the MIC. Where a sub-sub-system does not provide a sufficiently low level of breakdown the system difference code (SDC) enables differentiation across parts of the same sub-sub-system. The Disassembly Code (DC) is a two character alphanumeric code that identifies the breakdown of an equipment assembly. As such it both allows an identification of equipment parts and suggests a procedural ordering for equipment disassembly. The DC highlights the way in which a particular code can describe either the equipment or a process on that equipment, depending on the context of the user at the time.

Outside the scope of the current DMC but worthy of note are the additional zoning and access codes that assist in referencing assemblies, access panels and doors within data modules and technical publications. The DMC focuses on functional breakdown of military equipment, but these functional breakdowns may not be of particular use to a mechanic who is looking at a physical embodiment of that equipment. Zoning provides a standardised alternative index into relevant data modules by supplying a physical breakdown of a piece of equipment. Such an alternative view of the equipment is valuable, providing a variety of paths to the right information.

To this point all parts of the DMC describe the decomposition of the equipment it services. The Information Code (IC), however, is used to denote the type of information in the associated data module. The identification of both product and process is valuable in facilitating retrieval of information fit for a work-based purpose. A general description of a car tyre, for example, is ineffective in training the reader how to change that tyre. The IC encodes the way in which the DM content describes an operation on the equipment identified elsewhere in the DMC. Once fully defined for a particular data module, a DMC describes the low-level part it describes (product) and how it describes that part (process).

3.3 DEF STAN 00-60 Feasibility Study

This section describes the DS0060 feasibility study undertaken with Logistics Support Consulting (LSC, Tamworth, UK) and the UK Ministry of Defence. At the time of the feasibility study the DEF STAN 00-60 recommendations described in the previous sections had not been proven, and no DS0060 implementation had been produced. The aim of the feasibility study was to provide a technical proof of concept that would enable the MoD to undertake extensive user trials with its designers and engineers using production documentation for a single piece of defence equipment. A set of recommendations followed as a result of working practically with DS0060, which were passed onto the Electronic Documentation Sub-Group (EDSG) within the MoD.

The purpose of the software demonstrator was to complement LSC's existing Module Master product which manages the input phase of the DS0060 document process (figure 3-1). Module Master provides a mechanism for the addition of low-level data modules into the Common Source Database (CSDB) for further processing: the CSDB contains all the DMs for a particular military product. The software demonstrator extended the DS0060 support provided by Module Master, providing a method for producing the DMH, LOEDMH, Views and IETP layers of the architecture, and for viewing the resulting output on-screen. After iteration, the software demonstrator became a full-strength DS0060 product that is now being used to support a number of in-service military products. Its role was two-fold:

- To support the **automatic production of the DS0060 layered model** from input (DM) to output (IETP) phases
- To **provide a simple document browser** that enabled access to the resulting publications

Due to the involvement of a number of people and organisations in the feasibility study it is important to briefly highlight the author's contribution to the work detailed in this chapter. The DEF STAN 00-60 standard was developed by the Ministry of Defence (MoD), adapted from AECMA SPEC 1000D. The author's contribution, through Logistics Support Consulting (LSC) was three-fold:

- the development of software to prove the DS0060 **layered architecture** through the automation of data migration from lowest-level DMs to top-level IETP.

- the development of **viewer software** to navigate the resulting IETP, resolving pre-requisites and supporting hyperlinks between data modules.

- the delivery of a set of **technical recommendatio**ns to the Electronic Documentation Sub-Group within the MoD that were incorporated into a subsequent change of DEF STAN 00-60

3.3.1 Layer Generation

The layered structure of DS0060 architecture serves two purposes. First, it ensures insularity between the stored data modules and the final publication, reducing the impact of low-level changes. Second, it adds value by storing layer-specific information at the relevant layer. A publication's status or security classification, for example, may differ to that of the DMs it contains. With an estimated nine million data modules expected across the three military services, it is clear that manual production of the DS0060 layers is not feasible. Extraction of relevant information from lower-level DMs permits a high level of automatic layer generation:

- The **DMH** contains extracted information from its associated data module
- The **LOEDMH** contains an entry for each DMH
- The simpler **views** contain a list of available DMs and a standard structuring scheme

- The **IETP** layer contains much information that can be added at the time of IETP generation.

A software tool was developed, using an SGML parser to perform the migration from the DM layer to the IETP layer. Each of the layers, represented by one or more SGML files, was navigated and manipulated to extract relevant information and package that information into the layer immediately above. Each SGML file was parsed with an SGML API - Synex Viewport - that permitted dynamic SGML processing and assembly of the SGML instances.

Although DS0060 outlined the general structure of each of the layers in the form of SGML document type definitions (DTDs), the proposed DTDs were syntactically invalid and did not convey all the information required for adequate software production. The project findings passed to EDSG recommended corrected SGML DTDs that would more fully support the role of each of the DS0060 layers.

3.3.2 IETP Viewer

The second deliverable from the feasibility study was a proof of concept IETP browser for the DS0060 architecture. The task of the browser was to create an on-screen representation of a publication for the end-user to navigate[11]. The viewer was implemented using Synex Viewport and Borland Delphi, providing a Windows-based viewer with functionality similar to that of current web browsers, but unavailable at the time. The viewer implements the IETP layout semantic recommended by DS0060, providing a simple interface and a hyperlinks navigation model. A simple TOC (table of contents) view layer was implemented for the demonstrator, providing high-level entry points into the generated publication. Basic

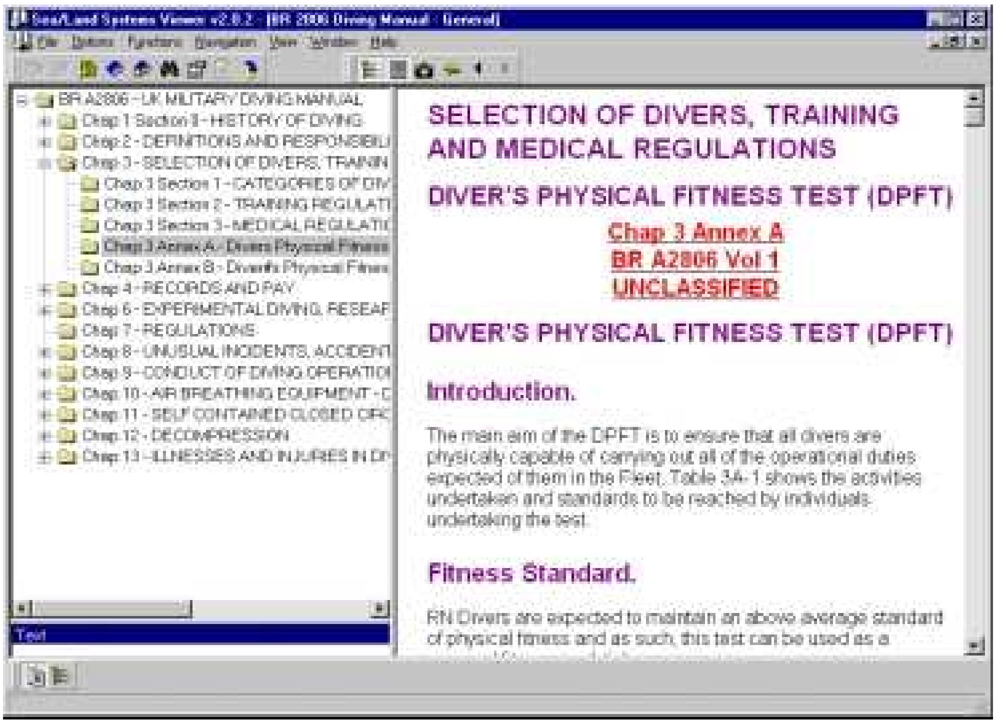


Figure 3-5: IETP Browser Display Showing TOC and SGML Views

1

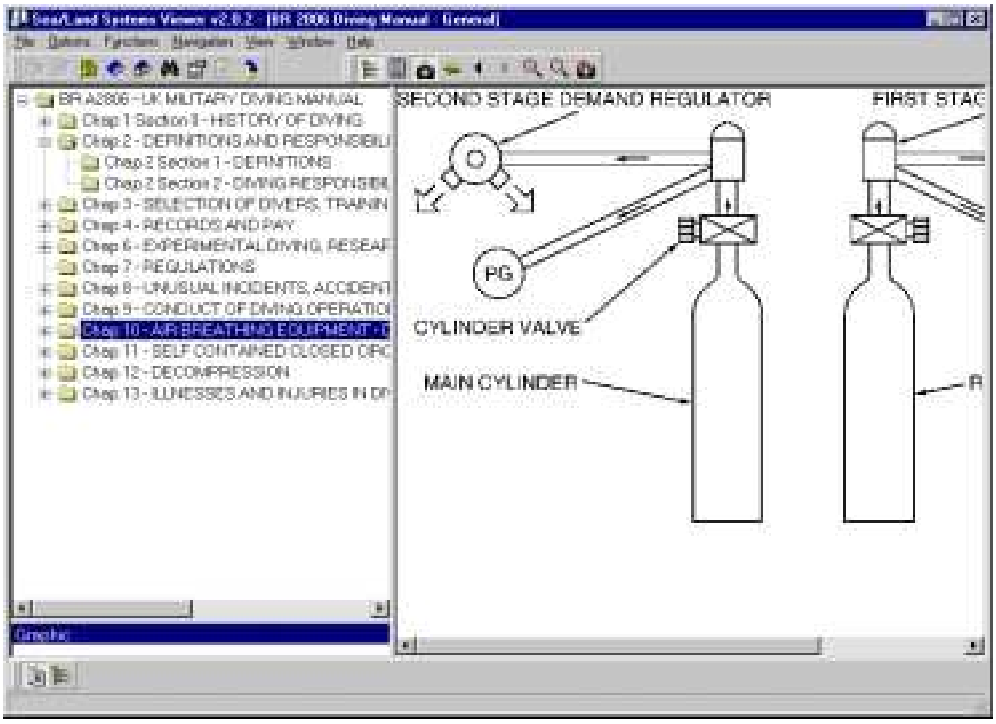


Figure 3-6: IETP Browser Display Showing TOC and Graphic Views

navigation functions provided within the IETP viewer are similar to those available from web browsers - frame forward/backward, simple search, bookmark, history, and navigation of hyperlinks. Further to basic browser requirements the viewer is able to view and navigate the automatically generated TOC at any time. For illustrative data the browser supports basic zoom and scroll functions. Example screenshots of the IETP browser are provided in figures 3-5 and 3-6. Both show the table of contents view on the IETP in the left pane of the display, and in the left pane show a rendered SGML document (figure 3-5) and an illustration (figure 3-6).

Of significant importance to the IETP viewer is a level of control or direction over the user. Whereas web browsers allow the user to freely navigate around the document space, this is undesirable and potentially hazardous in an equipment maintenance environment. Freeform delivery of an IETP is not suitable for task-based navigation, where sub-tasks must be completed in a specific order. Of specific concern to the proof of concept viewer were support for pre-requisites and warnings, cautions and notes.

Pre-requisites

Although the DS0060 architecture describes a data module as 'a self-contained unit of data', in practice this level of separation is very difficult to maintain. DS0060 realistically recognises that a particular DM may have one or more essential pre-requisites to enable its successful use, and provides an explicit mechanism for their representation within the data module DTD. Pre-requisites are created at the DM layer as a set of pointers. A data module with pre-requisites contains pointers to all its pre-requisite DMs by means of their unique data module codes. The storage of pre-requisites has only a minor effect on the storage model but at view-time these pre-requisites must be resolved to ensure all relevant information has been retrieved. Further complexity becomes apparent when the pre-requisite DMs themselves contain pre-requisites.

For example, a maintenance task to fit a new air filter may contain 'remove old air filter' as one of its pre-requisites. In turn, 'remove old air filter' may contain 'locate air pump' as a pre-requisite. The decision of whether a particular DM is essential enough to become a pre-requisite remains at the discretion of the engineer who designs a particular DMC scheme - DS0060 provides only the mechanism, not guidance on the use of such a mechanism. In the example provided, the user should be guided first to the 'locate air pump' DM, then to 'remove old air filter' and finally 'fit new air filter'. The user is provided with a 'Done' button which must be pressed to progress through the list of DMs. This level of control over the user is deemed necessary by the creators of DS0060 because of the potentially hazardous side-effects of freely navigating the publication.

The IETP viewer's task is to resolve all pre-requisites and display all relevant DMs in a suitable sequence. Due to the nested nature of the pre-requisites a doubly linked list was developed that included and ordered pre-requisites. Navigation of the linked list ensures that all pre-requisite DMs are viewed before the DM itself is displayed.

Warnings, Cautions and Notes

Due to the potentially hazardous nature of the procedures described within an IETP, DS0060 advocates three levels of caution that can be associated with any particular data module. These cautions form part of the DM, delimited from the actual DM content by SGML tags. At view-time the viewer application displays any warnings, cautions or notes in modal dialogue boxes before presenting the end-user with the DM content itself. The border colour of the dialogue box indicates whether the item is a warning (red), caution (yellow) or note (black). As with the pre-requisites, restriction of the user's browsing ability essentially highlights information that must be seen and acknowledged before the user is allowed to undertake the procedure described within the DM.

3.3.3 Evaluation

The feasibility study was delivered[12] to the Electronic Documentation Sub-Group (EDSG) - a sub-group of the Defence Technical Documentation Policy Committee. An evaluation of the demonstrator was undertaken by the Royal Navy, who created a large set of DS0060-compliant data modules for testing and

[12] on 3rd October 1996

operational use. Once the DMs had been added to the Common Source Database (CSDB) the following procedure was followed to produce an IETP for a particular need.

1. Using LSC's Module Master product, a set of DMs was extracted from the CSDB for a particular purpose.
2. Using the demonstrator's IETP layer generation facility, the DMH, LOEDMH, Views and IETP layers were generated for the DMs extracted at stage 1.
3. Using the demonstrator's IETP viewer the end-users were able to view the resulting publication, using the generated TOC and DM hyperlinks for navigation.

The demonstrator resolved all DM pre-requisites as the end-user browsed the publication, highlighting any warnings, cautions and notes as they arose. Simple hypertext navigation features were supported, enabling the user to re-trace his path through the publication.

The evaluation was overseen by LSC for a period of nine months, and was judged to have been successful in providing effective access and control over a range of publications generated from a core set of data modules. All parts of the layer production and viewer application scaled to the increased number of data modules in the test environment.

User feedback was mostly reactive, commenting on the functionality of the viewer rather than the method of access to relevant information. Comments on the layer model centred around the views layer where DMs relevant to a particular use are identified. Two views were identified as valuable to the end-users during the trials:

- the **maintenance schedule** that identifies and supports maintenance tasks
- the **illustrated parts catalogue** that provides a graphical breakdown ('zoning') of equipment.

The users commented that rather than searching the system for a technical description, they are far more likely to seek relevant information from the task they are currently undertaking[13]. This observation suggests that users view a publication in terms of the role it serves rather than the subject matter it objectively describes. It further suggests that the construction of descriptive information into task-specific procedures is a valuable exercise, providing a means of assisting the user in finding information relevant to the current need. Although only two layers were identified as particularly useful during trials, 14 different specifications for output have now been developed for various projects. The development of these output

---

[13] This observation was made by Paul Clark of LSC Consulting (Tamworth, UK) during an email discussion on 14th January 1999.

specifications points towards reuse of underlying DMs for more than one purpose, although significant reuse has yet to be realised across the limited domain used in the evaluation period.

3.3.4 Extensions to the Feasibility Study

The feasibility study provided a successful proof of concept of the DS0060 layered model and a mechanism for storing and retrieving data modules. During the evaluation period, the main concerns raised were the amount of time and expertise required to codify both the data modules added to the CSDB and the views layer that provided differing views over the DMs stored.

Automatic Views

Within the DS0060 layered architecture the views layer provides a work-focussed approach to data module collection, combination and customisation. Currently the DMC conveys the product decomposition view and the information view - how the DM describes the identified product or component. The DMC is intended to represent the purpose of a data module without explicitly representing the way in which it is eventually to be used. This inherently provides problems in automatically compiling publications for a specific work-purpose, because the work purpose has been deliberately omitted from the data module's classification. Currently each view is manually constructed, containing a list of relevant DMs to include in that view. The MoD is currently undertaking studies to develop automated methods for collecting DMs together for IETP production by dividing tasks into subtasks. Once a set of subtasks is resolved it is expected that these subtasks can be combined to describe a larger task. The additional layer of abstraction at the views layer is expected to yield additional opportunities for reuse, but this work does not form part of this thesis.

Automatic DM Classification

The 23 character data module code (DMC) is the key to successful DM storage and retrieval, and has been carefully constructed to cover all configurations of current military equipment. The correct DMC assignment to a particular data module is currently a manual task undertaken by experts with a knowledge of both the domain of military equipment and the structure of the DMC. With an estimated nine million data modules across the three military services the manual method of DMC allocation is a major process bottleneck. It would be advantageous if an automatic method could be found for allocating codes to data modules.

Chapter 2 described efforts by the information retrieval community to recognise clusters of documents based on their syntactic content, building a statistical correlation between sets of documents. In a small experiment of a similar kind, the feasibility study analysed a sample of 100 in-service data modules and a statistical syntax-based clustering was found. Word frequency tables were extracted for each data module and 100 *n*-term vectors were produced (where *n* was the number of unique terms across the test set). A graph was plotted with a single point representing a single data module, and a correlation sought between points. No correlation was found, indicating that on word-count alone data modules cannot be clustered. The research literature points to some reasons as to why such clustering may have been unsuccessful [Salton95] [Schweighofer95]. Particular to the DS0060 problem, each DM contains content that describes a similar domain - some engineering process over a piece of military equipment. Information retrieval techniques are most effective over larger document domains where the subject matter differs more widely. Salton suggests that a lower-level analysis of a document's sentence structure may be able to further aid distinction.

Even if DM clustering does improve, there exists the further complex task of actually classifying these clustered DMs. The allocation of a very specific DMC to a data module would require an understanding of both the surrounding document area and the detailed domain of military equipment. Such problems are non-trivial and are not further investigated as part of this work.

3.4 Extension to Other Domains

The document architecture described in this chapter serves a very restricted document domain - one in which documents describe a well-defined domain. The previous section described some of the problems facing the on-going development of DS0060. During the feasibility study a number of other problems became evident, which are also significant to the management of lesser-structured documents in lesser-structured domains. These problems may be seen as inputs into an extended document architecture, rather than problems for the MoD alone.

3.4.1 Change Within the Model

The DS0060 data module code represents the well-defined equipment breakdown and the operations that are possible on that equipment. If the configuration of any piece of military equipment changes there is a corresponding change required to the DMC. This change may be a minor change if, for example, a particular component is swapped for a similar component from a different supplier. Equally the change may be significant, requiring a re-evaluation of a larger part of the DMC. Currently there is no change management mechanism built into the DS0060 document architecture. Instead, the external Logistic Support Analysis Record (LSAR) manages the configuration of the DMC; the LSAR is the repository which controls the build standard, configuration management and design of the equipment. Therefore any change to the product through either amendment or design change will first be modelled in the LSAR and the necessary changes to the documentation produced via an automated link between the LSAR And CSDB[14].

[14] Elicited through Paul Clark (LSC Consulting), 23rd October 1997

The external influence of the LSAR is necessary to retain integrity between military equipment and its associated documentation, but the loose coupling between the document architecture and its configuration manager reduces the effectiveness of the document architecture. Only through very rigid audit control is the CSDB able to accurately represent the current equipment configurations. This rigidity is required within the mission-critical environment of the MoD, but not conducive to the lesser-structured domain of more general-purpose document management. It would be advantageous to be able to absorb enterprise change within the document architecture rather than manage it externally.

3.4.2 Document Classification

The allocation of data module codes to individual data modules is essential for successful storage and retrieval within DS0060, but this allocation is made either manually by experts or, less often, automatically through the LSAR. The problem of assigning the correct DMC is one of both scale and accuracy. With nine million data modules the manual addition of codes is costly and some parts of the DMC are subjective to the expert coding the modules; if a data module is poorly coded it may not be retrieved for an appropriate need. The monolithic nature of the DS0060 classification does not lend itself to localisation of a DM to a particular need, but this cross-enterprise coding scheme is regarded by the MoD as essential. More general and dynamic document management within enterprise may not be so well-suited to the application of a rigorous cross-enterprise classification scheme. The next chapter proposes an alternative model to the monolithic cross-enterprise coding scheme, in favour of more localised method.

3.4.3 An Emphasis on Output

Due to the bottom-up approach adopted by the DS0060 architecture, little effort has been made to date to develop output specifications at the higher layers of the DS0060 model. An effective output mechanism should be able to combine data modules for a number of specific work purposes, creating a high level of reuse of component data modules. The DS0060 community currently constructs very simple output views, listing sets of DMs that are required for a particular task, and then publishing the appropriate manual. There is currently a lack of automation at the views layer in constructing manuals for other as-yet undefined tasks. A more complex views layer is required to identify rules for DM combination into publications. The next chapter proposes a more complex output specification that allows views to be automatically combined into more powerful views.

3.4.4 Networked Document Management

The DEF STAN 00-60 document architecture makes no suggestion as to how IETPs are to be delivered to its end-users. The current IETP Viewer (LSC) is a stand-alone software application that accesses its component DMs locally. This is clearly contrary to the distribution of documents required for a cross-enterprise document management solution. The DS0060 architecture focuses on the production of publications, providing no recommendation of how these publications are to be distributed. For security reasons, the Internet is not a viable medium for military documents, but Internet technologies offer much even to closed networks (intranets). As an extension of the DS0060 ideas, this thesis describes an architecture that delivers documents to its users using a range of currently available network technologies.

3.4.5 Modular Writing Style

The final identified problem is not a technical one, but one of document authoring. Although DS0060 advocates the authoring of documents in a 'modular format', it gives no indication of how this should be done. Within the MoD product model environment documents are well-defined and breakdown is cognitively assisted by the natural breakdown of associated equipment, but other domains are less obviously divisible. The literary problems of dividing a text into a modular format are not addressed in this thesis, but mechanisms for supporting such a modularisation are described in chapter 4.

3.5 Conclusion

This chapter has described DEF STAN 00-60 - a document architecture designed to support a product-focussed manufacturing environment. It has detailed a feasibility study undertaken by the author, in conjunction with LSC, which provided a proof of concept of DS0060 principles and identified potential performance bottlenecks in the DS0060 recommendations. In order to extend DS0060 principles into more general and less restricted document domains, section 3.4 outlined some aspects of the document management process that were not directly addressed by the DS0060 recommendations: management of change within the organisation, the problem of scaling document classification to the enterprise level, focussing on the way in which document fragments may be built into a number of different publications and scaling to the networked multi-user environment.

The DS0060 study provided a technical proof of concept of the ability of SGML to represent both structured documents and the supporting data structures at the DMH, LOEDMH, View and IETP layers. It also provided conceptual validation of the core DS0060 idea – the division of documents into data modules and subsequent construction of publications from collections of data modules. The subsequent reuse at the data module layer provides 'store once, use often' functionality that fulfils a number of DS0060 aims: potentially reducing document production costs, avoiding duplication and increasing integrity of the overall document space.

The next chapter presents the Model-driven Reuse Architecture, based on the core DS0060 architecture and extended to support document management and reuse across a diverse and adaptive enterprise.

4 The Model-driven Reuse Architecture

4.1 Introduction

The previous chapters have described the requirements for an enterprise document management architecture, highlighting the demands that have brought about those requirements. Chapter 2 described a number of research and commercial projects that provide the background to these requirements. Chapter 3 described one such project in depth - the development of an implementation of the DEF STAN 00-60 (DS0060) architecture - and pointed towards some of the problems that were not solved by the existing architecture. The purpose of this chapter is to describe a new document architecture that puts a greater explicit emphasis on document reuse. The new architecture overcomes many of the deficiencies present in DS0060 and incorporates change, hypertext mechanisms and explicit support for document discovery and reuse within the enterprise. Table 4-1 reiterates the demands on enterprise document management from chapter 1 and the high-level response to meet these demands.

Demand	**Increased variation of products/services**
Response	Provide explicit support for the construction of a number of views on the document store
Response	Provide explicit facilities for reuse of existing documents for new purposes
Demand	**Distributed working**
Response	Provide distributed access to the document store
Demand	**The cost and limitations of paper**
Response	Focus on the specific demands of documentation
Demand	**Changes in working patterns**
Response	Provide an adaptive environment that is able to reflect organisational change
Demand	**Long-term storage**
Response	Provide technology-independent document storage models and methods
Demand	**Increased desire to manage knowledge within the enterprise**
Response	Provide a document-model based approach that recognises and manages the content and structure of the documents stored

Table 4-1: Demands on enterprise document management and high-level responses

As the architecture described in this chapter has been designed with document reuse as a main design goal, it is useful to provide a re-declaration of the dimensions of reuse. Chapter 2 recognised that documents comprise three components: content, structure and presentation. It further identified variance in all of the three layers: customisation of the presentation for a particular need, validation of the structure to enable automatic information navigation and extraction, and reuse of the content within those structures. Levy describes an encapsulation of reuse as a four-stage process: creation, collection, combination and customisation. The method by which new document content is created has clear implications on its potential for reuse, and is discussed in chapter 5. The purpose of the architecture described in this chapter is to support the remaining three parts of the process:

Collection To assist the user in finding document content of relevance to the current need, by providing improved access to documents across the enterprise.

Combination To assist the creation of the target document structure, by providing an awareness of the structure models of both the collected and target documents.

Customisation To apply one of a number of presentation schemes automatically, by providing a level of separation between the document content and its presentation.

This thesis recognises that document reuse currently takes place inside organisations today, but claims that reuse is often undertaken in an ad-hoc fashion. A user may know likely sources of useful information, or may be able to use current retrieval techniques to identify candidate resources. The combination and customisation of those resources is, however, likely to be informal; copied and pasted through the operating system's general-purpose buffers and then reformatted for the current need. The architecture in this thesis attempts to formalise these procedures and integrate the reuse process within the document management mechanism. The architecture is referred to as the Model-driven Reuse Architecture, emphasising reuse. The overall aim is to improve document discovery and facilitate reuse of discovered documents for a new purpose. Users are encouraged to reuse existing materials through effective document discovery and reuse mechanisms.

The first stage of the design was to capture the user requirements. These were captured and evaluated using a scenario-based approach. Four realistic scenarios have been devised - one extended from the ILS environment and three from the domain of learning in higher education. They are:

- An engineer performing a change procedure within the DS0060 environment
- An academic validating and producing resources
- A tutor assembling resources for a specific course
- A student following a pre-defined course

To better understand the requirements from the four identified scenarios, a comparison was made with the current VSP resource room provision for the four identified scenarios (see section 4.2.1).

The MRA is based on the Business Object Framework Architecture, which extends the more traditional storage/application/presentation three-layer model into a five-layer architecture (figure 4-1, described in [Schmid98]). Schmid divides the traditional storage layer into a data storage layer and a data abstraction layer, suggesting that a business application should not be concerned with the storage and retrieval of its underlying data. The data abstraction layer isolates the storage mechanism of the underlying data, analogous to the Data Module Header (DMH) layer in the DS0060 architecture. Schmid further suggests a separation of the application layer into a business process layer and an underlying business entity layer. The business entity layer provides a business context wrapper around the underlying data, and the business process layer performs operations on those business entities. Schmid argues that the business entity layer embodies business rules that are common to and used in different business functions, but which are not embodied in the lower data storage layer. The separation of data from its storage, and the separation of business information from business process make the five-layer Business Object Framework a suitable starting point for the Model-driven Reuse Architecture.

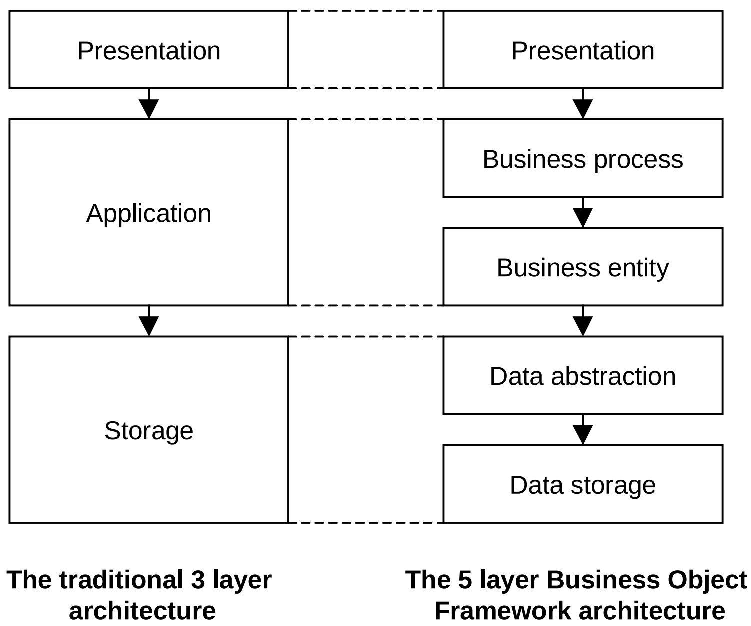


Figure 4-1: The Business Object Framework Architecture

4.2 User Scenarios and Requirements

The case study described in chapter 3 provided one example of document reuse through the use of a single document in multiple technical publications. The limited motivation of DS0060, however, restricts the possibilities for reuse. The DS0060 architecture provides a means to 'store once, use often'; the opportunities for reuse are well defined and manually added to the system. As such it overlooks the valuable mechanism of dynamic resource discovery as an enabler for reuse. DS0060 further excludes explicit support for change within the enterprise: change which is inevitable during the lifecycle of the military equipment it serves, brought about by advances in technology and changes to the equipment supply chain.

In order to extend the ILS ideas two domains are considered to determine requirements and evaluate an implementation of MRA. These are (i) an extended ILS scenario that incorporates change, and (ii) a document management problem within the domain of learning support in higher education. The extended ILS scenario makes explicit the inevitable change that an enterprise undergoes during the lifespan of the information and products it manages. The domain of higher education has been selected for three reasons. First, there are identified requirements to support increasingly diverse and personalised course materials that correspond to an individual's learning needs. Second, the domain of higher education is of a very different nature to the product-based domain of military documentation, adding contrast to the study. Third, the domain of higher education is local to the author's working environment: both human and material resources are readily available. This section describes the two domains and identifies a number of scenarios within each domain.

4.2.1 An Extension of DS0060 to Facilitate Change

When data modules are authored within the current DS0060 environment they are coded with an appropriate DMC, allocated a set of status data and stored in the Common Source Database (CSDB). At retrieval time a collection of DMs is extracted according to some pre-defined rules, built into a publication and delivered to the user; the whole process is managed through the DS0060 architecture. DMs are stored according to the DMC and retrieved according to some function (or list) of those DMCs for a work-focussed purpose. The output purpose is encoded at the views layer of the DS0060 model (figure 3-3 in the previous chapter), providing a mapping from storage (input) model to publication (output) model.

The existing DS0060 architecture relies on the external Logistic Support Analysis Record (LSAR) for its configuration management (see section 3.4.1). It is apparent that in order to absorb the change process internally the architecture must be able to support change to both the DMC coding layer and to all subsequently affected output specifications. Assuming that change control is not managed externally to the architecture, a typical change scenario is described:

User scenario 1

A set of coded documents has been collected in the common source database (CSDB) and a set of output specifications has been derived for a variety of purposes. Tim is an engineer responsible for replacing one brand of cooling fan with another that has different maintenance procedures. Tim analyses the current data module code (DMC) schema, removes the elements associated with the older cooling fan and adds new elements for the newer fan. All documents associated with the new cooling fan are added to the CSDB coded with an appropriate DMC. Tim then needs to find out which publications this change has affected, so asks the system to highlight all publications that reference the old cooling fan. For each of these affected publications Tim adjusts all references to point at the documents he has just added.

This scenario highlights the need to reconfigure the document classification scheme during system run-time and the knock-on effects within the architecture. Since this scenario differs from the DS0060 architecture in only the change management process, the scenario demonstrates how the DS0060 architecture may be implemented within the MRA. The scenario uses only a small part of the new architecture proposed in this chapter due to the limited focus of DS0060. It is clear that the requirements identified and resolved by DS0060 are limited by the need to provide a working methodology for controlled document storage and retrieval. DS0060 contains no notion of data module or publication ownership, and all data modules are of an identical type. The DS0060 architecture's focus on DMs as black-box objects further precludes the application of document techniques to the document store. For these reasons this thesis identifies a second problem domain that further develops issues of a similar nature but which is not limited by its rigid industrial positioning.

4.2.2 Supporting Learning in Higher Education

The research and commercial literature recognises the need to provide more flexible teaching models than traditional university practices currently allow. Current widespread practice in curriculum design is based on defining a set of courses. A course is traditionally a packaging of subject matter offered indivisibly to students over one or two semesters through a series of lectures and assessed through assignments and final examinations [Zarhan94]. It is a common belief ([Rowett98], [Zarhan94], [Lotus98]) that the university model of teaching cannot scale to the needs of lifelong learning and increasingly diverse needs of the student or professional customer. [Zarhan94] argues that the role of some tutors will change from that of imparting knowledge to that of facilitating the learning of the student. As learning facilitators, academics will increasingly spend their time helping students, individually or in groups, to develop and execute their learning programmes. Each learning programme may be abstracted as a limited view on a complete set of learning resources held within the document store. This view is limited by the learning objectives and the tutor's judgement as to what is appropriate to serve those objectives. Information technology is regarded as a potentially useful tool for assisting the tutor in providing such a set of restricted views, and for reducing costs in course material production and distribution.

[Mendes97] explains how education is concerned with transmitting information in such a way that allows the learner to construct a mental model of the information given; it is important to provide learners with information within a context. [Whalley93] warns that hypermedia were not purposefully designed for education: *'One of the great strengths of hypertext, but also a potential source of weakness, is that it is essentially a formless medium.'* Whalley claims that a direct consequence of the fragmentation effect in hypertext is that it is likely to make it more difficult for a learner to perceive the author's intended argument structure, unless certain linearity constraints are imposed on the hypertext form. This thesis argues that providing a layer of mediation between the document store and the student, provided through the tutor's experience and judgement, adds a form of purpose to the more 'formless medium' of a large document store.

The particular needs of supported learning in higher education are two-fold: the accurate collection of a set of candidate learning resources and the selection of a subset of these resources to satisfy a particular learning demand. The need to collect resources of a common nature is currently supported by a number of efforts. [Gehringer98] describes an attempt to collect together course materials on teaching Object-Oriented methods. Steve Beaty from the Metropolitan State College of Denver has developed a similar list of materials covering a range of computing-related courses[15]. These efforts provide a list of discovered course materials and attempt to classify them into broad categories (learning level and subject area), but provide little indication on how these materials may be used. The Instructional Management Systems (IMS) project [IMS99] attempts to specify a more rigid classification structure for learning materials through a set of meta-data schemas. IMS work is still in its early stages, however, and no recommendations have yet been made regarding further processing of the coded resources.

The use of documentation in supporting learning in higher education is very different to that within the MoD, but its management has similar characteristics:

- There is a desire to move away from paper-based resources towards electronic representation of those resources
- There exists a variety of end-users with differing requirements from such documentation
- There is a need, at least from an overall management perspective, to provide an integrated environment in which the organisational memory can be stored, retrieved and reused.

---

[15] http://lamar.colostate.edu/~beaty/ (Last verified 8th March 2000)

- There is a need to restrict the bounds of 'hypertext' to the end user to avoid confusion and to retain focus
- There is a constant need to reduce document preparation and distribution costs.

The domain of learning is less rigidly structured and more progressive than the military domain, requiring fewer security considerations and encouraging cooperation and openness. The learning domain therefore permits, and would hopefully encourage, the use of hypertext mechanisms within its documents. It also highlights the issue of document ownership, whereby value may be added by associating a resource with its author or owner; the ownership of documents within the DS0060 recommendations is mostly disregarded. The provision of increasingly varied delivery of educational courses calls for the reuse (or re-purposing) of course material, which universities regard as their intellectual property. The scenarios created in this section identify a number of roles performed within the domain of increasingly diverse educational course provision. These roles exemplify the more general roles of information provider, information mediator and information user that exist in many modern information systems, signifying the application of resulting methods to other user domains. Figure 4-2 shows the process of producing a diverse range of courses, and identifies three roles: an academic producing resources, a course tutor constructing a course and a student undertaking that course. Each of the three roles is explained in the three scenarios that follow.

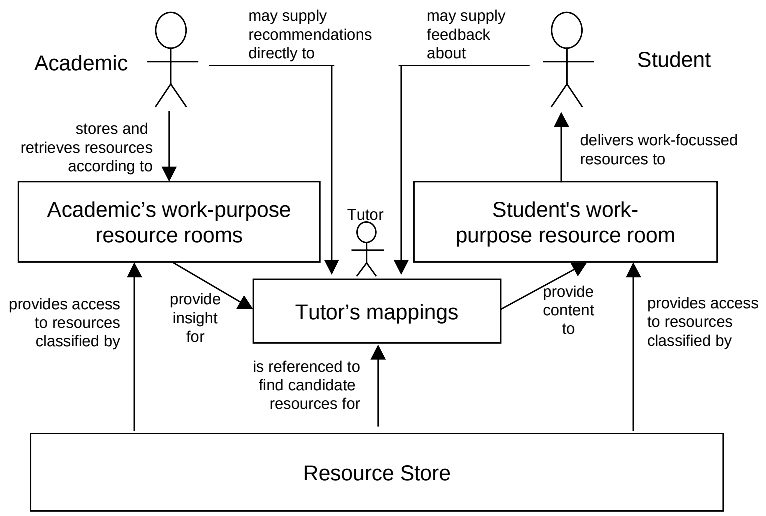


Figure 4-2: Supporting the Document Process for the Learning Domain

User scenario 2 – An academic validating and producing resources

As an academic, Paul produces resources for a particular reason – either a taught course or a non-teaching reason. As a resource provider he has a new resource that he would like to store and later access for maybe more than the current purpose. Paul stores the resource in his own workspace (resource room), which may or may not reflect his currently perceived usage of that resource. He may add keywords or a short description of the resource, and can access the resource from any web browser. Paul knows that this *resource is useful to a course that he is running, so makes it part of that course by pointing to it from the resource room for that course.*

Mid-year Paul decides that his personal classification scheme needs updating or refining. He would like all knock-on effects to be changed automatically. He also wants to know how often his resource has been seen and by whom, whether anyone has expressed an interest in his resource or any of his folders, and where else within the system his resource is being used. This information allows Paul to evaluate whether it is worthwhile keeping it up-to-date. Paul would like control over his resource, even though others may use it. Out of interest, Paul views what else has been viewed by the people who have seen his resource, and finds a document of interest from the resulting list.

User scenario 3 - Tutor assembling resources for a specific course

As a tutor you are in charge of a module called 'Advanced Computer Graphics' and a course called 'Business Information Systems'. First you would like to put together a set of learning resources for the graphics module from resources already available. You want to put together a set of materials on 'gourad shading', and so perform a search across the keywords of all the available documents. A list of relevant documents is displayed, and can be browsed for suitability. You copy a number of the candidate resources to your workspace for later reworking. You see that Phil Smith has authored a number of these resources and view the part of his resource room that these are stored in. After browsing Phil's classification you find another two resources that are of use to your current need and copy them to your workspace. You also see a folder within his resource room that may be of use to a future need and bookmark that folder for later inspection. You view Phil's personal details and contact him by email to ask for further assistance.

You want to include some materials on modern CAD/CAM methods, and know the name of an expert within the University – Chris Taylor. You search through the organisation for his name and find that there are two Chris Taylors, one belonging to the School of History and one belonging to the School of Mechanical Engineering. You search through the documents owned by the School of Mechanical Engineering and find another two resources that Chris Taylor has written that may be of use, and you copy these to your workspace.

You realise you have not yet discovered any resources about a particular transform that you would like to provide. There are no resulting documents from a keyword search, so you perform a more in-depth document search, discovering a single relevant resource. This resource is classified under two resource rooms – 'Tom Daly's Resource Room : Graphics Algorithms' and 'School of Computer Studies : Theory of Graphics'. Upon inspection of the 'Theory of Graphics' folder you decide that all the resources therein may be of interest to the course you are running, and create a pointer to that classification folder.

You return to your workspace and decide a structure for the resources you have collected. You would like to make a static copy of some of the resources and retain a live link to others. You apply an ordering to the collected resources, generate the resulting resource, apply your presentation rules and place it in the classification 'Advanced Computer Graphics : Lectures'.

Once you have constructed the Advanced Computer Graphics module, you would like to add it to the overall Business Information Systems (BIS) course. From the BIS resource room, you create a bridge to the Advanced Computer Graphics module, including all resources from the single module within the overall course resource room. You locate all the resource rooms owned by the School of Computer Studies and, after contacting the owners of three of the modules, you include those modules within your BIS course reading room.

This scenario moves away from a one-course-one-tutor principle towards a tutor-mediated model. The extended scenario highlights a number of features of the proposed architecture: search of the document/metadata space, navigation of the document model, registration of interests, navigation of organisational models, reuse and re-purposing of resources. The reuse of existing materials and application of new presentation rules highlights the ability to repackage resources for a new purpose.

User scenario 4 - A student following a pre-defined course
As a student Maria has agreed to undertake a particular course (Business Information Systems) which is delivered on-line. Maria's tutor has approved the particular course she is undertaking and her resource room has been prepared containing a number of modules and a personal working space. Maria is allowed to navigate around her resource room according to its hierarchical structure. There are occasional hyperlinks to other materials elsewhere. These links may have been applied by the original resource author, the tutor or another student - Maria chooses only to follow links that her tutor has added to the resource room. Maria also sees that other students have annotated the learning materials with points of clarification, and she views those annotations as she progresses. She chooses to add her own annotation to one of the documents and contact one of the students who has made an annotation to the same document. Maria also finds a web site that she thinks may be useful to others on her course, so places it in her work space with a description and creates a link from one of the existing resources to the web address.

This scenario highlights the end-user perspective, which provides a view on the resources collected and structured in scenarios 2 and 3. It introduces the use of hypertext approaches such as annotation and hyperlinking

4.2.3 VSP Resource Room Provision for Scenarios

VSP resource rooms, described in chapter 2, provide a suitable basis from which to evaluate an MRA implementation, for a number of reasons:

- It is representative of the current class of information management solutions that allow distributed access to document resources. Another example of such a system is BSCW.
- It supports the provision of information to a particular domain or organisation and its members, rather than providing a general-purpose user-unaware solution such as those proposed by Autonomy and Excalibur.
- The structure of VSP resource rooms represents an organisational purpose for the documents stored within that structure through its resource rooms. The existing resource room structure provides a classification scheme that the MRA can extend. Other architectures such as Microcosm do not provide such a tight binding with work purpose.
- Its architecture is Internet-driven, rather than Internet-enabled. Some document management products such as Lotus Notes provide only a subset of their total functionality through a web interface.
- Its designers and users are available locally, facilitating design feedback and evaluation

The original VSP resource rooms were not explicitly designed to support scenarios such as those described earlier in this section. It is useful, however, to assess how the VSP would best manage the problems highlighted in the scenarios to enable evaluation of the MRA. This section is a brief statement of which parts of the scenarios would be currently possible by using VSP resource rooms, and which would not. The current VSP resource room entity-relationship information model is shown in figure 4-3, using the IDEF1X notation[16].

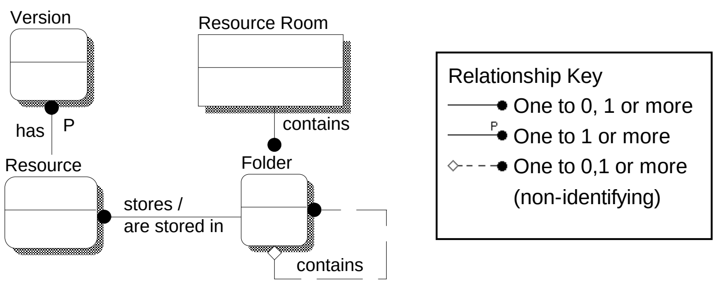


Figure 4-3: VSP Resource Room Information Model

[16] A description of the IDEF1X notation lies outside the scope of this thesis, but more information can be found at IDEF web site at http://www.idef.com/overviews/idef1x.htm (Last verified 8th March 2000)

User Scenario 1 - An extension of DS0060 to facilitate change

The VSP would find the whole DS0060 architecture difficult to implement with its current architecture. DS0060 relies heavily on the different views available over the complete document store. Although the VSP would be able to store the data modules according to the DMC scheme, mapping from input to output schemas would prove difficult due to the lack of inter resource room support. The change process described in scenario 1 could not be fully supported because there is no means of managing the multiple classifications associated with a particular document.

User Scenario 2 - An academic validating and producing resources

Current VSP resource rooms would enable Paul to add new resources (and associated meta-data) into his resource room, but would not be able to support the multiple classification of the resource in both his resource room and the course's resource room. Also Paul would not be able to find out who had registered an interest in a particular resource, nor be able to restructure his resource room.

User scenario 3 - Tutor assembling resources for a specific course

Although the VSP could support the creation of a resource room for a particular course, the reuse of existing VSP resources within that room would be very difficult using the current architecture. A keyword search is possible, but a full-text search through documents is not. The current resource room architecture does not support an association between a resource's author and the corresponding name within the VSP tenancy, so as a tutor you would not be able to find out who the author 'Phil Smith' was. The VSP does support the registration of interests in a particular resource room folder.

The most effective way you could find work by Chris Taylor in the VSP is by finding his details and checking his official publications. The list of publications is, however, unconnected to the resource rooms. The current resource rooms provide only a space for storage and retrieval of documents, so it contains no mechanisms for resource reuse. There is no inherent separation of document content, structure and presentation, so reuse is harder to automate.

User scenario 4 - A student following a pre-defined course.

The VSP can provide a course-specific resource room, which Maria would be able to use to undertake her course. Maria would not, however, be able to distinguish between hyperlinks added by her tutor from links added by other students or other users of the system. Neither is annotation supported in the current VSP implementation

4.2.4 Resource Rooms Shortfalls

The scenarios highlight a number of shortfalls of the current resource room implementation. Some of the shortfalls may be resolved by increasing functionality over the current architecture. More fundamentally, other shortfalls highlight architectural omissions in the VSP resource room architecture. These are:

- Within the VSP resource rooms, each resource is connected tightly with the folder it was originally stored in, tying the document to its original purpose. Support for pointing to a single resource from a number of resource room classifications is possible, but currently disabled due to the complexity of the process to the end-user. The nature of the resource room information model, however, hinders large-scale re-purposing of documents because of the tight binding of a document to a specific (primary) resource room folder.

- The VSP currently supports the addition of lightweight hyperlinks between resources and the registration of interests in a particular resource or folder. However, the VSP provides no architectural support for hypertext approaches such as document history trails, typed hyperlinks and annotations. These are fundamental omissions from the current architecture.

- The VSP resource room architecture does not provide a document-model based approach. Beyond the meta-data that the document owner must supply at document check-in, reading room resources are considered as Binary Large Objects (BLOBs). There is no distinction between a resource's content, structure and presentation. It is therefore difficult to promote reuse of the document resources for more than their original purposes.

- There is no direct link between the resources held within the VSP and the VSP information directory. It is impossible, therefore, to locate information about a document's author except through a disconnected text search across the information space. This connection is essential to facilitate collaboration once a document of interest is found.

It is the intention of the Model-driven Reuse Architecture to provide a document management model that retains many of the resource room facilities but provides an architecture that explicitly supports multiple classification of resources, associations between documents and their authors/owners, hypertext methods and reuse of existing documents.

4.3 MRA Models

Traditional management information system (IS) developments follow a software lifecycle that considers the domain to be supported at the requirements collection stage. Subsequent changes are incorporated into a feedback cycle, which revises the requirements and changes the designed architecture accordingly. IS support is often independent of other systems, creating a series of stand-alone products lacking integration [Amor97]. The model-driven approach described in this thesis differs in that it provides a domain-independent architecture inside which domain-specific models are placed to create an implementation for a particular purpose. The domain-specific models represent the business models of an enterprise, sensitive to current and changing needs within the enterprise. A change in business practice is represented as an adaptation of implementation models rather than a change in the architecture itself. The MRA contains two sets of models which enable an MRA implementation:

- a set of **enterprise-specific business models** that classify the entities and processes which the documents service within the enterprise

- a **domain-independent information model** that describes the entities and relationships within the document architecture.

4.3.1 Enterprise-Specific Models

The enterprise-specific models ('business models') provide a representation of the enterprise inside which an MRA implementation is created. The MRA recommends three different types of business model:

- **Classification models**, which provide a number of work-focussed views on the document store. A single classification model may resemble a traditional file structure or may be more complex, providing differing views on the same information. A single classification model is analogous to a VSP resource room structure.

- **Organisational models**, which provide a view of the organisation within which documents are managed. Organisational models provide a description of the people within an organisation and the parts of the organisation to which they belong. An MRA organisational model for the University of Leeds, for example, provides the structure of the University and its schools, and information about the people within those schools.

- **Document models**, which provide the structure of the document types used within an organisation. A document type may be as simple as an office memorandum, or may be more complex – a DEF STAN 00-60 IETP, for example. By making explicit the form of the document and providing a structure model of that form, the MRA is able to manipulate a document's content, structure and presentation aspects.

One of the key ideas within the MRA is the association of a particular document with parts of the MRA business models that best describe its purpose and position within the enterprise. The document at the centre of figure 4-4 is shown to be of a particular document type, belong to a particular part of an organisation and be classified according to some part of a classification scheme. The overall classification of that document describes not only the subject area or purpose of the document, but also the person and organisational element to which it belongs, and its document type. An extension of this key idea is of further significance - the ability to multiply-classify a document. A multiply-classified document retains only one document type and one owner/author, but may have a number of different classifications.

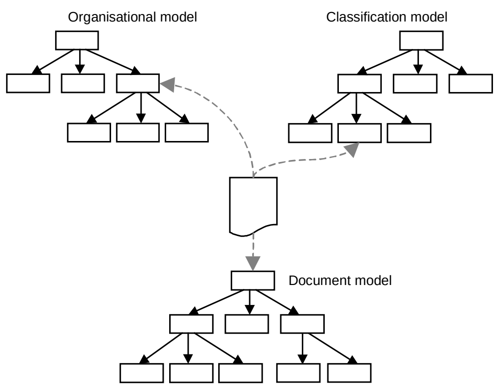


Figure 4-4: MRA Business Models

The user scenario described above will be used to illustrate the business models. First, the DEF STAN 00-60 domain provides a number of classification models against which its resources are stored and retrieved - a single 'input' model and a number of 'output' models. The input model is the classification according to which all data modules are primarily attached, specified by the data module code (DMC) schema. Figure 4-5 shows a small sample of the DMC scheme. It comprises two hierarchical decompositions: one of the equipment used within the MoD and the second of the functions that can be performed on such equipment. The DMC identifies the piece of equipment described and the procedure on that piece of equipment. As an example, the data module at the foot of the figure 4-5 is identified as describing the locking procedures on the nuclear propulsion of a naval ship.

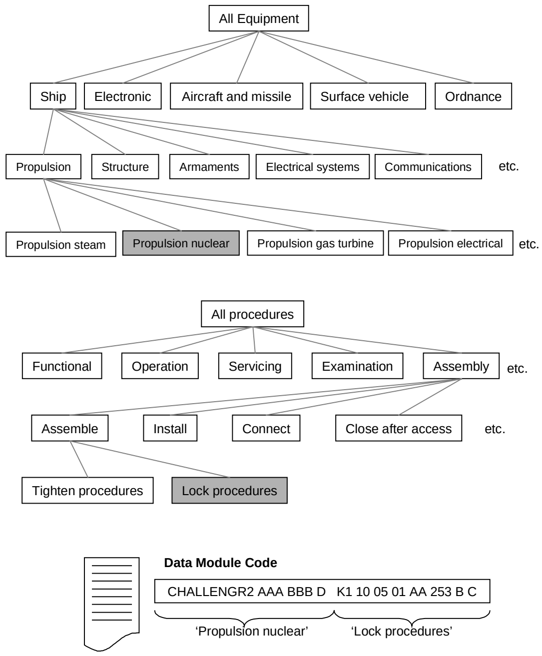


Figure 4-5: The Classification of a DEF STAN 00-60 Data Module

In contrast to the extensive input model, each output model within DS0060 represents the structure of a single technical manual that is published for specific purposes. Figure 4-6 shows a typical publication structure, which resembles a traditional book-like structure. If the data module coded in figure 4-6 is deemed to be useful to the 'Check structure' sub-section of the 'Monthly maintenance' section of the publication, a connection is made between the data module's DMC and the 'Check structure' element of the publication (output) classification model. The DM may be of use elsewhere in the same publication or in other publications: this is the essence of the 'reuse' made possible using DS0060.

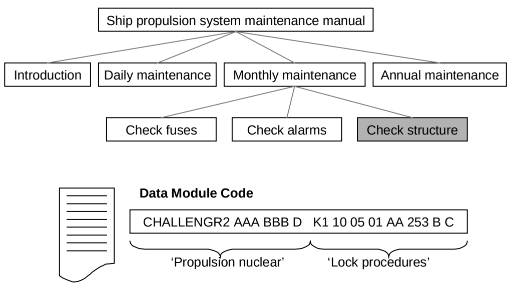


Figure 4-6 - Document Module use within a Publication

Whereas the classification models prescribed by DS0060 are extensive and complex, the work-based models for the supported learning in higher education scenario are typically much smaller-scale. Two example classification models are shown in figure 4-7 - one for an individual academic's resource room, and one to support a single taught course module. It is apparent that a resource stored in 'Dave's Resource Room : Microcosm' may be useful to 'VWE Module : Microcosm'. If a resource from Dave's Resource Room is found to be of use, it is further classified with a code representing 'VWE Module : Microcosm'. A single resource is now being referenced from two resource rooms - it has effectively been re-purposed.

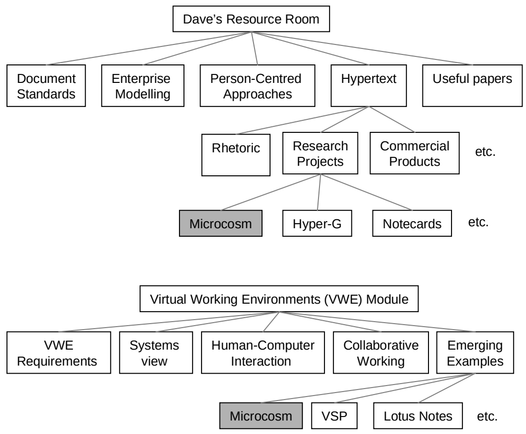


Figure 4-7: Two Classification Models within the Learning Domain

Classification models describe only one part of the overall model-driven nature of the MRA. Organisational models may also be constructed, representing the structure of the organisation within which the MRA implementation is to be used. A typical organisational model is shown in figure 4-8, showing the hierarchical decomposition of the University of Leeds. At the lowest level of the model are the people who belong to a particular part of the organisation.

Lastly, a representation of the document types within the organisation permit the manipulation of document structure and content. The representation of document models is essential for effective document manipulation and reuse. The form of these document models is described in the next chapter.

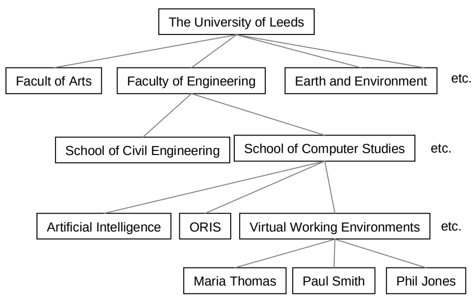


Figure 4-8: Part of the Organisational Model for the University of Leeds

4.3.2 The MRA Information Model

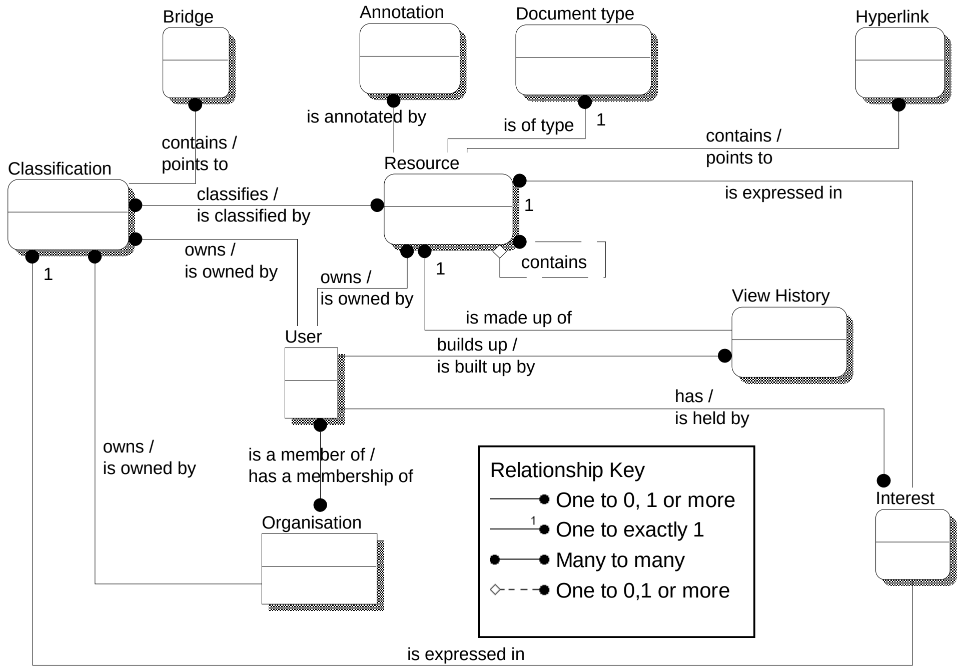


Figure 4-9: The MRA Information Model

The business models described in the previous section provide a domain-specific representation of the working practices and structures of an organisation, which may change as the enterprise develops and finds new ways of working. The underlying MRA information model, however, remains constant, providing a consistent means of navigating the enterprise-wide resource environment. The MRA information model, represented in the IDEF1X notation in figure 4-9, describes the relationships between document objects, the domain-specific business models and other document entities – hyperlinks, annotations and history trails. Notably, a resource may contain any number of sub-resources, each of which may possess its own classification, document type, view history, etc. Each of the entities shown in the E-R diagram is a first-order object with its own attributes and its own internal navigable structure. Although the business models are sensitive to change within the organisation, the information model that uses the business models remains unchanged, providing a stable means of navigation through the document space.

The information model captures aspects of modern hypertext and information systems with the aim to use these aspects (hyperlinks, bridges, annotations, history trails, personal interests) to both assist document discovery and add value to the resources in the work-focussed environment. Although the MRA information model need not change in response to changes within the enterprise, it is extensible, able to absorb new methods of information retrieval and hypertext support as they arise. The nature of the architecture design is such that new methods can be added and utilised without the need to alter the resources that are currently stored.

4.4 Description of the MRA

This section describes the Model-driven Reuse Architecture that provides improved enterprise document management and discovery, and a number of reuse mechanisms at a number of levels.

4.4.1 Overview

The high level requirements for the Model-driven Reuse Architecture are:

- **To support the association of documents with their work-based purposes** across an enterprise by supporting a number of localised classification schemes and by associating documents with their owner and organisation.
- **To support existing best-practice document management techniques** to improve document discovery and collaboration
- **To provide mechanisms to support change** within the document classification and across the enterprise
- **To facilitate reuse of documents through:**
  - Low coupling of documents and their classification, supporting multiple classifications of the same document
  - More effective support for document discovery
  - Open representation of document resources, facilitating document combination and customisation.

Figure 4-10 shows a high-level framework architecture that services these high-level requirements. Based around the five-layer architecture for business object framework described in [Schmid98], its main components are:

- A **document object** store that stores documents as authored
- A **document abstraction** layer that separates the physical location of the document from its identification and stores meta-data about the document itself.
- A **document classification** layer that models the domain within which documents are stored and used.
- A collection of **document services** that navigate and manipulate the business models and the documents within the document object store
- A **user interaction layer** that provides input from the user and presents resulting documents to that user.

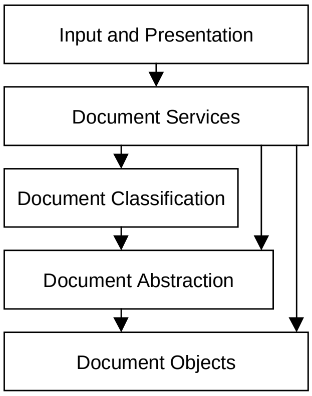


Figure 4-10: The MRA Framework Architecture

The connecting lines between the framework components represent the way in which higher layers provide functionality or abstraction over the lower layers. The value of such abstraction is explained as the rest of this section describes the form and purpose of each of the five layers of the MRA framework architecture.

Document Objects

The lowest level of the architecture is the distributed document object (DO) store that holds all documents in their original authored format, encoded in an open mark-up language. Each document is stored on any networked storage that can be uniquely addressed and accessed across the enterprise. Figure 4-11 provides an example implementation of the DO layer. There is no awareness of any document purpose at the document object layer, nor any connection between any two document objects. Document objects do, however, have a recognisable structure that may be navigated and used by higher layers.

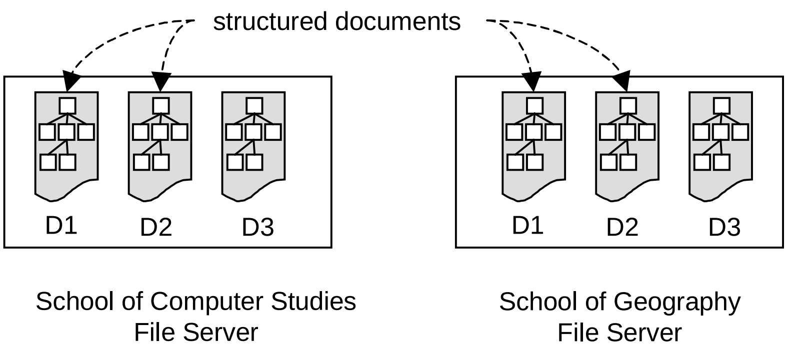


Figure 4-11: The Document Object Layer Implemented Over Two Servers

The DO layer notably allows the distribution of the document store across the organisation. The advantage of such a distributed system is the retention of local control over documents while providing global access across the enterprise. In figure 4-11, although the documents stored within the two file servers have the same file names (D1, D2 and D3), a namespace mechanism permits their unique identification across the enterprise. Document D1 on the School of Computer Studies file server is referenced externally by the storage identifier 'School of Computer Studies : D1'. By using such a mechanism the document owner retains ownership and responsibility over local file naming, while being able to supply its resources into a distributed environment. The open storage of documents at the DO layer promotes the ability to search the document and to provide further manipulation of document content and structure at higher layers of the architecture. The DO abstraction upholds the MRA as a document-model based approach and is critical to successful resource reuse.

Document Abstraction

The document abstraction (DA) layer provides two types of abstraction:

- the separation of document and system meta-data from the documents they describe
- the separation of the physical location of the document from the way in which it is referenced in higher layers.

The DA layer provides a number of different types of meta-data associated with a document at the DO layer. Document meta-data are added either manually or automatically when a document is first added to the document object store. Some meta-data, such as the document's owner, are implied from the user information of the person adding the document to the document store. Additional system meta-data, referring to external entities that relate to the document in some way, are added at system run-time. The MRA provides facilities for the additional storage of hyperlinks, annotations and interests at the DA layer, separating such information from the documents themselves. The MRA regards hyperlinks, annotations and interests as first-order objects, attributed with their own meta-data (creator and date) so that they may be selectively applied to the documents they are logically attached to. The following meta-data elements are supported for each document in the document object store.

Document meta-data added manually when a document is added at the DO layer

- Title, Description, Keywords, Author

Document meta-data added automatically when a document is added at the DO layer

- Version, Timestamp, Owner

System meta-data added automatically at run time by the document services

- List of 'hits', updated each time a document is viewed
- List of interested users, updated each time a user expresses an interest in the document
- List of annotations, updated each time an annotation is created for the document
- List of links, updated each time a link is created to or from the document

Further to the addition of meta-data, the DA layer also provides a pointer to the storage location of the associated document, providing a layer of abstraction between the actual storage location and the identifier by which all higher layers refer to that document. The storage location may change over time as storage methods and media improve in performance or cost [Reagan95], but the internal system identifier remains unchanged. The MRA provides this layer of separation at the per-resource level, eliminating the need to align the physical file structure with the semantic classification structure in any way. Figure 4-12 shows an example implementation of the document abstraction layer, which is implemented on an MRA server.

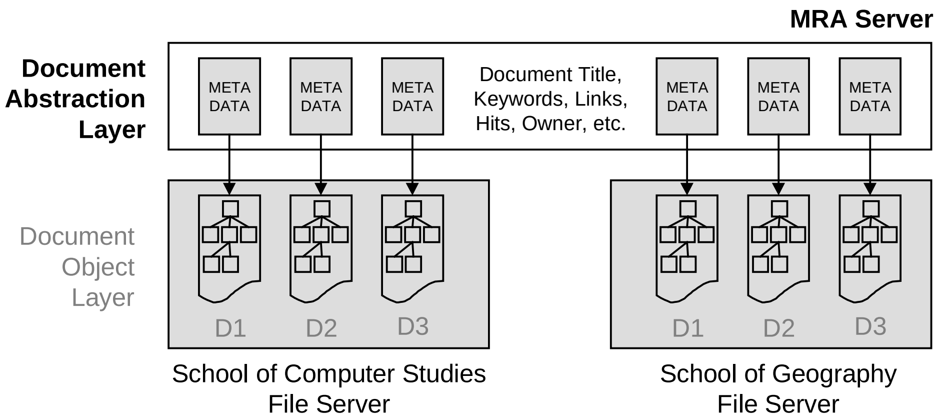


Figure 4-12: An Example Document Abstraction Layer Implementation

A number of different projects described in chapter 2 highlighted the value of storing and retrieving meta-data in order to store information about the document that may or may not be expressly provided in the document itself. In some systems (e.g. Lotus Notes, VSP) the extracted meta-data are the lowest level of searchable document-related data. In others (e.g. WWW) meta-data represent the emphasis or focus of the document that may not be discernible by a full-text search. The DA layer additionally provides a layer of separation between document content and its associated hyperlinks and annotations. This separation is regarded by the hypertext community as advantageous, and implemented in a number of current hypertext systems (e.g. Microcosm, Hyper-G). The separate storage of typed hyperlinks and annotations permits their selective application at run-time, providing a restriction on the hypertext capabilities if required.

The storage of a document's view history at the DA layer serves a number of purposes. It not only provides feedback to a document's owner, but the analysis of such trails has proved successful in identifying users with common interests for potential collaboration (e.g. the MEMOIR project [DeRoure98]). Furthermore, individual users may view their own document trail to return to documents already viewed – this provides similar functionality to that provided by current web browsers.

Document Classification

The document classification (DC) layer is key to the MRA architecture, providing a representation of the business models described in section 4.3.1, and associates the documents with these models. The association between a document and the business models describes not only the work-focussed purposes of the document, but also its owning organisational element and person. At its simplest, the document classification layer provides a single model by which all documents in the document store are classified, stored and retrieved. Alternatively, and more powerfully, the DC layer may store

- as many **classification models** as are deemed useful to the organisation
- any number of **organisational models** and their members
- any number of **document models**, each of which represents a particular document type

The MRA classification layer does not assert any necessary dimensions to either the number of business models or the nature of each of the models. Each business model is an entity-relationship (E-R) model that describes the enterprise entities that are captured within that view. Each E-R model may capture a different view of the same enterprise, or different parts of the same enterprise from the same view. Figure 4-13 provides an example of the document classification layer, providing a single document with two classifications within the MRA business models.

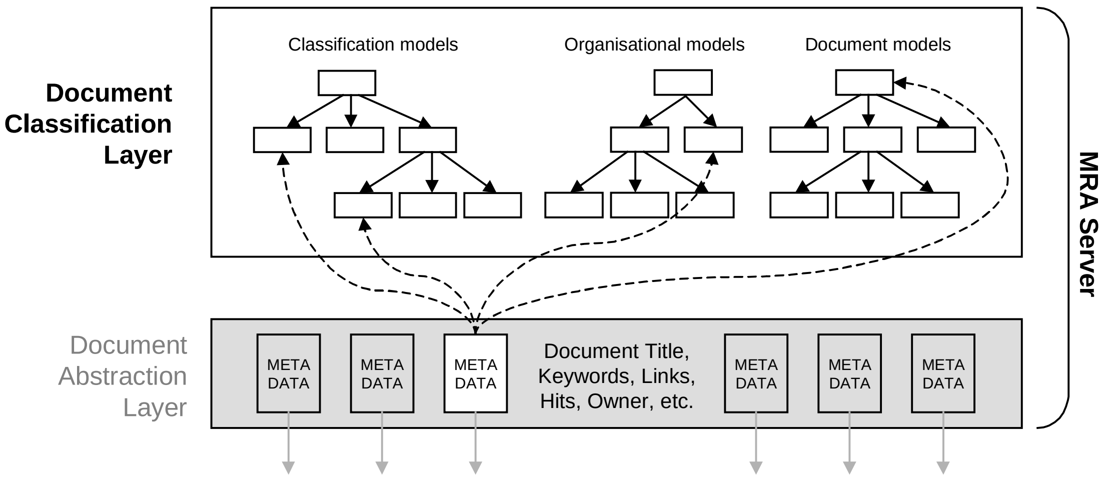


Figure 4-13: A Multiply Classified Document at the MRA Document Classification Layer

The DC layer attaches documents to nodes within the constructed E-R models and provides access to those documents by subsequent navigation of the E-R models. A multiply classified document is simply attached to two or more distinct parts of the classification models at the DC layer.

The classification of a document within the MRA implicitly adds meta-data that describe the document's work-focussed purposes or roles. These meta-data are added as a side effect of a user performing routine operations over the document store, requiring no additional user input. The meta-data are able to assist in the discovery of a document by its classification rather than by its content. This is a novel idea that adds value to the document discovery process, with no additional user intervention.

The MRA document classification layer is able to support a wide range of classification models, but some current projects advocate the use of a single unified classification structure. Some enterprise modelling efforts, discussed in chapter 6, recommend a single thoroughly considered enterprise-wide classification that encompasses all elements of the enterprise in a single model. There are, however, a number of problems with a unified model approach, identified in chapter 2. There exists the possibility that an ill-conceived cross-enterprise model may not effectively serve any single sub-group of people within that enterprise. Unified enterprise models are non-trivial to construct and are not suited to dynamic and virtual enterprises. Although DS0060 seemingly advocates a single classification - the DMC scheme – each of the IETP output specifications is a model in itself and in the MRA would be located at the DC layer.

One of the key ideas to cross-enterprise document reuse proposed in this thesis is the recognition of parts of one classification model that may be of use to another classification, and the construction of bridges between these classifications. For example (figure 4-14) a resource room that contains resources for a course module called IS14 may be of use to more than one course, each of which would have its own resource room. The inclusion of a bridge in both classifications to the 'IS14 Module Resource Room' classification would effectively supply both courses with the relevant resources. An implementation of the MRA could make this bridge appear to be part of each of the course classifications, effectively creating reuse at the resource room level.

The ability to adapt a classification at run-time reduces the risks of 'premature organisation' [Halasz88a] whereby a system becomes unusable due to the restrictions placed on classification structures at the system design stage. MRA support for analysis of knock-on effects is also valuable - to allow a user to change a particular classification without analysing knock-on effects would lead very quickly to a fragmented system, and the argued value of cross-linking would be quickly lost.

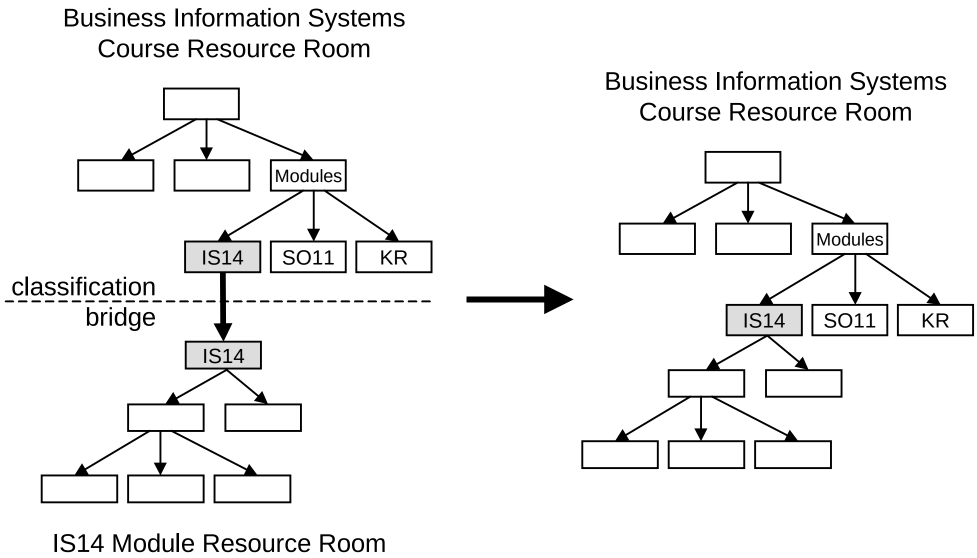


Figure 4-14: Resource Room Level Reuse in the MRA

Document Services

The document services (DS) layer contains services that operate on the lower layers of the MRA. The document services are key to the propagation of documents and associated information to relevant parts of the organisation. The DS layer is a process layer holding no information, but operating on the lower levels changing the state of the overall system. The primary service – the navigation service - operates over the entire MRA information model shown in figure 4-9. The navigation service provides the user with a means to navigate the information model, and ask a wide range of context-sensitive questions about the current state of the information model. Appendix A provides a full list of information that can be requested by the user and answered by the system. Apart from the navigation service, MRA document services divide into two classes:

- **entity services** that operate on the document abstraction and document objects layer
- **classification services** that operate on the document classification layer.

Table 4-1 shows the document services that are available at the entity and classification levels.

Entity services	**Classification services**
Document Check-in	Document Classification
Document Reuse	Classification Change
Search	Search
Hyperlinking	Bridging
Interest Management	Interest Management
Annotation	
Viewing History	

Table 4-1: Document Services within the MRA

Document Check-in

The document check-in service receives as input from the user the document to store, the manual document meta-data required (document title, keywords and description) and the primary classification under which the document should be stored. The check-in service stores the document at the document object layer and generates a corresponding set of meta-data at the document abstraction layer. It also calls the document classification service to associate the new document with the classification identified by the user, providing the primary point of access to the document from the document classification layer.

Search

A search service is provided at both the entity and classification levels. As an entity service the search operates at the document abstraction and document object layers of the model. The search service receives a search term from the user and returns a list of candidate documents, which may then be investigated further by the user. As a classification service the user is able to search all classifications (analogous to VSP resource room folders) at the document classification layer. The MRA advocates a set of hierarchical classification models at the DC layer comprising shorthand labels and more expanded descriptions for each label. By searching classification descriptions documents may be revealed that would not otherwise be discovered by an entity-level search. The search returns all classifications containing the search term, allowing the user to investigate any of the resulting 'folders' through the navigation service.

Hyperlinking and Bridging

The hyperlinking and bridging services allow the user to create links between documents and classifications that they need not necessarily own. Links between two documents are termed hyperlinks and links between two classifications are termed 'bridges', implying a flow of documents from one classification to another. Due to the separation of links from the documents and classifications that they connect, MRA hyperlinks and bridges are inherently two-way – both the source and destination anchors are equally aware of the created link. The hyperlinking service creates an entry in the relevant meta-data at the document abstraction layer and the bridging service creates entries at the document classification layer. Importantly each hyperlink and bridge added stores with it the name of the person who created the link, allowing an amount of user discretion when following such links.

Annotation

One of the identified weaknesses of the electronic document model over traditional paper-based documents is the relative lack of support for annotation [Levy93]. In response to this criticism, the MRA includes basic support for the addition of annotations to existing documents. The annotation process can be viewed as the creation of a new document and the creation of a hyperlink between the original document and the annotation. The annotation service performs this two-stage process within the MRA. As with hyperlinks and bridges, annotations are stored with the name of the person who created the annotation, allowing user-discretion when navigating the document store.

Viewing History

Whenever a document is viewed by a user of an MRA implementation, that 'hit' is recorded. The MRA history management service operates at the document abstraction layer, updating a document's system meta-data whenever that document is viewed. As with the hyperlinking, bridging and annotation services, each entry in a document's history trail is regarded as a first-order object, recording not only the fact that a document has been viewed, but also the name of the user who has viewed the resource, and a system-generated timestamp. A particular document's history can be viewed through the navigation service, showing not only how many times it has been viewed, but when and by whom.

Personal Interest Management

The MRA personal interest management service provides lightweight support for the registration of a user's interest in a particular document or classification, and automatic notification of relevant changes. An interest in a particular document is stored as system meta-data for that document at the document abstraction layer. An interest in a particular classification is stored at the document classification layer. The registration of personal interests and discovery of other users with particular interests may lead to an opportunity for collaboration.

Classification Change

The classification change service is responsible for managing updates to the document classifications in response to a user's request. The MRA supports the addition and removal of a classification 'folder' at the DC layer of the architecture, and further analyses knock-on effects to other classifications. The classification management service performs two roles:

- It allows a single classification to be changed in response to a change in business or culture of the owning organisation

- It highlights all affected classifications to either the owner of the updated classification or the owner of the affected classification. It further highlights any resources that are affected by the change in the classification model.

Document Classification

The document classification service is key to the Model-driven Reuse Architecture. It performs the association between a document and a particular classification at the DC layer which reflects the purpose and value of the document within the enterprise. Beyond a single classification, the service permits (and the

MRA encourages) multiple classifications of a single document, effectively re-purposing the document. It also supports the straightforward moving of a document from one classification to another. The service receives as input an existing pointer to a resource, a secondary classification and a choice of 'copy' or 'move'. If the user wishes to copy the resource to the new classification, this is an effective multiple classification as the resource itself is not copied – only a pointer to that resource. A 'move' option adds a document pointer to the secondary classification and removes the original association.

Document Reuse

The document reuse service provides explicit support for resource-level reuse of existing documents for a new purpose. The MRA promotes reuse by raising the level of abstraction of the documents stored so that they can be combined and repackaged for new purposes. The MRA document reuse service assists in both of these processes. The reuse service takes as input a collection of documents in their native document types and provides as output a single document of the target document type. Intermediate steps in this process permit the selection of resources or parts thereof, the ordering of those selected resources and the application of a new presentation to the resulting structure. The document reuse service makes use of the open storage of resources held within the MRA - as such the MRA represents a document model-based approach. The document reuse service provides only one of the three types of reuse available within the MRA, further described in section 4.5.

User Interaction

The top layer of the Model-driven Reuse Architecture is the user interaction (UI) layer that receives input from the user and returns results. User requests are mapped onto document service requests at the DS layer that in turn perform operations on each of the lower three layers. The UI layer is discussed in chapter 5.

4.5 Reuse in the MRA

Reuse is explicitly supported within the MRA at the three levels already discussed. They are:

- At the **resource level** the MRA supports the collection, combination and customisation of a set of existing resources for a new purpose. The resulting resource is a new entity that is classified according to the new purpose it serves. The MRA document model approach enables the analysis of resource content and structure and the application of a presentation specification to customise the resulting resource.

- At the **classification folder level** the MRA supports the re-purposing of a single resource, allowing it to be multiply classified within different classifications or resource rooms.

- At the **resource room level** the MRA provides classification bridges to make a collection of resources useful within a number of different work-focussed applications, enabling the construction of a number of views on document collections. Within the learning domain, for example, a collection of resources created for a module on 'Distributed Working' could be used within a number of different course-related resource room classifications.

4.6 Conclusion

This chapter has outlined the Model-driven Reuse Architecture, which incorporates documents and their classification, hypertext, resource ownership and reuse mechanisms. It emphasises the ability to multiply classify a single resource or collection of resources for more than one use, providing further mechanisms to create new resources through the collection, combination and customisation of resources that already exist. The selective application of hyperlinks and bridges provide a bounded document space for the end-user who may wish only to work within the confines of a subset of the available documentation. The next chapter describes an MRA implementation in pursuit of enterprise document management and reuse for learning resources.

5. Implementation

The previous chapter described a document architecture that supports a number of views on a common document store, allowing multiple classification of documents and providing mechanisms for the reuse of those documents for more than their original purpose. This chapter describes an implementation of the Model-driven Reuse Architecture described, using a client-server paradigm and a number of recent developments in document representation and manipulation. The implementation is delivered through a combination of web-server and web-client functionality, and through extensive use of the XML mark-up language. The Model-driven Reuse Architecture implementation aims to facilitate the effective use and reuse of its stored resources, not through the automation of document discovery techniques, but by supporting the recommendations of its users in their day-to-day work. The implementation, specific to the domain of supporting learning in higher education (HE), is named MRA-HE.

5.1 XML – A Mark-up Language

The Extensible Mark-up Language (XML) has been the subject of much on-going development, publicity and discussion in recent years. As a mark-up language, it represents a small subset of the Standard Generalised Mark-up Language (SGML) that became an ISO standard in 1986. Whereas SGML tools have been slow to emerge and restrictively expensive over the past 13 years, support for XML has grown rapidly. A number of active development programmes and freely available XML tools have improved its technical and commercial viability. Importantly, XML was released at a time when many organisations were recognising the inability of HTML to provide a solution to structured information distribution over the World Wide Web. Since the inception of the WWW in 1990, HTML has been its enforced document mark-up language. HTML offers a simple presentation-driven notation for marking up document content for presentation on a client browser. The simplicity of HTML has been a significant factor in the growth of the web to this point, but its limitations have hindered the use of the web for larger scale network-based information systems. [Bosak97] explains how HTML is failing in a number of important ways:

- **Extensibility** HTML does not allow users to specify their own tags or attributes in order to parameterise or otherwise semantically qualify their data.

- **Structure** HTML does not support the specification of complex data structures needed to represent database schemas or object-oriented hierarchies.

- **Validation** HTML does not support the kind of language specification that allows consuming applications to check data for structural validity on importation.

A number of current web-based information systems recognise and circumvent the loss of information semantic through the translation between structured information and web-friendly HTML (e.g. MetaCrawler[17] and RoboShopper[18]). Solutions lie in a painstaking extraction ('mining') of relevant information from HTML files. Additional to the computational load on extracting such information, this process has one obvious drawback, especially within the image-conscious world of web page design. Web site designers and owners change the design of their sites and pages frequently, and HTML structures change accordingly. Unless the 'mining' services are aware of upcoming changes to particular HTML formats, such methods immediately lose access to relevant information. XML aims to simplify the automation of such document mining processes by retaining the intended data structures for a particular application area. The ideas behind XML are not new, but widespread XML tool support and international interest have raised its profile far beyond that of other industrial-strength mark-up languages. As a language in itself, XML has expressive power but no processing ability. Its supporting standards - the Document Object Model (DOM), the Extensible Stylesheet Language (XSL) and others - facilitate manipulation and presentation of XML instances for an end-use. XML provides the rules by which its applications must adhere – HTML is one such application[19]. This thesis describes the way in which XML assists in the creation of a web-focussed MRA implementation, and so necessarily concentrates on the relationship between HTML and other XML applications.

The intention of XML (as it was with SGML) is to maintain a layer of separation between a document's content, structure and its eventual presentation. Through this separation an automated system has access to the underlying structure and content of an XML instance, while a stylesheet, coded in XSL, specifies the way in which such an instance may be displayed to an end-user for human inspection and interaction. A detailed description of the purpose and syntax of XML and XSL lies outside the scope of this thesis[20] but an example is given in Appendix B to make it understandable to the rest of this thesis. The example explains how an XML file that retains the semantic of the data it holds can be transformed into an HTML representation for display in a web browser. Figure 5-1 provides a simplified overview of the transformation process from XML to web-readable HTML.

---

[17] http://www.metacrawler.com/ (Last verified 8th March 2000)

[18] http://www.roboshopper.com/ (Last verified 8th March 2000)

[19] HTML is strictly an SGML application, but it could be easily re-written as an application of XML

[20] Many XML and XSL tutorials can be found on-line - http://www.w3c.org (Last verified 8th March 2000) provides a good starting point

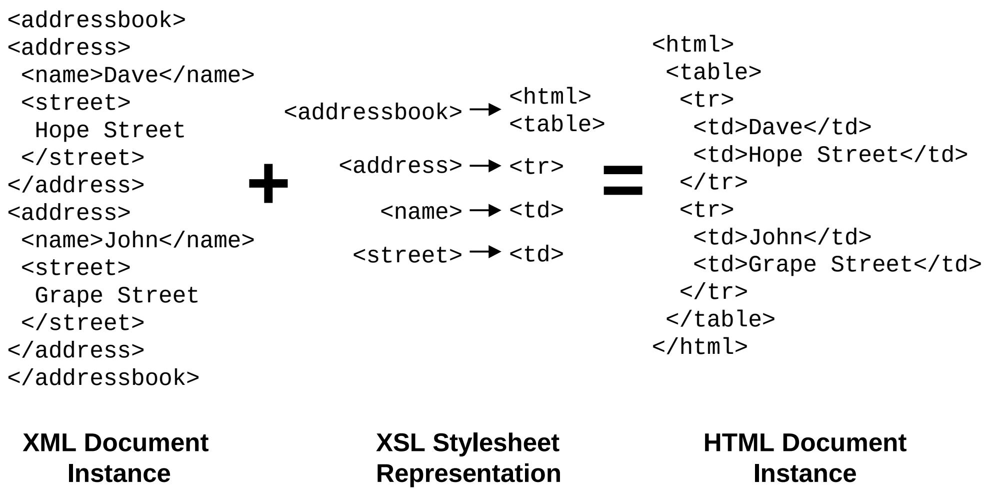


Figure 5-1: The XML to HTML Transformation Process

Importantly, the transformation from semantically rich XML to presentation-oriented HTML is performed at the very last moment, retaining the semantic until the user wishes to view the information. While the XML representation assists computational extraction and manipulation, a transformation to HTML ensures that the information can be viewed in a browser that does not hold an explicit understanding of how to render XML documents. The current generation of web browsers, however, understand and support the separation of XML and XSL, and are able to present the XML file according to the XSL file, while retaining the XML file as its source. This increases the ability of the browser to support behaviours and functionality that can be embedded within the XSL stylesheets and applied to XML files at the point of presentation. This emerging ability presents the opportunity to serve native XML files to the browser and rely on the browsers to apply the appropriate stylesheet as required. This has a number of advantages:

- **Document mining and manipulation** become simpler because the semantic of the document is retained

- **Functionality can now be divided** between client and server, allowing a much greater degree of client-side manipulation of the data. The XSL syntax permits reordering and alternative views on the source information.

- **’Publishing to the web’** becomes much simpler as the transformation from XML to HTML is automated and performed by the client on demand.

- **A** number of different **presentation styles** can be applied to the same XML document, providing dynamic views of the information stored.

- The **transmission size** of individual files is **much reduced**. Instead of passing a bulky HTML file from the server to the client, the server need only pass the XML file.

The power of XML and its related standards is illustrated throughout this chapter - its use is pervasive to the implementation discussed at all levels of the MRA.

5.1.1 Application of XML to the MRA

The applications of XML fall into two broad categories: document mark-up and information interchange. Karl Branting states that effective document reuse requires access to the original intentions underlying the document [Branting96]. The MRA assists this need by associating a document with the classification within which is it perceived to have value, and further providing a document model-based approach to enable manipulation of the documents themselves. As a document mark-up language XML facilitates the retention of more of the author's original intentions by enabling the mark-up that best captures the authored meaning. A document's 'social type' can be coded within the XML document, whereas its HTML equivalent is forced to degrade text descriptions to a single paragraph type. The development of document type definitions (DTDs) allows user communities to develop document styles that have common verifiable structures. It is this definition of styles through DTDs that permits the automation of reuse of document resources whose structure is understood by the system.

All layers of the MRA are implemented as XML files and basic operations over those files. Some represent recognisable document structures for the resources held, whereas others represent more informational aspects of the architecture. Further to the use of XML as a document mark-up language that better retains a document's semantic, XML can also be used as a basic open information modelling language. Each of the entities within the MRA information model (Figure 4-9) is represented as an XML document instance with a well-defined and navigable structure. XML manipulation methods, provided through the Document Object Model, are almost identical for both structured document mark-up and information modelling applications. In its basic form XML is able to model many kind of data structures easily, and the recent Document Content Descriptions (DCD) proposal submitted to W3C by Microsoft and IBM permits more enforceable data types definitions within an XML instance. The expressive power of XML and XML Data (which forms part of the DCD proposal) has proved sufficient for the needs of the demonstrator MRA implementation.

5.1.2 XML versus X500

Virtual Science Park resource rooms are currently implemented using an underlying X500 directory. X500 is a standard for directories which defines an information model, a functional model, a namespace, an authentication framework and a directory access protocol (DAP) with which to provide directory services [Drew97]. X500 was deemed to be suitable for the VSP implementation because of its decentralised model of storage, powerful search facilities, global namespace and structured information model. At first it appeared that the X500 directory structure might be useful for the implementation of the Model-driven Reuse Architecture. After investigation, however, it was decided that XML and its associated standards could provide a similar list of advantages and provide further support for the implementation requirements. In particular,

- The X500 design emphasises a tree-like structure to the domain of the entities it stores. However, the nature of hypertext and cross-enterprise links forms a focus on the 'interconnectedness of everything' [Whalley93] which does not naturally map into a tree structure. A network structure better supports such need and provides further support for incomplete and changing models. XML is more naturally able to support such a network structure with high connectivity and low dependency.

- Current work and opinion on structured documents is leading to a blurring of the distinction between database and document [Ressler95]. A person record, for example, can be considered as either a data record or a structured document. If a part-X500 implementation were undertaken, at some stage of development the designer would need to define which information types were to be regarded as data and which were to be regarded as documents. Person information, for example, may be regarded as data to be navigated and searched, but could also be regarded as a document to be viewed. If the person object were represented as an X500 model it would require explicit transformation on the server for viewing, but if stored as an XML file, that file could be navigated and searched using the Document Object Model and viewed directly in a web browser using an XSL stylesheet.

For these reasons, and to investigate the expressive and functional power of XML and its tools, it was decided to design an entirely XML-driven implementation.

5.1.3 The Document Object Model

The Document Object Model (DOM) is a platform- and language-neutral interface that allows programs and scripts to dynamically access and update the content, structure and style of documents[21]. The DOM provides a standard set of objects for representing XML documents as tree structures, a standard model of how these objects can be combined, and a standard interface for accessing and manipulating them. Its goal is to define a programming interface that operates over an XML instance regardless of the low-level parser used. The DOM represents a document as a hierarchy of 'node' objects, each of which contains properties that identify its type, content and attributes. The hierarchy of nodes can be navigated using straightforward DOM-specified API calls such as parentNode, childNodes and nextSibling. When called, these API functions move a pointer from the current node to one that is related in the specified way. The API also provides a number of methods by which to create new nodes and place them relative to the current node - insertBefore, replaceChild, appendChild, etc.

---

[21] Source: http://www.w3.org/DOM/ (Last verified 8th March 2000)

The DOM upholds the notions of a document object, a single node within that document, lists of nodes and node attributes - each can be accessed through a different API call.

The Document Object Model became a W3C Recommendation in October 1998. Although simple in its aims and recommendations, it was seen as vital to provide a standard to which document parser implementations should adhere. A number of DOM-compliant toolkit implementations are currently available free to developers and users. Microsoft's latest XML parser incorporates the DOM and was selected as the parser for the MRA implementation.

5.2 Client/Server Paradigm

Due to the increasing use and benefits of the World Wide Web client/server model for intranet and Internet applications, it was decided to develop the MRA implementation using a similar model. Web client/server technologies are now well-developed and freely available for experimentation and development. The HTTP protocol and CGI gateway provide a page serving protocol and more complex server-side computation, while developments in client-side web browsers are currently lessening the necessity to encode web-distributed information in the HTML document standard. The integration of information and user interface on the web browser permits the development and customisation of rapid and integrated user interfaces.

The architecture framework presented in the previous chapter is duplicated in figure 5-2 with the web client-server divide added. The figure shows that within the MRA implementation the client is responsible not only for providing interaction with the user, but also for providing a certain level of the functionality at the document services layer. It further identified two types of server - a proxy server and storage server. The storage server performs storage and retrieval of document resources themselves. In the implementation described the storage

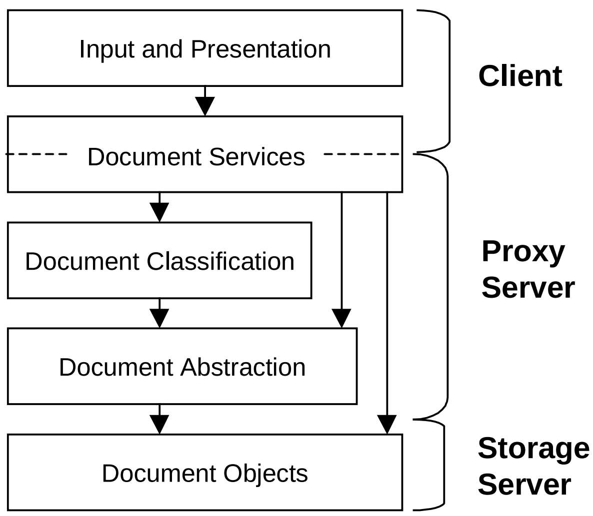


Figure 5-2: The MRA Client/Server Implementation

server is the same physical server as the proxy server, but this need not be so. A conventional proxy server is one that interfaces between the client and the storage server, often providing some level of caching or control. In the MRA implementation the proxy server performs a more significant role, controlling the classification and providing a layer of abstraction from the document resources themselves at the document abstraction layer. The proxy server performs a server role to the web client, and a smaller client role to the underlying storage server, or servers. The provision of a separate storage server layer provides the ability for a particular part of an enterprise to control and safeguard its resources on its own server, while being able to provide its resources to a wider community.

5.2.1 Client/Server Technologies

The MRA demonstrator implementation is based on a range of freely available World Wide Web technologies and open standards. A brief overview of the range of technologies used follows:

Web Server and Client

Microsoft Internet Information Server 4.0 (IIS4) with Active Server Page (ASP) extensions is used as the base web server for the MRA implementation, although any web server with ASP extensions can be used interchangeably. The web server base facilitates world-wide distribution over the Internet with its standard and extensible HTTP protocol. The use of a web client as the means to access the MRA implementation provides explicit support for cross-platform compatibility with no additional requirement from the user's machine. It was decided for the demonstrator implementation that an XML-aware web browser should be the baseline standard, although a standard HTML client implementation would be equally possible[22]. The use of an XML-aware browser provides additional client-side functionality while retaining storage and control at the web server. At the time of writing, Microsoft Internet Explorer 5.0 is the only XML-aware commercial web browser available, although the next major release of Netscape Navigator will be similarly XML-enabled.

Active Server Page Scripts

The use of server-side Active Server Page (ASP) scripts enables the embedding of procedural functionality within a web page held on an ASP-aware web server. ASP elements within a web page are executed in response to a user request for that page, and the results from the ASP are returned in-line with other mark-up elements. A small example is shown in figure 5-3. When requested by a web client, the ASP on the left is executed on the ASP-aware server and converted into an HTML instance, which is then passed to the client. ASPs are commonly written in either JavaScript or PerlScript. The MRA demonstrator uses PerlScript at its server-side implementation because of its powerful pattern matching and its ease of use.

---

[22] If the web browser is not XML-aware, the XML-to-HTML transformation would occur on the server rather than the client.

Server

```
<%@language=PerlScript%>
<html>
<head>
<title>asp page</title>
</head>
<body bgcolor="#ffffff">
<%$Response->write(’<p>Hello World');%>
</body>
</html>
```

Client

```
<html>
<head>
<title>asp page</title>
</head>
<body bgcolor="#ffffff">
<p>Hello World
</body>
</html>
```

Active Server Page —Web Server with ASP Extensions→ **Standard HTML**

Figure 5-3: An Example ASP Script and its Resulting HTML

XML, XSL and DOM

As noted earlier in this chapter, the XML family of standards has received much interest and support in recent years. The manipulation of structured documents permits the construction of documents and information on-the-fly from the current document store in response to user requests. The MRA implementation stores all its information models and business models as XML instances for information storage, manipulation and delivery. The availability of a number of royalty-free parsers permits a powerful and cost-effective document model-based approach to MRA document management. The simple integration of such parsers into the ASP architecture facilitates the use of such technology across the WWW. The parser selected for the MRA demonstrator is Microsoft's XMLDOM that is packaged with Internet Explorer 5.0, although any DOM-compliant parser would serve equally well. As a server-side parser XMLDOM is able to manipulate the XML instances that represent both the underlying MRA information model and the content and structure of the document resources that are stored and reused. The XMLDOM parser can also be used client-side, providing a consistent method for document manipulation across the client/server divide. The delivery of documents to the client as semantically rich XML instances rather than conventional HTML documents permits a level of client-side manipulation not before possible over the WWW. This chapter later explains how client-side XML manipulation is able to provide the user with a number of different views on the same XML document. Client-side JavaScript enables the XML parser to operate over a client-side copy of the transmitted XML document and dynamically apply a number of different XSL stylesheets.

5.3 The MRA Implementation

The complete software architecture for the MRA is shown in figure 5-4. The MRA client/server architecture provides a connectionless service to the user, based around user requests and system responses to those requests. This section outlines the major components of the MRA demonstrator implementation and their interactions.

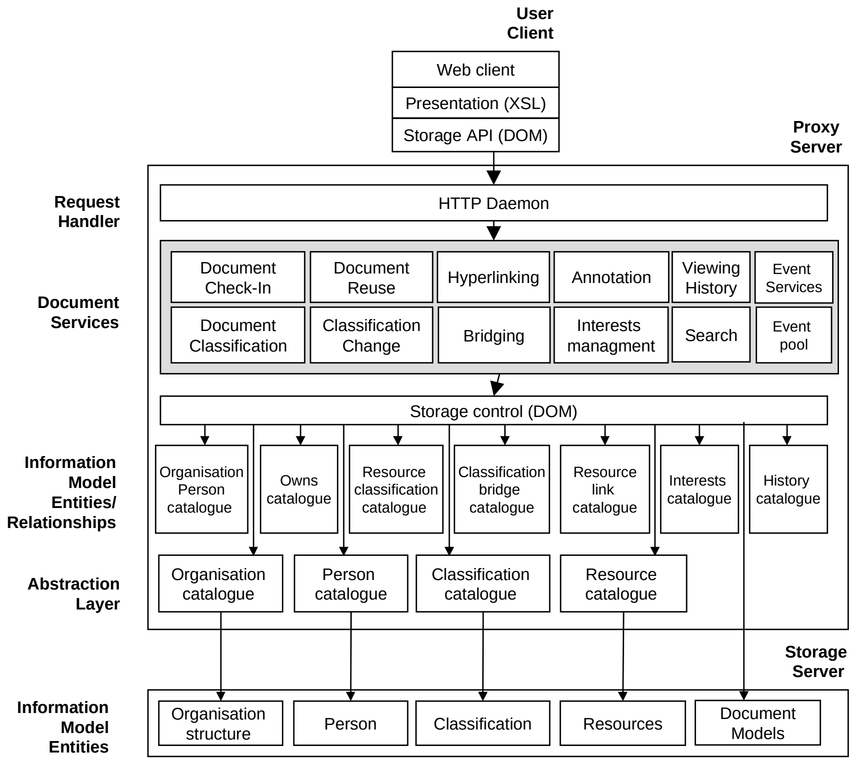


Figure 5-4: The MRA Software Architecture

5.3.1 Assumptions

A number of assumptions have been made about the working environment to provide a base-line implementation. The assumptions do not represent limitations of the MRA, but do provide a set of rules for the demonstration implementation. The assumptions are stated here, and form the bounds to the MRA demonstrator implementation.

- Every user must own one, and only one, resource room
- An organisation comprises any number of organisational elements. There may be as many organisations as are deemed necessary
- One organisational element may own one or more resource rooms
- One person may belong to more than one organisational element
- A resource must have a valid owner at all times, and must be linked to at least one classification. Every resource may be multiply classified.
- An enterprise's resources are the sum of all the resources that the people in those organisations own.
- When a user deletes a document, he does not mind others retaining a copy of that document

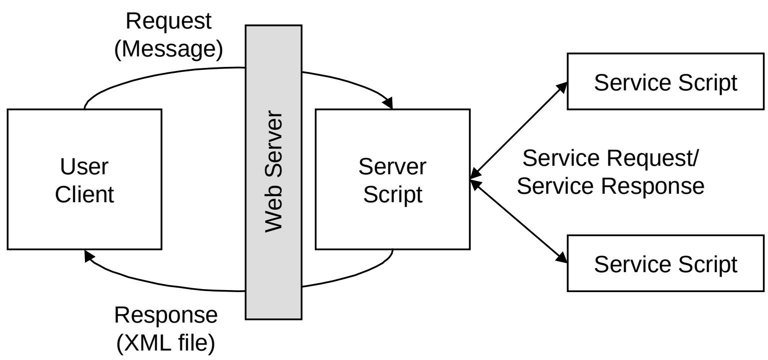


Figure 5-5: Request and Response Process

5.3.2 Request-Response Process

The request and response process, shown in figure 5-5, provides interaction between the client and the MRA proxy server. It involves the following steps:

1. The user client sends an HTTP request message to the proxy web server.
2. The proxy server decodes the request message from the client using a server-side ASP script (the 'server script').
3. The server script translates the request message from the client into a service request, which it passes to the relevant ASP service script.
4. The service scripts operate over the MRA information and business models and return a response to the server script.
5. Due to the connectionless nature of the WWW architecture on which the implementation is based, the server script adds a small amount of XML information to the file returned by the service script to remind the client of the request it has just made. The server script then returns the resulting XML file to the client via the server.

Although the MRA models provide a wide view of the enterprise's resources, classifications and organisational structures, the requests that a client can make at any particular time are limited by the client's current view on the information model. A full list of the allowable requests is listed in Appendix A, but some examples are given here to provide a flavour of the context-sensitive nature of the implementation. As an example, figure 5-6 reproduces the part of the MRA information model that involves the classification entity. It shows how a classification is linked to four other entities of the MRA information model. A user may request a particular classification or a list of classifications from each of the related entities. The following user scenarios all result in the return of a classification or list of classifications to the client:

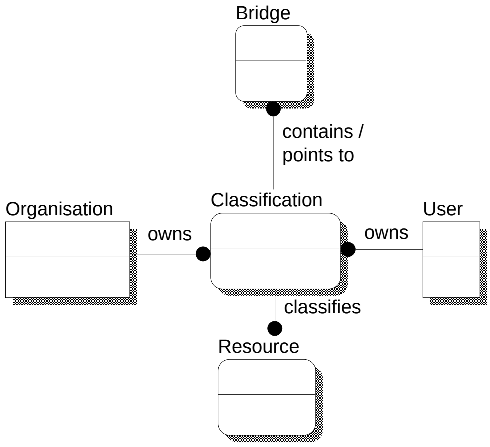


Figure 5-6: The Classification Entity of the MRA Information Model

1. A user is viewing a person's information and wishes to view the classification belonging to that person (e.g. Show me the resource room owned by Phil Taylor)

2. A user is viewing an organisational model and wishes to view the classifications belonging to a particular organisational element (e.g. Show me a list of resource rooms owned by the School of Computer Studies at the University of Leeds)

3. A user is viewing a classification and wishes to know what other classifications it has been linked to (e.g. Show me the resource room folders that have been linked to 'IS14 Resource Room : Research Architectures'?)

4. A user is viewing a resource and wishes to see where this resource is classified across the enterprise (e.g. Show me the resource room folders under which the resource 'Introduction to XML' is classified across the enterprise)

For scenario 1 the user client requests a classification, or resource room structure, from the server using the *ShowClassification* method. For scenarios 2, 3 and 4 the user client requests a list of relevant classifications using the *ShowClassificationList* method. In scenarios 2, 3 and 4 the server returns a list of classifications, from which a user may select one and make a subsequent *ShowClassification* request. As part of each request, the client also sends the context within which it is asking for a classification. The following four server requests correspond to the identified scenarios

Scenario	Request method	Context
1	*ShowClassification*	Person : Phil Taylor
2	*ShowClassificationList*	Organisation: Leeds University : School of Computer Studies
3	*ShowClassificationList*	Classification: IS14 Resource Room : Research Architectures
4	*ShowClassificationList*	Resource : Introduction to XML

The server script recognises each of the request methods within the contexts provided, and calls a service that can respond to that request. When the server script receives the response from the service script in the form of an XML file, it applies the initial context to the file and returns it to the client. The user is then able to navigate the model further.

The example given above has illustrated only the *ShowClassification* and *ShowClassificationList* request methods. The MRA demonstrator implementation recognises over 30 request methods that implement methods for information navigation, user navigation, search, addition and removal of resources and classifications, and hypertext and reuse mechanisms. They are listed in table 5-1 and explained in more detail in Appendix A. Although each request method may in turn invoke one of a number of document services, each returns to the client an XML file of the corresponding document type (either *classification* or *classificationList* in the example above) that is subsequently displayed in the client browser.

Information Navigation Methods		**Action Methods**	
ShowClassification	ShowPersonList	AddResource	DeleteFolder
ShowClassificationList	ShowPerson	MoveResources	CreateBridge
ShowResourceList	ShowOrganisationList	CopyResources	CreateLink
ShowResource	ShowAnnotations	DeleteResource	CreateAnnotation
ShowInterestList	Search	DeletePointer	AddInterest
ShowOrganisation		AddFolder	DeleteInterests

User Navigation Methods		**Resource Reuse Methods**	
ShowDropdownList	ChangeUser	SelectResources	ApplyPresentation
RememberFolder	ClearDesk	OrderResources	GenerateResource
RememberResource			

Table 5-1: All MRA Request Methods

5.3.3 MRA Models

Chapter 4 described the separation of the MRA models into a set of enterprise-specific business models and a more general set of information models that store entities and relationships that are manipulated by higher-level document services: hyperlinks, classification bridges, annotations, interests and document view histories. Within the MRA demonstrator each model is represented by an XML file of an appropriate document type. The business models describe the folder structures of each MRA resource room (classification), the organisational structure of the current enterprise, the personal details of the people within the enterprise and the structure models of the document types used by the enterprise.

A simplified classification model coded in XML is shown in figure 5-7, providing the folder structure for a single resource room. The classification model shown represents a three-folder resource room - a root folder

```
<classification>
 <entry id="id001">
  <subject>STEP Resources</subject>
  <description>STEP (ISO 10303)</description>
  <lastmodified>01/01/1999</lastmodified>
  <level>
   <entry id="id002">
    <subject>An Introduction to STEP</subject>
    <description>
     Introduction documents for the STEP standard
    </description>
    <lastmodified>01/01/1999</lastmodified>
   </entry>
   <entry id="id003">
    <subject>Application Protocols</subject>
    <lastmodified>22/09/1999</lastmodified>
    <description>
     Documents that describe STEP APs
    </description>
   </entry>
  </level>
 </entry>
</classification>
```

Figure 5-7: A Sample MRA Classification Model in XML

called 'STEP Resources' and two folders within that: 'An Introduction to STEP' and 'Application Protocols'. Each of the three folders has a title ('subject'), a description and a record of when it was last modified. Each node is identified by an 'id' attribute which is unique across each classification. The nested folder structure is represented by placing the 'Introduction to STEP' and 'Application Protocol' folder elements within the 'STEP Resources' element, delimited by a <level> element. The nesting of folders within a higher-level folder is allowable to any required depth. Similar files are constructed for each resource room structure within the current enterprise. Notably the classification models refer in no way to the documents that are classified within each resource room. A separate relationship file (the *resource classification catalogue*, described later) associates documents with the relevant resource room folder, or folders, creating a loose association with a single classification and permitting multiple classifications of a single document resource.

In a similar way, organisation structures are defined to represent the hierarchy of the enterprise that the MRA implementation supports. In the demonstrator implementation the organisational model represents part of the structure of the University of Leeds, shown in the figure 5-8.

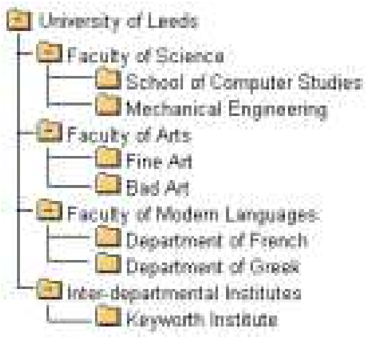
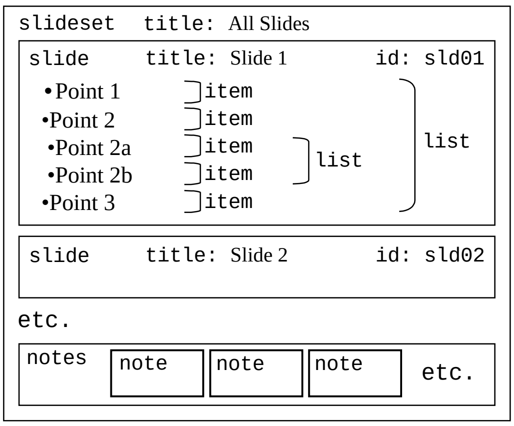


Figure 5-8: A Sample University of Leeds Organisation Model

The organisational structure provides a means to associate a document resource with the organisational unit within which it is created or owned. The organisation model describes the organisational units that comprise the enterprise but does not describe any person membership to those organisational units; it merely provides an organisational framework with which people and resources may be associated. Its aim is to assist with the storage and ownership of the implicit 'organisational memory' held within the document store. The organisational model is stored as a nested XML file with a structure similar to that described for individual resource room classifications. Information about each person within the organisation is stored separately in individual XML files, one for each person, containing contact details and a general text description. These details may be used by other users to make contact with document owners to enable collaboration or to elicit further information.

The final set of business models, referred to as document models, provide an explicit definition of the meta-structure of the documents created and used by a particular enterprise. The explicit representation of document models permits a level of automatic navigation over the documents stored, enabling a document model-based approach (described in chapter 2). To simplify the demonstrator implementation, only one document type is currently represented: that of teaching slides. The 'teaching slides' document type is shown in figure 5-9, both diagrammatically and as an XML document type definition (DTD). Each set of slides must conform to the rules laid down in the XML DTD in order to be considered valid.



XML DTD

```
<!element slideset (slide*, notes>
<!attlist slideset
          title    #REQUIRED
>
<!element slide    (list)>
<!attlist slide
          id       #REQUIRED
          title    #REQUIRED
>
<!element notes    (note*)>
<!element list     (item* | list>
<!element note     #PCDATA>
<!element item     #PCDATA>
```

Figure 5-9: The Meta-Structure of a Set of Teaching Slides

The previous chapter described the purpose and discussed the significance of the document abstraction (DA) layer in the overall framework architecture. To reiterate, the purposes of the DA layer are:

- to separate document and system meta-data from the documents they describe
- to separate the physical location of the document from the way in which it is referenced in higher layers.

The DA layer is implemented as a set of catalogue files stored on the proxy server that provide an overall view of the distributed business models. Each resource, classification, person file and organisation model is represented by one entry in the appropriate catalogue file. The resource catalogue, for example, stores meta-data for each document resource and a pointer to the physical location of the document file. Similarly, the organisation, person and classification catalogues provide an overall list of system identifiers and physical locations for each of the organisational models, person files and resource room structures. Each provides a layer of separation between the physical location of the files it describes and the way in which they are referenced by the document services.

To this point this section has described the construction of resource room structures, organisational structures, information files for each person within the enterprise and meta-structures for the documents stored within the MRA. Additional to the entities discussed, the following associations are made across the business models, representing relationships between entities in the MRA information model.

- A link between a document and its meta-structure (i.e. its document type)
- A link between a document and its owning person
- Links between an organisational element and its members (people)
- Links between a document and the classification folder/folders that classify the document

Due to the one-to-one mapping between a document and its type, and a document and its owner, the first two relationships could be stored as document meta-data stored in the resource catalogue at the DA layer. In practice, document ownership is separated from document meta-data for performance reasons, forming the *'owns catalogue'*. The third and fourth relationships are potentially one-to-many, and are stored separately so as not to impose any primacy of either document classification or organisational membership. The *'resource classification catalogue'* and *'organisation person catalogue'* held on the proxy server capture this information; one entry for each association. The following extract from the resource classification catalogue shows a single document multiply classified in two resource room folders:

```
<entry><resource id="res001"/><classification id="class001:id005"/></entry>
<entry><resource id="res001"/><classification id="class004:id012"/></entry>
```

Each entry in the resource classification catalogue associates a resource ID with a resource room classification. The resource ID, `res001`, provides a unique reference to a single resource within the MRA, and can be resolved to a storage location through the resource catalogue. The classification ID, `class001:id005`, represents the node with ID `id005` within the classification `class001`. Due to the use of the namespace mechanism, the ID `id005` only need be unique within a single classification. The classification referenced by ID `class001` is first resolved to a storage location through the classification catalogue; the relevant node within that classification is then identified within that single classification. Similar entries are made in the organisation person catalogue to associate people with organisational units. The ability to support the association of one person to more than one organisational unit allows for the representation of virtual teams and, in the case of the teaching model, inter-departmental centres and institutes.

The remaining MRA model implementations represent links between resources, links between classifications, document history trails and personal interests that have been registered by users in specific resources or resource room folders. Each entry represents a single link, document viewing or registered interest. Typical entries in the resource link catalogue, interests catalogue and history catalogue follow, with brief explanations:

<entry> <start id="res001"/> <end id="res105"/> <createdby id="person001"/> </entry>	This entry in the resource link catalogue represents a link created by person with ID `person001` between two resources with IDs `res001` and `res105`
<entry> <person id="person001"/> <resource id="res001"> <type>single</type> </entry>	This entry in the interests catalogue registers an interest of person with ID `person001` in the resource with ID `res001`. The user has requested update notification every time the resource changes (`type=single`)
<entry> <person id="person001"> <resource id="res001"/> <lastaccessed>22/09/99</lastaccessed> <hits>8</hits> </entry>	This entry in the history catalogue is updated every time person with ID person001 views the resource with ID res001. The last accessed date and the total number of hits by this user on this resource are stored.

Table 5-2: Example Entries in the MRA Information Catalogues

5.3.4 Document Services

Sections 5.3.2 and 5.3.3 have described the request and response process and the MRA models that provide storage for the MRA business and information models. When the user client issues a request to the MRA proxy server it translates that request into an appropriate service request, depending on the context within which the client has made the initial request. Each of the service requests is managed by the appropriate service script. The service scripts perform a range of operations on the MRA models described in the previous section and return an XML document of the relevant type. When called by the server script, a service creates a number of server-side XML parser instances - one for each file required to service the current request. Once loaded, the ASP calls a number of DOM methods on the XML parsers to extract or add relevant information, repackages relevant information into a new XML file and returns that file to the server script. Due to the connectionless protocol of the web server/client architecture the proxy server adds a small set of attributes attached to the root of the returned XML file that 'remind' the client of the context under which the request was made.

To illustrate the role of document services, this section provides a short example service that returns a list of resources hyperlinked to the currently selected resource. The full request-response process follows, including sample ASP script and fragments of the resulting XML responses to provide a flavour of the underlying programming mechanisms.

1. The user selects a resource and asks of the client 'What resources is this hyperlinked to?'.
2. The client sends a request to the MRA server: `'showresourcelist resource=res001'`.
3. The server script identifies the request as a request for hyperlinks and issues a subsequent request to the hyperlink service.
4. The hyperlink service loads the relevant XML file, the resource link catalogue:

```
$resLinkCat = $Server->CreateObject("Microsoft.XMLDOM");
$loaded=$resLinkCat->load($resLinkCatFile);
```

5. The hyperlink service makes a DOM request to locate all the entries in the resource link catalogue that have the resource `res001` as one of the link ends:

```
$resLinkCatRoot=$resLinkCat->documentElement;
$relevantEntries=$resLinkCatRoot->selectNodes(
 'entry[start/@id=$resourceToFind or end/@id=$resourceToFind]');
```

6. The hyperlink service creates a new document of type `resourcelist,` and for each of the identified entries creates an entry in the new resource list document.

```
<resourcelist>
<entry>
 <resource id="res002"><type>linksto</type><createdby id="person001">
 <metadata>
  [document resource's meta-data placed here - title, author, keywords, etc.)
 </metadata>
</entry>
...
</resourcelist>
```

7. This resource list is returned to the server script, which adds to the file the original context of the client request

```
<resourcelist resourceid="res001">
 <entry>...</entry>
...
</resourcelist>
```

8. The server script returns the XML file to the client for subsequent display, showing the user the list of resources attached to the resource `res001`, with their details. A typical response follows:

```
<resourcelist resourceid="res001">
 <entry>
  <resource id="res002"/>
  <type>linksto</type>
  <createdby personid="person001">Dave Small</createdby>
  <metadata>
   <owner id="person001"/>
   <title>Plan of the Lecture</title>
   <version>1.0</version>
   ...
  </metadata>
 </entry>
</resourcelist>
```

The client understands (via an XSL stylesheet) that this resource list has been returned because the user asked for a list of resources linked to resource `a001`. The list contains one entry, indicating that resource `res001` contains only one hyperlink to another document. The entry codifies the type of this link, the person who created the link and the meta-data associated with the target document. Although the range of MRA document services is extensive, the process for each is similar: decode request, manipulate the MRA models on the server and return a document of the relevant type to the user for display. More complex services not only access the MRA models on the server but also add, remove or update entries using the Document Object Model API.

5.3.5 Client Implementation

Until recently web browser clients have been capable only of displaying static HTML documents received from web servers. This has led to a static view of the data, requiring additional web-server intervention to provide different views on the same information. The current generation of web browsers is able to provide much more client-side functionality through the use of the JavaScript language to manipulate HTML client-side, changing

Figure 5-10: Web Browser Output from 'resourcelist' Server Response XML file

document display dynamically; this has become known as Dynamic HTML (DHTML). The MRA client implementation uses DHTML extensively; DHTML that is generated by the application of an XSL stylesheet to the XML files returned in response to client requests. In the example from the previous section, the client presents the user a list of however many linked documents have been identified. Figure 5-10 shows the web browser output with four links shown in the resulting resource list. This output is generated directly by applying an XSL file to the XML file from the previous section.

Client-side XSL stylesheets are able not only to wrap HTML elements around the XML content, but also to add client-side behaviour through DHTML. Due to the extensive nature of the MRA information model, the demonstrator implementation provides pop-up menus for each element that can be further investigated. The stylesheet identifies the allowable user operations within the current context and provides an appropriate set of pop-up menus, permitting the user to further navigate the MRA information model. Figure 5-11 shows the same results as figure 5-10, but further shows two sets of pop-up menus that are available. Each pop-up menu is activated by moving the on-screen mouse pointer over either a resource title or the name of the person who created the link. The menus contain a list of information that is available for the currently selected element. A selection from either of the available menus sends a new request to the MRA proxy server, enabling further navigation of the information model

The menus shown in figure 5-11 represent the full range of navigation capabilities available across the MRA information model. In order to support a range of different user types, both the displayed fields and the available menu options can be limited according to the type of the current user. Within the domain of higher education, for example, a tutor may wish to have full navigation over the information model, but restrict the options available to a student. This restriction is declared in the XSL stylesheets for each document type, simply but effectively removing fields and menu options for all student users, creating a customised view on the available information and a limit to the available navigation.

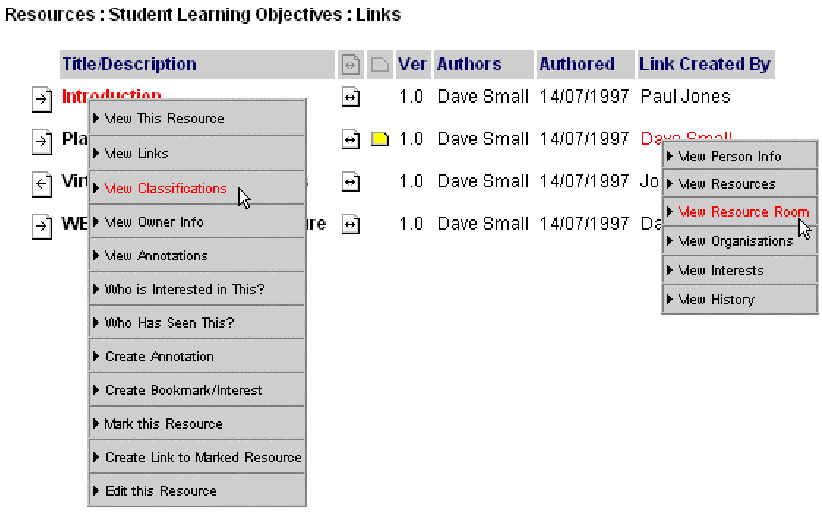


Figure 5-11: Context-Sensitive Pop-up Menus

The dynamic nature of the XSL stylesheet further permits the user to manipulate the current view on the XML information. The MRA implementation allows the user to order a list of resources alphabetically by title or author, or chronologically by date authored. Conventional HTML delivery cannot provide this level of client-side manipulation.

5.3.6 Trust Mechanisms

Chapter 4 discussed the need to restrict the bounds of hypertext to the end user to avoid confusion and to retain focus, especially within the domain of teaching and learning. The client implementation discussed in the previous section facilitates one type of restriction by providing users of type 'student' with limited information and navigation options. A second method of restriction is made possible through the selective retrieval of document hyperlinks and annotations. Annotations and hyperlinks that refer to a particular document are stored separately from the document, and are typed, storing not only the link or annotation, but also a pointer to the person that created the link or annotation. Within the learning environment, for example, the tutor is able to restrict the effect of hyperlinks and annotations on students by supplying a list of people whose hyperlinks and annotations should be trusted by the students as they navigate MRA resource rooms. The current simple trust model demonstrates the ability for the student to view only the annotations and hyperlinks created either by the tutor or any other student on the same course. The list of trusted people is stored in each user's personal XML record, and inspected each time a set of links or annotations is sought by a user. Only those created by users on the current user's trust list are returned to the client for display. Thus a document that is multiply classified across a number of resource rooms is able to contain many more links and annotations than are visible to any single user.

5.3.7 Event Pool and Services

Section 5.3.2 provided a description of the request-response mechanism that predominates throughout the MRA-HE implementation. The request-response mechanism is such that the system responds to user requests as each user navigates the information model and manipulates various aspects of the underlying information store. In order to maintain integrity across the document store and information models, and to trigger events such as regular update notifications, a persistent event system is provided to complement the request-response nature of the implementation. The event pool is implemented as an XML file which stores major system events as they occur. The event pool can then be interrogated by the asynchronous event services as required. There are two event services currently implemented for demonstration purposes:

- a notification service that provides email notification of changes in any interests specified by users
- an integrity service that ensures that documents, folders, links, etc. remain intact when changes are made elsewhere in the system.

A typical example of each service is described below to provide a flavour of the roles performed by each.

Notification service process

1. A user (Paul) registers an interest in a particular resource room folder, asking to be updated weekly on changes to the contents of that folder
2. The owner of that folder (Chris) adds a document some time later, creating an entry in the MRA-HE event pool
3. At a scheduled time the notification service inspects the event log and all users' interests
4. When the notification service finds the update event in the event log and Paul's interest, it sends an email to Paul notifying him of the change

Integrity service process

1. A user (Paul) points to a particular document from one of his resource room folders
2. Some time later the owner of the document (Chris) decides that it is of no further interest, so makes a request to delete the document
3. The document classification service removes the primary link to the document and creates an entry in the MRA-HE event pool
4. At a scheduled time the integrity service inspects the event log and all resource rooms
5. The integrity service finds a single folder that points to the 'deleted' document, creates a new copy of the resource, sets the 'owner' attribute of the new document to 'Paul' and re-links Paul's folder to point to the new resource
6. The integrity service sends an email to Paul informing him of the changes made.

The notification service is similar to the more thorough implementation provided for the VSP (described in [Drew97]). Through the provision of the integrity service it is hoped that users may be encouraged to use resources outside their own resource room, without the fear that a document which they may become dependent on may be lost. The event pool and services are by no means complete within the MRA-HE demonstrator, but their presence provides a proof of concept of the asynchronous capability of the MRA.

5.4 MRA-HE Reuse Mechanisms

The MRA operates beyond the bounds of many conventional document management systems by associating the document resources it contains with an owner and with at least one classification within which the resource is currently used. It is proposed that encoding a resource's additional work-focus and ownership will result in a system that can both manage resources and facilitate reuse of those resources for more than their original purpose. As previously discussed in this thesis, MRA supports reuse at the resource, folder and resource room levels.

5.4.1 Resource Level Reuse

The resource-level reuse process involves collection of documents, combination of parts of those document and customisation of the resulting document with a new presentation. To this point this chapter has highlighted opportunities for collection of existing materials. This section highlights MRA provision for resource combination and customisation.

The MRA-HE demonstrator introduces a concept of 'My Desk' to each of the personal resource rooms created. 'My Desk' is a resource room folder that has additional functionality to other resources rooms, providing the ability to select a subset of the resources it contains. During the collection phase a user browses the resource space by navigating the MRA information model. Upon the discovery of a resource suitable for reuse the user is able to copy a pointer to that resource to 'My Desk'. The desk metaphor provides a workspace separate from the work-specific resource rooms specifically for the purpose of reuse.

Once a set of resources is collected, the user is able to select an arbitrary resource order and view the combined resources. Combination relies on the XML structure of the collected resources, separating the individual resource content from its resource-specific structure. The generated resource is viewed in a client browser, where the user is able to refine the ordering of the component resources. The user is able to decide, for each component resource, whether a snapshot of the resource is required, or whether a pointer to the original resource would be beneficial. [Levy94] highlights the fixed/fluid problem within documents, arguing that the user may want to provide either a stabilised view on the document resource or a more fluid representation whereby the original resource is referenced for a 'latest view'. Research into virtual documents [Vercoustre97] highlights value in referencing up-to-date information from within a set of more static information. The MRA demonstrator supports both types of resource inclusion: a snapshot for 'fixed' information and a pointer mechanism for 'fluid' information, whereby the document component is added at view-time.

Once the user has selected the final ordering of the new resource and, for each component, whether a snapshot or pointer should be made of the resources, the system produces the resulting resource, adding it to the user's desk for subsequent classification.

5.4.2 Folder Level Reuse

Folder level reuse involves the multiple classification of a single document, effectively repurposing it within a new context. The re-purposing of a document is aided within MRA-HE by the loose coupling between a document and the resource room folders that classify it. In order to classify a document within more than one resource room folder, the relevant document simply has two entries within the resource classification catalogue, created by the document classification service.

5.4.3 Resource Room Level Reuse

The reuse of whole resource rooms, or parts thereof, within other resource rooms is facilitated by the ability to embed an XML fragment within another. Each resource room folder structure is represented as an XML file. To enable resource room level reuse, any single node from one resource room XML file is included as a pointer within a node from another resource room XML file. Upon presentation of the resulting resource room to the user, the navigation service resolves all pointers within the classification, providing a 'latest view' of the included resource room folder structure to the client

5.5 Conclusion

This chapter has described MRA-HE: an implementation of the Model-driven Reuse Architecture to support a range of activities within the domain of higher education. For a tutor it facilitates the discovery and reuse of existing documents for new teaching purposes, providing a rich information space and a number of reuse mechanisms. It also provides tutors with the ability to develop restricted resource rooms that students may use for a particular learning purpose. The ability to adapt the user interface and restrict the application of hypertext mechanisms limits the extent to which students are allowed to stray from the purpose for which the resource room has been created. Although the demonstrator has implemented a broad range of functionality, its key features are:

- the storage of structured documents that can be addressed and manipulated at the sub-resource level
- the classification and multiple classification of a single document to assist its use in more than one context
- the ability to build higher-level resource rooms (e.g. course rooms) from a set of lower-level resource rooms (e.g. learning module rooms)
- a range of hypertext mechanisms and navigation paths to promote document discovery (for tutors)
- the provision of work-focussed resource rooms that are able to provide a restricted view on the complete document store (for students)
- an adaptable interface to provide either the full range of MRA-HE functionality or a more restricted view

The next chapter assesses the MRA-HE in terms of how effective it is at providing for typical learning scenarios.

6. Evaluation

6.1 Methodology

The development of MRA-HE described in the previous chapter aims to provide the following task-based benefits to users within the domain of supporting higher education:

1. Improved organisation of stored documents across the enterprise
2. More effective document retrieval for specific work purposes
3. Increased ability to reuse documents in a new context for new learning purposes
4. Ability to focus document resources within a localised context for a learning purpose

The success of the MRA-HE implementation is measured in terms of how it achieves the above benefits, using current VSP Resource Rooms as the baseline against which to make comparison. The following evaluation process was used to appraise both technical and user objectives:

1. The scenarios identified in chapter 4 were used to identify key success/failure factors for key user groups

2. The MRA-HE implementation and alternative solutions were appraised in terms of their satisfaction of the success factors identified

3. Initial user feedback was obtained to determine whether achievement of stated user objectives resulted in user satisfaction with MRA-HE.

4. In light of the above appraisal of how well MRA-HE attains the task-related goals of users, the key design decisions that determined the character of MRA-HE's implementation were appraised.

6.2 Appraisal of MRA-HE and Alternative Solutions

In order to appraise MRA-HE in terms of its success in providing the task-based benefits listed above, this section returns to the use scenarios identified in chapter 4. A review of the teaching scenarios leads to the identification of critical success factors (CSFs) most likely to result in increased user satisfaction from an enterprise document management system. The tables on the next page list the CSFs identified for each scenario.

Critical success factor
1A) ability to manage more than one classification scheme
1B) ability to associate keywords and descriptions with resources when stored
1C) ability to multiply classify and manage the same document in a number of classification folders
1D) ability to copy resources between classification folders without duplicating documents
1E) ability to access folders, documents and management functionality from a web browser
1F) ability to automatic re-link resources in one folder when they are reclassified in a different folder
1G) ability to view who has seen a particular document
1H) ability to view who has expressed an interest in a document
1I) ability to find out under which classification folders a document is classified
1J) ability to control access permissions for reading and editing
1K) ability to find out what documents other people have seen

**Table 6-1: Critical Success Factors for Scenario 1:
An Academic Validating and Producing Resources**

Critical success factor
2A) ability to search the meta-data for all documents in the store
2B) ability to view information about the owner of a document
2C) ability to register an interest in a classification folder or document
2D) ability to search an organisation for document owners
2E) ability to perform a full-text search through all documents in the store
2F) ability to create pointers between classification folders
2G) ability to reuse fragments of existing documents to create a new document
2H) ability to retain a dynamic link between a fragment and documents that use that fragment
2I) ability to specify a new presentation style to an existing document
2J) ability to move resources between classification folders
2K) ability to include one folder structure within another classification folder
2L) ability to view information about the owner of a classification folder

**Table 6-2: Additional Critical Success Factors for Scenario 2:
A Tutor Assembling Resources for a Specific Course**

Critical success factor
3A) ability to create and follow links between documents
3B) ability to create and view document annotations
3C) ability to restrict the application of links and annotations
3D) ability to view information about the creator of links and annotations
3E) ability to adapt the interface for different user types

**Table 6-3: Additional Critical Success Factors for Scenario 3:
A Student Following a Pre-Defined Course**

6.2.1 MRA-HE Provision for Use Scenarios

The teaching scenarios in chapter 4 described three different user perspectives on the common document store and the ways in which each user type may wish to navigate and manipulate resource rooms and resources differently within those resource rooms. It identified a number of mechanisms that may assist navigation and create opportunities for collaboration and reuse across the enterprise. This section describes how MRA-HE mechanisms are able to realise the identified scenarios, and how they serve the identified critical success factors in tables 6-1 to 6-3. The following sections are detailed enough to describe a step-by-step MRA-HE walk-through for each scenario.

An Academic Validating and Producing Resources

As an academic overseeing the construction of a resource room for a particular course within MRA-HE, Paul will have access to two MRA-HE resource rooms – one for his own personal work purpose and one for the course he is running (CSF 1A). He is able to store his new resource in his own classification by selecting the relevant resource room folder and selecting 'Add a Resource' from the pop-up menu. When adding the new resource he is able to add a set of keywords and a description for that resource (CSF 1B). He is also able to point to this new resource from his course resource room by marking the document in his personal classification, selecting the relevant folder from the course resource room and selecting 'Copy Marked Resource Here' (CSF 1C). Rather than making a new copy of the document, MRA-HE creates a second classification for the single resource, effectively re-purposing the resource within a new context (CSF 1D).

When Paul perceives the need to reorganise his resource room he is able to move resources between folders in his own personal resource room without having to re-link the resources in the course resource room because of the loose coupling between resources and their classifications (CSF 1F). Through navigation of the MRA information model using pop-up menus in his web browser, Paul is able to find out who has viewed his resource (CSF 1G), how many times and how recently, and which other classifications reference his document (CSF 1I). He is also able to find out who has expressed an interest in his resource by selecting 'Who is Interested?' from the pop-up menu for the relevant resource (CSF 1H). From the list of people who have seen his resource, Paul is able to find out what other resources have been seen by these people by selecting 'View history' from the pop-up menu for a particular person (CSF 1K). Paul can switch between his two resource rooms quickly by adding a bookmark interest to each; this adds each classification to his personal bookmarks which are always available from a drop-down list across the top of his browser window.

A Tutor Assembling Resources for a Specific Course

As a tutor you have access to two resource rooms – one for the Advanced Computer Graphics module and one for the Business Information Systems course. Performing a search for 'gourad shading' from the top browser window, results in a list of resources whose meta-data contain that search term (CSF 2A); each can be viewed in turn in a new browser window. Upon finding relevant documents you are able to copy those resources to your personal desk space for later reworking by selecting 'Copy to My Desk' from the pop-up menu for the relevant resource. After noting that Phil Smith has authored a number of the resources (CSF 2B), you view his resource room by selecting 'Show Resources' from the pop-up menu for one of the relevant resources. By viewing the resource room folders under which the resource is classified you are able to view his personal resource room. By registering an interest in one of Phil Smith's resource room folders you can be notified automatically by email when any new resources are added to that folder (CSF 2C), or just add that folder to your bookmarks which are available from the drop-down list at the top of your browser window. By viewing Phil Smith's personal interests you discover his email address and contact him for further assistance.

By searching through the organisational structure of the University from the search form in the top browser window (CSF 2D), you are presented with information about two people with the name Chris Taylor. By viewing the organisational units to which each belong you ascertain that the Chris Taylor from the School of Mechanical Engineering, and are able to see a list of all the resources owned by that school. Of the resources listed, you select two and choose to 'Copy selected resources to my desk'. By further full-text document search (CSF 2E) and navigation you discover that all the documents within the Theory of Graphics folder of the School of Computer Studies resource room are useful to the module you are in charge of. By marking the Theory of Graphics folder from the pop-up menu, returning to your module resource room and selecting 'Add Bridge to Marked Folder' you create a link between your module resource room and the Theory of Graphics folder (CSF 2F). This allows anyone viewing the module resource room to see and navigate the link to the external folder. Although you have no control over the contents of the Theory of Graphics folder you regard it as valuable in some way to someone using your module resource room.

Once you have collected a set of resources (which are all of type 'lecture slides') in your desk space you click 'My Desk' from the top browser window to view your desk. To this stage you have performed a high-level collection process, but wish to reuse only certain slides from each of the collected slide sets (CSF 2G). You select from your desk space the slide sets from which you wish to create your new slide set and select 'Reuse Selected Resources'. A new browser window opens to guide you through the reuse process. First, you select the slides from the selected slide sets that you wish to reuse – this performs a lower level collection phase. You choose whether to include a snapshot of the current slide in your new resource or to include a pointer to that slide which will automatically update your resource when the source changes (CSF 2H). Pressing 'Next' moves you to the combination phase of the reuse process which allows you to order the slides for your new purpose. After ordering the resources within the browser window and pressing 'Next' you apply one of the pre-defined styles for your new slide set (CSF 2I) and add a title, description and keywords for the new resource. Upon completion a pointer to your new resource is placed in your desk space, which you can move to the Lectures folder within your Advanced Computer Graphics resource room by marking the resource and selecting 'Move Marked Resource Here' from the pop-up menu for that folder (CSF 2J).

In order to include the Advanced Computer Graphics (ACG) module within the Business Information Systems course, mark the root folder of the module resource room and from the pop-up menu for the Modules folder within the BIS resource room, select 'Add Bridge to Marked Folder'. Whereas the previous bridge you created only provided a pointer to the external resource room folder, you want to actually include the ACG module resource room within the Modules folder in the BIS resource room (CSF 2K). Once created, the link between the course room and the module room will appear almost seamless, allowing students on the BIS course to stay within the course room but access all the resources within the ACG module.

Finally, in order to locate all the resource rooms owned by the School of Computer Studies, you search through the organisation from the search form, locate the School of Computer Studies and from the pop-up menus select 'Show Resource Rooms'. Each teaching module has its own resource room listed, and you are able to find contact for each by selecting 'View Owner Info' from the pop-up menu for each relevant module resource room (CSF 2L). You include them in your BIS course room by creating bridges as you did for the ACG module.

A Student Following a Pre-Defined Course

As a student on the Business Information Systems (BIS) course, Maria is able to use the BIS course room to access documents that are particular to the course and the documents for each module contained therein. Each module has its own resource room, but the use of included bridges allows Maria to navigate the course resource room and access all relevant resources seemingly from one classification. The tutor has restricted access to MRA-HE functionality by defining which pop-up menus are visible to students using the system. Maria is presented only with the navigation routes that the tutor thinks suitable (CSF 3E). Although each resource may have annotations or hyperlinks created by many MRA-HE users, Maria is shown only those created either by the tutor or other students on the BIS course (CSF 3C). She can add an annotation to a document by selecting 'Add an Annotation' from the pop-up menu for the relevant resource (CSF 3B), create an entry for the web site she has found in her own workspace by selecting 'Add a Resource' from the relevant folder pop-up and create a link from a resource to the web site she has just added (CSF 3A). This becomes instantly accessible to others on the BIS course. Finally, Maria is able to select the 'View Person Info' option from the pop-up next to an annotation of interest to find out details about the creator of the annotation (CSF 3D).

6.2.2 Comparison with Other Systems and Methods

The previous section described how the MRA-HE implementation provides functionality to support the scenario CSFs. In order to assess the success of MRA-HE in terms of providing an improved document management environment over currently available methods, this section describes how existing methods may, or may not, achieve similar functionality and assist in the CSFs previously identified. Although MRA-HE was designed specifically to benefit the scenarios in chapter 4, there is value in assessing how other mechanisms may perform against the critical success factors identified through those scenarios.

Conventional Desktop Computing

Traditional office software products focus on support for the individual in accomplishing work-based tasks. They provide advanced functionality for document production, manipulation and storage on a per-document basis. General-purpose operating system buffers permit 'copying' and 'pasting' of existing materials into a new context. Mechanisms such as Microsoft's Object Linking and Embedding (OLE) protocol assist in the inclusion of one document inside another while retaining a live link to the original document. Software products such as recent versions of Microsoft PowerPoint allow the application of a 'master slide' that defines an overall presentation to a set of created slides, providing a limited separation between documents and their presentation. The resulting document is stored in a folder structure dictated by the owner of the file space; some office applications permit the addition of document meta-data and of annotations to the main document. Current file systems permit the definition of shortcuts or symbolic links that permit the definition of cross-folder relationships, allowing for multiple classification of documents and for the reuse of folders within a number of other folders. Users with access to the same local area network may access certain parts of a shared file space, or search across them, allowing them to access documents of relevance, and access to different drives provides a low level of support for multiple classification schemes. Across a local area network, permissions can be set for read and write access to individual files.

Conventional desktop mechanisms provide a means to implement a number of the features and qualities required by the use scenarios. There are, however, a number of fundamental drawbacks in using a traditional desktop solution

- Documents and 'resource rooms', which may be implemented as sets of work-specific folders, are accessible only over proprietary networks and proprietary operating systems, precluding the possibility for widespread access and discovery.
- Although documents and folders may be used elsewhere in the file structure, there are no rules by which to check the integrity of the file store, nor to ask the question 'where else is this document stored?'
- The provision of document meta-data is ad-hoc, and there is no explicit association between a document or folder and the person or organisational element to which it belongs

- There are no mechanisms for the separate application of document-related entities such as document annotations or hyperlinks.
- There is no notion of an event system to support the discovery of user history trails or to register interests and receive notification updates
- Although resource-level reuse is possible through 'copy and paste' and OLE methods, there is no structure or control over the widespread and distributed use of such mechanisms.
- There is no explicit support for different user types, providing the same interface to all users

Virtual Science Park Resource Rooms

The application of the current VSP Resource Room architecture to the identified use scenarios was discussed in chapter 4. The strength of current resource rooms in assisting the identified scenarios is the ability to define purposeful resource rooms for each VSP tenant that store documents in groupings that represent a valuable resource classification for a work purpose. The VSP shortfalls in pursuit of the use scenarios described in chapter 4 are reiterated below.

- Within the VSP resource rooms, each resource is connected tightly with the folder it was originally stored in, tying the document to its original purpose. The nature of the resource room information model hinders large-scale repurposing of documents because of the tight binding of a document to a specific (primary) resource room folder.

- The VSP currently supports the addition of lightweight hyperlinks between resources and the registration of interests in a particular resource or folder. However, the VSP provides no architectural support for hypertext approaches such as document history trails, typed hyperlinks and annotations. These are fundamental omissions from the current architecture.

- The VSP resource room architecture does not provide a document-model based approach. Beyond the meta-data that the document owner must supply at document check-in, reading room resources are considered as Binary Large Objects (BLOBs). There is no distinction between a resource's content, structure and presentation. It is therefore difficult to promote reuse of the document resources for more than their original purposes and to perform full-text searches across the documents.

- There is no direct link between the resources held within the VSP and the VSP information directory. It is impossible, therefore, to locate information about a document's author except through a disconnected text search across the information space. This connection is essential to facilitate collaboration once a document of interest is found.

Comparison using Critical Success Factors

Table 6-4 lists the critical success factors identified at the start of this chapter and shows those which are satisfied by the approaches discussed in this section – the MRA-HE implementation, conventional desktop computing and VSP resource rooms. In the following table, each CSF is given a three-stage rating according to how it is satisfied by the relevant approach.

○ – the approach is unable to satisfy the identified CSF
◐ – the approach is able to partly satisfy the identified CSF
● – the approach is able to more fully satisfy the identified CSF

Critical Success Factor (abbreviated)	MRA-HE	Desktop Computing	VSP Resource Rooms
1A) manage more than one classification scheme	●	◐	●
1B) associate keywords and descriptions with resources	●	●	●
1C) multiply classify and manage the same document	●	◐	◐
1D) copy resources without duplicating documents	●	●	◐
1E) access from a web browser	●	○	●
1F) automatically re-link resources	●	○	◐
1G) view who has seen a particular document	●	○	○
1H) view who has expressed an interest in a document	●	○	○
1I) find out where a document is classified	●	○	○
1J) control access permissions for reading and editing	○	●	●
1K) find out what documents other people have seen	●	○	○
2A) search the meta-data for all documents in the store	●	●	●
2B) view information about the owner of a document	●	◐	◐
2C) register an interest in a folder or document	●	○	●
2D) search an organisation for document owners	●	○	◐
2E) perform full-text search through documents	●	●	○
2F) create pointers between classification folders	●	●	◐
2G) reuse document fragments	●	◐	○
2H) retain a dynamic link to document fragments	●	◐	○
2I) specify a new document presentation style	●	◐	○
2J) move resources between classification folders	●	●	◐
2K) include one folder structure within another	●	○	○
2L) view information about a folder owner	●	○	◐
3A) create and follow links between documents	●	○	◐
3B) create and view document annotations	●	◐	○
3C) restrict the application of links and annotations	●	○	○
3D) view creator of links and annotations	●	○	○
3E) adapt the interface for different user types	●	○	○

Table 6-4: Satisfaction of CSFs by Three Approaches

Microcosm - A Modern Hypertext System

The Microcosm project and range of commercial products from the University of Southampton (UK) represent many years of research in the area of large-scale hypertext management. As such, it is able to offer robust solutions to a number of CSFs identified in this chapter. Microcosm started by focussing on the provision of a hypertext system that was able to automate linking between electronic documents within large heterogeneous document collections. In recent years, during the development of this thesis, the scope of Microcosm has extended to supporting more social aspects of a hypertext system - the mining of document trails (MEMOIR [DeRoure98], [Pikrakis97]) and the grouping of documents in application-specific collections (Microcosm TNG [Goose97]). Microcosm is able to provide functionality not available within MRA-HE, notably the construction of automated hyperlinks and other hypertext mechanism such as the provision of document 'tours' that end users are able to follow. It is clear that a system that incorporated all the ideas from Microcosm and its extended mechanisms would be able to serve many of the critical success factors from the scenarios identified in this thesis, but a few notable omissions follow:

- Microcosm provides an 'umbrella environment' for documents within its store, which may be of any type. As such, it provides only a limited document model-based approach, which does not emphasise opportunities for resource-level reuse.

- Microcosm provides no explicit association between a document and the organisational unit to which it belongs. Its emphasis is on making all information available, not on maintaining an ownership of that information or restricting access to unhelpful information such as inappropriate document links and annotations.

- Microcosm does not offer a working environment to support day-to-day working, although its work with Pirelli on specifically assisting maintenance tasks [Hall96] points to a single application for a work purpose. Microcosm's primary aim is on the development of large-scale hypermedia stores, not on supporting the day-to-day work of the individual system users.

- Microcosm is not a native Internet product, although a read-only option is available using a conventional web browser.

Due to the inability to fully assess a Microcosm product, and because it is not targeted as a document management system per se, it was neither possible nor advisable to rate its success against the identified critical success factors. Microcosm would rate highly on the hypertext CSFs, but the absence of support for focussed work and lack of emphasis on document classification and reuse renders a direct comparison meaningless. It is included here, however, as a well-developed hypertext system to highlight the way in which traditional document management and hypertext appear to be converging upon the same problem domain. Document management systems are developing understanding of document content, while hypertext systems are beginning to address the social aspects of document classification and use.

6.3 User Feedback

In order to assess the utility and usability of the MRA-HE implementation, a number of preliminary user trials were conducted towards the end of the research period. The aim of the user trials was to assess the ideas implemented in MRA-HE in terms of their usefulness and usability, so assessing the suitability for further future development. Due to the extensive scope of the demonstrator, it was decided that a baseline requirement for potential evaluators should be an understanding and experience of using current VSP Resource Rooms. This baseline requirement allowed the feedback sessions to build on the users' existing knowledge and enabled direct comparisons to be made with current resource rooms. Three types of users were approached, providing three distinct perspectives for evaluation

- members of the current VSP development team, who have spent three years working with the technical implementation of VSP resource rooms and their users.

- lecturers from the School of Computer Studies (University of Leeds) who experience in using existing VSP resource rooms 'in anger'[23] to provide supporting documentation for a number of undergraduate and postgraduate course modules.

- a member of the Centre for Virtual Working Systems responsible for usability testing and system evaluation activities for the Virtual Science Park.

Five feedback sessions took place, each lasting two hours and involving a presentation of the aims and scope of the MRA-HE implementation (reproduced in Appendix C), a guided tour around the MRA-HE system and a set of simple tasks to perform using the system (Appendix D). Throughout the feedback sessions the users were encouraged to feedback their thoughts on MRA-HE usability and wider issues. After the feedback sessions each user was asked to complete an evaluation form assessing the utility and perceived usability of the system, and the ways in which it performed better or worse than the existing VSP resource rooms. Lecturers were asked how they thought the MRA-HE system would better assist their specific needs in supporting their teaching activities, and the other user types were asked to assess how the MRA-HE would satisfy any outstanding requirements from current VSP tenants. The responses from the evaluation forms is reproduced in Appendix E. Although a full user evaluation of the MRA-HE system lies outside the scope of this thesis, early feedback from the surveys was encouraging.

---

[23] One of the tutors noted how a system can only be assessed by using it 'in anger' - a situation that could not be fully recreated within the limited evaluation sessions.

6.3.1 Value of Document Reuse

The first question on the evaluation form was intentionally open-ended: 'Is document reuse currently of interest to you/your customers?' The aim of this question was to ascertain each user's perceived value of document reuse. All responded positively, providing a number of examples. The lecturers saw 'many opportunities within teaching' for reuse at each of the resource, folder and resource room levels. At the resource level, both of the lecturers explained how they currently achieve a level of reuse through cutting and pasting existing materials into new contexts, but suggested that this was far from an ideal solution. One tutor cited the case of test papers that remain mostly static but need to include different code excerpts for different purposes. He noted how this is only currently possible through the creation of a brand new document each time, creating a great deal of duplication. The VWS research officer also cited a simple case whereby the VSP training manual is used across each tenancy but requires small modifications to contextualise the document within particular tenancies. Both situations call for a document-model based approach that can treat documents as more than objects to be stored and distributed. At the folder level, one of the lecturers noted that there is a need to teach 'the same material at different levels' noting there was 'substantial overlap' between the two levels. This represents a clear call to be able to use some materials in more than one context. The VWS research officer again cited the VSP training manual example, regularly wanting to know where across current VSP resource rooms the manual was being referenced. Both members of the VSP development team recognised the value in resource reuse 'particularly for the teaching (model), but also in project management and consultancy'.

6.3.2 Usefulness of MRA-HE Facilities

Having seen the presentation and demonstration, each user was asked to rate 13 facilities of the system that they had seen demonstrated in terms of how useful those aspects would be, either to themselves (as tutors) or to current or future VSP customers. Each user rated the following 13 facilities in terms of usefulness on a scale of 1 (very useful) to 5 (not at all useful):

a) Enterprise-wide access to documents
b) Internet access to resource rooms and document resources
c) Links between documents and their owners and organisations
d) Hypertext facilities (history trails, annotations and link creation)
e) Typed hyperlinks, trails and annotations for selective application
f) Registration of personal interests and notification of updates
g) Restricted view on documents through local 'trust' mechanisms
h) Different interfaces for different types of user (e.g. tutor/student)
i) Discovering new documents

j) Creation of a new document resource from existing documents
k) Multiple classification of documents
l) Reuse of one part of a resource room within another resource room
m) Ability to support change within the enterprise

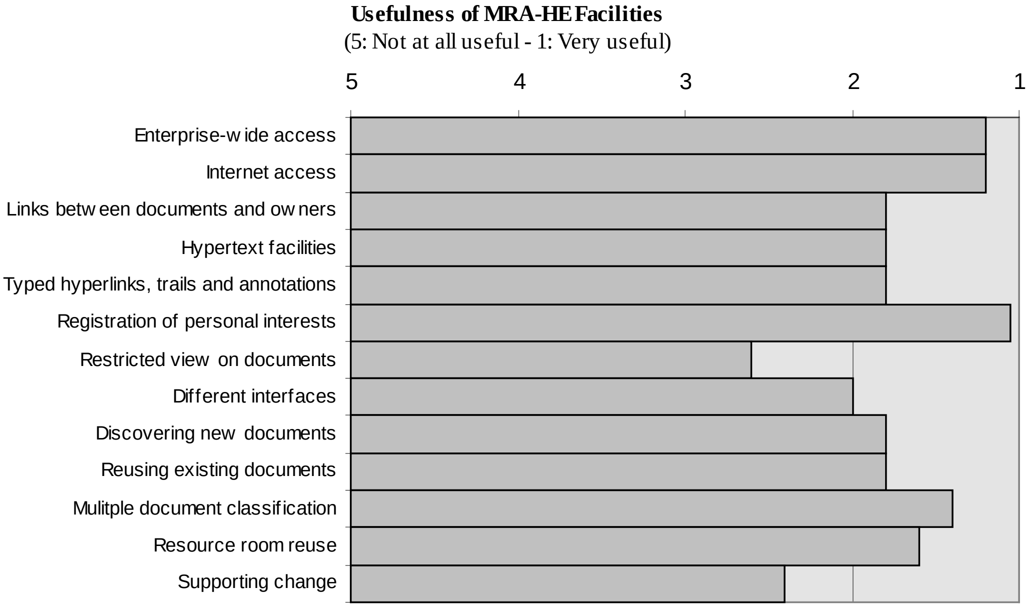


Figure 6-1

Figure 6-1 shows the average usefulness rated for each of the 13 MRA-HE facilities demonstrated. The average ratings represent an encouraging assessment of the usefulness of the key aspects of the demonstrator. All users rated enterprise-wide access and Internet access highly, highlighting the validity of the move away from localised storage into the ubiquitous networked environment. All users also rated the registration of personal interests as very useful, with one user explaining how, with the increasing amounts of available information, it is no longer possible to manually track every 'avenue of interest'. Hypertext facilities such as links and annotations received a more reserved response by users, with one of the tutors expressing concern not over the mechanisms for such approaches, but on the quality of the links and annotations. There was a sense of confusion over MRA-HE's ability to support change. The confusion is understandable with the limited scope of the demonstrator, and further assessment of the change mechanisms would be required if longer-term trials were conducted. The least useful facility was that of selectively applying only links and annotations created by people in a user's 'trusted people' list. The perceived usefulness of such a mechanism may increase as reuse mechanisms provide access to the same documents from different resource rooms - this is not very easy to show in a limited demonstration. The way in which the restricted view is currently implemented independently of other permission mechanisms made it appear trivial to one of the users; in a future implementation it would form part of the overall permissions functionality, and so may be of more obvious significance. More encouragingly, each of the three levels of reuse was accepted as being useful, especially the ability to multiply classify a single document.

6.3.3 Usability of MRA-HE Facilities

The usability survey attempted to gauge from the users how easy it was to perform the simple tasks outlined and demonstrated during the feedback sessions. In three of the cases the user's existing knowledge of VSP resource room functionality and a brief introduction to the differences between the VSP and MRA-HE provided enough cues to allow the user to perform the tasks themselves. Two users, however, were shown the tasks rather than performing them themselves, due to time restrictions, and asked to rate the perceived usability of the system rather than the outright usability. The users were asked to complete 6 tasks of varying complexity that used a range of different facilities of the MRA demonstrator.

1. Add a document to a resource room folder
2. Find a document and a contact person
3. Find out who has read a particular document
4. Prepare a resource room for a new teaching module
5. Prepare a resource room for a new teaching course, incorporating a number of modules
6. Use the resource room prepared in task 5 as a student with limited functionality

The tasks provided a structured demonstration of a number of MRA-HE facilities, showing the users or allowing them to discover how to achieve each of the tasks in turn. Users were encouraged to describe the process as they were performing the tasks to provide feedback on the interface and general usability of the system. After completing the tasks, each of the users was asked to rate the following facilities, in terms of their ease of use, on a scale of 1 (Very easy to use) to 5 (Very difficult to use):

a) Document addition
b) Construction and navigation of resource room folders
c) Navigation of the information model: history trails, document owners, organisations, interests, annotations, links and bridges
d) Creation of a new resource from existing documents
e) Multiple classification of a single document
f) Reuse of one part of a resource room within another resource room

Figure 6-2 shows the average usability for each of the six MRA-HE facilities used while completing the tasks. Again, the results are encouraging. While not targeting all of the MRA-HE functionality, the selected

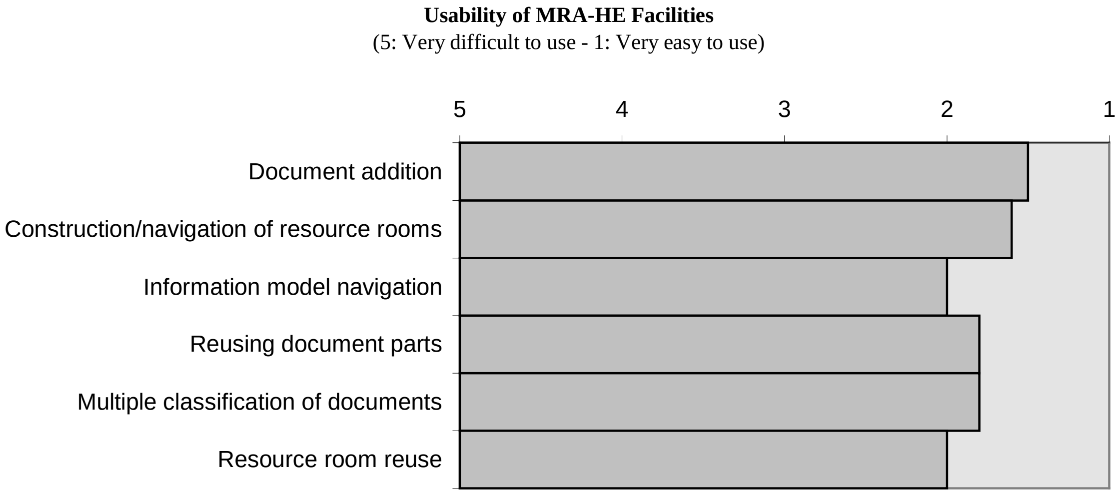


Figure 6-2

facilities aimed to highlight a typical range of user functions. The most common comment on system usability was the need to provide training to use the system, although a brief tour of the system and the demonstration of context-sensitive pop-up menus provided most of the users with enough cues to proceed. A noticeable problem with copying and moving documents was highlighted in two of the sessions: the MRA-HE notion of 'copy' was expected by some to create a new copy of the resource, whereas the system actually creates a link to the original document resource. It was suggested that a change in terminology would assist clarity on this point, but that the mechanism of multiply classifying a single document was valuable.

Of the six facilities rated, information model navigation and resource room reuse received the worst scores. One of the tutors suggested that the richness of the information model 'might take time to become comfortable with' and that it should be tailorable to avoid confusion. MRA-HE is already able to do this by removing navigation options from the relevant pop-up menus, but this could not be shown fully within the time scale of the feedback session. The resource room reuse process was confusing to most of the users, particularly the option to either point to one folder from another or actually include one folder within another, creating a more seamless integration. The confusion was similar to that encountered during implementation of current VSP resource rooms, but should become easier with an improved interface - from its usefulness rating, it is obviously useful, but not easy to implement easily. The resource-level reuse process was better received; its inherent complexity and novelty prevented it from being 'very easy to use', but an overall 'fairly easy to use' is an encouraging indication that the reuse process is understandable.

6.3.4 Overall User Ratings

At the end of the feedback session, each user was asked to provide an overall rating for a number of aspects of the MRA-HE implementation. Each user was asked to rate the following in terms of an overall rating on a scale of 1 (Excellent) to 5 (Very poor):

a) Ease of navigation
b) User interface
c) Presentation of information
d) Document reuse mechanisms
e) Ability to support distributed and diverse teaching needs

The overall average user ratings are shown in figure 6-3.

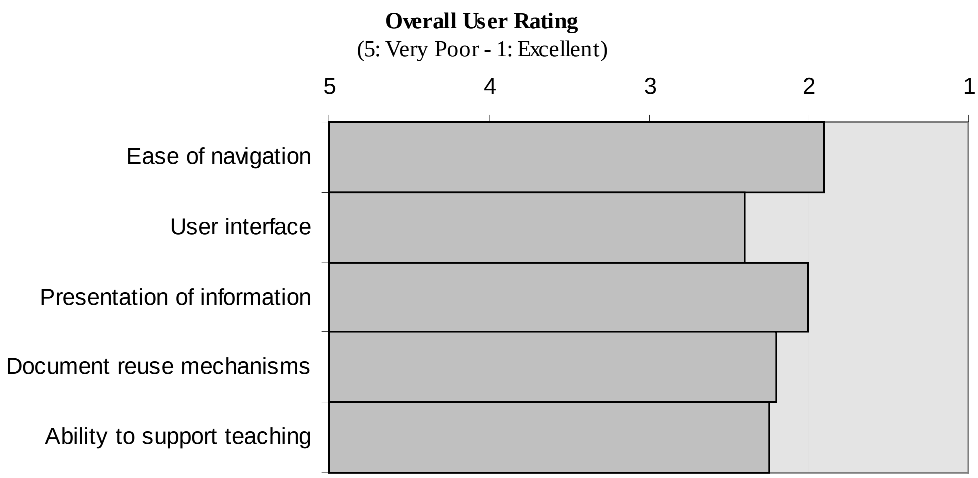


Figure 6-3

The overall user ratings are very encouraging, with the limited amount of time that the users had to become accustomed to the newer elements of the user interface and added functionality. The use of pop-up menus was popular, if a little confusing initially to one of the VSP development team. The separation of the resource room structure from the resources stored within each folder was seen as an improvement over current VSP resource rooms, providing a persistent context for the other information displayed. The final aspect, the ability of MRA-HE to support distributed and diverse teaching needs, was a little too open-ended to obtain a realistic score. One of the tutors rated it '?', highlighting the difficulty in rating a system on such a broad objective in such a short time. Comments on the teaching support aspect were, however, encouraging: 'the potential is clear', 'the potential seems to be there', 'it has the potential to be a component in a very powerful system'. On the three types of document reuse, comments were more concrete and positive. One of the tutors commented 'I can envisage using the reuse facilities at all three levels. They all worked as I would have expected'. None of the users were negative about the reuse mechanisms implemented, although one tutor noted that a move beyond supporting more than one document types would considerably complicate the reuse process. Overall, reservation was expressed over the amount of time it would take to realise the true benefits and pitfalls of MRA-HE's more advanced features.

6.3.5 Perceived Strengths/Weaknesses over VSP Resource Rooms

As regular users of current VSP resource rooms, each user was asked for their perceived strengths or weaknesses of MRA-HE. A list follows of the opinions expressed:

- The MRA-HE interface is less cluttered than the VSP equivalent
- Generally, the MRA-HE navigation is improved, although there are still times when a user might feel lost
- The information model may be too rich for some applications - perhaps it should be tailorable
- Links from documents to their owners provides functionality above that currently available
- The use of history trails satisfies a demand that has been made by some existing VSP users
- The simple personalisation of the interface is a strength over the VSP
- More use of newer technologies allows more to be done on the client without server intervention
- The ability to view documents in the browser reduces the need to wait for proprietary document viewers to start up on the client machine
- The possibility of addressing resources at the sub-resource level is a clear advantage
- The reuse of fragments and recombination into a single resource is a clear strength
- The management of documents that are multiply-classified within MRA-HE is easier
- MRA-HE is definitely more flexible than VSP resource rooms
- 'I believe it could open up new ways of doing things as well as making current tasks more efficient'
- 'The MRA supports not only the exchange of formalised and codified knowledge (e.g. in the form of a research report) but also promotes informal interaction and the exchange of informal knowledge... the opportunity to explore (information) is based not on location or content but on implicit recommendation, social factors, etc.'

The only common weakness identified during the user feedback was an inability to provide a proper evaluation in such a limited time scale. Doubtless other, more concrete weaknesses would surface during more extensive trials: one user declined to list any perceived strengths or weaknesses due to time limit. Overall, however, user feedback suggests that the ideas behind MRA-HE represent improvements over current resource rooms and that all three types of document reuse are well founded and potentially useful .

6.4 Implementation Architecture Assessment

This section briefly assesses some of the design decisions taken during the implementation of MRA-HE and the ways in which they help to realise the critical success factors identified at the start of this chapter. During the feedback trials the current VSP development team members were able to provide feedback on lower-level implementation issues, which have been incorporated into this discussion. The author is also well placed to assess some of the technical decisions made during design, having implemented the demonstrator system.

One of the lower level decisions made during the design of the implementation architecture was of using XML and its related standards throughout the MRA-HE implementation. Aside from technical difficulties from using new technologies, the decision to go all-XML was successful in providing the required functionality. XML is able to represent both the document files that form the document store and lower-level information such as personal contact details. At no time in the implementation process did a decision have to be made to convert from an underlying information model to a document-like representation, because the document/database divide was never established. Personal information from which relevant details could be extracted in a data-like manner was the same data that could be passed directly to the client for display. The manipulation functions provided by the DOM API were more than sufficient for the needs of the server in both storing the entities and manipulating the relationships of the underlying information model. The XML implementation of the MRA information model provides a loosely coupled network of entities with no enforced primacy. This facilitated the provision of any number of classification schemes (CSF 1A), the flexible classification of a single document entity (CSF 1C, 1D, 1F, 1I, 2F, 2J, 2K) and support for a number of socially-oriented document approaches: document ownership, view history, annotation, links and personal interests (CSF 1G, 1H, 1K, 2B, 2C, 2D, 2L, 3A, 3B, 3D). This loose coupling is well suited to the networked but dynamically evolving nature of MRA information model, and further facilitates the selective application of mechanisms such as annotations and links (CSF 3C).

As a client-side technology XML was also found to be both workable and beneficial, providing full functionality through a standard web browser (CSF 1E). Through the application of XSL stylesheets to basic XML documents, the client was able to apply presentation and behaviour to raw XML for the end-user. The simple definition of rules within the XSL stylesheet created opportunities for lightweight interface customisation (CSF 3E) without the need to adjust the MRA proxy server implementation in any way. Client-side manipulation of the XSL stylesheet also provided lightweight solutions to document ordering on particular attributes (author, title, date authored, etc.) and in providing effective ways of displaying documents, either all at once or one slide at a time.

All three levels of reuse were assisted by XML. At the resource level, the ability to refer to a sub-document fragment facilitated the construction of new document content and structure from fragments using the

Document Object Model (CSF 2G, 2H). The separation of document presentation from content and structure allowed the consistent application of document style (CSF 2I). At the folder level XML provided multiple entries in the resource classification catalogue for a multiply classified document - this was a very straightforward but powerful mechanism, allowing a user to find out what documents were classified in a particular folder and in which folders a particular resource was classified. At the resource room level, the ability to embed one XML object within another enabled the seamless embedding of one resource room structure within another. The XML namespace mechanism was used effectively to avoid address conflicts when one classification was being included within another.

The layered implementation architecture provided a manageable and extensible development platform for the MRA-HE implementation. Underlying entities and relationships were modelled in XML and manipulated by a set of services, each of which performed a small set of operations on the underlying information. The event services operated on the XML-encoded event pool that was appended by each of the significant document services, offering an asynchronous service additional to the synchronous document services. New services can be independently added to the document services layer and called by updating the XSL files that dictate user interface functionality. The overall loose coupling of entities and modularised document services makes the architecture extensible and suited to the dynamically changing needs of enterprise document management and to selective views on the common information store. For performance reasons it became necessary at times to create new relationship files such as the 'owns catalogue', which were generated automatically by mining the resource catalogue for the 'owns' meta-data elements and building them into a single file. In terms of supplying the task-based benefits described at the start of this chapter, the XML-enabled layered architecture provided the following mechanisms.

In terms of the scalability of the XML implementation, larger-scale tests have been conducted by importing existing VSP resource room structures and content into MRA-HE using the data integration tool described in the previous chapter. It is evident that loading XML directly from text files is sufficient at the client, but not so suitable for larger XML objects on the server, particularly the resource catalogue which contains an entry for each resource within the system. There is no need, however, to revert to a non-XML mechanism - merely to provide an XML database on the server that can more quickly parse larger XML structures. This is an interesting piece of future work, discussed again in the next chapter.

6.5 Conclusion

This chapter has described a range of evaluation methods used to assess the MRA-HE implementation of the Model-driven Reuse Architecture. As a preliminary implementation of the MRA, the MRA-HE system was tested against a number of critical success factors identified from a set of case scenarios, and proved able to provide solutions to almost all of the CSFs identified. A comparison was made with two existing environments to perform an assessment of how the MRA-HE implementation improves on currently available methods. A set of preliminary trials with existing users of VSP resource rooms provided encouraging feedback on the design decisions made and the usefulness of the system to their needs. Finally as assessment was made about the design and implementation of MRA-HE, highlighting the ways in which XML and the layered model-driven architecture assisted the scenarios. The next chapter builds on the success of the MRA-HE implementation to suggest ways in which MRA implementations may be able to satisfy domains other than higher education.

7. Conclusions and Future Work

This thesis has described an investigation into how documents may be better managed within an enterprise, placing emphasis on the importance and value of document reuse. Building on current thinking and existing systems, it has described the increasing value being placed upon an organisation's document resources and investigated ways by which organisations may capture that value to their benefit. The nature of document management is very different to that of more structured information management, containing less formal structure and implicitly encompassing more social aspects. This final chapter draws conclusions from the investigation undertaken in this thesis and points to ways in which this research may lead to further work.

7.1 Documents and Organisational Memory

Documents form a high percentage of the total amount of information held within an organisation. As organisations aspire to becoming learning organisations there are pressing demands for the effective management of those documents. Some office computer systems provide little more than the creation of page-oriented document files and limited distribution of those electronic files within the bounds of an isolated part of an overall organisation. More recently, corporate intranets and workflow tools have begun to assist the wider distribution of such documents, but the focus remains on the document as a black-box object that is stored, retrieved and passed around the organisation between people who understand the purpose of such documents. This thesis has argued that there is a need to move away from traditional integrated document management methods towards document-model based approaches. This will significantly assist in the retention and reuse of the overall organisational memory. While mechanisms such as the application of document meta-data and co-ordination technologies are valuable in potentially describing and transporting documents within the organisation, they are not sufficient. This thesis has argued that there are additional benefits in managing documents as complex objects with a clear separation of content, structure and presentational aspects. As structured objects, documents differ from structured information only in granularity, not in nature. Their suitability for automated manipulation has been investigated through an approach that permits both human consumption and automated approaches over the common document store. This is increasingly important with the volume of documents that form part of an organisational memory.

7.2 Influence and Motivation

Beyond an investigation into the nature of documents and how they serve organisations, this thesis has described a number of approaches that assist document management at different levels. Chapter 3 described the aims and development of the DEF STAN 00-60 (DS0060) document architecture for the UK Ministry of Defence (MoD), which formed a foundation for the Model-driven Reuse Architecture. The limited aims of DS0060 were to provide mechanical methods for the reuse of document parts across the military domain.

Its primary motivation was the reduction of costs by minimising duplication and the need to provide paper publications for specific purposes. While its reuse mechanisms were limited, the architecture developed an effective common source database from which publications (views) could be built for specific purposes. It also provided an emphasis on the construction of documents from sub-document components, each of which was coded according to the purpose it served. The DS0060 feasibility study reported in chapter 3 demonstrated the validity of the layered architecture and the potential for the construction of task-specific documents from collections of 'independent' data modules. The limited scope of DEF STAN 00-60 made it deficient in a number of ways that would make the architecture useful to a lesser-structured and more dynamic organisation. Specifically, DS0060 was designed for completeness and for a limited purpose, unable to explicitly support change within the military equipment it served.

In order to extend the ideas from DS0060 this thesis considered a second domain - that of supporting teaching and learning at higher education. An enterprise engaged in such a domain has much broader document management requirements and is much more person-oriented than the product-process focus of DS0060. The scenarios used to capture user requirements were based on the need to provide more flexible teaching models than traditional university practices currently allow. The university was identified as a suitable enterprise requiring improved document management to satisfy its varied teaching requirements. The particular user needs of supported learning in higher education identified were the accurate collection of a set of candidate learning resources and the selection of a subset of these resources to satisfy a particular learning demand. This provided the 'discovery and reuse' aspects that are examined throughout this thesis and new Model-driven Reuse Architecture proposal.

It became apparent from a review of the research literature that hypertext approaches such as those developed in the Microcosm product were suited to the document discovery process for a number of reasons, outlined in chapter 2. The Model-driven Reuse Architecture was designed to overcome many of the deficiencies present in DS0060 and incorporate change, hypertext mechanisms and explicit support for document discovery and reuse within the enterprise. The local availability of Virtual Science Park resource rooms provided an organisational focus to the architecture, emphasising document ownership within the enterprise.

7.3 MRA and MRA-HE

Chapter 4 described the design of Model-driven Reuse Architecture, based on the business object framework architecture. Its high level requirements were identified using an extended DS0060 scenario and scenarios that identified three person roles within the domain of learning in higher education. The MRA framework architecture in chapter 4 led to an implementation specific to the domain of higher education. MRA-HE (Model-driven Reuse Architecture - Higher Education) was implemented using a web server/proxy server/web client paradigm. With justification, XML and its related standards were used throughout the implementation to provide the storage, manipulation and delivery of documents and associated information. The MRA-HE demonstrator was evaluated in chapter 6 in terms of its success in fulfilling the identified scenarios and improving on existing methods. Utility and usability feedback was also obtained from existing users of VSP resource rooms, providing a valuable and positive critique of the system and architectural aims. The intention behind developing a domain-specific implementation of the MRA was to create a tangible instance of the overall architecture for evaluation purposes. While the evaluation focussed specifically on the MRA-HE implementation and the domain-specific scenarios, it is clear that many of the MRA ideas are applicable to other domains.

Although it has not been explicitly shown, the DEF STAN 00-60 document architecture is a subset of the more general MRA. The organisational and hypertext elements of the MRA would not be used, but the ability exists within the MRA to store low-level documents and construct publications from limited subsets of those documents. One resource room structure would represent the product and process breakdown of the data module code (DMC); others would represent existing output specifications in the form of chapter/section/subsection publication (IETP) structures. Data modules would be input via the DMC resource room and task-specific publications would be accessed from any one of the IETP resource rooms. An MRA implementation would provide users with a richer information environment if required and publications could now be viewed in a standard web browser rather than having to use an expensive SGML viewer.

7.4 The Role and Advantages of the MRA

The first two chapters of this thesis outlined some of the broader demands on organisations to better manage and reuse their document resources. This section returns to the demands from chapter 1, highlighting some of the specific ways in which aspects of the MRA may assist these demands.

Increased desire to manage knowledge within the enterprise

Organisation memories are motivated by the desire to preserve and share the knowledge and experiences that reside in an organisation [Sumner98]. Returning to Koulopoulos's definition of a document as 'a collection of information, authored for the purpose of transferring and preserving knowledge' [Koulopoulos95], documents can clearly assist in making explicit the implicit organisational memory. A number of features of the MRA make it potentially useful in assisting the process of organisational learning:

- the MRA is able to provide distributed access to an enterprise's document store, accessible from any point of the organisation with standard Internet access. Distinctively its use of XML/XSL for document mark-up and presentation permits advanced viewing behaviours in a standard web browser.

- the MRA is able to perform a search across all documents, organisational units and document classifications to locate resources relevant to a current need. The association between a document and its owner and multiple classifications presents new opportunities for collaboration and document discovery.

- the MRA provides localised classification views of documents, enabling the storage and retrieval of documents according to an understood classification scheme but not precluding the possibilities for cross-classification document discovery. The use of multiple resource rooms is similar to current VSP resource rooms, but the MRA's provision for multiple classification emphasises resource rooms as views on an enterprise-wide document store.

- the MRA supports documents marked up in semantically rich XML to retain a level of semantic meaning. Although MRA-HE provided only one document type for simplicity, an organisation would be able to define any number of document types that may represent document purpose. An explicit representation of document types enables a level of document content manipulation not available from more proprietary document formats.

- the MRA event service provides user notification of changes to the document store according to the user's expressed interests and ensures integrity of the document store.

Increased variation of products/services

Increasing variation across a range of products or services inevitably increases the amount of duplication that needs to be managed. The MRA assists in the reduction of such duplication through the provision of views on a common document store and the use of multiple resource room classifications. Only those document components relevant to the current need are displayed, localising the information for a particular need. At a lower level of abstraction, the MRA provides the ability to construct new documents from sets of existing sub-documents, while optionally retaining dynamic links to the source documents.

Distributed working

At a basic level, distributed working highlights the need to provide distributed access to the same document store. Such a need can be satisfied through the provision of wide-area networks or an intranet. There is also a more complex situation, where disparate teams working for the same enterprise will have their own working practices and ways of classifying and storing information. The MRA's use of XML provides a structured mechanism by which representational diversity may be managed across organisations. In order to support contextual diversity a system should support a local language and concept set, but recognise the relationships between these local languages. The MRA can assist this scenario in some way through the provision of a number of localised resource rooms that provide a view on the common document store. The multiple classification of documents or use of classification bridges would assist the availability of information across classifications.

Changes in working patterns

Modern organisations face constant change at all levels of the enterprise, creating the need to adjust information systems used to support the organisation. The MRA supports change by providing a model-driven approach whereby the underlying information model is able to adapt without needing to adjust the higher-level document services. The MRA is able to support change within any of the entities it models - the organisational structure, classification structure or the allocation of documents to a particular classification. Although predominantly a synchronous architecture, supplying responses to user requests, the event system is able to perform integrity checks across the underlying model. Two examples follow, illustrating the way in which the MRA event service is able to maintain document integrity across the enterprise.

- If someone leave the organisation, the MRA is able to adjust the ownership of the documents they currently own to retain the documents within the enterprise

- If a user deletes a document that a number of resource rooms point to, each affected resource room owner is able to make a copy of the document locally, taking ownership of the new document instance. This reduces the chances of a user losing an important resource, without incurring the need for each user to make their own copy from the outset.

Long-term storage

Some enterprises produce and support products or services that span many years. Military products, for example, have a typical lifecycle of 20 years [Brooks94] - far longer than the planned support for current versions of word processors. For such products, documents printed on paper have a potentially much longer useful life span than the equivalent electronic version. It is the aim of DS0060 to replace paper-based support with documents represented in ways that will outlast any single generation of word processors. In the same way the MRA supports documents marked up in XML, reducing the need to rely on any single technology with a limited shelf life.

7.5 Support for Reuse

The primary novelty of the Model-driven Reuse Architecture is its support for reuse at resource, classification folder and resource room levels. Using XML as the representational language for the documents it stores, the MRA is able to address document fragments and provide explicit support for document creation from existing fragments. Placing an emphasis on the multiple classification of documents across resource rooms permits a higher-level reuse whereby a document may be reused for more than its original purpose. At the resource room level the MRA is able to make whole classification structures and their documents available to other resource rooms. The use of XML to represent documents and resource rooms promotes the ability to represent networks of documents and resource room folders that are able to assist the diverse and changing needs of an organisation.

7.6 Future Work

This thesis has described the development of a document architecture that emphasises the reuse of documents, document parts and collections of documents. The loose coupling and abstraction of the MRA layers makes possible the future inclusion of other research or commercial document management and hypertext aspects. The ideas and aims underpinning the MRA also highlight a number of areas of future work which require further study if document reuse is to become an integral part of an organisation's document management strategy.

7.6.1 Extensions to the MRA

The extensible nature of the MRA enables a wide range of future work to be incorporated at all layers of the architecture. This section briefly highlights possibilities for extension of the MRA using methods or products currently available.

Extended search

The search service implemented within the MRA-HE demonstrator performed a basic syntax search across documents, organisations and resource room structures separately. Future MRA implementations may extend this basic search to combine searches across a combination of the business models to elicit more relevant results or results that may not be discovered at all using a basic search. For example, a set of teaching slides describing some aspect of virtual working environments may not contain the phrase 'virtual working environments', but may be stored under a classification that did contain that phrase. If a user was searching for 'virtual working environments user model' it may be useful for the system to return any documents that contain the word 'user model' within any classifications that contains the phrase 'virtual working environments'. This two-phase search may in turn uncover more appropriate documents by searching document content and the document context provided by resource room classifications. Similarly a search across document content and organisational models may reveal a more useful set of candidate documents by limiting the search to only those documents owned by a particular part of the organisation.

Future work in searching for relevant documents for reuse may also include the addition of information retrieval methods such as a dictionary or thesaurus-based search to elicit further suitable documents from the document store. More complex semantic searches currently performed by commercial systems such as RetrievalWare from Excalibur could also be incorporated as additional services at the document services layer of the MRA.

Extension of document mechanisms

While the MRA provides hyperlinks, annotations and document view trails, other available document mechanisms could be easily incorporated into the overall architecture to improve navigation or document discovery. Currently available examples include the document clustering techniques used in the Hyper-G research project, document tours such as those implemented in Microcosm and document recommendation such as that available with British Telecom's Jasper product. Any or all of these mechanisms could be added to the MRA, providing extra means of navigating the document store, pre-defining paths through documents, or assisting collaboration by recommending people or documents. Automated link mechanisms such as those available within the Microcosm architecture could also be added as a document service, adding links to the separate resource link catalogue based on similarities in document content.

The loose coupling between the stored meta-data at the document abstraction layer and the processing methods at the document services layer allows each element of meta-data to be used for more than one service purpose. Document trails, for example, are currently used to allow a user to backtrack through the documents viewed and to find out who else has viewed a particular document resource. A new service could automatically provide pointers to information about people who have seen a particular document by searching the same document trailbase, reusing the information originally collected for other purposes. Although some new document mechanisms would be able to directly manipulate the information models already defined within the MRA, others would require the creation of extra data structures at the document abstraction layer. If document clusters were to be implemented, for example, a new set of meta-data could be created that described which clusters a particular document was classified under. Each cluster would be represented as a first-order object with its own attributes, in the same way as document annotations and hyperlinks are first-order objects in the current MRA. Services that operate on the document clusters would be implemented at the document services level, retaining the abstraction layers of the current MRA.

Version control and permission model

It is clear that extensive version control and permission mechanisms would be required if a commercial implementation of the MRA were to be undertaken in order to maintain integrity across the distributed store. These mechanisms were deliberately omitted from the MRA demonstrator because they are already studied and represented in other systems such as the Virtual Science Park and because other features were considered more valuable to the thesis. Versioning and permission mechanisms could, however, be implemented within the MRA as new document services on the proxy server and called by existing services to complement their operation. The permission model would not only incorporate access rules for documents, but also for classification structures, annotations, hyperlinks and all other document mechanisms used, to limit the bounds of hypertext and restrict access to confidentially sensitive information. The versioning model could similarly be applied not only to documents but also to any other aspects of the MRA business models for which change management is considered valuable.

Extension of the event system

The current MRA event system permits an alert service to its users and periodic integrity checks across the document store. The event system represented within the MRA-HE implementation aimed only to demonstrate the ability to incorporate an asynchronous model within an essentially synchronous request-response architecture. The event system could be easily extended in the future to incorporate additional integrity checks or more detailed workflow models such as those available in current groupware products (e.g. Lotus Notes). It may be desirable, for example, to send a notification to a group of users when a document has been added to a particular resource room folder regardless of whether those users have explicitly expressed an interest in that folder. It would further be possible to create a rule that each user in that group should read the document before the document was considered 'accepted' – in this scenario the event service would access the view history to assess the documents readership to date by the group members.

The MRA integrity service demonstrated by MRA-HE is intentionally limited in its scope, but is currently able to highlight updates to documents and classifications that a user relies on in some way. A more thoroughly designed integrity service would be essential in enabling enterprise-wide sharing and reuse. Without such a service users may lose access to documents or document parts that their constructed resource rooms depend on. A suitably well-developed integrity service, for example, may retain a document marked for deletion until the owners of its dependent documents or classifications had acted upon the proposed deletion. For each user affected this may involve a search for a suitable replacement document or the creation of a local copy of the deleted resource for continued use. It is clear that an effective integrity service encouraging reuse would need to define a wide range of policies to retain integrity across the document space, while upholding the ultimate ownership of a document or classification.

Improved interfaces

Within the MRA-HE implementation the user interface is generated on the client machine using XSL stylesheets applied to XML information, retaining the semantic of that information throughout lower layers. This approach lends itself to the simple application of improved interface mechanisms as they arise. At a simple level, user interface components such as the Hyperbolic Tree from Inxight Software[24] would permit more than a one-dimensional hierarchical view of resource room structures. More interestingly, a virtual world environment could be applied to the MRA by converting its XML structures into VRML, much in

---

[24] http://www.inxight.com/ (Last verified 8th March 2000)

the same way as they are currently converted to HTML using XSL stylesheets. The generation of the interface at the highest layer of the architecture also prepares the MRA for the expected increase in the variety of user clients, as mobile business and digital television become a larger part of the overall web client market.

Collaboration

The MRA currently provides access to document content using a standard web browser and the HTTP1.1 web protocol. A common requirement is the ability to work collaboratively on such documents. On-going work on extending the HTTP protocol to support collaboration (WebDAV – [Whitehead98]) could add much value to the MRA, which currently has very primitive methods for editing a document on-line. The MRA could also include on-line collaboration tools such as those currently available within the Virtual Science Park, allowing users to communicate through third-party video-conferencing, shared whiteboards and application sharing.

Document type management

The MRA-HE demonstrator was implemented to manage only one document type – sets of teaching slides containing simple bullet lists. The use of only one document type permitted the implementation of the overall architecture without becoming involved with defining low-level translations between document types. The MRA recommendations, however, promote the use of as many document types as are necessary to support the work purposes of the user community. Document type definitions (DTDs) describe the composition and imply the purpose of a document, adding structure to the document content. An extended MRA should support a range of document types and, more powerfully, provide a mapping between document types so that information may be extracted from one document type and reused meaningfully in another.

Scalability

The MRA-HE demonstrator was implemented on a single server performing the roles of both the MRA proxy server and the storage server. It is evident that a system targeted at large distributed organisations may require a level of service not attainable from a single server, and that distributed storage and processing may be required. Although not proven in this thesis, the modular nature of the proposed architecture should enable a scalable implementation across a number of server machines. The use of standard XML files and the HTTP protocol throughout the MRA-HE implementation should allow distribution of both the underlying data and the document services that act upon that data. A scalable MRA implementation may, however, require an additional focus on data replication and coordination. It is also clear from the initial MRA-HE implementation that a more efficient storage mechanism would be required for increasing amounts of stored information, maybe through one of the recently-developed XML-enabled databases.

7.6.2 Related Future Work

During the development of the MRA it has become evident that a number of related areas of research may complement future versions of the architecture. This final section briefly outlines some research problems that could further assist the aims of the MRA.

Enterprise modelling

The research literature describes a great deal of activity in the area of modelling enterprise activities and processes in order to support cross-enterprise information systems. Enterprise modelling is currently undergoing investigation at various levels of granularity: Carole Goble's work [Goble94] is concerned with the low-level assumptions that can be made from the currently known information and rules of deduction that can operate over that information. Using the restricted domain of medicine the aim of Goble's work is to create more information than has been manually entered through deduction. At a higher level of abstraction a number of projects (e.g. the Enterprise Project [Stader96], the Toronto Virtual Enterprise (TOVE) [Gruninger96], the Holonic Enterprise Modelling Ontology (HEMO) [Presley97]) are concerned with enterprise modelling, taking an enterprise-wide view of an organisation. An integrated enterprise model provides a coherent picture of the enterprise and supports efficient management of inter-operations [Fillion95]. These projects aim primarily to model the enterprise for the purpose of holistic analysis and process re-engineering, but also point to the possible use of the resulting models in supporting enterprise-wide information systems.

By associating documents with appropriate parts of an enterprise model it may be possible to assign some level of cross-enterprise classification to documents, much as individual work-focussed resource rooms classify documents within the current MRA. The longer-term aim of enterprise models is to incorporate change into the models, creating a 'living enterprise model' [Whitman97] that can be used to support any number of enterprise information systems. This presents an exciting opportunity for the Model-driven Reuse Architecture – to use an enterprise model not only to store and access documents using a model that is common across the enterprise, but also to manage the effects on the document store of change within that enterprise.

Versioning and document ownership

Although the previous section claimed that methods for supporting document versioning are already studied and represented in current document management methods, this is only true for versioning over relatively fixed documents. David Levy, however, points to the inherent fluidity of documents as they change to describe an ever-changing world [Levy94]. Whenever a document changes it may be seen either as continuity of identity or as the creation of something new and different. When a document is constructed from a number of smaller documents, each of which may undergo versioning, this problem is exacerbated. A set of teaching slides, for example, may contain single slides from a number of other sources. If one of those sources is updated the change will impact all sets of slides within which that source is used. As a result the set of teaching slides has changed and may or may not be deemed to be a new version. Furthermore, if the changes are significant the new set of slides may or may not be considered a new document altogether rather than a new version of the same document.

It is clear that a new versioning mechanism is required in order to make document reuse at the resource level an attractive and controllable proposition. There is an apparent need to develop a new versioning model that is far more adaptable than existing mechanisms, able to model the subjective distinction between a new version of the same document and a first version of a new document, and able to support versioning at both the document and sub-document levels.

As documents become reused for new purposes and used as parts of other documents, document ownership also becomes a major issue to be considered and modelled. The explicit representation of document ownership has been considered a valuable aspect of the MRA throughout this thesis, providing a user with a wide range of information about a document owner. As discussed in the previous section, document permissions were intentionally omitted from the MRA implementation, so all users had full access to all documents within the demonstration environment. For a commercial implementation, however, uncontrolled access to the document store would probably not be desirable: users should only be given access to a restricted subset of information. This raises issues specific to the creation of documents from a number of document fragments when access to those fragments is restricted in some way. A set of slides in a resource room, for example, may be accessible to a group of students but a document fragment contained within those slides may have permissions set to prevent a student from viewing that fragment. Even if a document fragment has appropriate viewing permissions when it is selected for use, the owner has the right to change permissions on that fragment at any time.

It is clear that versioning and permission models currently suited to controlling and managing documents are not immediately applicable to a document-model based approach such as the MRA, where a user is able to access documents at the sub-document level. A great deal of future work lies in these areas to create and maintain the document integrity and security necessary for a commercially acceptable document reuse system.

Changing work practices

The final area of future work lies outside the technical architecture proposed in this thesis. The DEF STAN 00-60 document architecture described in chapter 3 highlighted the need to change working practices as well as technical practices and to perceive the enterprise's documents in a different way. When describing the conversion of existing learning materials into electronic form, the Open University (UK) has also recognised added value in rearranging its course structures and creating more customised courses

[Lincoln97]. During user feedback on the MRA-HE implementation, one of the lecturers noted independently that change would be required to fully take advantage of the potential benefits of the system. Not the least of these changes is the need for authors to regard documents at lower levels of granularity – the sub-document level – so that document fragments may be used within new contexts. The changes in work practice required to more fully enable document reuse presents a very difficult problem that David Levy suggests may be assisted by the disciplines of textual criticism and scholarship [Levy93]. The challenge is upon software architects and business managers alike to recognise electronic documents as more than a replacement for paper documents and change working practices to take advantage of new document management methods.

7.7 Conclusion

The MRA supports the localisation of documents as instances of recognisable social types using a document model-based approach that places documents within work-specific resource room classifications and within the organisation they serve. It enables the synthesis of information, people and the work that people perform, assisted by documents. The separation of a document's content and structure from its presentation and the association of a document to its work purposes facilitate document discovery, re-purposing and reuse. This thesis has presented an architecture that supports the collection, combination, customisation and classification of document resources for new purposes, so serving the increasingly diverse needs of learning organisations. Using a case study of supporting learning in higher education to create a domain-specific implementation, the MRA-HE implementation of the more general MRA has been shown to be both useful and usable. Although there exists a number of obstacles to overcome before a commercial implementation of the MRA can be built, this thesis has argued that the need to reuse documents across an enterprise is sufficient to warrant further investigation and development. It is clear that the implementation of a successful reuse strategy across an enterprise places demands on both the technical infrastructure and the working practices of the people that together form an organisation and its memory. Only through the careful implementation of properly designed systems within well-managed enterprises may much-needed reuse become a central theme in a successful corporate document management strategy.

References

[Amor97] R Amor and M Clift. Documents as an Enabling Mechanism for Concurrent Engineering in the Construction Industry. *Proceedings of the 1st International Conference on Concurrent Engineering in Construction (CEC97), London, UK*, pages 151–162, July 1997.

[Apte94] Chidanand Apte, Fred Damerau and Sholom Weiss. Automated Learning of Decision Rules for Text Categorisation. *ACM Transactions on Information Systems*, 12(3):233–251, July 1994.

[Argamon98] Shlomo Argamon, Mashe Koppel and Galit Avneri. Routing Documents According to Style. *Proceedings of the 1st International Workshop on Innovative Internet Information Systems (IIIS-98), Pisa, Italy*, June 1998.

[Ariadne98] Alliance of Remote Instructional Authoring and Distribution Networks for Europe (ARIANDE). ARIANDE's April 98 Provisional Recommendation. Working Document, Alliance of Remote Instructional Authoring and Distribution Networks for Europe (ARIANDE), April 1998.

[Autonomy98] Autonomy Inc. Autonomy Technology White Paper. Technical report, Autonomy Inc, 1998. Available on-line at http://www.autonomy.com/tech/wp.html (Last verified 8th March 2000).

[Balasubramanian94] V Balasubramanian and Helen Ashman. Long Distance Perspectives on Hypermedia: Keynote Address at the 1994 European Conference on Hypermedia Technology (ECHT94) by Douglas C. Engelbart and Christina Engelbart, Bootstrap Institute, September 1994.

[Balasubramanian98] V Balasubramanian and Alf Bashian. Document Management and Web Technologies: Alice Marries the Mad Hatter. *Communications of the ACM*, 41(7):107–115, July 1998.

[Bass98] Len Bass, Paul Clements and Rick Kazman. *Software Architecture in Practice*, chapter 7. Addison Wesley Longman Higher Education, February 1998.

[Bentley97] R Bentley, W Appelt, U Busbach, E Hinrichs, D Kerr, K Sikkel, J Trevor and G Woetzel. Basic Support for Cooperative Work on the World Wide Web. *International Journal of Human Computer Studies: Special Issue on Novel Applications of the World Wide Web*, 46(6):827–846, June 1997.

[Berghel99] Hal Berghel and Douglas Blank. The World Wide Web. Marvin Zelkowitz, editor, *Advances in Computers Volume 49*, pages 178–218. Academic Press, October 1999.

[BernersLee89] Tim Berners-Lee. Information Management: A Proposal. Technical report, World Wide Web Consortium, March 1989. Available on-line at http://www.w3.org/History/1989/proposal.html (Last verified 8th March 2000).

[Bosak97] Jon Bosak. XML, Java and the future of the Web. White Paper, Sun Microsystems, March 1997. Available on-line at http://metalab.unc.edu/pub/sun-info/standards/xml/why/xmlapps.htm (Last verified 8th March 2000).

[Branting96] L Karl Branting and James C Lester. Justification Structures for Document Reuse. *Proceedings of the Third European Workshop on Case-Based Reasoning (EWCR96), Lausanne, Switzerland*, November 1996.

[Branting97] L Karl Branting, James C Lester and Charles B Callaway. Automated Drafting of Self-Explaining Documents. *Proceedings of the 6th International Conference on Artificial Intelligence and Law (ICAIL97), Melbourne, Australia*, pages 72–81, July 1997.

[Broadbent98] Marianne Broadbent. The Phenomenon of Knowledge Management: What Does it Mean to the Information Profession? *Information Outlook On-Line*, May 1998. Available on-line at http://www.sla.org/pubs/serial/io/ (Last verified 8th March 2000).

[Brooks94] G Brooks. Changes in the Documentation Engineering Process. *Electrical Communications*, pages 173–180, 2nd Quarter 1994.

[Brophy97] Peter Brophy, Shelagh Fisher and Geoffrey Hare. Towards a National Agency for Resource Discovery. *British Library Research and Innovation Report 58 [RIC/G/364]*, July 1997.

[Brown96] John Seely Brown and Paul Duguid. The Social Life of Documents. *First Monday*, 1(1), May 1996. Available on-line at http://www.firstmonday.dk/ (Last verified 8th March 2000).

[Bush45] Vannevar Bush. As We May Think. *The Atlantic Monthly*, pages 101–108, July 1945.

[Chase94] Scott Chase. Three Prototype CDDM Systems. *Electronic Journal of Information Technology in Construction*, August 1994. Available on-line at http://itcon.org/ (Last verified 8th March 2000).

[Conklin87] Jeff Conklin. Hypertext - An Introduction and Survey. *IEEE Computer*, 20(9):17–41, September 1987.

[Croft91] W B Croft, Howard R Turtle and David D Lewis. The Use of Phrases and Structured Queries in Information Retrieval. *Proceedings of the Fourteenth Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval (SIGIR91), Chicago, Illinois*, pages 32–45, 1991.

[Cunningham93] Donald J Cunningham, Thomas M Duffy and Randy A Knuth. The Textbook of the Future. C McKnight, A Dillon and J Richardson, editors, *Hypertext - A Psychological Perspective*, pages 19–49. Ellis Horwood Series in Interactive Information Systems, 1993.

[Davies98] John Davies and Ralph Cochrane. Knowledge Discovery and Delivery. *British Telecommunications Engineering*, 17:25–35, April 1998.

[Davies98b] John Davies, Scott Stewart and Richard Weeks. Knowledge Sharing over the World Wide Web. *Proceedings of WebNet98, Florida, USA*, November 1998.

[Davis92] Hugh Davis, Wendy Hall, Ian Heath, Gary Hill and Robert Wilkins. MICROCOSM: An Open Hypermedia Environment for Information Integration. Computer Science Technical Report - CSTR 92-15, University of Southampton, UK, 1992.

[Demarest97] Marc Demarest. Understanding Knowledge Management. *Long Range Planning*, 30(3):374–384, June 1997.

[DeRoure96] David C DeRoure, Wendy Hall, Hugh C Davis and Jonathon Dale. Agents for Distributed Multimedia Information Management. *Proceedings of the First International Conference on the Practical Application of Intellgient Agents and Multi-Agent Technology (PAAM96), London, UK*, pages 91–102, April 1996.

[DeRoure98] David C DeRoure, Wendy Hall, Sigi Reich, A Pikrakis, Gary J Hill and M Stairmand. An Open Framework for Collaborative Distributed Information Management. *Proceedings of the 7th International World Wide Web Conference (WWW7), Brisbane, Australia*, volume 30, pages 624–625, April 1998.

[Dix97] Alan Dix. Challenges for Cooperative Work on the Web: An Analytical Approach. *Computer Supported Cooperative Work: The Journal of Collaborative Computing*, 6:135–156, 1997.

[DOD96] US Department Of Defense. The Integrated Data Environment - Continuous Acquisition and Lifecycle Support. White Paper, US Department Of Defense, October 1996.

[Drew97] Richard Drew. *Integrated Information Directory Services to Support the Innovation Process.* PhD thesis, School of Computer Studies, University of Leeds, UK, 1997.

[Esposito94] Floriana Esposito, Donato Malerba and Giovannni Semeraro. Multistrategy Learning for Document Recognition. *Applied Artificial Intelligence*, 8:33–84, 1994.

[Excalibur99] Excalibur Knowledge Retrieval Solutions. Knowledge Retrieval - The Critical Enabler of Knowledge Management. White paper, Excalibur Technologies, 1999. Available on-line at http://www.excalib.com/pub/KMKR-rwdm.pdf (Last verified 8th March 2000).

[Fillion95] Florence Fillion, Christopher Menzel, Thomas Blinn and Richard Mayer. An Ontology-Based Environment for Enterprise Model Integration. *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), Montreal, Canada*, August 1995.

[Fox92] Mark Fox. The TOVE Project: Towards a Common-Sense Model of the Enterprise. *Proceedings of the 5th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems (IEA/AIE92), Paderborn, Germany*, pages 25–34. June 1992.

[Gehringer98] Edward F Gehringer. Classroom Reuse Experiences: What Makes Course Materials Reusable? *Educators' Symposium, Object-Oriented Programming Systems, Languages and Applications (OOPSLA98), Vancouver, Canada*, October 1998.

[Goble94] Carole Goble, Sean Bechhofer, Danny Solomen, Alan Rector, Anthony Nowlan and Andrzej Glowinski. Conceptual, Semantic and Information Models for Medicine. *Proceedings of the 4th European-Japanese Seminar on Information Modelling and Knowledge Bases, Stockholm, Sweden*, June 1994.

[Goble97] Carole Goble, Jacki O'Neill and Joe Bullock. Authoring Using a Terminology Service. Poster and Report from the 8th ACM International Hypertext Conference on Hypertext (Hypertext97), Southampton, UK, Department of Computer Science, University of Manchester, UK, April 1997.

[Golovchinsky97] Gene Golovchinsky. What the Query Told the Link: The Integration of Hypertext and Information Retrieval. *Proceedings of the 8th ACM International Hypertext Conference (Hypertext97), Southampton, UK*, pages 67–74, April 1997.

[Goose97] Stuart Goose, Jonathon Dale, Wendy Hall and David DeRoure. Microcosm TNG: A Distributed Architecture to Support Reflexive Hypermedia Applications. *Proceedings of the 8th ACM International Hypertext Conference (Hypertext97), Southampton, UK*, pages 226–227, April 1997.

[Goranson92] H T Goranson. Some Initial Results from Enterprise Integration Studies. Reports of the International Conference on Enterprise Integration Modeling Technology (ICEIMT) Special Interest Groups, Deparment of Defense, USA, August 1992.

[Greenwood95] R M Greenwood, I Robertson, R A Snowdon and B C Warboys. Active Models in Business. *Proceedings of the 5th Conference on Business Information Technology (CBIT95), Manchester, UK*, November 1995.

[Gruninger96] M Gruninger and M S Fox. The Logic of Enterprise Modelling. P Bernus and L Nemes, editors, *Modelling and Methodologies for Enterprise Integration*. Chapman and Hall, 1996.

[Halasz88a] Frank G Halasz. Reflections on Notecards: Seven Issue for the Next Generation of Hypermedia Systems. *Communications of the ACM*, 31(7):836–852, July 1988.

[Halasz88b] Frank G Halasz. "Seven Issues" Revisited. *Keynote Address at the ACM Conference on Hypertext (Hypertext91), San Antonio, Texas, USA*, December 1991.

[Hall96] W Hall, R Crowther, I Heath and R Bernard. The Application of Large-Scale Hypermedia Information Systems in the Manufacturing Industry. *1995/6 Research Journal, Multimedia Systems Group, University of Southampton, UK*, 1996.

[Heeman92] Frans C Heeman. Granularity in Structured Documents. *Electronic Publishing*, 5(3):143–155, September 1992.

[Hitchcock98] Steve Hitchcock, Les Carr, Wendy Hall, Steve Harris, Steve Probets, David Evans and David Brailsford. Linking Electronic Journals - Lessons from the Open Journal Project. *D-Lib Magazine*, December 1998. Available on-line at http://www.dlib.org/ (Last verified 8th March 2000).

[Honkela95] Timo Honkela, Ville Pulkki and Teuvo Kohonen. Contextual Relations of Words in Grimm Tales, Analyzed by Self-Organising Map. *Proceedings of International Conference on Artificial Neural Networks (ICANN95), Paris, France*, pages 3–7, 1995.

[IMS99] Instructional Management Systems Project. IMS Meta-Data Specification. Draft Technical Report, Instructional Management Systems (IMS), 1999. Available on-line at http://www.imsproject.org/metadata/ (Last verified 8th March 2000).

[Kaminski95] Paul G Kaminski. The Financial Year 1996 Department of Defense Research, Development, Testing and Evaluation Program. Statement of the Under Secretary of Defense for Acquisition and Technology, US Department of Defense, March 1995.

[Kappe94] Frank Kappe, Keith Andrew, Jorg Faschingbauer, Mansuet Gaisbauser, Michael Pichler and Jurgen Schipflinger. Hyper-G: A New Tool for Distributed Hypermedia. *Proceedings of the International Conference on Distributed Multimedia Systems and Applications, Honolulu, Hawaii*, pages 209–214, August 1994.

[Karben99] Alan Karben. News You Can Reuse: Content Repurposing at The Wall Street Journal Interactive Edition. *Markup Languages: Theory and Practice*, 1(1):33–45, 1999.

[Kawalek94] P Kawalek and D G Wastell. The Development of a Process Modelling Method. *Proceedings of the 2nd BCS Information Systems Methodology Conference, Edinburgh, Scotland*, 1994.

[Koulopoulos95] Thomas M Koulopoulos and Carl Frappaolo. *Electronic Document Management Systems - A Portable Consultant.* McGraw Hill, 1995.

[Lassila96] Ora Lassila. Web MetaData: A Matter of Semantics. *IEEE Internet Computing*, 2(4):30–37, July/August 1998.

[Levy93] David M Levy. Document Reuse and Document Systems. *Electronic Publishing*, 6(4):339–348, December 1993.

[Levy94] David M Levy. Fixed or Fluid? Document Stability and New Media. *Proceedings of the 5th ACM Conference on Hypermedia Technology, Edinburgh, Scotland*, Sept 1994.

[Levy95] David M Levy and Catherine C Marshall. Going Digital: A Look at Assumptions Underlying Digital Libraries. *Communications of the ACM*, 38(4):77–84, April 1995.

[Levy99] David M Levy. The Universe is Expanding: Reflections on the Social (and Cosmic) Significance of Documents in a Digital Age. *Bulletin of the American Society for Information Science*, 25(4):17–20, 1999.

[Li92] Zhuoxun Li, Hugh Davis and Wendy Hall. Hypermedia Links and Information Retrieval. *Presented at the British Computer Society 14th Information Retrieval Colloquium, Lancaster University, UK*, April 1992.

[Lincoln97] Ches Lincoln. Webbing Course Units. EMERG Project Report, Electronic Media in Education Research Group (EMERG), The Open University, UK, November 1997.

[Lotus98] Lotus Development Corporation. LearningSpace Solutions for Anytime Learning. White Paper, Lotus Development Corporation, 1998.

[Malcolm91] Kathryn C Malcolm, Steven E Poltrock and Douglas Schuler. Industrial Strength Hypermedia: Requirements for a Large Engineering Enterprise. *Proceedings of the 3rd ACM Conference in Hypertext (Hypertext91), San Antonio, Texas, USA*, pages 13–24, December 1991.

[Manola98] Frank Manola. Towards a Web Object Model. Technical report, Object Services and Consulting, Inc, February 1998. Available on-line at http://www.objs.com/OSA/wom.htm (Last verified 8th March 2000).

[Matthews98] Brian Matthews and David Johnston. Using Parameterised Templates for Reusable Documentation. *Presented at the European Product Data Technology Conference, Cranfield University, UK*, March 1998.

[McCarty97] Willard McCarty. The Shape of Things to Come is Continuous Change: Fundamental Problems in Electronic Publishing. *Presented at Electronic Publishing - A Day Conference, University of London, UK*, January 1997.

[Mendes97] Maria Emilia Xavier Mendes and Wendy Hall. The SHAPE of Hypermedia Authoring for Education. *Proceedings of the World Conference on Educational Multimedia/Hypermedia and Educational Telecommunications (ED-MEDIA/ED-TELECOM), Calgary, Canada*, June 1997.

[Merkl95] Dieter Merkl. A Connectionist View on Document Classification. *Proceedings of the 6th Australasian Database Conference (ADC95), Adelaide, Australia*, pages 153–161, January 1995.

[Merkl97] Dieter Merkl. Lessons Learned in Text Document Classification. *Proceedings of the Workshop on Self-Organising Maps (WSOM97), Helsinki, Finland*, June 1997.

[MOD99] United Kingdom Ministry of Defence. Defence Standard 00-60. The United Kingdom MoD Standard for Integrated Logistic Support. Defence Standard, The United Kingdom Ministry of Defence, 1999. Available on-line at http://www.demon.co.uk/ilsuk/ds60hp1.html (Last verified 8th March 2000).

[Multicosm98] Multicosm Ltd. Knowledge Management. White Paper, Multicosm Ltd, UK, 1998.

[NATO96] The North Atlantic Treaty Organisation (NATO). The NATO CALS Handbook, 2nd Draft. Technical report, The North Atlantic Treaty Organisation, March 1996. Available on-line at http://cals.debbs.ndhq.dnd.ca/english/natobook/toc.html (Last verified 8th March 2000).

[Neves97] Fernando Das Neves. The Aleph: A Tool to Spatially Represent User Knowledge about the WWW Docuverse. *Proceedings of the 8th ACM Conference on Hypertext (Hypertext97), Southampton, UK*, pages 197–207, April 1997.

[Nurnberg97] Peter J Nurnberg, John J Leggett and Erich R Schneider. As We Should Have Thought. *Proceedings of the 8th ACM Conference on Hypertext (Hypertext97), Southampton, UK*, pages 96–101, April 1997.

[Pikrakis97] A Pikrakis, T Bitsikas, S Sfakianakis, M Hatzopoulos, David C DeRoure, Wendy Hall, Sigi Reich, Gary J Hill and M Stairmand. MEMOIR - Software Agents for Finding

Similar Users by Trails. *Proceedings of the 3rd International Conference and Exhibition on The Practical Application of Intelligent Agents and Multi-Agents (PAAM98), London, UK*, pages 453–466, March 1998.

[Pinfield98] Stephen Pinfield, Jonathan Eaton, Catherine Edwards, Rosemary Russell, Astrid Wissenburg and Peter Wynne. Realizing the Hybrid Library. *D-Lib Magazine*, October 1998. Available on-line at http://www.dlib.org/ (Last verified 8th March 2000).

[Powell98] Andy Powel. Metadata for the Web: RDF and the Dublin Core. *Proceedings of the 20th Birthday UK Online User Group Conference - New Networks, Old Information (UKOLUG98), Manchester, UK*, July 1998.

[Presley96] Adrien Presley and K J Rogers. Process Modelling to Support Integration of Business Practices and Processes in Virtual Enterprises. *Proceedings of the International Engineering Management Conference (IEMC96), Vancouver, Canada*, August 1996.

[Presley97] Adrien Presley. A Multi-View Enterprise Modeling Scheme. *Proceedings of the 6th Industrial Engineering Research Conference (IERC97), Miami Beach, Florida*, pages 610–615, May 1997.

[Quintas97] Paul Quintas, Paul Lefrere and Geoff Jones. Knowledge Management: A Strategic Agenda. *Long Range Planning*, 30(3):385–391, 1997.

[Rada90] Roy Rada. Hypertext Writing and Document Reuse: The Role of a Semantic Net. *Electronic Publishing*, 3(3):125–140, August 1990.

[Raymond96] Darrell Raymond, Frank Tompa and Derick Wood. From Data Representation to Data Model: Meta-Semantic Issues in the Evolution of SGML. *Computer Standards and Interfaces*, 18(1):25–36, January 1996.

[Reagan95] P Reagan, C Holtan and K Lorenz. Components and Uses with Three Case Studies. Available on-line at http://www.documentmanagement.org.uk/articles/EDMS_Comp_Uses.htm (Last verified 8th March 2000), 1995.

[Reinhardt94] Andy Reinhardt. Managing the New Document. *Byte Magazine*, 19(8):91–104, August 1994.

[Ressler95] Sandy Ressler. Network Based Documents. *OII Spectrum Management Handbook*. January 1995. Available on-line at http://www.itl.nist.gov/iaui/ovrt/people/sressler/netDocs/netDocs.fm.html (Last verified 8th March 2000).

[Rezgui96] Yacine Rezgui and Philippe Debras. An Integrated Approach for a Model Based Document Production and Management. *Electronic Journal of Information Technology in Construction*, 1:1–21, 1996.

[Richardson96] John Richardson. Application Analysis and Cost Justification. Presentation at Integrated Computer Aided Technology Exhibition (ICAT96), Birmingham, UK, Project Blue Lion, 1996.

[Riloff94] Ellen Riloff and Wendy Lehnert. Information Extraction as a Basis for High -Precision Text Classification. *ACM Transactions on Information Systems*, 12(3):296–333, July 1994.

[Rowett98] Steven Rowett. *Effective Communications and Information Technology Support for Professional Higher Education and Development.* PhD thesis, School of Computer Studies, University of Leeds, UK, 1998.

[Rusbridge98] Chris Rusbridge. Towards the Hybrid Library. *D-Lib Magazine*, July/August 1998. Available on-line at http://www.dlib.org/ (Last verified 8th March 2000).

[Salton95] Gerard Salton. Selective Text Utilisation and Text Traversal. *International Journal of Human-Computer Studies*, 43:483–497, 1995.

[Schmid98] Hans Albrecht Schmid, Matthias Riebisch, Torsten Heverhagen and Harald Liessmann. A Business Object Framework Architecture. *Presented at the ACM Conference on Object-Oriented Programming, Systems, Languages, and Applications (OOPSLA98) Business Object Workshop, Vancouver, Canada*, October 1998.

[Schuldt95] Ronald Schuldt. What is the Role of Universal Data Element Framework in the Envisioned Integrated Data Environment? *Presented at CALS Expo International 95, Long Beach, California*, October 1995.

[Schwartz95] Gary Schwartz. Everything You Ever Wanted to Know About Client/Servers. Information Technology Article, April 1995. Available on-line at http://magi.com/ mmelick/it95aprl.htm (Last verified 8th March 2000).

[Schwartz98] David G Schwartz. Towards the Use of User-Centric Meta-Knowledge in Applying Organizational Memory to Email Communications. *Interdisciplinary Workshop on Building, Maintaining, and Using Organizational Memories (OM98), Brigthon, UK*, August 1998.

[Schweighofer95] Erich Schweighofer, Werner Winiwarter and Dieter Merkl. Information Filtering: The Computation of Similarities in Large Corpora of Legal Texts. *Proceedings of the 5th International Conference on Artificial Intelligence and Law (ICAIL95), Maryland, USA*, pages 119–126, 1995.

[Sloan98] Bernie Sloan. Service Perspectives for the Digital Library: Remote Reference Services. *Library Trends*, 47(1), Summer 1998.

[Sprague95] Ralph Sprague. Electronic Document Management: Challenges and Opportunities for Information Systems Managers. *Management Information Systems Quarterly*, 19(1):29–49, March 1995.

[Stader96] Jussi Stader. Results of the Enterprise Project. *Proceedings of Expert Systems 96, the 16th Annual Conference of the British Computer Society Specialist Group on Expert Systems, Cambridge, UK*, December 1996.

[Sumner98] Tamara Sumner, John Domingue and Zdenek Zdrahal et al. Enriching Representations of Work to Support Organisational Learning. *Interdisciplinary Workshop on Building, Maintaining, and Using Organizational Memories (OM98), Brigthon, UK*, August 1998.

[Swaby98] Mike Swaby. *A Model-Based Approach to Construction of Integrated Internet CSCW Systems*. PhD thesis, School of Computer Studies, University of Leeds, UK, 1998.

[Trevor97] Jonathan Trevor, Thomas Koch and Gerd Woetzel. MetaWeb: Bringing Synchronous Groupware to the World Wide Web. *Proceedings of the 5th European Conference on Computer Supported Cooperative Work (ECSCW97), Lancaster, UK*, September 1997.

[vanHeijst96] Gertjan van Heijst, Rob van der Spek and Eelco Kruizinga. Organizing Corporate Memories. *Proceedings of the 10th Banff Workshop on Knowledge Acquisition for Knowledge-Based Systems, Banff, Canada*, November 1996.

[Vercoustre97] Anne-Marie Vercoustre and Francois Paradis. A Descriptive Language for Information Object Reuse through Virtual Documents. *Proceedings of the International Conference on Object Oriented Information Systems (OOIS97), Brisbane, Australia*, pages 299–311, 1997.

[W3C98] The World Wide Web Corsortium. HTML 4.0 Specification. W3C Recommendation, The World Wide Web Corsortium (W3C), April 1998. Available on-line at http://www.w3.org/TR/html4/ (Last verified 8th March 2000).

[Wang95] Yih-Chen Wang, Martha Evens and James Vanderdorpe. Relational Thesauri in Information Retrieval. *Journal of the American Society for Information Science*, 36(1):15–27, 1995.

[Watson98] Hugh J Watson and Barbara J Harley. Managerial Considerations. *Communications of the ACM*, 41(9):32–37, September 1998.

[Weibel97] S Weibel. The Dublin Core: A Simple Content Description Model for Electronic Resources. *Bulletin of the American Society for Information Science*, 24(1), October/November 1997.

[Whalley93] Peter Whalley. An Alternative Rhetoric for Hypertext. C McKnight, A Dillon and J Richardson, editors, *Hypertext - A Psychological Perspective*, pages 7–17. Ellis Horwood Series in Interactive Information Systems, 1993.

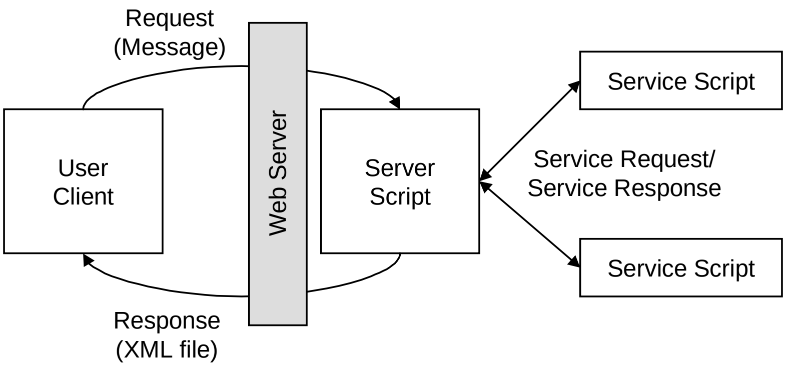
[Whitehead98] E James Whitehead Jr and Meredith Wiggins. WEBDAV: IETF Standard for Collaborative Authoring on the Web. *IEEE Internet Computing*, 2(5), September/October 1998.

[Whitman97] L Whitman and Brian Huff. A Living Enterprise Model. *Proceedings of the 6th Industrial Engineering Research Conference, Miami Beach, Florida*, pages 598–603, 1997.

[Zarhan94] H Zarhan and R Zaidah. Learning Units: A Framework for Implementation Strategies. *CAPABILITY: The Higher Education for Capability (HEC) Journal*, 1(1), 1994.

Appendix A: MRA-HE Server Functions

The following pages contain a full list of the MRA-HE server functions that are called by the server script upon receipt of a request message from the user client. Each request message contains two parts: a request method and a request context. The server script decodes the request message and calls the relevant service script to complete the request. The service script returns a document of the specified type to the server script, which returns a contextualised response to the user client. The request-response process is duplicated below for convenience.



Request and Response Process

Request Method	Returned Document Type
Request Context	Service Request

Navigation Functions	
ShowClassification	<classification>
classification=X:Y	Return the structure of resource room X with folder Y highlighted
person=X	Return the structure of the resource room belonging to person X
ShowClassificationList	<classificationlist>
<none>	Return the list of all
resource=X	Return the resource room folders under which resource X is classified
classification=X:Y	Return the list of resource room folders to which folder Y within resource room folder X is bridged
organisation=X:Y	Return the list of resource rooms belonging to organisational unit Y within organisation X

Request Method	Returned Document Type
Request Context	Service Request

Navigation Functions contd.	
ShowResourceList	<resourcelist>
classification=X:Y	Return the list of resources classified under folder Y within resource room X
resource=X	Return the list of resources hyperlinked to resource X
person=X, extra=history	Return the list of resources that have been viewed by person X
person=X, extra=owns	Return the list of resources owned by person X
organisation=X:Y	Return the list of resources owned by organisational unit Y within organisation X (inferred by summing the documents owned by people who are members of organisational unit Y within organisation X)
ShowResource	<resource>
resource=X	Return the content and structure of resource X
ShowInterestList	<interestlist>
person=X	Return the list of interests that person X has expressed
ShowOrganisation	
organisation=X	Return the structure of organisation X
ShowPersonList	<personlist>
organisation=X:Y	Return the list of people who are members of organisational unit Y within organisation X
classification=X:Y	Return the list of people who have expressed an interest in folder Y within resource room X
resource=X, extra=history	Return the list of people who have viewed resource X
resource=X, extra=interest	Return the list of people who have expressed an interest in resource X
ShowPerson	<person>
resource=X	Return information about the person who owns resource X
classification=X	Return information about the person who owns resource room X
person=X	Return information about person X
ShowOrganisationList	<organisationlist>
person=X	Return the list of organisational units of which person X is a member

Request Method	Returned Document Type
Request Context	Service Request

Navigation Functions contd.	
ShowOrganisation	<organisation>
organisation=X:Y	Return the structure of organisation X with organisational unit Y highlighted
classification=X	Return the strucutre of the organisation that owns classification X
ShowAnnotations	<annotationlist>
resource=X	Return the list of annotations for resource X that were created by people on the user's trust list
Search	<searchlist>
searchterm=X, searchresources=yes/no, searchmetadata=yes/no, searchclass=yes/no, searchorg=yes/no	Return pointers to the resources/metadata/resource room folders/organisational units that contain the search term X

User Navigation Functions	
ShowDropdownList	<dropdownlist>
person=X	Return the list of resources and classifications that person X has expressed an interest in, for display in the top pane
ClearDesk	<resourcelist>
<none>	Remove all secondary pointers to resources from the user's desk space and return the list of resources remaining

Reuse Functions	
SelectResources	<selectresources>
resourcelist=X	Return the list of all the resources in string X for sub-resource selection
OrderResources	<orderresources>
resourcelist=X	Return the list of all the sub-resources in string X for sub-resource ordering
ApplyPresentation	<applypresentation>
resourcelist=X	Return the list of all the ordered sub-resources in string X and all presentation styles for presentation selection and meta-data addition
GenerateResource	<generateresource>
resourcelist=X, presentation=Y, metadata=Z	Create the new resource made up of the ordered sub-resources in string X with presentation Y and new meta-data Z, and return confirmation or otherwise

Request Method	Returned Document Type
Request Context	Service Request

Action Functions	
AddResource	<addresource>
classificaton=X:Y	Return a request for new resource information to be stored in folder Y within resource room X
classification=X:Y, file=Z, metadata=W	Upload file X, add metadata W, store within folder Y of resource room X, and return confirmation
MoveResources	<html>
resources=Z, classificationfrom=X:Y, classificationto=A:B	Remove all resources in string Z from folder Y within resource room X, add all resources in string Z to folder B within resource room A, and return confirmation
CopyResources	<html>
resources=Z, classificationto=X:Y	Add all resources in string Z to folder Y within resource room X, and return confirmation
DeleteResource	<html>
resource=X	Mark resource X as ’deleted’ and return confirmation
DeletePointer	<html>
resource=Z, classification=X:Y	Remove pointer from folder Y within resource room X to resource Z, and return confirmation
AddFolder	<html>
classification=X:Y	Return a request for new folder information to be added to folder Y within resource room X
classification=X:Y, title=Z, description=W	Add a folder with title Z and description W to folder Y within resource room X, and return confirmation
DeleteFolder	<html>
classification=X:Y	Remove folder Y within resource room X, and return confirmation
CreateBridge	<html>
classificationfrom=X:Y, classificationto=A:B	Add a bridge from folder Y within resource room X to folder B within resource room A, and return confirmation
CreateLink	<html>
resourcefrom=X, resourceto:Y	Add a hyperlink from resource X to resource Y, and return confirmation

Request Method	Returned Document Type
Request Context	Service Request

Action Functions contd.	
CreateAnnotation	<html>
resource=X	Return a request for new annotation information to be attached to resource X
resource=X, annotation=Y	Add annotation Y to resource X, and return confirmation
AddInterest	<html>
resource=X	Add an interest for the current user to resource X
classification=X:Y	Add an interest for the current user to folder Y within resource room X, and return confirmation.
DeleteInterests	<html>
interestlist=X	Remove all the interests in string X for the current user, and return confirmation

Appendix B: An XML/XSL Example

Consider a small address book that stores the name, address and email addresses of two people. An XML representation of that information follows[1].

```
 D1: <addressbook>
 D2:  <entry>
 D3:   <name>Dave Smith</name>
 D4:   <address>9 Thorpe Street, Leeds</address>
 D5:   <email>dave@dave.com</email>
 D6:  </entry>
 D7:  <entry>
 D8:   <name>Paul Jones</name>
 D9:   <address>12 Cardigan Rise, Leeds</address>
D10:   <email>paul@email.com</email>
D11:  </entry>
D12: </addressbook>
```

The XML syntax is similar to HTML in that it contains textual content between mark-up tags, but the tags used are not limited to those provided by HTML – an XML tag may have any name.

In order to view this XML instance in a web browser a corresponding XSL instance describes the HTML formatting that should be applied to such an XML instance. XSL provides a language that serve two purposes: transforming XML documents from one form to another, and specifying formatting semantics to those transformed XML documents. The following XSL sample demonstrates only the first of these - the transformation of an XML file into an HTML file for presentation.

```
 S1: <xsl:stylesheet>
 S2: <xsl:template match="/">
 S3:  <html>
 S4:  <head><title>Address Book</title></head>
 S5:  <body>
 S6:   <xsl:apply-templates select="addressbook"/>
 S7:  </body>
 S8:  </html>
 S9: </xsl:template>
S10: <xsl:template match="addressbook">
S11: <h2>This is your address book</h2>
S12:   <ul>
```

[1] Line numbers do not form part of the file - they are added for clarity

```
S13:   <xsl:for-each select="entry">
S14:   <li>
S15:    Name : <xsl:value-of select="name"/><br/>
S16:    Address: <xsl:value-of  select="address"/><br/>
S17:    Email : <xsl:value-of select="email"/>
S18:   </li>
S19:   </xsl:for-each>
S20:   </ul>
S21: </xsl:template>
S22: </xsl:stylesheet>
```

The XML stylesheet contains a set of tree construction rules. The tree construction rules have two parts: a pattern that is matched against elements in the source tree and a template that constructs a portion of the result tree. In the above example, line S2 shows the base template that is matched by default. Upon activation the code in lines S3-S8 are output. Additionally, line S6 activates the second template (line S10) that attempts to match any <addressbook> tags within the XML document. Line D1 of the XML file matches that template, and further HTML is output (lines S11-S20). An additional rule is matched at line S13 that loops through lines S14-S18 each time it is activated (twice in this case at lines D2 and D7). Lines S15-S17 specify, for each entry in XML file, a very brief description of the data and the data itself. For completeness the resulting HTML is shown below - this is processed by an HTML-aware web browser for display to the user.

```
<html>
 <head><title>Address Book</title></head>
 <body>
  <h2>This is your address book</h2>
  <ul>
   <li>
    Name : Dave Smith<br/>
    Address : 9 Thorpe Street, Leeds
    Email : dave@dave.com
   </li>
   <li>
    Name : Paul Jones
    Address: 12 Cardigan Rise, Leeds
    Email : paul@email.com
   </li>
  </ul>
 </body>
</html>
```

Appendix C: Feedback Session Slides

What is the MRA?

- The Model-driven Reuse Architecture
- Extension to current Reading Rooms
- Work purpose and 'organisational memory'
- Documents as manipulable objects
- Improved document discovery
- Focus on reuse
- Supports change across the enterprise

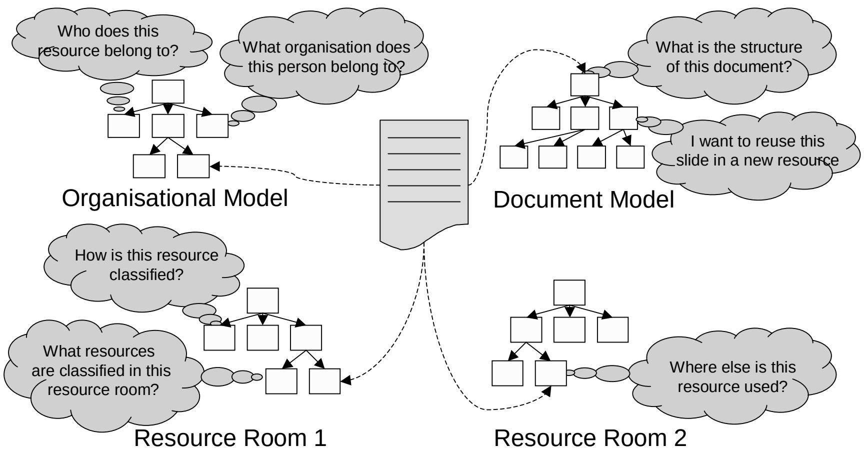
Document Discovery

- Information model encompasses
  - documents and their classifications
  - hypertext mechanisms
    - history trails, annotations, links between documents (hyperlinks), links between folders ('bridges')
  - document models for document manipulation
    - one for each type of document in the store
  - people (resource owners) and organisations
  - registration and investigation of personal interests
- Full information model navigation and search

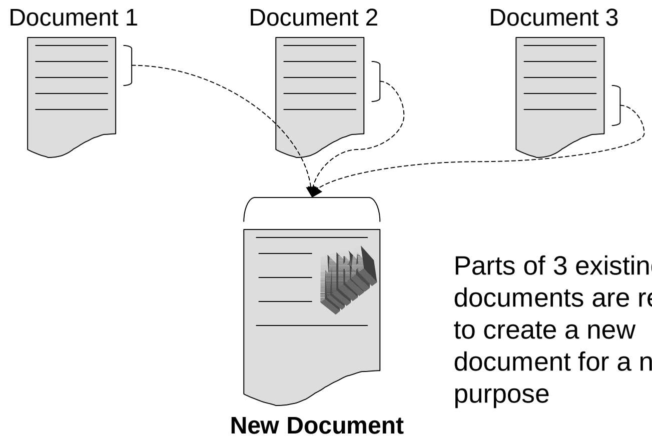
Hypertext Approaches

- typed links between documents and between resource room folders
  - answers the question ‘who created this link’?
  - separation of links, permitting selective application
- typed view history
  - explicitly managed document trail mechanism
  - answers the question ‘who has seen this, and when?’
- typed document annotation
  - permits annotation of any document
  - answers the question ‘who made this annotation’

Model-driven Document Classification

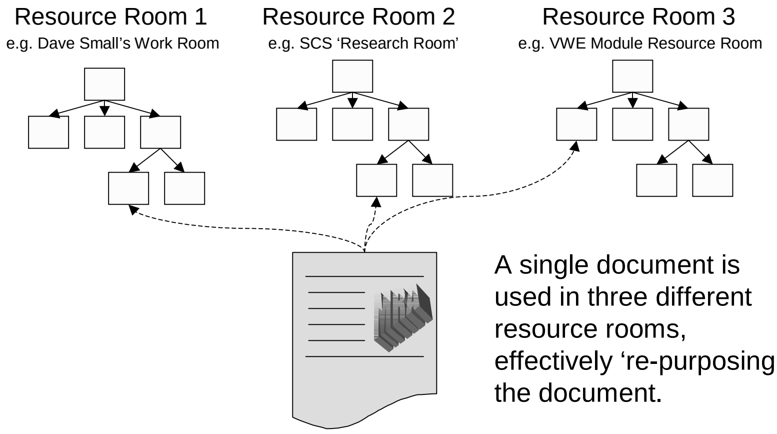


Reuse 1 - Resource Level Reuse



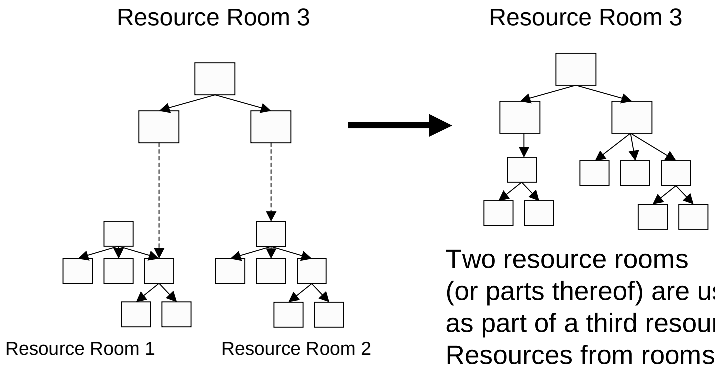
Parts of 3 existing documents are reused to create a new document for a new purpose

Reuse 2 - Folder Level Reuse



A single document is used in three different resource rooms, effectively 're-purposing' the document.

Reuse 3 - Resource Room Level Reuse



Two resource rooms
(or parts thereof) are used
as part of a third resource room.
Resources from rooms
1 and 2 are immediately
accessible to resource room 3

The Reuse Process

- 1. Collection
  - of existing document resources
- 2. Combination
  - adding an ordering to collected resources
- 3. Customisation
  - application of a new presentation style
- 4. Classification
  - into the new purpose

MRA Implementation - MRA-HE

- Focused on supporting learning
  - focus on an overall ‘enterprise’ to be served
  - supports person-focused course provision
  - currently understands ‘lecture slide’ resources
- Provides views on document store
  - power to tutor, restriction to students
- ‘Trust’ mechanism for annotations and links
  - limits external influence, permits local ‘collaboration’
- Functionality for tutors, ease of use for students
  - adaptable ‘low cost’ interface via XML/XSL

Technical Advances

- Increased client-side functionality
  - using new-generation web browsers
- Fully implemented in XML
  - overcomes the document/database divide
  - more natural support for the ‘interconnectedness of everything’
- Explicit support for resource-level reuse
  - using document models and applying new structure and presentation
- Resources viewable/editable in browser
  - standards-based document mark-up
- Links between resources and their owners
- Two-way typed links, history trails, annotations
- Full text search through resources and change management

'Next Generation'

- Extensible implementation model
  - permits the addition of new services, such as automated clustering of documents, relevance user-feedback
  - more document types as required
- Scalability experiments
  - potential to use database as back-end rather than XML
- Document permissions and versioning
  - facilitated through proxy-server architecture
- Lower-level hyperlinking (when XLL is ready)
- Improve interface to 'drag and drop'
- Application of new interfaces (e.g. DiMe, VRML)
  - facilitated through the wide use of semantically rich XML

Extensions to the Current VSP RRs

- Reuse mechanisms
  - new resource-level reuse, improved folder-level and room-level reuse
- Business models connect more tightly to the enterprise
- Increase client-side functionality
- Typed hyperlinks ('who'?)
- Typed bridges between resource rooms ('who?')
- History trails ('who?', 'when?' and 'how often?')
- Link resources to their owners
- Resource annotation ('who?')

Appendix D: Feedback Session Tasks

Model-driven Reuse Architecture Evaluation

An MRA implementation for supporting learning in higher education aims to support the daily fulfilment of learning requirements and the cross-enterprise retention, management and reuse of learning resources. Currently learning requirements are course-based - a set of students undertakes a single pre-determined course. The mechanisms provided by the MRA should, however, facilitate more student-focussed course provision, whereby courses are constructed from a customised set of resources

The success of the MRA implementation is measured in terms of how it assists the following objectives within the domain of higher-level education.

1. Improved ability to store and locate relevant documents across the enterprise for specific work purposes

2. Ability to reuse located documents in a new context for new learning purposes

3. Ability to focus document resources within a localised context for a learning purpose

The aim of the MRA is to extend the document management techniques currently available within VSP resource rooms. If such a system is to be useful in its aims of reuse, it must both provide improved methods for the management of document resources across the enterprise (i.e. the University) and encourage use of the system for day-to-day work. The following tasks are somewhat artificial due to the lack of real document information within the MRA, but it provides examples of a range of MRA functions. For the purposes of the current implementation each document resource within the document store is a set of sub-resources, each of which may be regarded as a lecture slide. Although the range of MRA functionality can be more clearly shown using only one document type, one of the overall architecture's aims is to store and operate over a range of different document types.

Objective 1: Improved ability to store and locate relevant documents across the enterprise for specific work purposes

Objective 1 represents the day-to-day work scenario of a general user of an MRA resource room; Two example ‘general users’ are

- an academic maintaining a resource room for storage and retrieval of her resources
- a course tutor producing materials for a course she is running

The following introductory tasks show how an MRA user can store and retrieve documents for her own purposes outside the strict scope of producing targeted learning resources for a course.

Task 1: Store a document

You have created a document about reuse in an XML-compliant document editor and stored it temporarily on your machine's hard drive. Store it in the MRA in your own resource room under the classification 'Document Reuse'.

Task 2: Find a document and a contact person

You want to find a document in the document store about web servers. Perform a search across all resource room folders and document meta-data (description, keywords, etc.). Ideally you'd like to find something about web server administration, and someone who knows about web server administration. Locate both a document and a person to contact.

Task 3: Find out who has read a document that you have just read

You have just read a document on Neural Networks in the School of Computer Studies reading room (SCS Resource Room : Pattern Recognition) and want to find out who else has read it or expressed an interest in it. Locate the document and find out who has either viewed the document or expressed an interest in it. This may lead to further work or collaboration.

Tasks 1, 2 and 3 show that the MRA can be used as a basic distributed store-and-retrieve document management system, with additional links to document owners, tracking of other users through the document space and investigation of interests expressed.

Objective 2: Ability to reuse located documents in a new context for new learning purposes

You are responsible for creating a new learning module and a mini-course partly based on that module. Task 4 shows how the MRA may be used to find documents that you may be able to use for the module you are preparing.

Task 4: Preparing a resource room for a 'Knowledge Management' module

An empty resource room has been created for the new on-line Knowledge Management (KM) course module that you are responsible for. You wish to populate the resource room with suitable documents. You have created some documents specifically for the module, and others you may be able to use from elsewhere. A set of sub-tasks follows:

Step 1. Create the following resource room structure

- Document Approaches
  - Hypertext
  - Document Reuse
- Web Information Systems
  - HTML and XML
  - Commercial Examples

Step 2. Copy the document about reuse that you created in task 1 into the 'Document Reuse' folder. This involves first locating the resource, then the new folder and copying the resource into the folder.

Step 3. Find a document about the Excalibur document management product and copy it to 'Web Information Systems : Commercial Examples'

Step 4. You need to find a document to describe HTML and the advantages of XML over HTML. A search yields no directly applicable document, so you decide to combine two documents for the purpose. Find two documents, one that outlines the HTML syntax and one that highlights the advantages of XML over HTML. Make sure you bookmark the HTML document for later use, by registering an interest in the document. Copy the two documents to your desk space and create a new document from them (i.e. reuse them) with a 'KM Module' look and feel. You should not include all the detail from each of the resources - only a couple of slides from each. Copy the resulting document to the 'Web Information Systems : HTML and XML' resource room folder.

Step 5. Create a link from the new resource to the original document about HTML.

By the end of task 4 you should have a new resource room containing three documents that are representative of the content of the KM module. You have reused parts of two documents to create a new document, and reused one of your existing documents (i.e. reclassified it for more than its original purpose)

Objective 3: Ability to focus document resources within a localised context for a learning purpose

The next task shows how a number of modules can be combined to provide a mini-course. Someone else has created a resource room for the IS14 module, and you want to provide a resource room for use by students registered on the 'Business Information Systems' mini-course that includes both the KM module you have created and the IS14 module resources.

Task 5: Create a resource room for the 'Business Information Systems' (BIS) course
You already have an empty resource room, and want to include the two module resource rooms within a single resource room and add an overview document. Create links (or bridges) to the KM and IS14 resource rooms from the 'Modules' resource room folder. In order to create a bridge (link) between two folders, you first need to 'Remember' one of the folders and 'Create a bridge' from the other folder. Finally, add a note to the 'Introduction' folder explaining what BIS stands for.

Task 6 highlights the student perspective. It provides a limited view on the collected documents, including only hyperlinks and annotations created by the tutor or other students on that course.

Task 6: Use the 'Business Information Systems' (BIS) course resource room
As a student, you have access to all the resources within the BIS resource room. Navigate these resources, create an annotation to one of the documents and view the annotations that have been added by other students.

Appendix E: MRA-HE Evaluation Forms

User Type:
University Lecturer

MRA-HE Evaluation Form

Is document reuse currently of interest to you/your customers? Please give reasons why or why not.

Yes, there are many opportunities for re-using teaching materials which are currently achieved mainly through cut and paste. A good example arises from the need to teach the same material at different levels (eg undergraduate or taught MSc), the documentation supporting such teaching would differ, but might overlap substantially.

The MRA-HE demonstrator aimed to show an extension of the existing VSP Resource Room implementation to incorporate:

- increased hypertext and document discovery mechanisms
- reuse mechanisms at the resource, folder and resource room levels
- an interface that permits or restricts access to functionality
- 'trust' mechanisms to restrict the affect of user-applied hyperlinks and annotations

The overall aims of the system to discover documents that are suitable for reuse through a number of discovery techniques (e.g. 'who owns this document?', 'who has seen this document?', 'who has expressed an interest in this document?') and to explicitly support reuse of resources and resource rooms. The support for manipulable document models allows the construction of a new document from parts of existing resources.

How would you rate MRA-HE on the following features:

1) Excellent 2) Fairly good 3) Average 4) Fairly poor 5) Very poor

a) Ease of navigation	Rating:1
b) User interface	Rating:2
c) Presentation of information	Rating:1
d) Document reuse mechanisms	Rating:2
e) Ability to support distributed and diverse teaching needs	Rating:?

Comments on answer a) The navigation 'tree' helped me to keep track of where I was in the system There were times when a 'back' button would have helped.

Comments on answer b) Generally this seemed fine; it relies to some extent on knowledge of similar interfaces (for example the ability to re-order lists by clicking on a column heading is not something every user would 'guess'). There seemed to be fewer mouse-clicks required than for the original VSP interface. I was not overly keen on the use of 'marked' documents, folders etc, because it was necessary to read quite a lot of text to be sure that the task (of copying, for example) was being done correctly. The 'copy/paste' metaphor might have been used instead perhaps (there may be a good reason why not)

Comments on answer c) I was particularly pleased to see the different document types being displayed in the browser itself, rather than having to start up another application; this gave a uniformity to the presentation. The 'down' side is that editing in place is not available.

Comments on answer d) I can envisage using the re-use facilities at all three levels. They all worked as I would have expected. The re-use of slides is, of course, easier to manage than re-use of elements of documents that contain many paragraphs, diagrams, tables, images etc.; I am not sure that the interface would be so successful in general.

Comments on answer e) I would have to use the system in anger to be sure on this. The potential seems to be there, but a lot would depend on how easy it would be to work with the underlying model of (e.g.) the trust mechanism. My guess is that it could be very useful, but that if piloted a number of changes might be needed.

Please rate the following aspects of MRA-HE in terms of how useful you think they are to your/your customers' needs:

1) Very useful 2) Fairly useful 3) Average 4) Not very useful 5) Not at all useful

a) Enterprise-wide access to documents	Rating:1
b) Internet access to resource rooms and document resources	Rating:1
c) Links between documents and their owners and organisations	Rating:3
d) Hypertext facilities (history trails, annotations and link creation)	Rating:3
e) Typed hyperlinks, trails and annotations for selective application	Rating:3
f) Registration of personal interests and notification of updates	Rating:1
g) Restricted view on documents through local 'trust' mechanisms	Rating:3
h) Different interfaces for different types of user (e.g. tutor/student)	Rating:3
i) Discovering new documents	Rating:3
j) Creation of a new document resource from existing documents	Rating:2
k) Multiple classification of documents	Rating:1
l) Reuse of one part of a resource room within another resource room	Rating:3
m) Ability to support change within the enterprise	Rating:4

Comments on any of the above points

Difficult to answer any of these with precision because there are many applications for this type of system and the answers differ between them. For example, the management of School-wide access to documents for office staff (and beyond) is very important. B) is important because we run different hardware platforms and need to give access to students from outside the university. The importance of some features (eg e) and c) depends on the quality of the data.

During the demonstration, you saw the following tasks being performed. Please rate the following aspects of MRA-HE in terms of how easy or difficult you thought they were to use:

1) Very easy to use	2) Fairly easy to use	3) Average	4) Fairly difficult to use	5) Very difficult to use

a) Document addition	Rating:?
b) Construction and navigation of resource room folders	Rating:2
c) Navigation of the information model: history trails, document owners, organisations, interests, annotations, links and bridges	Rating:3
d) Creation of a new resource from existing documents	Rating:2
e) Multiple classification of a single document	Rating:2
f) Reuse of one part of a resource room within another resource room	Rating:3

Comments on answer a) :I don't particularly remember this which suggests it could not have been complicated

Comments on answer b) Again I don't remember the construction of a folder, but navigation was straightforward.

Comments on answer c) : I didn't have time to look at all of these in great detail, but these appeared to intuitive and made use of current interface design techniques. The richness of the information model means that it might take some time to become comfortable with all of these.

Comments on answer d) : This is certainly in line with (for example) creating shortcuts in folders using Microsoft Explorer for example. The inability to 'see' more than one resource room at a time makes 'copying' from one resource room to another a little hard.

Comments on answer e) :

Comments on answer f) : There was one aspect to this that I didn't fully understand the need for due to lack of time (had to do with whether it was done seamlessly or not).

What are your perceived strengths/weaknesses of MRA-HE over current VSP resource rooms?

I didn't notice any weaknesses. The less-cluttered interface and use of the directory tree are definite strengths, as too is the ability to view power point slides (and presumably word documents etc) in the browser itself. The management of documents that are multiply-classified is also much easier. The ability to address a sub-document item is a clear advantage, although I'm not sure if this is available for anything other than slides at the moment.

Generally the navigation is improved, although there are still (I think) times when a user might feel lost.

Do you have any suggestions for improving MRA-HE?

A suggestion which may or may not work would be to have a graphical representation (as a stack?) of multiple resource rooms (similar to the 'photoshop' stack of layers perhaps). This would be inorder to return to where you last were in a resource room simply by clicking on the appropriate room in the stack. Another way of doing it might be to use tabs.

The information model is very rich; perhaps it should be 'tailorable' (for example ability to switch off features such as who has seen a document, where these are not needed.

Are there any additional features or services you would like to see from MRA-HE?

I would like to see automatic assignment of documents to classes based on document properties; I don't think the system allows this at the moment.

Do you have any other comments on what you have seen of MRA-HE?

I would like to have spent more time with it; I think like most things of this nature it would need an investment of time to re-organise the way I and my colleagues work before the full benefits could be realised, but it is certainly an improvement over the previous VSP implementation and I believe it could open up new ways of doing things as well as making current tasks more efficient.

User Type: University Lecturer

MRA-HE Evaluation Form

Is document reuse currently of interest to you/your customers? Please give reasons why or why not.

In some way, I am already reusing some of my own documents & documents which are provided by others on-line.

The MRA-HE demonstrator aimed to show an extension of the existing VSP Resource Room implementation to incorporate:

- increased hypertext and document discovery mechanisms
- reuse mechanisms at the resource, folder and resource room levels
- an interface that permits or restricts access to functionality, depending on the user type
- 'trust' mechanisms to restrict the effect of user-applied hyperlinks and annotations

The overall aims of the system to discover documents that are suitable for reuse through a number of discovery techniques (e.g. 'who owns this document?', 'who has seen this document?', 'who has expressed an interest in this document?') and to explicitly support reuse of resources and resource rooms. The support for manipulable document models allows the construction of a new document from parts of existing resources.

How would you rate MRA-HE on the following features:

1) Excellent 2) Fairly good 3) Average 4) Fairly poor 5) Very poor

a) Ease of navigation	Rating:3
b) User interface	Rating:4
c) Presentation of information	Rating:3
d) Document reuse mechanisms	Rating:4
e) Ability to support distributed and diverse teaching needs	Rating:3

Comments on answer a) still rather easy to get a bit lost

Comments on answer b) needs getting use to the environment & where the boxes are …..
but rather easy to learn.

Comments on answer c) find it hard to rate …. trying hard to remember the kind of information presented.
In some way, every screen is full of information !

Comments on answer d) a step further from cut + paste ; but still a lot of issues to be resolved (ownership, version control, etc.)

Comments on answer e) Have not really try out in `real' so cannot give a fair assessment.

Please rate the following aspects of MRA-HE in terms of how useful you think they are to your/your customers' needs:

1) Very useful 2) Fairly useful 3) Average 4) Not very useful 5) Not at all useful

a) Enterprise-wide access to documents	Rating:2
b) Internet access to resource rooms and document resources	Rating:2
c) Links between documents and their owners and organisations	Rating:2
d) Hypertext facilities (history trails, annotations and link creation)	Rating:1
e) Typed hyperlinks, trails and annotations for selective application	Rating:1
f) Registration of personal interests and notification of updates	Rating:?
g) Restricted view on documents through local 'trust' mechanisms	Rating:4
h) Different interfaces for different types of user (e.g. tutor/student)	Rating:3
i) Discovering new documents	Rating:2
j) Creation of a new document resource from existing documents	Rating:3
k) Multiple classification of documents	Rating:2
l) Reuse of one part of a resource room within another resource room	Rating:1
m) Ability to support change within the enterprise	Rating:4

Comments on any of the above points

During the demonstration, you saw the following tasks being performed. Please rate the following aspects of MRA-HE in terms of how easy or difficult you thought they were to use:

1) Very easy to use	2) Fairly easy to use	3) Average	4) Fairly difficult to use	5) Very difficult to use

a) Document addition	Rating:2
b) Construction and navigation of resource room folders	Rating:1
c) Navigation of the information model: history trails, document owners, organisations, interests, annotations, links and bridges	Rating:1
d) Creation of a new resource from existing documents	Rating:1
e) Multiple classification of a single document	Rating:2
f) Reuse of one part of a resource room within another resource room	Rating:1

Comments on answer a) as it was the first task, took a while to get use to …. to reposition oneself in a `virtual' document environment.

Comments on answer b)

Comments on answer c)

Comments on answer d)

Comments on answer e)

Comments on answer f)

What are your perceived strengths/weaknesses of MRA-HE over current VSP resource rooms?

Strength :- more options in viewing the documents; more associated info to documents.

Weaknesses :- not as polished (still looks like a toy system)

Do you have any suggestions for improving MRA-HE?

(See the comments I made , which you have noted down)

Are there any additional features or services you would like to see from MRA-HE?

As commented during demo ……

Do you have any other comments on what you have seen of MRA-HE?

MRA-HE Evaluation Form

User Type:
VSP Research Officer

Is document reuse currently of interest to you/your customers? Please give reasons why or why not.

Within the context of my own function (generally client support) I can see various uses for the MRA throughout the VSP including: easier for VSP Tenancy Information Officers to customise and manage VSP user guides and training documentation for use with their own client, easier for me to maintain VSP user guides and training documentation that is located in several different places in the VSP (also sometimes I don't know where the document are located – e.g. a tenant may have downloaded a copy of a user guide to their own resource room not told me .. some people don't like using the links facility, i.e. the ability to link from 1 folder to another), The MRA supports not only the exchange of formalised and codified knowledge (e.g. in the form of a research report) which we do well already, but also promotes informal interaction and the exchange of informal knowledge, e.g. social navigation of the form `who else has looked at this' – the opportunity to explore info is based not on location or content but on implicit recommendation, social factors, etc. This we don't do so well yet

The MRA-HE demonstrator aimed to show an extension of the existing VSP Resource Room implementation to incorporate:

- increased hypertext and document discovery mechanisms
- reuse mechanisms at the resource, folder and resource room levels
- an interface that permits or restricts access to functionality, depending on the user type
- 'trust' mechanisms to restrict the effect of user-applied hyperlinks and annotations

The overall aims of the system to discover documents that are suitable for reuse through a number of discovery techniques (e.g. 'who owns this document?', 'who has seen this document?', 'who has expressed an interest in this document?') and to explicitly support reuse of resources and resource rooms. The support for manipulable document models allows the construction of a new document from parts of existing resources.

How would you rate MRA-HE on the following features:

1) Excellent 2) Fairly good 3) Average 4) Fairly poor 5) Very poor

a) Ease of navigation	Rating:1
b) User interface	Rating:2
c) Presentation of information	Rating:2
d) Document reuse mechanisms	Rating:2
e) Ability to support distributed and diverse teaching needs	Rating:2

Comments on answer a) I like the dropdown menus – this makes navigation very easy as all the options are listed right next to one's mouse pointer

Comments on answer b) The interface looks like it's in a development environment. It's on a 17 inch monitor and the text is very small. Try viewing that on my 14 inch monitor and you'd soon be doing too much scrolling. Ability to resize LHS/RHS frames is great – something sadly lacking and much requested in the VSP!

Comments on answer c) Seems ok.

Comments on answer d) Seems ok but would really need to use it myself for a period of time to properly evaluate it.

Comments on answer e) Same comment as in (d).

Please rate the following aspects of MRA-HE in terms of how useful you think they are to your/your customers' needs:

1) Very useful 2) Fairly useful 3) Average 4) Not very useful 5) Not at all useful

a) Enterprise-wide access to documents	Rating:1
b) Internet access to resource rooms and document resources	Rating:1
c) Links between documents and their owners and organisations	Rating:1
d) Hypertext facilities (history trails, annotations and link creation)	Rating:1
e) Typed hyperlinks, trails and annotations for selective application	Rating:1
f) Registration of personal interests and notification of updates	Rating:1
g) Restricted view on documents through local 'trust' mechanisms	Rating:3
h) Different interfaces for different types of user (e.g. tutor/student)	Rating:2
i) Discovering new documents	Rating:1
j) Creation of a new document resource from existing documents	Rating:1
k) Multiple classification of documents	Rating:1
l) Reuse of one part of a resource room within another resource room	Rating:1
m) Ability to support change within the enterprise	Rating:1

Comments on any of the above points

(e) why selective applications – don't know what this means

(g) not sure I understand what local `trust' mechanisms are

(h) but not too different

(i) but we have a monitoring facility in the VSP – how is your different/better?

(m) anything that can do this is good!

During the demonstration, you saw the following tasks being performed. Please rate the following aspects of MRA-HE in terms of how easy or difficult you thought they were to use:

1) Very easy to use	2) Fairly easy to use	3) Average	4) Fairly difficult to use	5) Very difficult to use

a) Document addition	Rating:2
b) Construction and navigation of resource room folders	Rating:2
c) Navigation of the information model: history trails, document owners, organisations, interests, annotations, links and bridges	Rating:2
d) Creation of a new resource from existing documents	Rating:2
e) Multiple classification of a single document	Rating:2
f) Reuse of one part of a resource room within another resource room	Rating:2

Comments on answer a)

Comments on answer b)

Comments on answer c)

Comments on answer d)

Comments on answer e)

Comments on answer f)

What are your perceived strengths/weaknesses of MRA-HE over current VSP resource rooms?

Mostly as identified as above. I would need to see more of it in detail.

Do you have any suggestions for improving MRA-HE?

My assessment has been based on a short demo. To make suggestions for improvements I would want to be using the system myself for a sufficient length of time.

Are there any additional features or services you would like to see from MRA-HE?

Do you have any other comments on what you have seen of MRA-HE?

MRA-HE Evaluation Form

User Type:
VSP development team member

Is document reuse currently of interest to you/your customers? Please give reasons why or why not.

Our product provides support tools for people who work in a number of areas (teaching, project management, consultancy) both with people outside the VSP environment, and other people within the environment. For this reason, the ability to share and reuse documents in different contexts is particularly valuable. The teaching environment is particularly in need of reuse functions, as materials are often used in other contexts with little or no change.

The MRA-HE demonstrator aimed to show an extension of the existing VSP Resource Room implementation to incorporate:

- increased hypertext and document discovery mechanisms
- reuse mechanisms at the resource, folder and resource room levels
- an interface that permits or restricts access to functionality, depending on the user type
- 'trust' mechanisms to restrict the effect of user-applied hyperlinks and annotations

The overall aims of the system to discover documents that are suitable for reuse through a number of discovery techniques (e.g. 'who owns this document?', 'who has seen this document?', 'who has expressed an interest in this document?') and to explicitly support reuse of resources and resource rooms. The support for manipulable document models allows the construction of a new document from parts of existing resources.

How would you rate MRA-HE on the following features:

1) Excellent 2) Fairly good 3) Average 4) Fairly poor 5) Very poor

a) Ease of navigation	Rating:2/3
b) User interface	Rating:2
c) Presentation of information	Rating:2
d) Document reuse mechanisms	Rating:1
e) Ability to support distributed and diverse teaching needs	Rating:2

Comments on answer a) the interface is quite difficult to use for a novice, mainly due to the confusion between the use of popup menus using left/right buttons. With use, this becomes better. The differing views (structure plus `content' view, as well as the bookmarking facilities make navigating easy.

Comments on answer b) see (a) for some user interface issues, it is quite clean and uncluttered, although it Demands a lot of screen real-estate.

Comments on answer c) the presentation is clear with good use of more advanced browser features to enable more to be achieved on the client side, without the need to make trips back to the server, which can be frustrating for the user.

Comments on answer d) the reuse mechanisms are excellent, especially the ability to `split' a presentation into components and reuse individual slides. This is a very common requirement.

Comments on answer e) With the addition of other tools (for example, supporting interaction between Students and their peers and tutors, it has the potential to be a component in a very powerful system.

Please rate the following aspects of MRA-HE in terms of how useful you think they are to your/your customers' needs:

1) Very useful   2) Fairly useful   3) Average   4) Not very useful   5) Not at all useful

a) Enterprise-wide access to documents	Rating:1
b) Internet access to resource rooms and document resources	Rating:1
c) Links between documents and their owners and organisations	Rating:2
d) Hypertext facilities (history trails, annotations and link creation)	Rating:3
e) Typed hyperlinks, trails and annotations for selective application	Rating:3
f) Registration of personal interests and notification of updates	Rating:1
g) Restricted view on documents through local 'trust' mechanisms	Rating:2
h) Different interfaces for different types of user (e.g. tutor/student)	Rating:1
i) Discovering new documents	Rating:2
j) Creation of a new document resource from existing documents	Rating:1
k) Multiple classification of documents	Rating:2
l) Reuse of one part of a resource room within another resource room	Rating:2
m) Ability to support change within the enterprise	Rating:1

Comments on any of the above points

A lot of these are generally useful requirements for most domains, however the way that support for them has been implemented distinguishes the system from others.

(m) needs a bit more expansion, however I assume this is the reuse facilities specifically.

The potential for the system will become a lot more apparent with the addition of the access control facilities as this helps to enforce the user views and capabilities, which is very important within a teaching environment.

During the demonstration, you saw the following tasks being performed. Please rate the following aspects of MRA-HE in terms of how easy or difficult you thought they were to use:

1) Very easy to use	2) Fairly easy to use	3) Average	4) Fairly difficult to use	5) Very difficult to use

a) Document addition	Rating:1
b) Construction and navigation of resource room folders	Rating:1
c) Navigation of the information model: history trails, document owners, organisations, interests, annotations, links and bridges	Rating:2
d) Creation of a new resource from existing documents	Rating:2
e) Multiple classification of a single document	Rating:2
f) Reuse of one part of a resource room within another resource room	Rating:2

Comments on answer a)

Comments on answer b)

Comments on answer c) : some confusion between bookmarks, links and bridges confused at times. Maybe the links and bridges should be merged to `link' with resource rooms as being just a different type of target?

Comments on answer d) : this is easy – would be nice if you could upload a `standard' document (such as powerpoint doc) and split into fragments automatically.

Comments on answer e)

Comments on answer f) : the two options `link' or `copy' are a bit confusing (exactly the reason why we've shied away from this in the VSP!).

What are your perceived strengths/weaknesses of MRA-HE over current VSP resource rooms?

Lots more use of newer technologies allows more to be done on the client side, and for the interface to be lot less cluttered. VSPs relative simplicity makes it easier to use, at least initially. The reuse fragments and recombination into a single resource is definitely something that has been asked for in the VSP.

Do you have any suggestions for improving MRA-HE?

1. different views of the interface which add/remove functionality, e.g. a `common' view that doesn't have things like bridge options on, and a simple browse view, which hides a lot of the functionality, but lets you access all of the resources easily.
2. A graphical `graph' view of the structure (as opposed to the hierarchy view) is far more important when there is a lot of reuse going on as hierarchies become counter intuitive here.
3. Drag and drop functionality (activeX/java controls?)

Are there any additional features or services you would like to see from MRA-HE?

Do you have any other comments on what you have seen of MRA-HE?

Great – we'll definitely be absorbing some of the features ☺

User Type:
VSP development team member

MRA-HE Evaluation Form

Is document reuse currently of interest to you/your customers? Please give reasons why or why not.

Yes. Primarily for the creation of customised training material from existing training material.

The MRA-HE demonstrator aimed to show an extension of the existing VSP Resource Room implementation to incorporate: • increased hypertext and document discovery mechanisms • reuse mechanisms at the resource, folder and resource room levels • an interface that permits or restricts access to functionality • 'trust' mechanisms to restrict the affect of user-applied hyperlinks and annotations The overall aims of the system to discover documents that are suitable for reuse through a number of discovery techniques (e.g. 'who owns this document?', 'who has seen this document?', 'who has expressed an interest in this document?') and to explicitly support reuse of resources and resource rooms. The support for manipulable document models allows the construction of a new document from parts of existing resources.

How would you rate MRA-HE on the following features:

1) Excellent 2) Fairly good 3) Average 4) Fairly poor 5) Very poor

a) Ease of navigation	Rating:2
b) User interface	Rating:2
c) Presentation of information	Rating:2
d) Document reuse mechanisms	Rating:2
e) Ability to support distributed and diverse teaching needs	Rating:2

Comments on answer a)

I would prefer the default action on clicking links to be to view the folder/resource and not produce a menu. The currently marked folder/document areas could be links.

Comments on answer b)

The interface is nicely laid out, but the top toolbar is perhaps a little "busy".

Comments on answer c)

Nice!

Comments on answer d)

I loved this feature. Support for a number of different document types would be superb, but as a prototype this really shows the potential.

Comments on answer e)

Again, the potential is clear. Not sure how the system would cope with synchronisation issues with multiple parties updating the same resource.

Please rate the following aspects of MRA-HE in terms of how useful you think they are to your/your customers' needs:

1) Very useful 2) Fairly useful 3) Average 4) Not very useful 5) Not at all useful

a) Enterprise-wide access to documents	Rating:1
b) Internet access to resource rooms and document resources	Rating:1
c) Links between documents and their owners and organisations	Rating:1
d) Hypertext facilities (history trails, annotations and link creation)	Rating:1
e) Typed hyperlinks, trails and annotations for selective application	Rating:1
f) Registration of personal interests and notification of updates	Rating:1
g) Restricted view on documents through local 'trust' mechanisms	Rating:1
h) Different interfaces for different types of user (e.g. tutor/student)	Rating:1
i) Discovering new documents	Rating:1
j) Creation of a new document resource from existing documents	Rating:2
k) Multiple classification of documents	Rating:1
l) Reuse of one part of a resource room within another resource room	Rating:1
m) Ability to support change within the enterprise	Rating:2

Comments on any of the above points

During the demonstration, you saw the following tasks being performed. Please rate the following aspects of MRA-HE in terms of how easy or difficult you thought they were to use:

1) Very easy to use	2) Fairly easy to use	3) Average	4) Fairly difficult to use	5) Very difficult to use

a) Document addition	Rating:1
b) Construction and navigation of resource room folders	Rating:2
c) Navigation of the information model: history trails, document owners, organisations, interests, annotations, links and bridges	Rating:2
d) Creation of a new resource from existing documents	Rating:2
e) Multiple classification of a single document	Rating:1
f) Reuse of one part of a resource room within another resource room	Rating:2

Comments on answer a)

Comments on answer b)

Pretty straightforward after training.

Comments on answer c)

Pretty straightforward after training.

Comments on answer d)

Pretty straightforward after training.

Comments on answer e)

Comments on answer f)

Pretty straightforward after training.

What are your perceived strengths/weaknesses of MRA-HE over current VSP resource rooms?

Strengths:

Resource reuse. Potential simple personalisation of the interface. Much cleaner interface. History trails. Link from document to author/owner.

Weaknesses:

Doesn't support a wide range of documents. Performance issues due to file based backend database.

Do you have any suggestions for improving MRA-HE?

Use a database for the backend.

Are there any additional features or services you would like to see from MRA-HE?

Support for standard XML documents as they become available.
Collaboration support.

Do you have any other comments on what you have seen of MRA-HE?