Managing long-term access to Digital Data Objects:
a metadata approach

by

Alison H. Chilvers

A Doctoral Thesis
Submitted in partial fulfilment of the requirements
for the award of
Doctor of Philosophy of Loughborough University

May 2000

Abstract

As society becomes increasingly reliant on information technology for data exchange and long-term data storage the need for a system of data management to document and provide access to the 'societal memory' is becoming imperative. An examination of both the literature and current 'best practice' underlines the absence to date of a proven universal conceptual basis to digital data preservation. The examination of differences in nature and sources of origin, between traditional 'print-based' and digital objects leads to a re-appraisal of current practices of data selection and preservation. The need to embrace past, present and future metadata developments in a rapidly changing environment is considered. Various hypotheses were formulated and supported regarding; the similarities and differences required in selection criteria for different types of Digital Data Objects (DDOs), the ability to define universal threshold standards for a framework of metadata for digital data preservation, and the role of selection criteria in such a framework. The research uses Soft Systems Methodology to investigate the potential of the metadata concept as the key to universal data management. Semi-structured interviews were conducted to explore the attitudes of information professionals in the United Kingdom towards the challenges facing information-dependent organisations attempting to preserve digital data over the long-term. In particular, the nature of DDOs being encountered by stakeholders, the reasons, policies, and procedures for preserving them, together with a range of specific issues such as; the role of metadata, access to, and rights management of DDOs. The societal need for selection to ensure efficient long-term access is considered. Drawing on SSM modelling, this research develops a flexible, long-term management framework for digital data at a level higher than metadata, with selection as an essential component. The framework's conceptual feasibility has been examined from both financial and societal benefit perspectives, together with the recognition of constraints. The super-metadata framework provides a possible systematic approach to managing a wide range of digital data in a variety of formats, created/owned by a spectrum of information-dependent organisations.

Table of Contents

List of Figures

List of Tables

Acknowledgements

First and foremost I would like to thank my supervisor Professor John Feather, Pro-Vice-Chancellor, for being intellectually challenging and for his advice and guidance throughout the course of this research. I would like to thank Professor Cliff McKnight, my Director of Research, for his advice and administrative work while I have been at Loughborough.

I wish to express my sincere thanks to all stakeholders, albeit they remain nameless, who gave freely of their time and valuable information, without whom this research would not have been possible.

My thanks go to all staff members of the Department of Information Science for their guidance and advice, particularly Dr Ann O'Brien for her insight into metadata issues. I thank Professor Ron Summers and Mr Ian Murray for their evaluation and comments on the super-metadata framework. I am grateful to Professor Charles Oppenheim for his advice on copyright and legal deposit issues and Dr Paul Sturges for information on digital preservation. My thanks also go to Mrs Inese Smith for her valuable advice on bibliographic matters.

I owe a great deal to my parents for their continued support and encouragement over the years, and to my husband for his patience and support.

Chapter 1

Introduction

1.1 General background

The increasing reliance which society is placing on digital data is leading to an increasing urgency for strategies to be developed to manage digital materials deemed to be of long-term importance. Ignoring the development of long-term strategies could lead to the loss of unique data and also create expense in terms of time and human resources in recreating data where this is possible. Cave paintings such as those at Lascaux, France, bear witness to the importance of preserving the whole entity. The product of the creator's labours can be seen as a physical expression of images of animals and symbols, but we can only guess at the significance of these to their creator(s). Context has been lost and with it our ability to understand their true significance. In the modern world there is the possibility that the data context, content and structure of objects which have only ever existed in digital form 'born digital' may also be lost. The ability to even read the information is dependent on careful long-term management.

Meadows (1996) highlights the growing amount of scholarly material which is produced in solely digital form, which because of its interactive nature, cannot be adequately transferred to a static medium such as paper.

> This applies, for example, to the large volumes of information found in databases, to information sources that involve interactivity (including hypertext links), and to multi-media publications. (Meadows, 1996, p.216)

New developments are occurring rapidly and technology is becoming more sophisticated, reliable and economical. Increasing numbers of businesses, institutions and even private individuals are creating digital data. This in turn has

implications for the management of information, both in terms of the magnitude of data involved and the increasingly sophisticated means of interpretation and access.

The majority of Digital Data Objects (DDO) (see Section 1.2.3) discussed in this thesis are scholarly because the issues are mainly arising with this category of material and this category is the main concern of the majority of stakeholder organisations that were interviewed. However, DDOs are being produced for a range of non-scholarly reasons, such as for entertainment[1], administrative[2] or legal[3] purposes. This research therefore seeks to develop a framework flexible enough to cater for DDOs irrespective of the reason for their creation. The original purpose for which the DDO was created may not be the only reason for its long-term preservation and future use. As the government of the United Kingdom has noted that "Family historians, for example, use records of army service for purposes which have nothing to do with the state's need to document an individuals time in uniform." (*Government Policy on Archives*, 1999, p. 8). The adaptations in the super-metadata framework would need to take account of the differences in the electronic characteristics between DDOs created for different purposes and be capable of being tailored to an organisation's remit.

---

[1] Digital broadcasting is creating a plethora of DDOs some of which may have long-term value and to manage these objects some repositories have set up initiatives. For example the National Archives and Records Administration has a pilot database, the NARA Archival Information Locator (NAIL), which has 89,000 searchable moving image records some of which are digital copies (URL: http://www.nara.gov/nara/nail.html).

[2] A national census taken at regular intervals is used as an administrative tool by many countries in the world for example the 1999 French population census (URL: http://wcatinsee.jouve.fr/recens/index.htm). A second example which is of both administrative and legal importance is the National Street Gazetter created by the Ordnance Survey working in partnership with local authorities (URL: http://www.nsg.org.uk/whatis/what1.htm). The database is useful to public utilties companies such as water, gas, electricity and telecommunications together with local authorities who need to have detailed plans of where pipes, cables and sewers are located underground.

[3]Her Majesty's Land Registry has developed an initiative to modernise the conveyancing process at a national level (URL: http://www.landreg.gov.uk). The National Land Information Service (NLIS) anticipates the creation of computerised records with each property being assigned a Unique Property Reference Number (UPRN).

This chapter details the major events and developments in the preservation of 'traditional' library materials, against which developments in the field of digital data preservation can be set. The Chapter also explores the concept of intellectual preservation in more depth in order to illustrate some of the issues involved in the area of preservation, defines terms and illustrates the nature of Digital Data Objects (DDOs).

As a result of researching the background to digital data preservation, the issues of integrity, authentication and access to DDOs led to a realisation that there is a need to re-evaluate selection criteria. This led to an overall aim to the research, the formulation of hypotheses, and a series of objectives designed to attain this aim (as detailed in Sections 2.2 to 2.2.3). At this stage (May 1997) it was felt necessary to examine the types of DDOs which information professionals were encountering in their work and the challenges facing them during the process of selecting these objects for preservation. To achieve this, a postal questionnaire was developed and pilot tested (Appendix A). The results of this pilot questionnaire gave some insight into current practices (as demonstrated in Section 2.4.2). However, in the light of these findings the research methods were re-evaluated. It was apparent that conducting interviews with key stakeholders would be better suited to providing detailed information of methods that were currently employed for selection, reasons why organisations undertake certain procedures and an examination of whether these same organisations have thought through the reasons for adopting a particular strategy. The interviews with stakeholders focused on organisations identified through the literature review which had relevant experience in preserving the intellectual content of DDOs. To examine a potentially wider spectrum of approaches to the selection for preservation, the variety of organisations was enlarged to cover commercial companies, publishers, specialist archives, universities, a national library and a government agency.

Prior to undertaking the interviews it was deemed necessary to examine existing criteria for the selection of digital data content to set the findings in context. Some of the existing selection criteria do not focus solely on information content, as described later in this Chapter, and further work in this area led to the formulation

of the hypotheses stated in Section 2.2.3. This Chapter also focuses on advances, to date, in frameworks for managing DDOs and developments of metadata formats and interoperability.

This research is concerned with the intellectual preservation of DDOs which encompasses managing the following three aspects of the DDO (1) the substantive content, (2) the context relative to other DDOs and (3) the organisational structure of the DDO. Although all the research interviews have been conducted in the United Kingdom the super-metadata framework that this research proposes has been developed to be as flexible as possible and to take into account prominent international publications. As such, it is hoped that the framework will be equally applicable on a global basis, taking into consideration adaptations required for administrative differences between stakeholder organisations in different countries and even between stakeholders with a similar remit within the same country. This research is aware that differences in the legal practices between countries also needs to be considered (Hofman, 1998). To aid understanding of the complex issues that the digital environment creates, a brief examination of the preservation of both print-based and digital resources has been undertaken within this chapter.

1.2 Definition of terms

As the preservation of digital data is a relatively new field, some definition of terms is necessary to avoid ambiguities.

1.2.1 Preservation

Preservation of digital data in the context of this research is concerned with the financial, managerial and technical aspects involved in the preservation of DDOs in information-dependent organisations, with the aim of prolonging the object's useful life. The aim of this preservation is linked to an organisation's mission statement.

1.2.2 Long-term

The meaning adopted in the context of this research is the preservation of digital data for as long as is required. The life cycle of a DDO is highly dependent on the 'value' society attaches to it, that is whether a DDO is of permanent importance and if so why, and to whom.

1.2.3 Digital Data Object (DDO)

> Is defined as computer-based, machine-readable resource (such as a Web page or an electronic journal), the information content of which can be stored and accessed independently of the form in which it was originally created. Whether or not an information resource is formally 'published' is irrelevant to this definition. (Chilvers & Feather, 1998, p.365)

1.2.4 Interoperability

> The ability of different vendor devices to transmit data and exchange information, while having the total capability to process and act upon such information independently. This relies heavily on international standards. (Nader, 1998, p. 351)

1.2.5 Metadata

> Data about data, includes information about the context of data and the content of data and the control of or over data. (Pasquinelli, 1997, section vi)

1.2.6 Super-metadata

Super-metadata is defined as data assigned to each DDO and would contain the following information:

1. The name of any metadata scheme relating to the DDO, to enable the metadata to be read.

2. Any information needed to manage the DDO that may not be contained in its metadata, such as the expected life span of a DDO which is to be reviewed or superseded by a later version.

A catalogue of catalogues is an obvious example of super-metadata.

1.3 Print-based preservation

This section serves to demonstrate briefly the fundamental differences and similarities between print-based and digital preservation. The ethos of preserving print-based objects centres around the management of the physical object, as the intellectual content and the medium which supports it cannot be separated. This, as will be discussed later in this chapter, is the major difference between print-based and digital objects. Print-based preservation revolves around conserving the tangible object, including providing a controlled environment for unique items to protect them from the effects of man-made or natural disasters. Retrieval and storage systems are often labour intensive, such as the need to shelve and re-shelve material in a specific order. In addition, the need for large-scale storage has cost implications.

Access and preservation are inextricably linked in digital data preservation (see Section 1.6.1), however, there are differences in emphasis. Print-based access systems often focus on the location and description of the physical object as much as the intellectual content. With digital objects the emphasis is much more biased towards documenting their intellectual content (see Sections 1.6.6 and 1.6.7). Access to print-based resources can be governed by limitations on the number of copies of a published object that are held by repositories. Moreover, heavy usage of unique items can be detrimental to the physical condition of such items. Therefore, to ensure access, organisations preserving unique print-based objects,

are in some cases using digitization as a means of ameliorating this situation, as we shall see in the following section.

1.4 Digitization

Although the technology to support digitization had been in existence in the USA since the early 1980s, it was not widely adopted until the 1990s. This was because of the technical advances which "...led to improved image capture, lower costs, and greater accessibility" (Kenney and Chapman, 1996, p.2). This process of digitization refers to objects which have previously existed in other media and the digitization process can be seen as a transitional preservation method. In a situation where objects have only existed in digital form this process of digitization becomes unnecessary.

Programmes to model the process of converting microfilm to digital imagery have been undertaken, as evidenced by Yale University Library's *Project Open Book*, which is examining the feasibility and costs of large-scale conversion. The goals of this project are to create a digital image library, which aims to improve intellectual and physical access, the enhanced intellectual access being achieved by creating a document structure and page number indexes for each volume and physical access improved by providing distributed access through a campus wide network. Manuals have been produced to aid digitization projects, such as Kenney and Chapman's (1996) *Digital imaging for libraries and archives*.

Developments in the preservation of traditional 'paper-based' materials within the United Kingdom have been influenced greatly by the British Library Research and Development Department's investigation *Preservation policies and conservation in British Libraries*, generally known as the Ratcliffe Report, published in 1984. The report's recommendations centred around raising the profile of preservation, linking it into organisational policy within individual libraries and the development of co-operative activities on a nation-wide scale.

The first comprehensive national survey since the Ratcliffe Report (Ratcliffe and Patterson, 1984) funded by the Leverhulme Trust, was undertaken in 1993 by Feather, Matthews and Eden (1996). This survey looked at the role of preservation as an important aspect of resource management in the current context of technological change and financial constraints facing libraries. The survey's overriding conclusion was the need for a national access policy, but it also concluded that:

> Access, retention and preservation cannot be separated from each other. Moreover none of them is specific to any one sector of libraries, nor to particular categories of materials or media, although the problems are, of course, more acute in some areas than in others. (Feather, Matthews and Eden, 1996, p. 155)

The above events and reports have served to heighten the profile of preservation of traditional print-based library and archive materials. Swartzburg (1995) has reported that it is only in the last two decades that librarians have become fully aware of the seriousness of the physical preservation problem.

1.5 Digital preservation

The increasing use of computers, and the consequent increase in the amount of information held solely in machine-readable form by organisations, generated some academic interest in the preservation of electronic information as early as the 1970s. Bell (1979) displays this interest when he states that, "...there is a growing amount of information held by governments and others in machine-readable form, and increasingly only in that form, which imposes on us as archivists an obligation to have regard to the archival value of that information" (Bell, 1979, p. 85).

With the development of the Internet, growth in electronic documentation has been greatly increased, and new technology has been seen as a panacea for recording this newly created information. Lesk (1997) identifies this trend, but also states in the abstract of *Preserving digital objects: recurrent needs and challenges*, that new technology can create new problems.

> But digital technology seems to come to the rescue, allowing indefinite storage without loss. Now we find that digital information too, has a dark side, and although it can be kept without loss it can not be kept without cost. (Lesk, 1997)

There is a growth in interest in the preservation of digital data as an increasing number of documents are produced solely in this format.

> Because of the immense size of the preservation problem in libraries and the enormous resources required to preserve all items in their original form, distinctions are made between preservation of the item itself (the artefact) and preservation of the intellectual content of the item. Preservation of the artefact is warranted when the item has some special value as a physical object; preservation of only the intellectual content is justified in other cases. (Harvey, 1993a, p. 309)

During the 1990s several organisations have produced reports/workshops which have had significant impact on the long-term preservation of electronic materials. The JISC/British Library Workshop *Long-term preservation of electronic materials* (1995), provided a wide range of presentations covering areas such as collection policies, legal deposit, and strategies for managing electronic archives. The purpose of the report *Preserving digital information* (1996), was "...to investigate the means of ensuring 'continued access indefinitely into the future of records stored in digital electronic form'" (Task Force on Archiving of Digital Information, 1996, p. iii). Also in 1996 the British Library commissioned CIMTECH to undertake an examination of the preservation of digital materials. The report findings pointed to the need for new legal deposit legislation to cater for the deposit of electronic resources. The Follett Report (Joint Funding Councils' Library Review Group, 1993) has also emphasised in its conclusions the need to utilise the benefits of new technology without infringing copyright laws. The report also emphasised resource sharing and the need to increase access to collections.

The preservation of traditional library and archive materials has centred around the physical preservation. This has been inevitable as the *medium* and the *message* are inseparable (Hildreth, 1996), but with Digital Data Objects this separation is

possible and this itself presents new challenges. Rothenberg (1995) makes the point that the impact of information technology on record management is as great as the introduction of printing and perhaps even writing.

As early as 1978, Dollar identified a significant difference in the approach to archiving machine-readable records, as opposed to textual records, in the American context, and suggested that with advances in computer technology these differences would become more pronounced. However, some parallels exist between print-based and digital resources as evidenced by the comments of Hendley (1996) in the CIMTECH report for the British Library, which regarded many issues as common to both print and digital publications. These include, "...understanding the nature of the media, monitoring the storage environment, handling and use, re-formatting, disaster planning and preservation treatment" (Hendley, 1996, p. 5-6).

As with conventional copyright material, digital copyright materials will have to be kept for the use of future generations. There are, however, a number of aspects related to the nature of DDOs that are important for their management and these are described in the following paragraphs.

1.6 Nature of digital data objects

1.6.1 Preservation through access

In the digital environment Hedstrom (1995) identified a closer relationship between preservation and access than is the case with printed documents. "In the past, archivists argued that there was no point preserving records if one could not provide access; in the digital environment, without access there will be no preservation" (Hedstrom, 1995, p. 189). The need to ensure long-term access to DDOs is the essential reason for preserving the information. The achievement of such long-term access requires a system which ensures that the various issues surrounding access are addressed. These issues are detailed in the following points.

- Ensuring readability of these computer based machine-readable resources by human beings. This involves a process of monitoring DDOs to ensure that the data is supported by the medium on which it is stored. If this is not the case, processes such as migration need to be undertaken and checks made to ensure the data has not become corrupted as a result of these processes.

- Balancing the preservation of resources with originator's rights, and user access to them, has been the concern of the information community in the pre-digital environment and this has been catered for in a regulated manner. Those now concerned with preserving digital data are faced with similar challenges of balancing the interests of stakeholders but without the support of a regulated environment as exists for traditional paper-based materials. The stakeholder interests in the digital environment include creators of digital information, who may wish to restrict access to the information content of DDOs, for example, for financial or security reasons, whilst maintaining access to their data for their organisation's use in a data archive environment. The needs of creators of DDOs can cause a conflict of interest with users who may require access to the same resources for a variety of uses.

- Development of a rigorous system of documentation and enforcement of any access restrictions which have been imposed by the creator of the resource. This entails maintaining documentation and developing control mechanisms that function by computer-assisted means.

- The need for the development of a system to retain access to the intellectual content of remotely held DDOs is becoming increasingly important; this in an environment where the information community is increasingly subscribing to electronic journals and acquiring interactive resources in growing numbers. Subscription to electronic journals raises the issue of access to back runs of a journal if it is held remotely on a publisher's server, in the event of the publisher increasing subscription charges significantly or ceasing trading.

The evanescent nature of DDOs and the insidious way that data can become corrupted (as highlighted in Sections 1.6.8) means that it is even more important, in the digital environment, to create an effective framework to ensure its preservation. This framework would need to facilitate the uniform capture of the essential content, context and structure of a DDO. This is achieved by the use of metadata, of which there are various formats, the diversity of which is discussed in Section 1.12. There are also many methods for achieving interoperability between these formats in networked environments (see Section 1.13).

The development of metadata (data about data) standards and the capability of access systems are seen as vital components in creating online storage systems. The form these metadata standards should take varies as Heery (1996) displays in the article *Review of Metadata Formats,* as does opinion on the phase in an object's life cycle that metadata should be created. Wallace (1995), takes the view that metadata should be created at the beginning of an object's life cycle in order to capture "...crucial contextual information..." (Wallace, 1995, p. 106), that is, at the system design stage. However, MacNeil (1995), suggests that providing the objects are well managed during their 'active phase', post hoc description of documents using archival description is possible. How an organisation defines the concept of 'long-term intellectual preservation' is felt to be heavily influenced by the collection management policy of an organisation and ultimately by its overall aims and objectives, its mission statement. This concept of life cycle may vary along a continuum from 'as long as it is needed', to 'in perpetuity'. The definition of life cycle is thought to have a major impact on the types of resources held and the methods of selection and retention; this is examined in greater detail in Chapter 3.

1.6.2 Copyright

Until 1977, there was not a perceived need for specific protection for computer software, as most DDOs were in-house creations and as such, not transferable for other uses, or to other organisations. Additional legal protection became necessary when software piracy became an issue for the computer industry in the late 1970s. Arising from the White paper *Intellectual Property and Innovation* (1986), the

Copyright Designs and Patents Act 1988, aimed to address these concerns by implementing legislation to cater for future change. Since 1988, extensions to copyright legislation have occurred to cover databases[4] and specific areas such as crown copyright[5]. The lack of focus on the implications of copyright in the networked environment is identified in the following quotation.

> There has been insufficient focus on the existence of the intangible right that is copyright, and the advent of the Internet freeing the copyright from its medium, like releasing the genie from the bottle, will have interesting, and probably unexpected, consequences. (Bainbridge, 1999, p. 245)

The legislation which has been developed for copyright and rights issues in databases is concerned with measures relating to the prevention of unauthorised extraction and re-utilization of the contents of databases. As Bainbridge (1999) notes, this exempts public repositories lending a copy of a database that is not for direct, or indirect commercial gain. The new instrument protects databases on two levels as given below.

- Copyright protects material that is the author's own intellectual creation.

- Protects databases where substantial funds are expended in finding, establishing the authenticity and displaying the contents.

The new rights in databases stems from a desire to harmonise the protection of databases within the European Community.

Within the academic community, initiatives are being undertaken to establish copyright clearance centres, namely Higher Education Resources ON-demand (Heron)[6] and the National Electronic Site Licence Initiative (NESLI)[7]. The first of

---

[4] *The Copyright and Rights in Databases Regulations 1997.*
[5] *Crown copyright in the information age.*
[6] URL: http://www.stir.ac.uk/infoserv/heron/home.htm
[7] URL: http://www.nesli.ac.uk/index.html

these offers an income to rights-holders as a by-product of the creation of an on-demand publishing and electronic reserve (OD/ER). The second promotes the widespread delivery and use of electronic journals in the United Kingdom higher education research community and offers libraries the choice of purchasing electronic journals separately from their print equivalent, and unlike the Pilot Site Licence Initiative (PSLI) it aims to be self-financing.

However, differences relating to copyright between print and digital publications have been highlighted by Hendley (1996) with respect to remote access to digital publications raising concerns with publishers (an issue which is considered in Section 6.3.2).

> Digital copyright material (like all copyright material) must be made freely available to researchers and readers in our reading rooms. For non-digital material we also provide very limited remote document supply services in that we allow a limited number of photocopying for remote supply (within the limits prescribed by the Copyright Act). The latter may be a more problematical issue with non-print material, as we fear that remote access will from the start be vetoed by publishers. (Hendley, 1996, p. 17)

There is a general trend within the European Community to harmonise aspects of copyright with other legislative rights, an aim which forms part of the philosophy behind the development of the 'information society', as evidenced by Directive COM(97) 628 (1997). Included in this is the protection for electronic 'rights management information', with a compliance date of June 2000.

1.6.3 Legal deposit

The previous sub-section indicates that existing copyright legislation with modification translates into the digital environment, however, this is not the case with legal deposit. New legislation is required to resolve the rights issues that the long-term management of DDOs presents. The British Library has been spearheading initiatives in the United Kingdom to recommend changes to existing legal deposit legislation. As part of this process it commissioned CIMTECH to

conduct research into the preservation of digital materials (Hendley, 1996). This re-assessment was felt necessary, as existing legislation is only applicable to non-digital publications, with the potential danger that unless there is a change to incorporate digital media into the collections held by legal deposit libraries, these collections would become less representative. The resultant analysis of the situation detailed in the report concentrates on published material which has been 'born' digital, and details 'best practice' in other national libraries and specialist archives, such as the Data Archive at Essex University[8]. Although, as stated in the report, it was not possible within the time frame to create a strategy for long-term preservation, it did raise a number of crucial issues:

- Requirement for selection
- Definition of a digital publication
- The importance of documentation
- An assessment of costs based on the life cycle of an object
- Absence of standards between the metadata and the data content.

This report also stated that the new legislation required by the British Library to expand legal deposit to cover DDOs would comprise four elements:

> i A statement of the objectives of legal deposit.
>
> ii A definition of the material which is eligible to be designated for legal deposit.
>
> iii Identification of the authority empowered to designate eligible material as subject to, or not subject to, legal deposit and for each class of eligible material subject to legal deposit, the person(s) responsible for depositing it, the agency with which it is to be deposited and any special conditions relating either to its deposit or to its treatment by the designate agency.
>
> iv A definition of the criteria to be taken into account by the designating authority.
>
> (Hendley, 1996, p. 19)

---

[8] URL: http://dawww.essex.ac.uk/

In 1997 the Library published a consultation paper entitled, *Legal deposit of publications: a consultation paper* (Department of National Heritage, 1997) of how it saw legal deposit moving forward. This is based in part from the recommendations made by Sir Anthony Kenny who stated that:

> In an environment of constantly extending networks, it is important that any extension of legal deposit should go hand in hand with safeguards for the rights of producers and publishers. After extensive consultation with right holders, the library has concluded that the most acceptable form of deposit for electronic texts is the one which most closely parallels the system for printed text. (Kenny, 1996, p. vii)

The response to the consultation document was positive and in agreement with progressing with this plan. However, this document had several drawbacks, firstly it focused primarily on extending cover to CD-ROMs[9] and secondly reducing the burden of responsibility on publishers to maintain an electronic archive of their publications. The product of the consultation procedure led to the formation of a working party that advocated increasing the range of digital materials provided for by new legislation (Working Party on Legal Deposit, 1999). The report outlined the following as the main features of a system for compulsory legal deposit:

- The extension to cover new media as it becomes available.

- Provision granted to a national archive of published material by rights owners to hold the item for archiving and access to the archive.

- The fulfilment of the publisher's obligation to deposit an item by depositing only in a single medium. The choice of medium being made by the repository.

- Access to deposited items for authorised users of the repository, with the exception of cases where negotiated temporary access restrictions apply.

---

[9] No specific reference for the provision of recordable compact disk (CD-R) is indicated in this consultative paper.

- Placing of restrictions on the whole or partial dissemination of items only when, either the copyright has expired, or with the permission of rights holders.

- The right of the national archive to exempt certain categories of material that they feel would place an unfair burden on publishers.

- The right of publishers to embargo or exclude additional material and the creation of a standing committee responsible to the Secretary of State to resolve disputes arising from the interpretation of legal deposit legislation.

The broad theoretical nature of these principles enables the proposed legislation to apply to publications in a range of media, including media that do not yet exist. The Working Party's proposals include online material, where this is the primary form of publication. However, such online publications would initially be restricted to certain monographs and journals, pending the results of further examination of the technical and economic factors of depositing material as online databases. The ability to access material at more than one location was identified by the Working Group as a potential problem in terms of potential loss of control over the data and loss of revenue for publishers. The Working Party's report sees physical location of deposited material as a matter for negotiation between repositories.

Any of the proposed changes to legal deposit legislation will require a revision in selection criteria as part of the British Library's overall strategy for the long-term management of DDOs.

1.6.4 Cost

The preservation of digital data can incur high costs and this, as Hendley (1996) notes, has resulted in many digital archives having to adopt a more selective digital preservation policy. The cost of preserving digital materials has largely been estimated on extrapolations of traditional paper-based material, as in the case of the CIMTECH report commissioned by the British Library (see Section 1.6.3). These

may be sufficient for materials that most closely resemble traditional resources, but are insufficient for those which are interactive (as examined in Sections 5.2 and 5.3). These dynamic and living DDOs (see Section 1.6.5) require much greater monitoring in their 'active phase' and more critical management once they are received by repositories.

Costing models for library acquisition and storage have been developed in the United States by Yale University Library, to estimate the differences between DDOs and traditional materials. The findings demonstrated that storage and access costs for DDOs decreased over time, whilst the opposite was true for traditional resources, as seen in Table 1.1.

Table 1.1 Hard-copy *v* electronic library costs

	Year 1	Year 4	Year 10
Hard-copy library			
Storage cost per volume ($)	0.24	0.27	0.34
Access cost per volume used ($)	3.97	4.46	5.64
Electronic library			
Storage cost per volume ($)	2.77	1.83	0.80
Access costs per volume used ($)	6.65	4.76	2.70

Source: Meadows (1996, p. 217)

This does not signify that electronic storage is cheaper than traditional forms of storage, (as is shown by the higher costs in the early years of storage). There will be more digital materials requiring retention that are constantly being acquired, together with hidden costs involved in staff training and re-training in these new methods.

1.6.5 Diversity of types of DDOs

This research has identified a number of taxonomies that have been mooted to explain the diversity of digital data for specific purposes (Solbakk, 1995; Hendley, 1996), which concentrate on a differentiation based on the physical medium. The works of Ross (1997) and Barker (1997) mark a departure from this line of thinking, and concentrate on the functional features (or publication medium) of a digital object.

Ross divides DDOs into three categories as shown below.

1. static documents which are composed of such elements as text, tables, and images;

2. multimedia or data-rich documents such as the kind of documents that we encounter in the networked environment using such technologies as the world-wide web or using www-based corporate intranets;

3. dynamic documents which are dependent upon data that might have variable instantiations. The increasing use of voice mail, digital video records and virtual reality alongside these other document types adds an additional dimension to the document management issue.

(Ross, 1997, p. 332)

This taxonomy is useful in providing a division based on function and providing examples of each type. Barker (1997), in his classification of electronic books examines the type of information embedded in the book.

> Depending upon the types of information that they embed, the basic properties that they exhibit and the functions that they have to perform, electronic books can be classified into ten basic types. (Barker, 1997, p. 125)

The diversity of types of DDOs can be demonstrated by the taxonomy of electronic books shown in Table 1.2. The book types display a greater degree of interactivity, the lower their position in the table. These ten book types can be assigned to the three categories of Table 1.3; 'static', 'dynamic' or 'living'.

Table 1.2 A simple taxonomy of electronic books

Book type	Characteristic feature
Textbooks	Contain linear text/search engine
Picture-books	Embed various sorts of static pictures
Talking books	Use audio narrations and sound effects
Moving picture books	Embed animation and/or video
Multimedia	Combine text, pictures, and sound
Polymedia books	Use electronic and nonelectronic media
Hypermedia books	Involves nonlinear information structures
Intelligent electronic books	Can adapt to given user's behaviour
Telemedia books	Embed telecommunications facilities
Cyberbooks	Involve the use of virtual reality techniques

(Barker, 1997, p. 125)

Within the same work Barker develops the construction of types a stage further by suggesting that electronic books fall into three distinct categories based on publication medium as seen in Table 1.3.

Table 1.3 Categories of electronic books based on publication medium

Book types	Basic properties
Static	Fixed pages Fixed information
Dynamic	Reconfigurable pages Fixed information
Living	Reconfigurable pages Changing information

(Barker, 1997, p. 128)

This taxonomy has been adapted and used as a starting point to generate discussion in stakeholder interviews, as shown in Table B.1, Appendix B. This table shows that the traditional media of paper, cassette and video are all static and contain few book types. DDOs on the other hand contain a diversity of 'book' types as well as being subdivided into 'dynamic' and 'living' categories. Clearly in this scenario the traditional selection techniques applied to paper, cassette and video media, static

media, are not adequate for 'dynamic' or 'living' DDOs. Later in this chapter these types will be discussed in greater detail with regard to the way each influences the development of selection criteria (see Section 1.10).

1.6.6 Relationship to medium

Unlike printed materials, DDOs can be viewed separately from their medium. (Graham, 1996; Mandel, 1996). Shepherd (1996), when referring specifically to electronic records, emphasises that both the medium and the message "...must be captured and preserved if the record is to be preserved" (Shepherd, 1996, p. 2-3). However, Shepherd's (1996) findings do not take into account future technological advancements which may enable the message to continue to exist when it fails to be supported by its original medium.

The findings in this thesis agree that the medium and the message can be divorced but while both need to be considered in long-term preservation the medium is of secondary importance to the message.

1.6.7 Reliance on printed support documentation

The way in which some DDOs are produced by manufacturers at present has led to a reliance on printed support documentation. If in the future instructions are incorporated into the DDO itself, reliance will be negated.

1.6.8 Fragility of DDOs

Unlike printed materials, where there is hard evidence of their existence over centuries the decay rates of digital media are based on extrapolations of accelerated testing. In reality DDOs may decay at a slower or faster rate than the extrapolations predict. As the decay which occurs to digital objects is likely to be invisible to the human eye it is likely that specialist technologies will be required to detect it and indicate when it is necessary to refresh the media or migrate the substantive content to a new media. Continuous management of DDOs is therefore

vital for their continued existence. Monitoring mechanisms need to be linked to the individual digital resources. As stated in Section 1.6.6, the medium and the message can be separated, however the continued existence of DDOs cannot be achieved unless the message continues to be accessible in a digital state. "An archival record must remain inviolate but remain accessible, and so its format must be capable of change" (Murdock, 1997, p.118). Although the medium is important in order to maintain the existence of the message, as section 1.6.6 concludes, it is subsidiary to the message itself, a view articulated by Murdock (1997).

> The real issue in preserving electronic records is identifying when the decision to migrate (and the most suitable approach to migration) has to be made. In managing electronic records we are seeking to preserve access to information by ensuring that records remain legible, rather than preserving the media on which records are held. (Murdock, 1997, p. 121)

This is where physical preservation, the management of the physical entity, becomes an important part of intellectual preservation, the management of the information content. The DDO is therefore dependent on some form of monitoring such as migration, a process which itself has inherent dangers. Rothenberg (1995) suggests that although digital data itself is immune to decay, this is not true of the physical storage media and that digital documents are highly susceptible to corruption during the process of migration.

> If digital documents and their programs are to be saved, their migration must not modify their bit streams, because programs and their files can be corrupted by the slightest change. If such changes are unavoidable, they must be reversible without loss. (Rothenberg, 1995, p. 29)

However, there have been calls for research "We have little concrete research that can guide decisions about the possible loss of meaning and integrity of a digital objects if their physical structure is altered during the migration process" (Hedstrom, 1997).

There is therefore a need to break out of the technology cycle to prevent being driven by technological obsolescence. It is hoped that the development of

intellectual preservation strategies to monitor the information content mooted in Chapter 7 will aid this process.

The twin challenges of overcoming fragility of DDOs and technological obsolescence of their storage media are required for their successful long-term management. A number of strategies are being put forward to achieve the physical preservation of DDOs:

1. Conversion to paper or microfiche
2. Migration of digital information
3. Technology emulation

The first of these, would only be feasible for static DDOs (defined in Section 1.6.5). The migration of data from one hardware platform to another has the advantage of harnessing the capabilities of advances in hardware, although some loss of functionality or presentation is possible. Data that is being re-worked, or accreted, may benefit from this 'periodic upgrading'. According to Kranch (1998), by necessity migration requires periodic selection of resources, which he suggests could cause problems such as changing the object physically as it is migrated from one medium to the next, and could lead to little material surviving long-term if it goes thorough a process of selection each time migration is undertaken. He therefore advocates the use of *digital tablets*.

> For a substantial preservation effort to be launched that can be considered in any way permanent, it must use an alternative to migration that does not include continual selection or run the risk of permanently altering the original as it travels through time. What is needed is a preservation technology that includes both hardware and software and eliminates costly migration and storage costs. One alternative could be a digital preservation tablet that contains both the hardware and software needed to preserve the documents it contains. (Kranch, 1998, p. 143)

The third option, is emulation, which involves developing programs that mimic the characteristics of the original software but allow it to run on future hardware. Proponents of emulation (Rothenberg, 1999) suggest that it allows replication of

the original 'look and feel' of a DDO, which may be of value to, for example, a humanities scholar wishing to preserve the DDO in its original form.

Important though the preservation of the medium is to convey the digital information content into the future, it is fundamental to develop an all embracing framework that will provide for the continued existence of the message, and by necessity, its supporting medium. Existing selection mechanisms need to be re-assessed (Section 1.10) and integrated into this new framework. This subject is explored in detail in Chapter 5.

1.6.9 Hardware and software compatibility

At present it would appear that system design is being driven by the technology cycle, a cycle which is not catering for the long-term management of DDOs. The creation of software that considers forward compatibility and hardware that facilitates compatibility between systems, are seen as vital to the continued access of DDOs. Strategies to create interoperability have been mooted (see Section 1.13) and these will be examined in detail in Chapter 5.

1.6.10 Technological obsolescence

Obsolescence of technology is seen to be more problematic than the physical decay of the storage medium. Graham (1996), in a paper given at the JISC/British Library Workshop held at the University of Warwick, saw the need to take positive action to prevent this obsolescence. This action, he felt, should be the migrating of data from old technologies, which vendors and users cease to support, to new technologies as they emerge. However, those supporting emulation would favour the ability to mimic the characteristics of older software.

1.6.11 Integrity and authentication

The growth in volume of the digital data and its fluid nature, or *malleability* as Neavill (1984) defines it, are seen as major challenges facing digital preservation.

The ease with which a networked DDO can be altered either intentionally, or unintentionally, challenges our faith in the object displayed on the screen being a 'true copy' of the original. This issue is unique to DDOs because with print-based publications there are tangible indicators which point to the authenticity of the object, such as; the publisher and the Cataloguing In Publication information.

Due to the fluid nature of DDOs it becomes important to put in place mechanisms which allow the authentication of objects. Graham (1993) advocates the process of time-stamping as, "...a means not only of authenticating a document, but its existence at a specific time" (Graham, 1993, p. 90).

The main focus of the literature in the 1990s has to date been concerned with the physical preservation of Digital Data Objects, a continuation in the approach used for the preservation of traditional library and archive materials. However, as stated earlier, digital data raise many issues which are unique to this form of material and it may be that we are progressing down an unsuitable route. Once solutions have been found to overcome these physical preservation problems, the unanswered questions of intellectual preservation remain. Rather, there is a need to re-evaluate our reasons for preserving Digital Data Objects long-term and focus on intellectual preservation issues before re-evaluating physical solutions to the preservation problems which this digital environment creates.

Bearman (1996a), in the introductory section of *Virtual archives* suggests, that there are two critical barriers to the ubiquitous creation and management of electronic records.

> First, the assurance that such records satisfy the requirements for evidence and
> Second, methods by which records can be made available over time without constant re-presentation and migration of their intellectual contents. Unless we can ensure that these two requirements are satisfied in electronic environments, no electronic records system can be archival. (Bearman, 1996a)

Extensive research has focused on the developing strategies to represent the intellectual content of DDOs through the use of metadata (see Section 1.12). Various formats have been mooted, these include the Encoded Archival Description (EAD) and the Universal Preservation Format (UPF). These metadata formats increase in complexity, precision and cost from full-text indexing as utilised by harvesters such as Alta Vista; unified surrogates, minimally fielded surrogates (this includes the Dublin Core Minimalist approach); qualified surrogates (this includes the Dublin Core Structuralist approach); through to richly-structured surrogates for example MARC and the Text Encoding Initiative. How search engines identify Digital Data Objects has been only one strand to the Internet resource discovery problem. Alternative approaches have centred on encouraging people to select resources against criteria, for example, the NISS and BUBL services. This lack of standards is highlighted by Lynch (1997) when he states that, "...something very much like traditional library services will be needed to organize, access and preserve networked information" (Lynch, 1997, p. 44). Brake (1997) echoes this view when he states, "Just how much information is out there is impossible to say, not least because anyone with a computer attached to the Internet can publish Web pages" (Brake, 1997, p. 12). Many traditional methods used for archiving printed materials, according to Rothenberg (1995), are unsuitable for preserving digital data, yet preservation is essential if our heritage is not to be lost to future generations.

Feather (1996) sees ownership as the key to addressing the long-term preservation and access to digital resources. "The diversity of the origin of digital information, and the consequent diversity of ownership of that data, is actually the key to understanding and addressing the issue of its long-term preservation and availability" (Feather, 1996, p. 59).

1.7 Aspects of intellectual preservation

Whether the original image is required in the creation of an electronic object is determined by the purpose and origin of the DDO. For example some DDOs that

are 'born' digitally as in the case of three-dimensional digital brain scans[10] do not have a non-digital equivalent. This can be an advantage as it can extend the possibilities for manipulating and displaying the image but also a disadvantage as there is no hard copy to recourse to if the DDO becomes unreadable.

1.7.1 The importance of preserving the intellectual content of DDOs

In a paper presented to the Seminar *Digital Archiving - Emerging Practice*, Elkington (1997)[11] described the fundamental drivers which underline the need for intellectual preservation of the content of digital information:

- Current or anticipated use
- Research interests
- Legal imperatives
- Business imperatives
- Public mission
- Social mandate.

(Elkington, 1997)

Proposed strategies to manage DDOs over the long-term need to take into consideration the existence of a range of DDOs created for different purposes. The super-metadata framework proposed by this research aims to accommodate scholarly and non-scholarly DDOs (see Chapter 7).

The current or anticipated use of digital data acts as a driver by governing the types of resources that are created and felt to be required in the future.

Research has shown that the intellectual preservation of digital data is important for the needs of future research interests, as seen by the work of Gould and Pearce (1991) *Information needs in the sciences: an assessment.* This research consulted

---

[10] These characteristics are exhibited by resources such as the three dimensional reconstruction of the human occipital lobe displayed on the following site (http://spectrum.vislab.ucl.ac.uk/WEB/imgbrain.html).

[11] *Integrated international access to distributed archives.* Paper presented at the *Digital Archiving - Emerging Practice seminar*, held at the Public Record Office, Kew, in July 1997.

131 people who were involved in teaching, or research, or closely connected to scientific research. The findings suggested that scholars in eight different scientific disciplines value the access not only to the most up-to-date information but also less current information. The exception to this pattern being computer science, a discipline whose scholars are concerned with the latest and most current information.

As Digital Data Objects are beginning to be put forward as legal evidence in courts, there is a growing requirement for records to be auditable and accessible, thus acting as a driver for the preservation of the intellectual content (see Section 1.8.2). Similarly, commercial organisations require access to their records, for example a bank may require access, for tax purposes, to records dating back over the previous six years. A more detailed discussion of an initiative from the banking sector is detailed in Section 1.8.8.

The social mandate of some organisations may also act as a driver. Small museums or non-government repositories may wish to preserve the intellectual content of DDOs but be prevented from doing so by lack of finance or infrastructure to cope with this goal.

The increasing importance which the United Kingdom government is attaching to records in electronic form, is acting as a major factor in promoting the need to preserve the information contained in Digital Data Objects. The Public Record Office is developing an Electronic Records in Office System (EROS) with government departments (see Section 1.8.5).

1.7.2 Influence of organisational policy on which DDOs are kept long-term

The range of DDOs an organisation expects to preserve is influenced by internal and external factors. A lack of expertise of the staff, and/or financial resources at the disposal of an organisation, may mitigate against the preservation of data, for example, the preservation of complex structures of hypermedia documents can be costly to maintain. External influences can be exerted by outside agencies who create the DDOs.

This influence can be exerted by imposing restrictions on access written into licensing agreements. The aims and objectives of organisations themselves can exclude some materials from preservation.

1.8 Organising the management of DDOs

As digital information is becoming widespread, and networking more pervasive, more sectors of the information community are seeking to develop solutions to challenges presented by the long-term retention of digital data, as evidenced by the strategies developed by archivists, records managers, librarians and commercial enterprises. Sections 1.8.1 to 1.8.9 do not attempt to provide an exhaustive discussion of such initiatives, but to demonstrate the strengths, weaknesses and diversity of characteristics of the initiatives, together with challenges they are designed to surmount. The issues addressed by these initiatives are drawn upon later in this thesis.

1.8.1 Use of diplomatics and archival theory to manage authenticity

The University of British Columbia (UBC) research project entitled *The preservation of integrity of electronic records* ran from April 1994 to March 1997 and stated the following objectives in Section three of the project overview.

> - to establish what a record is in principle and how it can be recognised in an electronic environment;
> - to determine what kind of electronic systems generate records;
> - to formulate criteria that allow for the appropriate segregation of records from all other types of information in electronic systems generating and/or storing a variety of data aggregations;
> - to define the conceptual requirements for guaranteeing the reliability and authenticity of records in electronic systems;
> - to access those methods against different administrative, juridical, cultural and disciplinary points of view.
>
> (Duranti, Eastwood and MacNeil, N. d.)

The methodology on which this research is founded, centres around the use of concepts and methods developed in the seventeenth and eighteenth centuries for demonstrating the validity and authenticity of documents, known as diplomatics. Diplomatics, describe records purely as individual entities, and to gain a broader approach to record management, this was combined with the collective approach advocated by archival science.

As a result of this research, specific methods were developed to ensure the validity and authenticity of electronic records, these being achieved by integrating documentary and business procedures and embedding procedural rules into the record system. Additionally, a second category of findings centred on management issues, such as the life cycle of managerial activity and responsibility for the integrity of records. The findings pointed to the division of the life cycle of managerial involvement in the preservation of the integrity of electronic records into two, one for active and semi-active documents, and the other for inactive documents. In terms of division of responsibility the research concluded that:

> ...ultimately, different parties are and should be accountable for different functions: the creator should be accountable for its action through its records, the preserver should be accountable for those records. This division of responsibility is what ensures the authenticity of inactive records and make of them the impartial sources that society needs. (Duranti, 1997, p. 64)

The research devised a number of templates that detail; record creation, completeness, reliability and guarantee proof of authenticity. The templates were developed for each of the above in both the electronic and traditional environments. The research provides a comprehensive framework for the authentication of records based on societal warrant. However, there would be a need to expand these templates if they are to be used to select a particular record for long-term preservation, by examination and inclusion of selection criteria.

1.8.2 Evidential approach to managing authenticity

A research project carried out at the University of Pittsburgh School of Library and Information Science, between 1993 and 1996, examined the variables which influence the satisfaction of archival requirements for electronic records management. The project originated from the general shift in attitude towards electronic records management in the United States, and the development of priority areas arising from the 1991 research conference on electronic records management, organised by the Minnesota Historical Society, and funded by the National Historical Publications and Records Commission. A team of library, information and archival science experts examined the following issues:

> (1.) Recordkeeping functional requirements for electronic information systems.
>
> (2.) Variables in organizations that affect the way in which both software and hardware are utilized and which may affect the degree to which archival functional requirements can be adopted.
>
> (3.) Technical capabilities or organisational software products to satisfy archival requirements.
>
> (4.) Other means, such as policy and standards, to satisfy archival functional requirements.
>
> (5.) Effectiveness of technology and policy strategies to ensure that archival interests can be met.
> (Cox, 1994, p. 283)

The tangible outcome of the research was to produce a "...set of research products that address the viability and recordkeeping functional requirements..." (Cox, 1994, p.283). To achieve its goal the research project aimed to test the following hypotheses:

> (1.) The functional requirements for the archival management of electronic records are the same as for traditional records, although many functional requirements will not be satisfied by traditional record systems.

(2.) It will be possible to satisfy each of these functional requirements following any or all of the four tactics, although many requirements will be more fully satisfied for electronic records than they could for paper records.

(3.) Different business applications will share different sets of archival functional requirements, and differing degrees of risk are associated with the non-satisfaction of these requirements in different business applications.

(4.) Different software applications will not dictate different recordkeeping functional requirements, but different packages within the applications categories will satisfy the recordkeeping functional requirements to different degrees.

(5.) Recordkeeping functional requirements will be the same for each business sector, and different sectors will not determine the choice of tactics as much as different corporate cultures.

(6.) The best way to satisfy functional requirements will depend heavily on an organisation's corporate culture, but the technological capabilities of the archives and its agents will be less critical in satisfying archival requirements than will be the acceptance of archival responsibility by managers throughout an organisation.
(Cox, 1994, p. 283-284)

The set of 'functional requirements' for electronic records generated by this research are regarded by Bearman (1996b) as equally applicable to manual or hybrid systems. A hybrid system, being one which is expected to cope with both electronic and manual records. These functional requirements are concerned with the organisational structure and the design of systems architecture to facilitate 'best' electronic records management practice, and whilst comprehensive in these respects does not attempt to address the issue of selecting records for long-term retention. These functional requirements may provide an overlying structure into which specific selection criteria for the long-term preservation of these records could be constructed and implemented.

According to Bearman (1996a) there is an expectation from users, information technology staff, regulators, auditors and lawyers of records managers providing guidance on the management of electronic records. One of the basic problems associated with the management of electronic records is, defining what constitutes a

record. In Bearman's view, for data to be classed as a complete record, it must "...contain the content, structure and context generated by the transaction they document" (Bearman, 1994, p. 300). Content refers to the message that is conveyed by the record. Structure refers to the relationship between the storage of different blocks of the data within a record, within the application it is running on and even across computer systems. The final characteristic, context, provides a framework for determining provenance, the reason for the creation of the record, together with information on current and future access.

The nature of communication in the electronic environment creates problems for record management precisely because it is virtual. The sending of a message does not necessarily leave a trace, its receipt does not necessitate a copy, and preserving the physical entity, the bit stream on a diskette does not ensure its survival. Once data has been identified as having the three characteristics of '*recordness*', to be a true record of a communication according to Bearman (1997)[8] it must exhibit the following:

> Must contain the data that was actually communicated.
> Must contain evidence of the process by which the data was obtained.
> Must document the transaction which they are recording.
> Each transaction generates a new record.
> (Bearman, 1997, p. 5)

By stating that a record is a communication it excludes data created for other purposes. It should also be noted that this definition specifically focuses on records created as a result of conducting business or records received as a communication from other organisations which can provide legal evidence if required. For this reason the super-metadata framework detailed in Chapter 7 proposes to be flexible enough to cater for information created in digital form regardless of whether it was intended as a communication or not.

---

[8] Workshop hand-out *Strategies for the management of organisational electronic records*: Workshop on Electronic Records, held at the Public Record Office, Kew, 12-13 September 1997.

To provide evidence that the authenticity and validity of a record has remained intact, Bearman (1997) suggests that records are required to be captured as they are sent. This capture being performed by the record passing over a logical switch in a computer network. Once a record has passed over this switch a copy of the record is sent to the intended recipient and the original document is electronically stored in a records management repository. These stored records would enter the repository, which may reside within an organisation, or an outside agency/organisation. These stored records would enter the repository with an electronic seal which would remain unbroken until such a time as the document is required as evidence of a transaction.

The development of Bearman's electronic records management, the Reference model for Business Acceptable Communications, revolves around the development of an organisational structure in the form of a policy framework, capture guidelines and the positioning of an archival records management unit within an organisation, together with the development of a systems structure to consider metadata requirements, architectures and migration strategy. The basis for this model is to provide a recordkeeping structure for the electronic environment, and with a focus on the provision of evidence in the business environment.

The Reference Model for Business Acceptable Communications outlines the importance of designing systems that define records for both people and systems. In creating such systems, human barriers such as the "habit of preserving the medium, the 'original'..." (Bearman, 1997, p. 8), need to be surmounted. Additionally, systems barriers such as the "'recordness' must be machine recognisable" and "Metadata captured for recordness must conform to standards" (Bearman, 1997, p. 10) need to be addressed. This approach is fundamentally different to the metadata standards being created (Section 1.12) which focus on the system mechanics of record management rather than focusing as Bearman (1994) does on developing a model to perfom a specific function, provide evidentiality.

The UBC and Pittsburgh research projects provide procedural rules and management structure to facilitate authentication, which is seen as vital for

preserving electronic records. Both research projects demonstrate a management structure for intellectual preservation storage, but do not provide record managers with the tools to select resources for preservation.

1.8.3 Open Archival Information System (OAIS)

The Consultative Committee for Space Data Systems (1999) has outlined *A reference model for the Open Archival Information System (OAIS)*. As with the model developed at Pittsburgh it caters for both digital and traditional resources (including artefacts), and its major purpose is to promote "...a much wider understanding of what is required to preserve and access information for the long-term" (Consultative Committee for Space Data Systems, 1999, p. 21), by serving the requirements of one or more designated archival communities. An integral part of the model is its extensive use of metadata as a mechanism to aid management throughout the life of a resource. The concept of the OAIS model has been adopted by the Arts and Humanities Data Service (AHDS) as stated in *Digital collections: a selection policy framework for creating and preserving digital resources* (Beagrie & Greenstein, 1998). The executive summary and recommendations of the above document states that, the framework aims to provide "...strategic guidelines to stakeholders involved with digital resources at various stages of their life cycle" (Beagrie & Greenstein, 1998). The AHDS framework centres around the selection of 'appropriate and cost-effective practices'. The study also notes the importance of selection in the cost considerations and a cost structure.

1.8.4 Preserving and Accessing Network DOcumentary Resources of Australia (PANDORA)

The data model defines the logical entities that the National Library of Australia require to develop policies and procedures for preservation and access, and is the basis for the Digital Service Project (DSP). This initiative has identified a number of issues requiring resolution for their objectives, including the need for agreed arrangements for recording metadata to aid the preservation of publications.

1.8.5 Electronic Records in Office Systems (EROS)

In the United Kingdom the Public Record Office (Macfarlane *et.al.*, 1998) has developed a series of *Guidelines for the management and appraisal of electronic records* as part of the EROS project. These guidelines were developed from a request from the public sector for advice on managing electronic records in the electronic office environment. These organisations need to keep records of business decisions to meet the demands of corporate accountability. This also links into the new demands for effective record keeping generated by proposed legislation, such as the draft Freedom of Information Bill[9]. A major driver in the programme's development has been the government's adoption of the electronic rather than printed document as the definitive version. This programme aims "...to ensure that electronic records of long-term value, created across government, are available for future access" (Blake, 1997[10]). This vision also accommodates records kept for interim deposit to which government requires access. The components of this programme are; the development of guidelines on 'best practice', case studies, pilot projects to trail processes for archiving, and the creation of a long-term transfer and access strategy. The early phase of the system ran from May 1995 to March 1997, and during this time an Advisory Programme Board has surveyed 106 government departments.

The resulting guidelines are based on the premise that the need for electronic records within this environment are the same as for conventional records, although the discussion which follows shows that differences in the management of these records do occur. Similarly selection criteria "...applied to paper records series ought also to apply to the electronic records, unless the branch and function has changed significantly", (Macfarlane *et al.*, 1998, p.56).

---

[9] Your right to know: the Government's proposals for a Freedom of Information Act.
[10] *Electronic documents - the only way to go.* Paper presented at the seminar entitled "Digital Archiving - Emerging Practice", held at the Public Record Office, Kew, in July 1997.

The Public Record Office aims to achieve the widest possible use and dissemination of this guidance within both the public and private sectors. The structure of these electronic records is determined by the metadata that the guidelines recommend should be attached at creation.

> In short, the information with which documents are assembled and located is metadata and if it is not present there is a significant loss of meaning and the records become difficult to manage. (Macfarlane *et al.*, 1998, p.8)

The metadata format adopted by the Public Record Office incorporates elements of Dublin Core and in-house formats, but is seen by some to be too sector specific to be applicable to data creators across the information community in general. It should be noted however, that records would only exist as static and dynamic DDOs as, within an archival context, living records have to be converted to one of these forms to be non-amendable. Although within the guidelines there is an acknowledgement that living records will be updated annually, there is no mention of how they are to be managed in their active phase, or how storage mechanisms for time dependent DDOs would operate, only that it is under consideration as the following quotation shows.

> There are a number of document types that these provisional standards do not address and more work needs to be done on those areas. Examples of those are multi-media formats – where, as present, there are no definitive standards – intelligent documents and the documents using dynamic links. The PRO will be considering these formats in future. (Macfarlane *et al.*, 1998, p. 47)

The strategies considered, up until this point, have been developed in the archival community, and have resulted in the creation of domain specific structures which would not necessarily serve the needs of the wider information community. This apart, they all have a common thread, that is, the essential role metadata plays in developing a long-term management strategy for DDOs. The sections which follow examine developments which have occurred in other sectors of the information community, namely libraries and the commercial sector.

1.8.6 The Alexandria Digital Library (ADL)

This project is an example of the construction of a meta-information framework. Six libraries funded by the National Science Foundation (NSF), the Department of Defence's Advanced Research Projects Agency (ARPA) and the National Aeronautics and Space Administration (NASA), under the Digital Libraries Initiative (DLI) aim to deliver access to library services and collections. The project is founded on a traditional library paradigm and as such uses traditional forms of metadata such as thesauri and subject headings. In order to create a framework that caters for the digital environment, in which there is little direct involvement from librarians, extensive metadata was developed which comprised of different components each catering for different areas of the libraries' operations. To achieve this meta-information framework a number of challenges were identified which needed to be overcome.

> (1)The characteristics of IBOs[11] (2) Characteristics of a digital library itself, and (3) the expanding demands of users in a digital environment. (Beard and Smith, 1998, p. 344)

As with the archival frameworks, this is domain specific, but as can be seen from the identified challenges, users figure more prominently, highlighting the influence organisational aims can have on the framework structure.

1.8.7 The Internet Archive

The increase in the volume of digital data on networks has occurred with the rapid growth of the Internet. In an attempt to archive the data on the Internet, Brewster Kahle (1997) developed the Internet Archive in 1996, which is an attempt to store the entire Internet resources and provide a searchable archive which can be accessed by researchers and others. This archive held seven terabytes of data on disks and tapes as of November 1998. The initiative has raised a number of important issues.

---

[11] Digital Information Bearing Objects.

- Where publications (either in print or electronic form) refer to Internet pages quite often access is denied due to broken links (the average life span of an Internet page being only 75 days). An Internet Archive may allow these links to be maintained.

- Although data on the Internet conforms to a limited number of markup languages, such as SGML, HTML and XML, if data is to be selected for long-term preservation, this research suggests the attachment of additional metadata to aid the management of future access.

- The potential difficulty of initiatives such as the Internet Archive to keep up with the exponential growth of World Wide Web, underlines the need to carefully utilise resources to ensure the long-term survival of evaluated resources, rather than the demise of substantial numbers of DDOs through the inability to manage them effectively.

- Material being stored, in some cases, without the permission of the owner could lead to rights management issues in terms of who can manipulate and store their data.

1.8.8 The International Bank for Reconstruction and Developments Document Management System

Established in 1944, the 'World Bank' has a goal of reducing poverty and improving living standards through sustainable growth and investment in people. The Bank has adopted an information technology strategy which requires the use of proprietary software, and as a result of this a document management systems workshop was held in April 1997. As part of the transition from paper to digitally-based document management system, final drafts of documents are scanned, indexed and stored in a document imaging system (IRIS/imageBank). As a result of implementing this digitization programme it has been decided that effective management of documents requires capture of the documents at creation and

subsequent revision stages. To meet the Bank's requirements the following would need to be fulfilled.

- document content and metadata can be captured at appropriate stages of the document creation and revision stages of the life cycle for storage in the document management system;
- documents are in easily reusable format (i.e., can be readily repackaged and republished in forms suitable to a variety of audiences);
- documents are retrievable via common business applications and delivery mechanisms such as *Lotus Notes* and Web browsers;
- documents can be preserved independent of a specific hardware/software platform; and
- the normal records management controls can be efficiently and cost-effectively applied at the back-end of the life cycle process.

(Lawrence, 1997, p. 2)

The above points show the importance the Bank attaches to the development of accessible resources, which can be reused and preserved in a system which is independent of proprietary hardware or software, although the Bank does have reservations about the security of its documents when using a proprietary system. The system will be dealing with a range of DDOs, including text, data, graphics and images, each of which are managed by different systems at the moment and it is envisaged that these will be integrated into a new system. This Document Management Architecture (DMA) is based on the following design, architectural and technical principles.

- responsiveness to business needs as expressed through the user requirements;
- document locations independence;
- adherence to open and *de facto* industry standards;
- consistency with IMA and integration with key Bank repositories;
- open, distributed, client/server computing architecture;

(Lawrence, 1997, p. 3)

Within the summary of main management requirements, the management of DDOs throughout their life cycle using metadata is recognised, as is the need to document access requirements. It is also envisaged that metadata would allow the automatic application of retention and disposal rules.

1.8.9 Control and identification of electronic business records – Astra Pharmaceuticals

The framework developed by Astra Pharmaceuticals recognises the need to develop the concept of information being managed as a 'corporate resource' which includes ensuring uniformity of the physical and logical structure of the data and contextual information is achieved. Metadata is seen as the vehicle to make this possible. Recorded metadata it is stated would support maintenance, reusability and the 'evidential value' of an organisations records. An electronic archive is also required to have the same reliability and legal status as archives concerned with traditional resources. The report identifies three important factors which are seen to be significant for the successful management of a DDO during its life cycle.

1. Defining data structures which allow description of the wider context of the data relate to the mission to be documented. This concept refers and take advantage of Swedish administrative tradition by defining the "avenda". The term is not translatable directly into English but in some areas case means the same. Those structures (information packages) must contain the necessary amount of metadata to allow full understanding. They shall maintain their evidential value even after EDI to another system.

2. Information in databases must be archived as flat ASCII files together with the information models and process model also documented in flat ASCII files. The syntax must be in accordance with established standards.

3. The base for maintaining text or multimedia is simple pieces, individually maintainable, linked to achieve the full presentation.

(Andersson, 1997, p. 10)

It is envisaged that the format used for storing the data would be changed through migration, but that the structure, context and content would not be changed by this process.

1.8.10 Common trends from the initiatives

The importance of metadata in aiding records management has been highlighted by these frameworks. The metadata created to aid management of the storage of electronic data may, with enhancements, aid the selection of these resources for preservation. At present this metadata exists in a number of formats used mainly for bibliographic control (see Section 1.12).

1.9 Source of origin

The origin of DDOs may have an influence over the availability of resources for long-term preservation and also their accessibility for use. As the diversity of organisations and an increasing number of individuals create DDOs and give credence to the digital form as the definitive version, so the diversity of sources of origin increase. Governments, commercial organisations, educational establishments, charitable organisations, societies, institutions, and individuals are all potential creators and are therefore sources of origin of DDOs. The diversity of origin of DDOs is outlined by Feather (1996) in the following quotation.

> The network environment is generating data at a unprecedented rate: some of this is essentially private, such as electronic mail; some of it is semi-public, such as the contents of e-mail discussion groups; some of it is clearly in the public domain, and to that extent 'published', such as the contents of sites on the World Wide Web. (Feather, 1996, p. 62)

1.10 The importance of selection and evaluation of some existing selection criteria for long-term preservation

In the pre-digital environment the selection of materials has been seen as a crucial phase in any preservation program (Harvey, 1993b). "In an ideal world deposit libraries might preserve everything, but operating in an environment of technical and resource constraints means that priorities must be set" (Webb, 1999, p.12).

Constraints aside, it is not necessarily desirable to preserve all DDOs as not all have long-term value.

It would appear that there are two distinct approaches to the creation of selection criteria emerging from the literature. The first is concerned primarily with the physical nature of the object, often translating criteria applicable to traditional paper-based material into the digital environment. For example, "publication medium" and "publication type" (Owen and Walle, 1996, Appendix 2, p. 1[12]). The construction of the typology of electronic types is seen by some as the basis for the formulation of guidelines for the management of these 'dynamic' types of electronic documents according to Owen and Walle (1996). Johnson (1996) in developing selection criteria applicable to both print-based and electronic objects, is less concerned with physical characteristics and more concerned with the management of information content to aid the development of collection management policy to facilitate acquisition and access, as evidenced by the following breakdown. She lists these common selection criteria as:

- *relevance of material to program needs
- *scope of treatment (refers to depth and breadth of the material)
- *ability to fill gaps
- *uniqueness of content or treatment
- *quality of scholarship
- quality of physical product
- *currency of information and frequency of updating
- *accessibility of information (how it is organized and retrieved)
- *language
- *cost (immediate and continuing)
- *reputation of author/reliability of publisher
- special features.

(Johnson, 1996, p. 15-16)

The researcher considers the starred items from the above criteria to be appropriate criteria for the intellectual preservation of DDOs. However, criteria such as quality

[12] This typology was formulated as a result of the CoBRA (Computerised Bibliographic Record Actions) instigated study, which was funded by the European Commission and was aimed at national deposit libraries. The work expands the criteria into guidelines for the selection of electronic documents.

of physical product are inappropriate to intellectual preservation. Johnson acknowledges that the nature of electronic resources demands additional criteria, these being:

> - network, hardware, and software compatibility
> - availability of network, hardware, and software resources
> - availability of electrical and telecommunication lines
> - quality of interface (ease of use for library users and staff)
> - quality of retrieval/search engine
> - training implications
> - potential use (size of user community and frequency of use)
> - reliability of vendor and availability of vendor support
> - availability of documentation
> - licensing considerations
> - treatment of graphics, formula, and other nonstandard characters.
>
> (Johnson, 1996, p. 16)

These additional criteria link to Johnson's aim of developing a collection policy for electronic resources, but are not necessarily designed to cater for the selection for long-term preservation.

The second approach to selection has been to assess the value of the information content as a first stage, rather than concentrate on the operational issues. Harvey (1993b) identified four elements considered to be important criteria for the selection of resources, namely; "...the physical condition of the item; how heavily it is used; its rarity; and its economic, aesthetic, historical or other value" (Harvey, 1993b, p. 162). The criteria, he suggests, can be applied to individual items as well as whole collections. Harvey's comment on the applicability of his criteria raises the question of what are the aims of the selection frameworks? For example, the aim may be to select for collection development and users needs, or preservation, or even a combination of these, as determined by the organisation's aims and objectives.

Criteria based on the selection of resources for long-term preservation have been generated and are evident in the work of Feather (1996) and the Steering Committee for the Study on the Long-term Retention of Selected Scientific and Technical Records of the Federal Government (1995). Feather (1996) examines the

issue of long-term preservation identifying established principles which can be applied to the selection of digital material for long-term preservation:

- is the information of legal or administrative significance?
- is the information of permanent historical, literary or scholarly significance?
- is the information updated in a way which makes earlier versions of lesser or no long-term importance?
- is the information more easily accessible in electronic format than any possible alternative? (Feather, 1996, p. 59)

According to Feather (1996), each of the above needs to be considered in turn, and this he makes possible through the following table of preservation options.

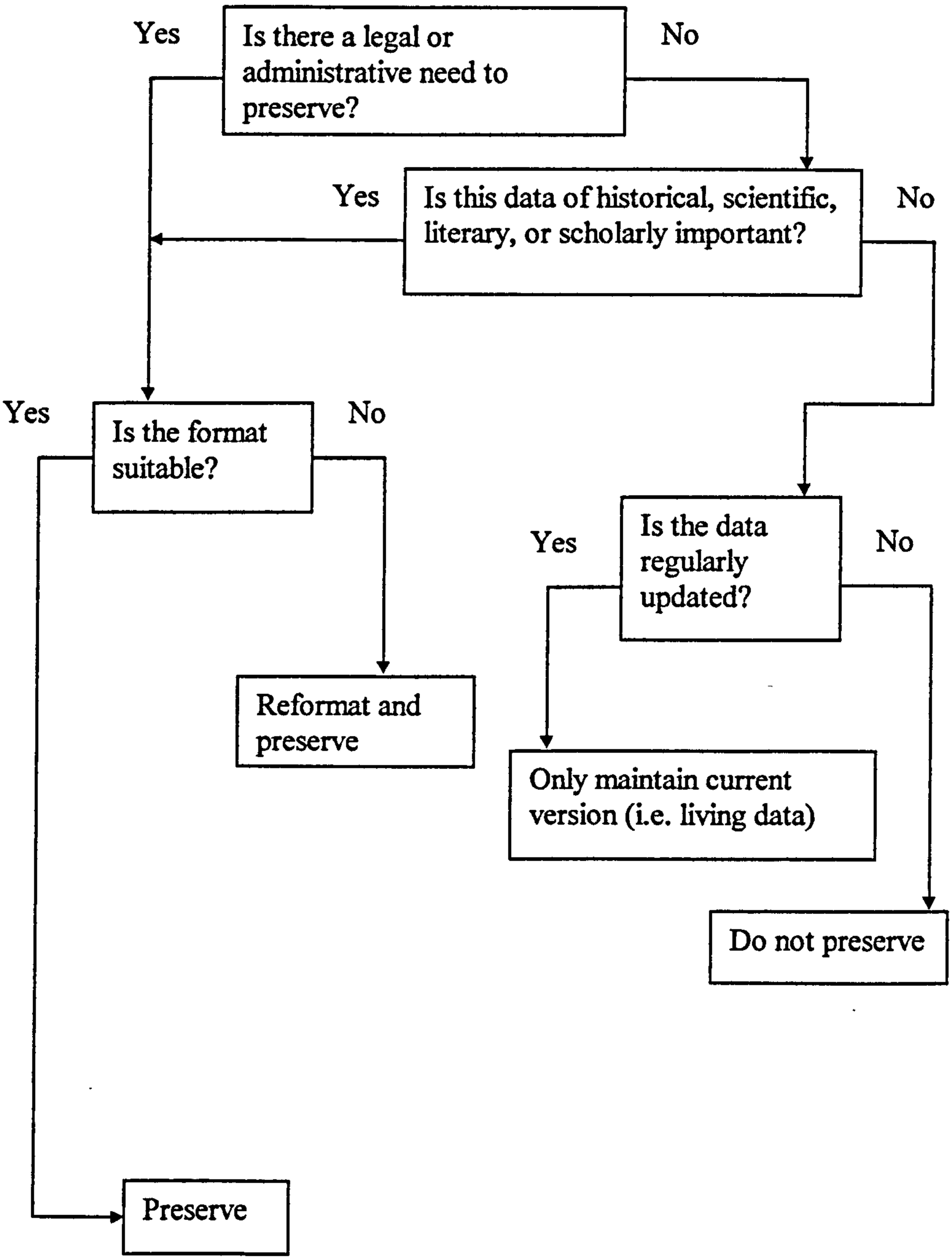
Table 1. 4 Preservation options: electronic data

	Legal or administrative significance	Scholarly importance	Regularly updated	Suitable format	Action
1	Y	Y	N	Y	Preserve
2	Y	Y or ?	Y	N	Preserve in a better format
3	Y	Y, N or ?	Y or N	Y	Preserve
4	N	Y	N	N	Preserve in a better format
5	N	Y	N	Y	Preserve
6	N	N	Y	Y or N	Maintain current file

(Feather, 1996, p. 60)

From Table 1.4 a flow chart can be produced (Figure 1.1), which highlights two core areas where selection criteria are required, namely (a) is the data of legal or administrative significance, and (b) is the data of permanent historical, literary or scholarly significance.

Figure 1.1 Flow chart based on Feather's (1996) preservation options for electronic data



The steering committee, mentioned previously, advocated the adoption of a subject based approach to the archiving of scientific information resources. These resources being observational and experimental data in the physical sciences. The steering committee developed the following retention criteria, which were felt to be common to both observational and laboratory sciences:

- uniqueness of data
- accessibility - adequacy of documentation
- accessibility - availability of hardware
- the cost of replacement.

The first of these criteria, the uniqueness of the data, attempts to eliminate duplication of holdings by determining whether another institution holds the same data, which also meets the required permanence and security standards and is adequately archived. Criteria relating to accessibility are subdivided into the adequacy of documentation and the availability of hardware. Adequacy of documentation refers to the comprehensiveness of the metadata that accompanies the data. The steering committee takes the following view.

> If documentation is lacking or is so poor that a data set is not likely to be of value to someone interested in the data of that type, or the data are more likely to mislead than to inform, that the data set should have a low priority for archiving, or perhaps should not be archived even if resources are available. (Steering Committee for the Study on Long-term Retention of Selected Scientific and Technical Records of the Federal Government, 1995, p. 34)

However, data sets should not be excluded from preservation solely on this basis, but they should be assessed against other criteria before the decision to exclude is made. Accessibility in relation to the availability of hardware is concerned with technological obsolescence (see Section 1.6.10). The preservation of such data requires an assessment of the cost effectiveness of emulating, or acquiring the required hardware to migrate the data, measured against the value of the data set.

The final criterion, cost of replacement, addresses the question, "Could the data be reacquired if a future national need for the data were to arise? If so, would reacquisition of the data be more costly than their preservation?" (Steering Committee for the Study on Long-term Retention of Selected Scientific and Technical Records of the Federal Government, 1995, p.34). The Steering Committee held the view that, the preservation of observational sciences was almost always more cost effective than reacquiring the data. Palaeontology, was cited as an exception, as data on the fossil record could be reacquired at a later date without incurring excessive costs.

In addition to the two general trends indicated at the start of this section, it would appear that selection criteria vary according to the subject matter they are attempting to preserve and the aims of the funding agencies or organisations instigating their creation. Many of these criteria modify existing standards or contain criterion concerned with physical preservation.

1.11 The importance of metadata in digital data management

The development of metadata (see Section 1.2.5) for describing and managing collections of information has been evolving for centuries but with the development of global networks, the pace has accelerated. There is a need to manage data held in electronic form, as Phillips (1995) states:

> To manage information seamlessly, without regard to physical format, requires developing a generic means of describing information. This "information about information" serves as a tool for managers and users of data to get a handle on the raw materials with which they work. (Phillips, 1995, p. 52)

The Steering Committee for the Study on Long-term Retention of Selected Scientific and Technical Records of the Federal Government, (1995) acknowledge the important future role that metadata will play in the maintenance of access to DDOs. They also identified the need to involve originators of digital material in the

formulation of this metadata at the creation stage in the DDOs life cycle. In addition, for Internet resources the creation of an index is seen as an essential part of providing access. Heery (1996) in her review of metadata formats states that:

> Control of the vast number of resources of the Internet requires an appropriate record format (or formats) which will enable the resource to be adequately described and easily located; records must be compatible with an appropriate search engine which in turn would ideally be compatible with a search and retrieval Internet protocol and all components should conform to international standards. At present there are a number of formats which meet at least some of these criteria, each of which has its own strengths. (Heery, 1996, p. 345)

1.12 Diversity of metadata formats

Through detailing a number of current formats, Table 1.5, demonstrates the diversity of current approaches to managing data using metadata. There is a range of views on what constitutes metadata, how it should be classified, what constitutes an appropriate format for a specific domain or a wider field and how it should be managed and applied. This section does not attempt to provide a detailed account of metadata formats which are documented in the literature (see Heery, 1996; Heery, Powell and Day, 1997; Taylor, 1999) rather to illustrate by example the complexity of the current environment, and the diversity of formats that need to be considered in developing a theoretical basis for the long-term management of digital data:

- Purpose for which they were constructed
- Characteristics exhibited
- Standards to which they comply

Dublin Core is specifically designed for the description of networked resources and aims to develop a core of elements that can be applied to resources by providing 'simple' guidelines for its use. The framework purports to be domain-independent,

Table 1.5 Diversity of characteristics and underlying principles exhibited by a selection of existing metadata types

Name	Dublin Core (DC)	Federal Geographic Data Committee for Digital Geospatial Metadata (FGDC)	Universal Preservation Format (UPF)	Encoded Archival Description (EAD)	MAchine Readable Catalogue Records (MARC)	Text Encoded Initiative (TEI)
Purpose	Designed to describe networked electronic information.	To create common terminology and definitions for metadata about geospatial data.	Concentrates on solving technological obsolescence and subsidiary interests such as metadata sets for data integrity and resource sharing	Development of a non-proprietary encoding standard for machine-readable finding aids	Catalogue record exchange mechanism for use between co-operating libraries.	Generic guidelines for representing textual information in electronic form to facilitate retrieval and re-use.
Characteristics	Creation of simple records to describe networked resources through a defined simple set of data elements.	Defines but does not format the content.	Focuses on non-interactive data.	Headers are based on the TEI header, whilst the content is based on the standard for describing archival materials Archives, Personal Papers and Manuscripts (APPM).	Record structure is highly regulated	Use of headers which are either attached to the document or point to remotely stored objects.
Standards conform to	Discussions to link with the Resource Description Format (RDF)	Spatial Data Transfer Standard (SDTS)	Open standard	SGML DTD	ISO 2709: 1981 (BS 4748) format for bibliographic information exchange on magnetic tape AACR2 and ISBD(G)	SGML DTD

although some have suggested that its roots in the library community focus the format on Document Like Objects (DLOs). In contrast, FGDC is a domain-specific format designed to cater for spatial data, for example, fields allow domain specific information such as cloud cover, vital for satellite images. The focus being on creating standardisation of definitions and terminology within communities concerned with geospatial data.

Resource discovery and re-use are seen to be the primary motive behind the development of these two metadata formats. Formats such as UPF concentrate on the preservation of DDOs however, this focuses on preserving the information carrier, favouring the emulation strategy for digital preservation, and largely concentrates on non-interactive objects. A different approach has been adopted by the EAD format where the development of a non-proprietary encoding standard for archival finding aids is the primary focus, conformant to archival standards, and is therefore focused on satisfying the needs of the archival community. Similarly, MARC was developed within the library community and has a highly regulated structure compliant with AACR and International Standard Bibliographic Description (General) (ISBD(G)). Including the MARC format in Table 1.5, raises the issue that some metadata formats are founded on principles designed to cater for older technologies, in this case tape-storage, which influences its structure and ability to function in a networked environment. TEI on the other hand has been developed with interoperability as a principle feature, however, this focuses as its name suggests on textual information. A number of formats conform to international standards as Table 1.5 illustrates.

1.13 Mechanisms for interoperability between metadata formats

The diversity of approaches to metadata has led people to consider ways of linking formats, that is, developing interoperability. There are a number of developments occurring in different sectors of the information community which aim to achieve interoperability either as a central feature or as a by-product of the system or

project, one example is the Metadata Interchange Specification (MDIS) developed by the Metadata Council in 1995.

> MDIS consists of components that represent the minimum set of metadata elements and the minimum integration points that must be incorporated into database tools for compliance. (Gardner, 1997, p.47)

This Council consists of six companies that are attempting to "...standardize metadata interchange among products from diverse vendors", (Gardner, 1997, p. 47). As this is a commercially based system, it could lead to standardisation of metadata types that are linked to the technology cycle. This dependence on the technology cycle could prevent data in legacy systems and future formats being incorporated into the standardisation process, issues discussed in detail in Chapter 5.

Interoperability is being encouraged by the development of encoding mechanisms such as the Standard Generalized Markup Language (SGML)[13]. This provides a structure of codes to which descriptive content can be added. Rather than attempting to standardise on one metadata format, a common syntax is used, allowing the co-existence of several domain specific formats. Thus enabling metadata created once to be used throughout global networks such as the Internet. Formats such as UKMARC and the TEI (detailed in Table 1.5) are examples of this approach to constructing metadata. An alternative approach is to establish the descriptive content and then encode this content.

Applications of SGML such as Hypertext Markup Language (HTML) were developed to structure documents on the World Wide Web by providing labels for the constituent parts of a document such as headings, paragraphs and allows the inclusion of images and sounds and hyperlinks to other documents. As the Web developed HTML has, according to Moody (1998), been forced beyond its original limits by competing software companies. This has led to the development of

---

[13] A set of rules, given in ISO 8879-1986, for designing markup languages that describe a document's structure and aid the interchange of DDOs between data systems.

eXtensible Markup Language (XML) a second generation Web language designed to allow greater manipulation of data content together with the ability to include musical notation, mathematical and chemical symbols in documents.

The development of a standardised metadata format is just one view of how widespread interoperability can be achieved. Proponents of the Warwick Framework (also known as Container Architecture)[14] suggest that it is possible to achieve interoperability between formats within a domain specific area, for example the development of standards for geospatial metadata. Similarly, the matching of elements or attributes within different metadata formats is seen as another means of achieving interoperability. This matching process, referred to as mapping or the creation of crosswalks, allows simple and more complex metadata formats to be linked. All of these methods, although purporting to give some measure of interoperability will be discussed in greater depth in Chapter 5, to examine the merits and challenges each presents to the long-term management of DDOs.

1.14 Summary

As this chapter has shown, the field of digital data management is still in its infancy and many terms have yet to be established. It is evident from the literature that metadata has emerged as a vital agent in the long-term management of DDOs. Similarly, there is a requirement for the development of a conceptual framework that will sustain digital data management over the width of the information community, the justification for this is detailed in Chapters 5 and 8. This chapter has also demonstrated that existing selection criteria do not adequately meet the needs of dynamic and living DDOs. Set against these piecemeal developments, there has been substantial research undertaken in the area of authenticating electronic records, which provides a framework for their storage, but not a means of selecting them for long-term preservation. What is required are selection criteria for the

[14] "At the time of writing the Framework has not been implemented, and it appears that the RDF is taking over the functions of the Warwick Framework." (Taylor, 1999, p. 74).

intellectual content of authenticated records, using metadata. This research aims to achieve this by evaluating the literature from various countries and examining stakeholder views within the United Kingdom, then developing a conceptual framework which caters for DDOs with long-term value irrespective of the purpose for which they were originally created.

Chapter 2 details the methodological approach to fulfil the objectives of this research. The analysis of the findings from interviews with stakeholders in the United Kingdom are detailed in Chapters 3 and 4. Chapter 5 draws on this data to test the hypotheses developed by this research. Chapter 6 explores the development of possible solutions to the research problem using conceptual modelling, whilst Chapters 7 and 8 progressively define the proposed conceptual super-metadata framework in order to support its conceptual feasibility.

Chapter 2

Research Methodology: 'hard' and 'soft' systems approaches

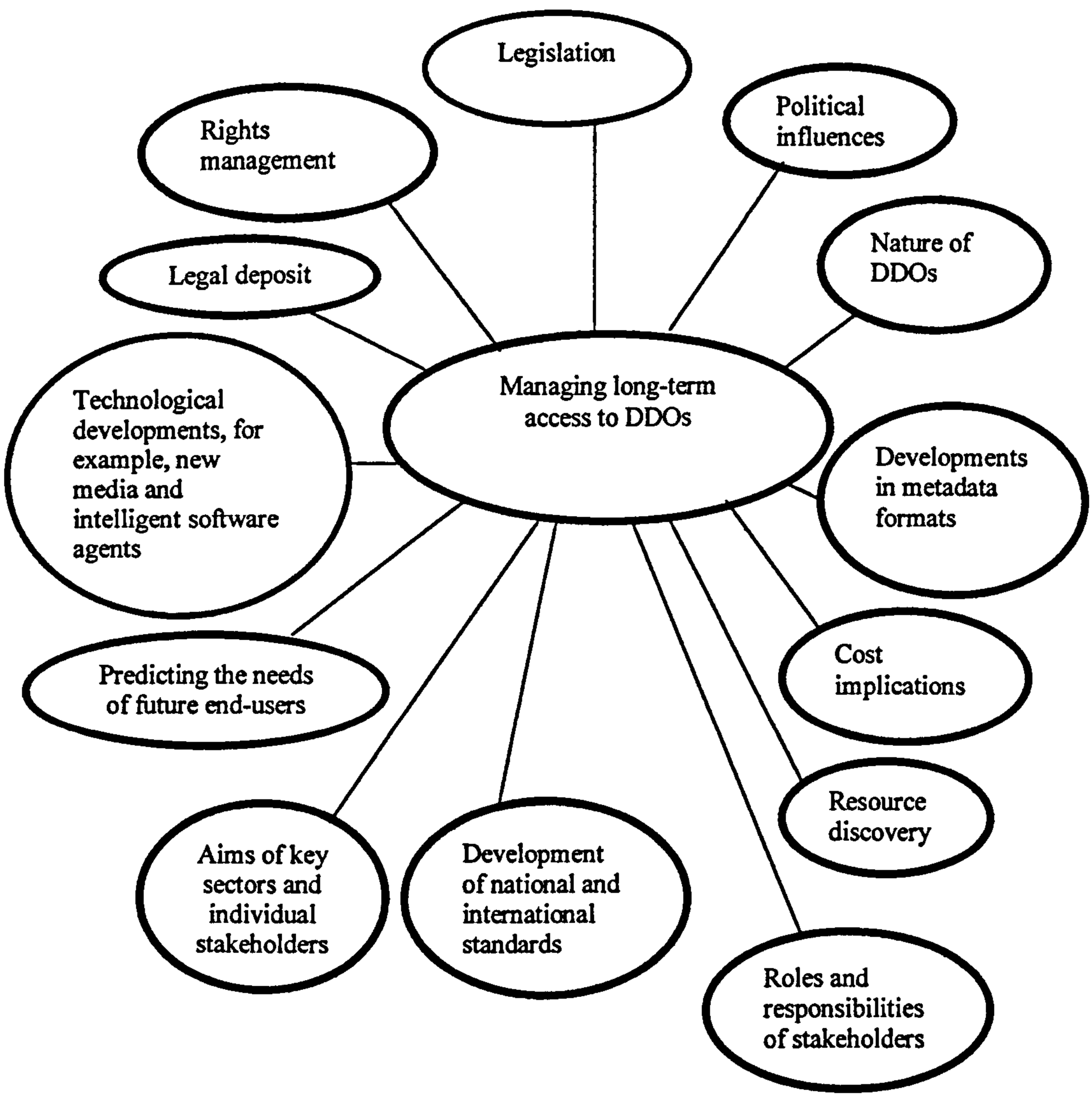
2.1 The nature of the 'research problem'

The literature review suggests that it is not necessarily desirable to preserve every DDO and that there is a need to select DDOs for long-term intellectual preservation. The current climate is characterised by the existence of a number of metadata formats and initiatives attempting to manage DDOs. They focus on particular aspects of digital data preservation, or are tailored to the aims of specific organisations or sectors of the information community. The nature of DDOs (Sections 1.6 to 1.6.11) appears to challenge the ability of society to preserve the digital data on which it is becoming increasingly dependent. This brings into question the adequacy of existing digital data preservation practice to preserve the 'societal memory'. The digital environment would appear to be challenging existing relationships between stakeholders and the ability to manage DDOs.

Although there is a need to preserve DDOs over the long-term, little is known about the interplay between the human centred and technological centred issues that have been identified through a review of the literature (as shown in Figure 2.1). In order to define the 'research problem', further objective research in this field focusing on current policies and procedures and challenges facing stakeholder groups appears to be necessary. As stated in Section 1.1 the in-depth interviews were undertaken in the United Kingdom due to the time and resources at the disposal of the researcher. However, every effort has been made to address global issues by consulting prominent literature from various countries. The research aims to create a conceptual framework that has technical universal applicability which is flexible enough to be adapted to accommodate differences between countries and even between stakeholders within the same country. Similarly it hopes to be adaptable enough to be used within different

legal frameworks. This chapter is divided into three parts concerning the frame of reference, the methods, and the data collection techniques used in this research.

Figure 2.1 Areas of concern identified in the literature



2.2 The frame of reference for the research

2.2.1 Aim

To use metadata as an approach to addressing the issues which information-dependent organisations face as they seek to preserve data in a changing technological environment.

2.2.2 Objectives

1. Examine stakeholders' attitudes towards intellectual preservation of DDOs.

2. Examine the current use (including developments over the next three years) made of DDOs.

3. Explore the reasons why the intellectual preservation of DDOs is undertaken.

4. Examine how intellectual preservation of DDOs is achieved at present.

5. Explore possible ways forward to achieve long-term preservation of DDOs.

2.2.3 Hypotheses

A. The selection of 'dynamic' DDOs for long-term preservation requires the use of similar criteria to those used for 'static' resources.

B. The selection of 'living' DDOs for long-term preservation requires different selection criteria from those used for 'static' resources.

C. It is possible to define universal threshold standards for a framework of metadata for long-term digital data preservation.

D. A universal conceptual framework for the long-term access to DDOs at a level higher than metadata is possible.

E. Criteria for the selection of DDOs for long-term access is a fundamental component of such a universal framework.

2.3 Research methodology

2.3.1 Literature review

As evidenced in the previous Chapter, it has been necessary to examine developments in a number of branches in the literature of the information community in several countries which have been actively engaged in preserving digital data over the long-term and determine the adequacy of existing frameworks and selection mechanisms.

2.3.2 Soft Systems Methodology

This section examines the benefits of using Soft Systems Methodology (SSM) to investigate information science challenges concerning human and technology issues, which at first sight are not clearly defined. The section briefly explores the origins and principles of the methodology and the combining of 'hard' and 'soft' systems approaches.

In trying to establish the current position on digital data management from the stakeholders' perspectives, SSM was viewed as a potential methodology that would allow an in-depth investigation of the human dimension of the problem. Its emphasis on the process of enquiry as a system, runs counter to the hard systems tradition of taking a systematic view of the world (see Jenkins, 1969). Challenges arising in the technology field often demand the creation of clearly defined objectives at the start of a proposed system and are therefore suited to a hard systems approach. Where systems are human-focused this is difficult to achieve (Checkland, 1984). The

complexity of issues facing those responsible for digital data management has been highlighted within the field (Beagrie, 1998; Saunders, 1998). SSM is part of a continuum of soft systems approaches, which can be grouped according to the assumptions each makes about the problem context, such as whether these contexts are simple or complex and what level of consensus is exhibited by the people being consulted (Flood and Jackson, 1991).

2.3.2.1 The principles of Soft Systems Methodology

The methodology, stemming as it does from a dissatisfaction of 'hard' systems to cope with problematic situations, does attempt to develop a means of clarifying such situations.

> Essentially, SSM has been developed for use in ill-structured or messy problem contexts where there is no clear view on what "constitutes the problem", or what action should be taken to overcome the difficulties being experienced. In fact, SSM in action should prevent decision makers from rushing into poorly thought-out solutions based on preconceived ideas about an assumed problem. (Flood and Jackson, 1991, p. 168)

At the core of SSM is the construction of conceptual models that are compared against the real world (Checkland and Scholes, 1990). A brief explanation of the seven stages of the methodology is given below (see also Figure 2.2). It is worth noting that the research process can start at different stages.

- At stage one, there is not a clear definition of the problem situation, but a greater understanding of it is obtained through involving stakeholders in the resolution process.

- All of this information is then brought together to form stage two, a *rich picture*. A "rich picture" being defined as "The expression of a *problem situation* compiled

by an investigator, often by examining elements of the *structure*, elements of *process*, and the situation climate" (Checkland, 1984, p.317). Elements of structure change slowly with time, for example the physical layout of a building, whilst elements of process include rapidly changing organisational activities, for example office procedures.

- Stage three involves the development of the root definition from the rich picture. The root definition being defined as, "A concise, tightly constructed description of a *human activity system* which states what the system is..." (Checkland, 1984, p. 317).

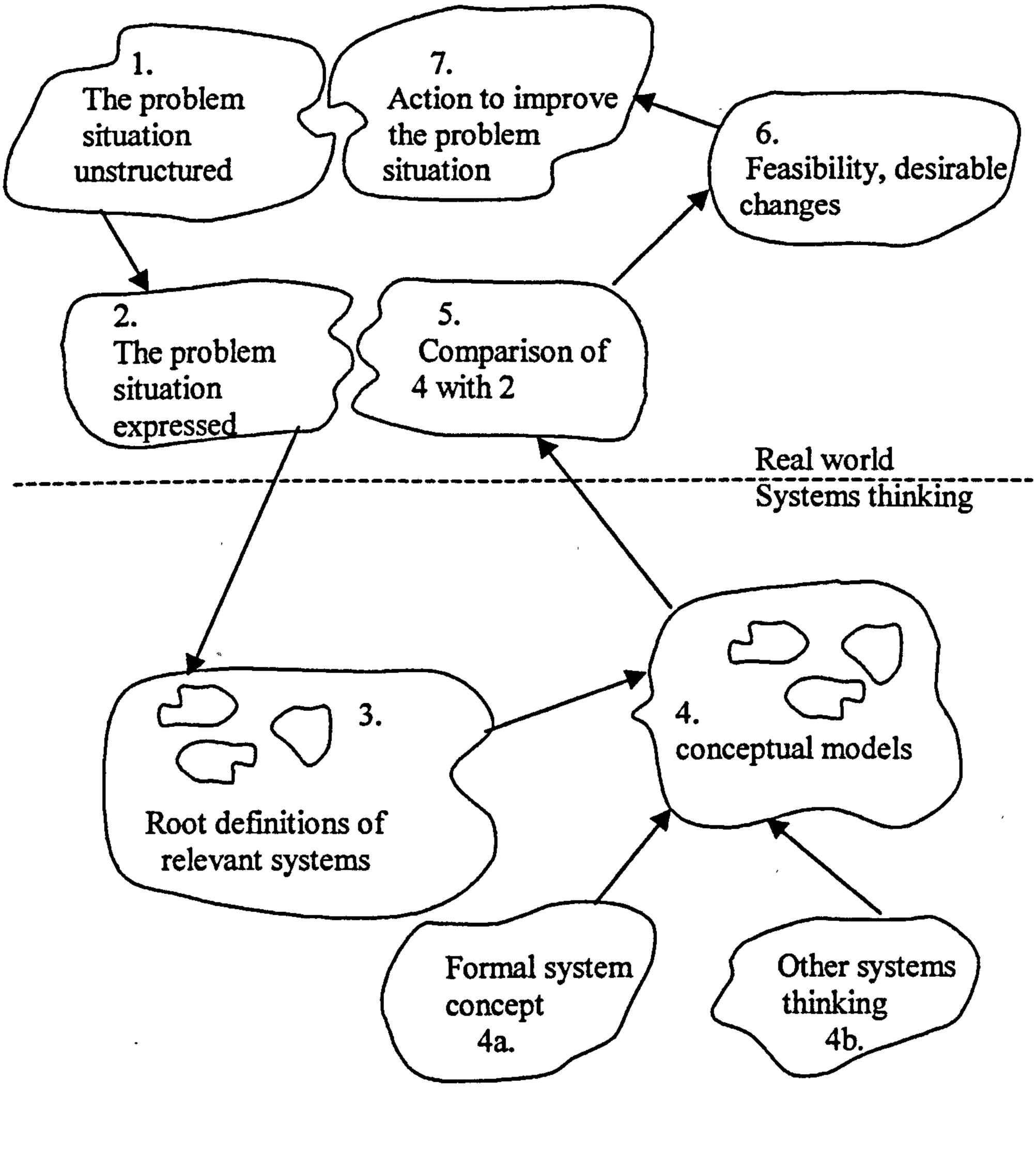
- Stage four is the construction of conceptual model(s) of the problem situation by either (a) determining the minimum set of activities satisfying the root definition or (b) by listing the activities, arranging them in a logical order and drawing a diagram to represent this order.

- The fifth stage concerns the comparison of model(s) against the expression of the problem situation.

- Stage six tests the model against the problem situation, and the possibility of implementing the changes suggested in the model.

- Finally stage seven involves the implementation of the changes to improve the problem situation. As the purpose of this research is to develop a conceptual framework and to test its conceptual feasibility, the practical implementation is outside the scope of this research.

Figure 2.2 Soft Systems Methodology in summary



Source: (Checkland, 1984, p. 163)

2.3.2.2 Benefits and drawbacks of Soft Systems Methodology

The methodology does impose limitations on sample size, through its requirement for an in-depth analysis of the problem situation based on stakeholders' first-hand accounts, a process that can be costly in terms of time and resources. As the problem being investigated involved the collection and analysis of the stakeholders' views and opinions, there was a need to compare the results, where possible, with relevant literature in the field in order to assess the reliability of the findings (Glazier, 1992).

2.3.2.3 The difference between 'hard' and 'soft' systems

Two complementary traditions exist within systems thinking. The 'hard' tradition which takes a systematic view of the world and the 'soft' tradition with its emphasis on the process of enquiry as a system (Checkland, 1984). Both hard and soft system traditions question underlying assumptions, but soft systems research does not define objectives, or ways of achieving and comparing them before the problem has been analysed. Unlike hard systems, SSM allows what Checkland (1984) terms a 'comparison stage', where different possible outcomes can be assessed and it concentrates on the collection and analysis of qualitative rather than quantitative data.

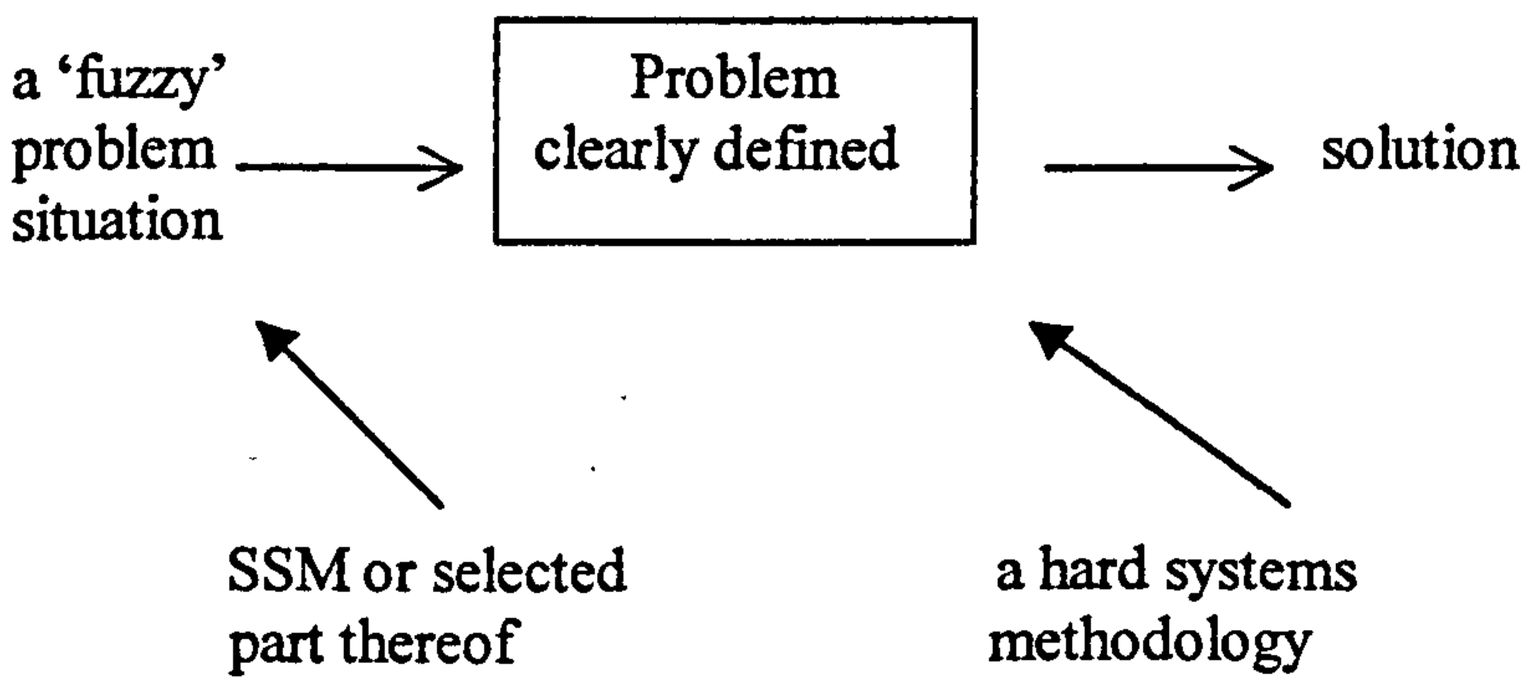
Although SSM belongs to the 'soft' systems tradition, it also makes use of systems models. "It thus subsumes the hard approach, which is a special case of it, one arising when there is local agreement on some system to be engineered" (Checkland and Scholes, 1990, p. 25). Despite this observation by Checkland and Scholes it is not necessarily a simple task to combine 'hard' and 'soft' methodology together in the same research, a point made in the following quotation.

> Unlike the 'hard' systems tradition, the 'soft' systems approach enables progress to be made without the prerequisite of defining the problem in terms of a means-end schema. Therefore to combine both perspectives into some kind of 'hard-soft' methodological framework of practical utility and consistency is certainly not a straightforward matter. (Miles, 1988, p. 55)

However he sees SSM, or a selected systems methodology, as a tool to clarify a 'fuzzy' problem situation. The application of hard systems can help to develop a systematic solution to the initial problem.

Miles (1988) illustrates the process of using 'soft' systems to clarify complex problem situations, and with the aid of 'hard' systems methodology, the development of a systematic solution to the initial problem. This process is as illustrated in Figure 2.3.

Figure 2.3 The 'defuzzification' process



Source: (Miles, 1988, p. 57)

This combination of approaches detailed above was utilised in this research. SSM was used to clarify the situation of stakeholders managing digital data within the United Kingdom. Three hypotheses were generated from the literature, semi-structured interviews conducted and the data analysed, a rich picture constructed and a further two hypotheses generated as a result. To support these hypotheses it was necessary to create conceptual models, and then to develop a potential conceptual framework (hard system) influenced by the SSM. Sections 2.5.3 to 2.5.5 describe this process.

The justification of such a framework is drawn from both the literature examined in Chapter 1, and the analysis of the findings from the interviews with stakeholders (Chapters 3 and 4). At the heart of this framework is the concept that there is a need to develop a system to manage the full diversity of metadata formats, past and present, namely super-metadata (as defined in Section 1.2.6). Such a system would allow for data in legacy systems to be accessed, metadata systems catering for both a broad range and also specific data types to co-exist without creating too rigid a structure, which would stifle the development of future formats. Chapter 7 details the structure of the super-metadata framework.

2.4 Applying the research methodology

2.4.1 Pilot questionnaire

The findings from the literature review underline the need for empirical research into the intellectual preservation of digital data. The initial approach was to construct a pilot postal questionnaire (see Appendix A) which, it was hoped, would provide information on current working practices, in particular the types of DDOs, the use made of them, and the methods of preservation within public and academic libraries within the United Kingdom. This questionnaire was designed to be as clear and concise as possible. Opportunities for respondents to expand answers were built into the questionnaire. To aid the collation of the data gathered, responses to questions, where possible, were pre-coded. To aid respondents, some of the terminology used was defined at the beginning of the questionnaire and to set the research in context the aim was explained in a covering letter.

The chief librarians of ten institutions were contacted by post and asked to complete, evaluate and return the pilot questionnaire. The chief librarians were selected on the basis of the type of organisation they represented, the numbers in brackets show the number of institutions in each category. These categories were; old university (3),

new university (former polytechnic) (2); English shire county (2) and city or London borough (3). The selection was made to deliberately include respondents from all of these groups to ensure the pilot sample was as representative as possible of the nation-wide population of senior managers within academic and public libraries of the United Kingdom. In some cases the task of completing and evaluating the questionnaire was delegated to a senior librarian or senior library manager. The results from the pilot questionnaire are discussed in the following section.

2.4.2 Findings from the pilot questionnaire

Eight out of the ten chief librarians approached completed the pilot questionnaire and the findings from these responses were useful in pointing to areas where improvements could be made to the structure of the questionnaire. These areas included the need to define a greater number of terms to aid understanding. Respondents suggested that this was necessary in an area where terms are not widely established and difficult to comprehend. Comments also highlighted the need to restructure some questions to lay greater emphasis on dynamic resources, and others to enable distinctions to be drawn between the medium on which DDOs are stored and the type of DDOs. This could be achieved by developing a simple matrix.

The answers gained from the pilot questionnaire showed that all eight respondents were concerned with static resources such as computer software and CD-ROMs. The same broad pattern was displayed irrespective of whether a library stocked or owned the resource. This pilot survey conducted in May 1997 shows that organisations had not really developed their collections of DDOs to include 'living' resources (as defined in Section 1.6.5) although there are indications that this situation may be changing rapidly. In addition, the question related to user access showed that some libraries provide access to living resources such as the Internet and databases.

The use of DDOs for income generation, reference work and loans to users, applied to both public and academic libraries. DDOs for teaching and research use were, as

expected, mainly confined to the academic libraries. The answers also revealed that CD-ROMs, computer software, together with electronic books and journals were common to the majority of respondents, but apart from these response options no new types of DDOs were cited.

Difficulties arose in attempting to obtain conclusive answers to the quantities and the time scale over which libraries preserve DDOs. The difficulties related to respondents being unsure of the definition of 'preserve' being used in this research. Collectively the libraries that preserve DDOs did so over a wide subject range and mainly to meet user demand (seven respondents cited user demand as a reason for preserving DDOs).

The pilot questionnaire responses were important in providing some insight into current professional practice. The responses showed that only one respondent from a city library authority did not experience problems when preserving DDOs, despite the fact that five respondents created in-house DDOs, five respondents indicated the absence of a written collection management policy and seven respondents stated that their organisation did not belong to any consortium that shared DDOs.

The use of a postal questionnaire directed at senior library management influenced the data that was obtained in a number of ways. Firstly the results of the pilot questionnaire suggested that targeting senior library management revealed current working practices in the management of digital data over the long-term to be very much in their infancy. Secondly it highlighted the limitations of the methodology. These limitations prevented the interviewer gaining knowledge about organisations through face to face discussion, did not enable respondents to clarify questions which were unclear and did not enable the researcher to uncover the reasons behind respondents views.

The data gathered from the pilot questionnaire did not provide sufficient information on current working practices and pointed to the need to focus more closely on key players in representative organisations, identified through the literature review. In

order to examine a potential diversity of opinions and approaches, it was felt necessary not to confine the research to the traditional library and information sector, but to include some of the major data creators and users in other sectors.

2.4.3 Stakeholder organisations

Senior managers from 12 stakeholder organisations were interviewed and included senior technical people and senior information professionals. These in-depth interviews were conducted between October 1997 and June 1998 and ranged in duration from approximately 60 to 90 minutes. The number of people interviewed at each organisation varied and frequently reflected the internal management within the organisation. For reasons of retaining confidentiality, these people and the organisations they represent can only be identified in general terms.

As the findings of the pilot questionnaire suggests that the experience of managing digital data was not extensive within libraries, it was therefore a conscious decision to focus on a small number of key players who had an interest in developing digital archives and/or had substantial experience in meeting the challenges digital data management presents. This approach allowed the researcher to examine human centred issues such as ownership from the stakeholders' perspective, in addition to gaining facts on the technological issues. The effect of the factors of small sample size, variations in the duration of the interviews, the number of people interviewed at each stakeholder organisation and the focus on the United Kingdom on the results is examined below.

The aim of the research was to gain in-depth qualitative information rather than obtain a statistical sample. Hence, the number of stakeholder organisations consulted was sufficient to gain a range of stakeholder views on their roles and responsibilities in the digital environment. Similarly the duration of the interviews was governed by stakeholders' responses and their areas of expertise.

The stakeholder organisations determined who they elected to provide the corporate view and in all cases these were senior managers. As the corporate view was sought the effect of differences in the number and status of respondents was reduced. Where it was not possible to gain a corporate view the respondents personal view was recorded and in some cases this indicated an absence of organisational practices.

Due to the reasons stated in Section 2.1 the geographic scope was restricted to the United Kingdom, therefore future comparative research will need to be undertaken to verify the applicability of the recommendations of this research to other countries. Factors such as the administrative procedures and the legislative structure operating in the United Kingdom and their influence on the framework need to be considered when translating the research recommendations to a global context.

The interviews were felt to be the most appropriate method for gathering comprehensive data on the attitudes of stakeholders to the long-term intellectual preservation of DDOs. It was hoped that the use of these interviews would enable the researcher to gain insight into the practices and procedures stakeholders adopt, why these are undertaken, their opinions on specific issues and general concerns presented by digital preservation. The questions were used to guide discussion during the interviews (see Appendix B) and some of these were tailored as necessary to be applicable to each stakeholder. The research data from these interviews was analysed to identify variables, determine inter-relationships that existed between these variables, and to test the hypotheses generated from the literature review (see Chapter 5).

2.4.4 Structure of the semi-structured interviews

Specific organisations were identified by means of the literature review which included searching leading journals in addition to the Internet. As the findings of the pilot questionnaire showed, these organisations could not be confined to the traditional

library and information sector, but had to include some of the major data creators and users in other sectors. The key sectors considered were:

- Libraries
- Record office
- Data archives
- Banks and similar organisations
- Science-based industry
- Publishers
- Government agencies.

The choice of institutions was made from as wide a variety of organisations as possible within the United Kingdom which had an established expertise in the long-term preservation of DDOs, or identified as potential stakeholders. The sample encompassed institutions dealing with data in different forms such as numbers, text and images. This deliberate selection of a wide range of institutions was devised to identify as many DDO preservation schemes as possible.

The 12 stakeholder organisations interviewed included: two publishers, three data archives, one record office, two libraries, a representative from science-based industry, a government agency, an advisory body dealing with the preservation of data, and an advisory body concerned with networking issues. The stakeholders were kept anonymous in order for honest corporate views to be given and in some cases personal views. Although banks and similar financial institutions were approached no response was gained from this sector.

At the start of each interview stakeholders were given a brief explanation of the background to the research and presented with a paper copy of Barker's (1997)

typology of electronic books adapted for the purposes of this research (see Table B.1, Appendix B). To guide the discussions the interviewees were asked a series of questions that increased in specificity as the interview progressed. The answers given by respondents to early questions influenced some of the questions asked later in the interview. Due to the diversity of stakeholders interviewed some additional questions were asked that were specific to a particular organisation.

The purpose of these interviews was to collect information on practices and procedures being undertaken by stakeholders and their opinions on why these were undertaken, together with an awareness of challenges and specific issues presented when attempting to manage DDOs over the long-term. The subjects defined in the questions were issues common to information-dependent stakeholders in several countries identified from the literature review and were used as starting points for discussion.

To aid the interviewees, the researcher opted to use an existing term to describe digital objects, namely *electronic records*. The reasons for this decision were twofold, (a) it was intended to use a known term so that the interviewee would be able to relate it to their organisation and (b) to test the need for the new term Digital Data Object, as defined in Section 1.2.3.

2.5 Applying Soft Systems Methodology

This research combines the analysis of secondary source material from a literature review with empirical data obtained by conducting interviews. SSM was seen as a particularly appropriate methodology because its focus on a complex yet ill defined problem in which there is a large human component, as the following quotation demonstrates.

> SSM is best suited to situations where there is a coalition of organisational stakeholders and the need to create, temporarily at least, a shared appreciation among these stakeholders of what is the best way forward from a given problem situation. It is extremely adept at

> providing creative solutions that enable organisational actors to escape the "traps" into which their current thinking has led them. (Flood and Jackson, 1991, p. 190)

The analysis of this data was used to inform the construction of a framework for the long-term management of digital data, based on hard systems modelling. A 'soft' systems approach thus informs the proposed 'hard' systems solution.

2.5.1 Techniques for analysing the interview data

Although some of the results could be tabulated, some of the data was unique to individual stakeholders; therefore, it was considered more productive to determine interconnections between stakeholders by creating a rich picture. SSM can help to pinpoint areas of conflict or consensus between stakeholders that require resolution. Such areas may not have been highlighted by using a purely hard systems approach and, unresolved, could lead to the potential failure of a newly developed system.

2.5.2 Rich picture and root definitions

The drawing of a rich picture establishes what structures exist and the interconnections and relationships that exist between various stakeholders. The data for the rich picture was obtained from the semi-structured interviews, which, where possible, have been tabulated to draw out similarities and anomalies. The extraction of critical trends from this tabulated information, together with information unique to individual stakeholders, was combined and placed in a succinct form in the rich picture, which is discussed in more detail in Chapter 4. This picture enabled the researcher to form a composite impression of the research problem from stakeholder opinions expressed at a particular point in time and significantly for this research the stakeholder interactions. From this, root definitions were formulated to clarify the 'research problem' by stating what is to be achieved, how to achieve it and why this is necessary. The rich picture and the root definition were shown to selected stakeholders as detailed in Section 2.5.5.

2.5.3 Conceptual modelling

The development of possible solutions involved the use of conceptual modelling, the results of which are presented in Chapter 6. This modelling provided a method for identifying issues that have a major impact on digital data preservation.

> The basic shape of the approach is to formulate some models which it is hoped will be relevant to the real-world situation, and use them by setting them against perceptions of the real world in a process of comparison. (Checkland and Scholes, 1990, p. 6)

The process enabled an exploration of alternative potential solutions to the research problem.

2.5.4 Hard systems modelling

As stated earlier in this chapter, hard and soft systems are underpinned by different philosophies of what constitutes the 'system' (see Section 2.3.2.3). The results of the SSM approach provided vital information on the nature of relationships between stakeholders, identification of key issues to be addressed, and the analysis of 'grey areas'. Although the two approaches are underpinned by different philosophies, the combination of the two, in this research, combined the flexibility of SSM enquiry with the translation of the findings into a tangible system using hard systems modelling (see Chapter 7). Such hard systems modelling has been said to be a special case of SSM that can be pursued if sufficient consensus exists between stakeholders on a system. This is evidenced by the need for guidance and expression of challenges faced when trying to preserve DDOs over the long-term, expressed by many stakeholders (see Chapter 3).

2.5.5 Testing the conceptual feasibility of the proposed framework

In order to assess the applicability of any system(s) created by this research it was decided to test the applicability of the framework against the published literature and to consult selected stakeholders (see Section 2.5.6). Having constructed a conceptual model it was felt necessary to demonstrate that the framework could be managed in such a way as to keep the costs within reasonable limits (see Sections 8.2 and 8.3). This aided the researcher to establish some of the critical parameters affecting the functioning of the system as a whole, and to discover some of the steps necessary to ensure a system could function effectively. This step of testing the conceptual feasibility aimed to draw out the interrelationships from a cost perspective and to compare these findings with those of the conceptual modelling (see Section 8.3).

2.5.6 Evaluation of the proposed framework by selected stakeholders

To test the conceptual framework against the problem situation the proposals were taken back to selected stakeholders for their evaluation. This was achieved by selecting three of the original stakeholders. The same people were interviewed to maintain comparability of views between the first and second round of interviews which took place between March and April 2000. The stakeholder organisations were selected so that they included one representative from the following three groups, a public sector organisation (a national library), a commercial organisation (a journal publisher) and a public sector organisation concerned with higher education (a national archive).

Prior to the interviews each stakeholder received the background information detailed in Appendix E, together with a covering letter explaining the main areas the interview was expected to cover. These interviews aimed to generate in-depth discussion of the evaluation of the framework and cost model assumptions, therefore the structure of the interviews was kept to a minimum. They varied in duration between 45 and 55 minutes.

These in-depth interviews focused on the following areas:

- The rich picture and root definition.

- How far the super-metadata framework meets the needs of a particular stakeholder's organisation, including the feasibility and desirability of the framework.

- The suitability of the recommendations to aid the implementation of the framework.

- Parameters within which costs and standard assumptions are considered to be realistic.

- The degree to which the costs are inclusive.

At the beginning of each interview respondents were shown the rich picture (Figure 4.1) and given a verbal explanation of what the rich picture aimed to show, particularly for the specific category to which they had been assigned. Respondents from each stakeholder were then asked to evaluate how accurately the picture represented their position, bearing in mind that this position could have changed since the initial interviews were conducted (1997 to 1998) and asked if, from their organisation's perspective, it was an accurate reflection of the 'problem situation'. Similarly they were shown the root definition and asked if, from the point of view of their organisation, it was an accurate description of the 'problem situation'. The main discussions focused on how far the framework met the stakeholder organisation's needs, together with their evaluation of how realistic and inclusive the cost model assumptions were.

2.6 Summary

This chapter has detailed the frame of reference, methods and data collection techniques used. Chapter 3 analyses the results of the 1997 to 1998 interviews to identify trends and anomalies and Chapter 4 takes this a stage further by constructing and analysing a rich picture. Chapter 5 tests and supports the hypotheses developed from the literature review. However, supporting hypothesis C led to the generation of two further hypotheses D and E (see Section 2.2.3), the support for which necessitated the conceptual modelling of the 'research problem' (Chapter 6), creation of a framework (Chapter 7), a cost model, an evaluation of these recommendations by selected stakeholders and then setting the research in a wider context (Chapter 8).

Chapter 3

Analysis of findings from the stakeholder interviews

3.1 Introduction

The interviews held with stakeholders covered a wide range of issues that have a bearing on the management of DDOs including the diversity of DDOs, selection policy, metadata issues, authentication, integrity and access. It was necessary to explore these diverse issues to evaluate their relative importance and interaction with each other in order to determine which are critical to the long-term management of DDOs. The interview questions are given in Appendix B.

Not all questions were applicable to all twelve stakeholders, since for example, the two advisory bodies do not archive DDOs. The tables shown in this chapter display the old term 'electronic record' or 'data set' rather than DDO, since these were the terms used when conducting the semi-structured interviews with stakeholders. This was for the reasons given in Section 2.4.4. However, for the reasons stated below the new term Digital Data Object was developed (as defined in Section 1.2.3) and is used in the discussion of the tabulated results.

3.2 Nature of electronic records

3.2.1 Definitions and terminology

In order to create a universal framework there is a need for a universal set of definitions concerning the material that is being preserved. The findings that follow show that a range of terms are used.

3.2.1.1 Electronic records and DDOs

The findings show (Tables 3.1A and 3.1B, row A) that some stakeholders incorporate digitized resources or digital surrogates within their definition of an electronic record. Variations between definitions do exist between organisations and even between different groups within an organisation as in the case of one government agency:

> ...electronic record for us, is a term that means different things to different people... the precise definition of what an electronic record is changes as to who you are talking to and what data they happen to be most familiar with. (A government agency)

A national archive suggested that the essence of a record was defined by the presence of content, contextual and structural information. Moreover, it would only be classed as a record if it had metadata detailing the place of a document in the wider picture, together with how it was created, used and archived (see Table 3.1A, cell 6A). In addition a record should be non-amendable at the point it is finalised. The issues of metadata and authentication will be explored in greater detail in Sections 3.6.1 and 3.7.

Other organisations within the archival and librarianship fields articulate much broader definitions, for example a national library stated that for them an electronic record was "...any set of structured data meaningful to a particular user group". A respondent from a national archive remarked that in their personal opinion the term *electronic record* has a specific meaning within the archival community and would confuse the issue if the term was used to refer to published material. These results concur with the findings of Kandur (1992) who identified differences in the application of terms including 'record', and emphasising that terminology was an important issue for archivist and record managers to address[1].

This comment together with the diversity of definitions given by stakeholders underlines the need for a new term to be established, a common term that

[1] The need to re-examine existing definitions of a 'record' in the digital environment is documented by Erlandsson (1997).

encompasses the essential characteristics of digital data, namely a DDO (as defined in Section 1.2.3).

A stakeholder within science-based industry defined a record as, a "...collection of files which meets some specific business need..." and also stated that they were concerned with data rather than records. Data being defined as, "...any information which is created or originated electronically". It would appear that archives lay most emphasis on the scholarly value of an item followed by the available documentation and the format of an item. This raises the issue of who decides what is 'scholarly'. One respondent from the archival community felt assessment of scholarly content was best left to experts in the particular field, stating that an archive would be on 'dangerous ground' if it attempted to assess whether the intellectual content of a DDO was 'scholarly' or not. Instead the specialist archive concentrates on pragmatic criteria, such as the value of a DDO for secondary use, a view which could reflect an appreciation of the importance of selection and an awareness of resource limitations.

It would appear that the role of an organisation aiming to manage the long-term retention of DDOs which it has not created itself, would entail assessing the accuracy and completeness of the metadata and ensuring it faithfully mirrors the DDO rather than attempting to assess the DDO itself.

3.2.1.2 The longevity of DDOs

Responses to the permanence of an organisation's records demonstrated stakeholders' concerns about how to develop criteria to select DDOs for long-term retention. The appropriateness of the term 'permanent' was also challenged by one stakeholder whose personal opinion was that it should be replaced with a less emotive phrase, 'long-term' (defined as a life span of 20 to 50 years).

Table 3.1A Analysis of the nature of electronic records/ data sets	**1** Publisher	**2** Specialist archive (Respondents: 1 & 4)*	**3** Company from science-based industry	**4** Distributed data service	**5** Advisory body concerned with networking issues (Respondents: 1 & 2)†	**6** A national archive
A What is an electronic record?	Piece of information in digital form.	Resp. 1. Anything that exists/created in digital form readable by a machine currently call a computer.	Deal with data not documents. Data = any information created or originated electronically. Record = collection of files which meets some specific business need.	Main criteria = scholarly -documentation available -format (huge range of data types).	Resp. 1. (Personal view) something specified by archival community good definitions of this e.g. projects of Universities of Pittsburgh & British Columbia.	Includes, content, contextual and structural information. Only a record if it has metadata that tells you the place of document, how used, created/archived.
B Does the organisation make a distinction between 'dynamic' and 'living' electronic records/data sets?	Have 'dynamic' recently 'living' Not thought about preservation of 'living' parts at present.	Resp. 1. A lot of material falls into dynamic category. 'Living' material is constantly added to, but the original material is not changed.	No.	Might be in certain circumstances.	Resp. 1. No. Deal mostly with 'living' electronic resources – Web material.	None living.
C Preserve electronic records/data sets?	Careful to preserve published material and internal material.	Yes.	Yes.	Yes.	N/A	Yes.
D Reason for preserving electronic records/data sets over the long-term?	• Financial • Provide a history of the organisation. • Re-use. • Commercial	Resp. 1. Technical obsolescence – preserved to ensure availability over long-term.	See responses 3F, G & H this table.	• Scholarly value • Re-use	N/A	Driven by government departments bringing in systems for efficiency. Non-amendable records more widely available.
E Legal requirement to preserve electronic records/data sets?	Yes. Accounts records, board minutes.	Resp. 1. No not a legal requirement to preserve anything in way there is to deposit paper books.	Accounts, regulated documents.	No.	Resp. 1. No.	Government departments and archive responsible for records in any media
F Administrative requirements to preserve electronic records/data sets?	Yes. strong requirement to keep internal correspondence: a. never know when need again b. if staff change better records more chance of following through.	Resp. 1. No, but – funding bodies recommend /require grant holders to offer materials for deposit. Not everything created electronically is worth preserving. Resp. 4. Preserve to prevent loss rather than retention for a specified time.	Number of drivers influencing electronic archive project e.g. legal protection, re-use of information in order to mine the data.	No.	Resp. 1. Only in general archiving sense, not necessarily for long-term. May become an issue in the future. At moment everything on the Web server dynamically updated old versions sometimes kept - no policy.	Re-use by government departments. Targets to put out finding aids plus indexes on Internet.

*** In this case respondents are differentiated as they were interviewed separately (this applies to all tables in this chapter).**
† In this case respondents are differentiated as they were interviewed separately. Where no respondent number is shown respondent number one applies (this applies to all tables in this chapter).

Table 3.1A Analysis of the nature of electronic records/ data sets cont.	**1** Publisher	**2** Specialist archive (Respondents: 1 & 4)	**3** Company from Science based industry	**4** Distributed data service	**5** Advisory body concerned with networking issues (Respondents: 1 & 2)	**6** A national archive
G Are electronic records/data sets of permanent importance and if so why and to whom?	No overall organisational strategy, hope keeping things of permanent interest.	Resp. 1. Yes. All material = available in paper form somewhere having it in electronic form a. increases numbers who have access to it b. can manipulate material in a way couldn't otherwise. Resp. 4. Yes, methodology use aims to keep permanently & securely for a variety of users that require data.	Some of data regulated = kept for defined periods could be very long time period. Why – regulated Important for commercial reasons, failure to find information could lead to delays = cost money, bad or corrupt information = individual liability.	Yes. Why = of scholarly value and expected to remain so to scholars. Trying to predict future is a problem = requirements change.	Resp. 1. Depends on definition of permanent – certain things of interest to information field e.g. some projects could have long-term importance but is not preserved elsewhere. Personal definition of permanent = emotive term would use long-term in preference & take this to mean 20-50 years.	Records taken into archive essentially kept permanently – could be a review of an area of records and decided that they are not kept – in a way saying that original decision was wrong. Reason for keeping perm. Show policy and actions of government and accountability reasons + provide information for researchers plus general public.
H Are electronic records/data sets short-term in nature and if so why?	See answer to cell 1D.	N/A	Depends on the sort of information. Sensitive, regulated information kept the longest.	Yes unless ensure otherwise. Short-term life expectancy of media and access compared to print.	Resp. 1. Some are Reason: difficult to define. Web pages few of permanent importance. Resp. 2. (Personal view) assess on record by record basis.	No long-term as legislation requires them to be. Records created electronically kept in electronic form for storage, accessibility and use of people's time
I Is the Internet short-term in nature & if so why?	See answer to cell 1B.	N/A	Something like the Internet will be around for a long time. Why = another way to present information	As pres. medium would be 'living' record – a lot of material on Internet short-term & difficult to fix.	Resp. 1. Yes as mechanism for long-term storage & delivery not worked out yet = Net designed for publishing not pres. Web = short-term information should not be. Resp. 2. Personal view – No. Internet may change networking become more pervasive everything become informationalized	Short-term through lack of security and integrity. Akin to published information. Organisation interested in the electronic records behind the Internet.

Table 3.1A Analysis of the nature of electronic records/ data sets cont.	**1** Publisher	**2** Specialist archive (Respondents: 1 & 4)	**3** Company from Science based industry	**4** Distributed data service	**5** Advisory body concerned with networking issues (Respondents: 1 & 2)	**6** A national archive
J Preserve 'dynamic' electronic records/data sets?	N/A	N/A	CD-ROMs + produce training materials. These stay static until revised. Fuzzy boundary between 'dynamic' and 'living'.	Yes.	N/A	Taxation records continually added to, archive takes snapshots & audit trails. Having brought it to archive = non-amendable and authentic. Cannot continue to be dynamic if to interrogate a set of electronic records.
K Preserve 'living' electronic records/data sets?	N/A	N/A	Yes. Research and development. Intranet and Lotus Notes database 'Living' different set of issue have to face as changing all time, taking management decisions on them need audit trials to prove decision making process – same is true of information on Web. Regulatory Document Management System records all data that goes to regulators in document form – like an archive as information never deleted except under controlled circumstances.	Often host and provide access to 'living' records – provide snapshot which is regularly updated. Frequently organisation which invested in records also investing in its preservation	N/A	No.

Table 3.1B Analysis of the nature of electronic records/ data sets	**7** Government Agency	**8** Advisory body for data preservation	**9** A national library	**10** Spokesperson from HE sector (personal view)	**11** Journal publisher	**12** Specialist archive
A What is an electronic record?	Precise definition changes depending on who talking to in the organisation and data most familiar with – Personal view = anything that is dynamic or living as long as it is not on anything static or fixed.	Material produced electronically originally and analogue to digital, so a surrogate medium as well.	Any set of structured data meaningful to a particular group.	Electronic + paper equivalent and those only in electronic form.	Data stored electronically not necessarily only electronically.	Not like conventional library with catalogue records & resources – rely on electronic metadata attached to item (stored with item and is electronic title page)
B Does the organisation make a distinction between 'dynamic' and 'living' electronic records/data sets?	Personal view = distinction between dynamic & living does not happen, distinction between dynamic and static (paper) sort of happens	N/A	No. Some people in organisation may not be aware of distinction, certain amount of education needs to be done.	No. Only just becoming aware of problems of digital archiving in general.	Distinction between CDs and online databases.	At moment not much dynamic data. In theory way preserve should not make any difference. What is contained in file of less importance than metadata.
C Preserve electronic records/data sets?	Yes and no. Never any formal guidelines for archiving & preservation = preservation patchy. Where has existed deposited with a specialist archive Some can't be deposit due to poor documentation	No not collection based. Help people with collections to think about implications of digital media.	Preserve some e.g. traditional key areas such as catalogue records. Other areas not as advanced Areas where electronic function succeeded traditional paper-based function – policy not caught up with practice.	Academic archive - 5 years ago started keeping electronic copies of printed course material . Administrative archive – paper based.	Yes. E-mail – in process of expanding into true electronic filing system plus variety of databases.	N/A
D Reason for preserving electronic records/data sets over the long-term?	Supply research community with data. Reasons 1. exists 2. someone requested it 3. remembered to do it 4. stable enough to be sent outside. 5. Chance card element.	N/A	Definitive bibliography of British publishing Decision made to rely on electronic catalogue record as definitive record.	Main reason = ability to re-use material Drivers: -access to it -demand for material from outside organisation	N/A	In past = act of good will. Now parent organisation mission statement Some material deposited with archive not been used but kept as may be in future.
E Legal requirement to preserve electronic records/data sets?	No. – exception is registration. Secondary obligation to comply with a national archive.	N/A	No. One reason interested in extending legal deposit.	No. Television programmes. created with third parties cannot keep for copyright reasons. Ways manage internal use be addressed when create digital archive.	Yes, personnel records people have to be advised on what held.	N/A

Table 3.1B Analysis of the nature of electronic records/ data sets cont.	**7** Government Agency	**8** Advisory body for data preservation	**9** A national library	**10** Spokesperson from HE sector (personal view)	**11** Journal publisher	**12** Specialist archive
F Administrative requirements to preserve electronic records/data sets?	Present some recs. as time series in this respect an admin requirement otherwise figures don't mean anything.	N/A	Must preserve electronic cataloguing records and electronic journals. Even if legal deposit does not develop as acquire material through purchase need emphasis on preservation material. Administrative obligation to keep financial data.	Have to keep administrative archive but whether legal requirement don't know.	Yes, -easier to search, -takes up less space -commercial reasons.	No. Idea that once material is deposited with one of service providers will be catalogue & documented & displayed on service wide catalogue.
G Are electronic records/data sets of permanent importance and if so why and to whom?	Permanent preservation limited to books at moment Permanent importance = subjective thing.- how decide what is important for selection. Everything organisation does goes forward for national importance.	N/A	Electronic catalogue records – next generation OPAC = First line of access to collections. Responding to duty of maintaining some kind of record of UK publishing output as move from print to electronic.	Of permanent importance - within academic archive = of historical interest.	Journals stored electronically = permanent. Access to these increasingly made available to third parties for document delivery and subscriptions.	Yes as educational resource of value & continue to be. Don't try to second guess what will be important in future. Same text can be used by different users.
H Are electronic records/data sets short-term in nature & if so why?	No, without computers would cease to function as viable unit delivering at present pace.	Some likely to be, others intellectual content great & want them accessible over time. Problem = long-term value but created to meet short-term ends.	Some are others are not. Focus has to be on selection for retention at earlier stage – acquisition. E-mail no formal policy cheaper in short-term to archive everything.	Very difficult with electronic records, new formats, new standards 5 years on. Policy changes.	No. Need long-term methodology in place to prevent data getting corrupted through use and time.	In sense that don't create a separate records from the data.
I Is the Internet short-term in nature and if so why?	Important thing is the data itself not how delivered. – Looking at Internet as a way of keeping delivery costs down.	Have to be selective. Some sites that are of permanent use and value.	Depends what mean by the Internet the information or the functionality? Most of material short-term. Library strong tradition of keeping ephemera.	Internet itself may not live for another 10-15 years but the idea of a global network that remains to be seen.	Quite the opposite – the growth in numbers proves this point. Excellent medium for access and dissemination.	Do not think Internet itself short-term, but way people use it at the moment is going to change.
J Preserve 'dynamic' electronic records/data sets?	Yes, give copy to British Library & similar orgs. 1 of areas difficult as rules have not been updated as technology has developed Like to lodge all d. products this way.	N/A	CD-ROMs have seen as immediate resource. Current journals in both electronic & paper. Finding out operational & financial issues of managing and making available.	No overall policy for multimedia, working on one.	Yes.	Have couple of texts deposited on CD-ROMs – don't do anything other than copy CD-ROMs. In past haven't deposited that kind of material now lots interest in this.
K Preserve 'living' electronic records/data sets?	No. involved in project to preserve living. Not unwillingness to, no means to do it.	N/A	Databases living in one direction –expand not deleted No policy for Internet resources.	Not systematically. Depends on person maintaining server. No policy for Web pages.	Yes. Databases + use third parties to provide material.	Only to extent people re-edit work & re-deposit. (mechanism in metadata of records different versions).

With regard to the longevity of DDOs there was a spectrum of opinion and this is governed by organisational aims together with the nature of the intellectual content of the record. There is a consensus amongst stakeholders that the content of at least some DDOs is of long-term value. Three of this same group stressed that to ensure the long-term existence of these records active intervention is necessary. The view that a DDO is the definitive version of a record is gathering momentum as evidenced by the following statement:

> ...both the Public Record Act and the forthcoming Freedom of Information Act require the records to be long-term...archives in the United States have to specifically maintain the electronic form, because it is the actual record that was created and used, therefore it is the genuine record, and it is not good enough to put it into a hard copy format and destroy the electronic record. (A national archive)

3.2.1.3 Longevity of the Internet

The Internet in its present form is regarded by many stakeholders as short-term, although the concept of a global network is seen as having greater longevity (see Tables 3.1A and 3.1B, row I). Some stakeholders attributed the lack of security as a problem that needs to be resolved. A spokesperson from an advisory body concerned with networking issues identified that the Internet was designed with publication rather than preservation in mind. Within this Web publishing environment a specialist archive identified, "...a tension between academics wanting to own material and being seen to own it and distribute it themselves and giving it to a place like us" (A specialist archive).

3.2.2 The range of DDO types

A table showing the categorisation of electronic books based on the work of Barker (1997) (see Section 1.6.5) was used to generate discussion, and to aid understanding of how the terms 'static', 'dynamic' and 'living' relate to DDOs. Several respondents commented that the term *electronic record* was imprecisely utilised and one respondent felt it should be substituted with the term *electronic resource*. One stakeholder from a national archive saw disadvantages of the table (See Table B.1, Appendix B) expressing the view that it only catered for published

material and was too library based. This stakeholder also stated that in his organisation there was a very precise meaning of what constituted an electronic record (see Table 3.1A, cell 6A) and the organisation dealt mainly with original material.

Table B.1 did prove to be beneficial in allowing respondents to identify whether any of their organisation's resources correspond to the categories given in the constructed table. Firstly, it identified that the majority of stakeholders' resources were confined to the static and dynamic categories, with a minority (four out of the ten who were collection based) concerned with preserving living DDOs[2]. Secondly, the boundaries between these categories were not as clear-cut as the initial table suggested (as identified in Table 1.3). The static category could include objects that have existed in traditional paper based form and have been digitized or resources created purely in digital form[3]. By analysing the results from the interviews it is possible to deepen our understanding of the relationship between the categories of DDOs (see Table 3.2). Dynamic resources acquired or produced by organisations can be sub-divided into two, (a) DDOs created and designed to be dynamic entities and (b) living DDOs converted to dynamic objects to enable long-term management to be undertaken.

---

[2] This number does not include stakeholders who are creating living DDOs but who do not at present have a preservation policy for this category (see Table 3.1A & 3.1B, row K).
[3] These have been referred to as resources that have been 'born' digital (see Hedstrom, 1998).

Table 3.2 Adaptation of Barker's table of book typology to more closely represent the electronic environment

Static DDOs	Dynamic DDOs	Living DDOs
Digitized resources OR Resources created solely in digital form but which closely resemble traditional paper-based resources.	Created in this form, for example CD-ROMs with interactive capabilities.	Born digital resources in their 'active' phase, for example Lotus Notes databases, or the United Kingdom Post Office database of postcodes. OR Working drafts of electronic files, for example spreadsheets. The archived form of living DDOs become static or dynamic when they are deposited in an archive to ensure their authenticity. How a DDO is likely to be used influences whether it is maintained as a static or dynamic object assuming the costs attached to its maintenance are sustainable.

Legend:

Table 3.2 allows for the possibility that DDOs can exist in more than one category. Similarly, depending on the way DDOs are used they can span the division between the living and dynamic categories, for example CD-ROM encyclopaedias either with or without Internet links respectively (as detailed in Section 7.2.6). All DDOs are living as they are created, but the majority are active only for a limited amount of time and then become static or dynamic.

All of the respondents (except for the two advisory bodies) are responsible for preserving static and dynamic DDOs. Living resources are held by six stakeholders and are identified as the most difficult to manage. These factors, combined with their relative scarcity has resulted in virtually no preservation policies existing for these living DDOs, except within the archival community

where some organisations are attempting to develop policies. This trend is not confined to the United Kingdom as evidenced by the initiatives outlined in Sections 1.8.1 to 1.8.9. A common view expressed by stakeholders from archives is the requirement for the substantive content of DDOs to be non-amendable. The interview data revealed that within the living category there are different types of living DDOs as detailed below.

Type 1

Living DDOs where the substantive content is added to but not deleted over time. These additions can occur periodically or continuously. An example of this type of living DDO is the Retail Price Indices produced by the ONS on a monthly basis[4].

Type 2

Living DDOs where the substantive content is overwritten. As with type 1 this overwriting can occur periodically or continuously. An example of this type is the Web page of NASA's Global Hydrology and Climate Centre[5].

It would appear that different strategies are required to ensure version control of each of the above types. Type one may require preserving at a less frequent interval than type two living DDOs[6].

Stakeholder attitudes to the archiving of living DDOs shows a concern, particularly within the archival profession, trying to ensure the authenticity of the object being preserved[7]. The commonly held view amongst archivists was that to vouch for the authenticity of a DDO the substantive content must be made non-amendable[8]. This

---

[4] URL: http://dawww.essex.ac.uk/online/ons/titles.html
[5] URL: http://www.ghcc.msfc.nasa.gov/GOES/
[6] These differences, which only apply to living DDOs, need to be considered as part of the selection criteria, an issue that is discussed in Section 5.3.
[7] One specialist archive stated that they only have living DDOs in the sense that some material is constantly being added to, but not changed (see Table 3.1A, cell 2B) a view confirmed by another respondent from the same organisation (see Table 3.8A, cell 2D).
[8] The importance of preserving the integrity of digital data material is supported by the work of Duranti (1997), see Section 1.8.1, and the development of the evidential based approach to managing authenticity (Bearman, 1994), see Section 1.8.2.

would involve changing living DDOs to either a static or dynamic DDO. The way DDOs are used, that is how much an end-user can select what is viewed, could influence whether the repository converts the DDO to either a static or dynamic object[9].

When stakeholder organisations were asked whether they make a distinction between the way they preserve 'dynamic' and 'living' electronic records, the answers again showed a diversity of attitudes to the nature of the electronic records that their organisation held. The answers given by some stakeholders suggest a recognition of issues to be addressed. Some respondents expressed a personal rather than corporate view to this question which revealed varying degrees of provision to cater for different types of DDOs (see Table 3.1B, cells 7B and 10B). A distributed data service indicated that there might be a distinction between the two in certain circumstances. At the other end of the spectrum a national archive does not hold any 'living' electronic records. A record, for this stakeholder, must be held in a non-amendable form, or as in the case of a specialist archive most material that is living is added to but not deleted. Three stakeholders identified reasons why the distinction between the dynamic and living DDOs has not occurred in their organisation to date. The first, a spokesperson from the higher education sector, when stating a personal view, highlighted the fact that his/her organisation is only just becoming aware of the problems of digital archiving in general. Similarly, the second stakeholder, a national library, is aware that some people in the organisation are not aware of the distinction between categories and that there is a certain amount of education required to address this problem. The third, a specialist archive, expressed the view that although their holdings did not contain much dynamic or living data in theory it should not make any difference to the way they preserve material. Few stakeholders are concerned with creating living DDOs or are actively pursuing strategies to facilitate the retention of these within their own organisations.

---

[9] There is seen to be a need to find a balance between establishing and maintaining the authenticity of a DDO and retaining the functionality of a living DDO. An issue that is examined where it impacts on the development of selection criteria (see Sections 5.2 and 5.3) and on the development of mechanisms to ensure authenticity (see Section 7.2.5).

From the archival community it is becoming clear that to ensure indefinite future access to these resources there is a need to retain them by converting them to dynamic or static entities, but policies to accommodate these resources have yet to be formulated.

To summarise, the following points have been drawn from the data.

- A range of DDOs are being preserved that broadly correspond to Barker's categories. However, some stakeholders require that the DDOs they acquire are retained in a non-amendable form. For living DDOs this would entail fixing their substantive information content so that they become static or dynamic.

- The stakeholders do not necessarily distinguish between the types of DDOs with respect to the preservation strategy they use.

- There is a recognition, by stakeholders creating and/or retaining living DDOs, that this category presents the most challenges.

- There appears to be a lack of a co-ordinated policy for the management of a range of DDOs and a need for guidance particularly for dynamic and living DDOs.

3.3 Challenges identified by stakeholders to the long-term management of DDOs

The following findings relate to Tables 3.3A and 3.3B. The findings point to a concern for the longevity, that is the life span of DDOs, and a major concern of the majority of stakeholders would appear to centre on physical preservation issues. This is supported by the preliminary findings of the Research Libraries Group survey (Hedstrom and Montgomery, 1998) of the needs and requirements of member institutions in the United States and the United Kingdom conducted in

1998. The survey indicated that the respondents saw physical condition as the largest threat to preservation followed by insufficient planning, technological obsolescence and insufficient resources.

The findings of the survey detailed above are mirrored by the empirical research detailed in this thesis. In addition this empirical research has highlighted the fact that many organisations possess staff who have expertise in particular subject disciplines and records management/archival training, but who would need to acquire additional skills to meet the new challenges presented by DDOs. Therefore, there is seen to be a need for the involvement of senior management at least in the initial stages of developing a framework for the selection of DDOs. The selection processes operating at present in some organisations would appear to cater for dynamic DDOs, whilst others only cater for static DDOs.

This lack of development is possibly due to the fact that many stakeholders are only just starting to create or receive dynamic and/or living DDOs from other organisations. In addition, the quantities being received at present are low and the issues they present unfamiliar to some stakeholders.

Dynamic/living DDOs were highlighted as records which cause problems when selecting them for long-term preservation, the former because of trying to maintain the 'look and feel' of the application whilst migrating it to new software platforms, and the latter because of the difficulty in determining a policy to accurately record their fluid nature and to allow for periodic retention to be undertaken.

3.4 Reasons for retaining DDOs over the long-term

3.4.1 Preservation policies and procedures

Written policies for the preservation of DDOs have been developed by stakeholders in the archival community, whilst some stakeholders in other sectors of the information community are still formulating such policies. The retention of DDOs is influenced by a range of drivers, which as this analysis suggests, are

strongly linked to organisational aims. These different organisational aims highlight a diversity of reasons for achieving the goal of long-term access to DDOs, irrespective of whether long-term management is undertaken by the organisation itself or a third party.

The findings (as shown in Tables 3.4A and 3.4B) indicate that management and monitoring procedures vary between stakeholder organisations and highlight differences in the internal structure of organisations together with the influence external pressures exert on this process. Some stakeholders already use, or would envisage, using different methods to preserve different types of DDOs, whilst others would adopt the same method for all DDOs regardless of the nature. The stakeholders' responses to the question of how the information content of DDOs such as Web pages, databases and multimedia resources are preserved, demonstrated that some stakeholders focus on physical preservation issues (Table 3.3A cells 1A and 2A), whilst others highlight the intellectual preservation issues related to the selection and responsibility for managing DDOs (Table 3.3B cells 7A and 10A).

The form in which records are preserved varies between organisations and in some instances demonstrates the organisational ethos towards digital data. These views ranged from the DDO being viewed as 'the' definitive record[10], to a lack of confidence in the electronic version and as a consequence there is a reliance on dual systems, one in paper form and the other in electronic form. Human and financial resources that stakeholder institutions have at their disposal together with storage considerations influence their ability or willingness to hold data in digital form over the long-term. In addition some organisations are dependent on the requirements of their customers/users, as is the case of journal publishers involved with electronic publishing, who to date continue to produce journals in print form as the majority of customers subscribe to this service (see Table 3.4B, cell 11D).

[10] This statement was made by a national archive charged with the responsibility to preserve government records. At the time of writing, this concurs with trends in United Kingdom government policy.

Table 3.3A Analysis of challenges identified by stakeholders and selection issues	**1** Publisher	**2** Specialist archive (Respondents: 1 & 4)	**3** Company from science-based industry	**4** Distributed data service	**5** Advisory body concerned with networking issues	**6** A national archive
A Challenges organisation encounters.	Ability to verify data. Physically refreshing and migrating of data.	Resp. 4. Finding a media that has a proven track record and will maintain life span of data. Employ tape/media refresh may migrate to new media in future.	See answer to cell 3E this table.	Little available data on long-term preservation of many digital record types. Technology evolving quickly + new and untested few people have experience of archiving + potential costs and strategies.	Depends on definition of long-term. Not central to purpose of organisation – if took seriously it would mean changes or challenges to organisation's practices.	Maintaining look and format of documents through migration & in future carrying out migrations on a massive scale.
B Training/experience person/people who take(s) decision to select electronic records/data sets have?	Not applicable (see answer to cell 1C below).	Resp. 1. Qualified subject specialists with computing background.	None. Have professional records managers/ archivists.	Subject specialists within their own discipline areas experienced in dealing with digital records.	Not applicable as not collection based.	Experience in records management and archivist experience of paper records + guidance and tutoring and eventually training on electronic records.
C Level of seniority at which electronic records/data sets selected.	Directoral strategy implemented at senior management level. Internal material not enough of a policy to answer.	Resp. 1. Head of service ultimate decision. In practice joint decision with colleagues Resp. 4. = Consultation amongst middle/senior managers.	Middle management level or below. If had system would control access to initiating & archiving event & middle/senior managers would be responsible.	Decisions made at a number of levels. Service providers have management advisory board. Steering committee for whole organisation.	N/A	Inspection Documentation Officers – junior managers.
D Does selection process cater adequately for 'dynamic' electronic records/data sets?	Yes.	Resp. 1. Yes. Less problems with 'dynamic' - finished product, know it is not going to change much/ minor changes.	At moment yes. No multimedia.	Not sure have any of these, apart from those buying in from elsewhere.	N/A	See answer cell 6J Table 3.1A.
E Does selection process cater adequately for 'living' electronic records/data sets?	No not in control of information customer's add to living resource. Never talked through whether it is important to archive this information.	Resp. 1. 'Living' create far more problems. -managing updates & changes (updates can change nature of resource creates problems for input & usage of material.	No, have problems with these – not easy to deal with & all other materials – can be volatile e.g. Lotus Notes database.	See answer cell 4K Table 3.1A.	N/A	N/A
F Factors considered important in determining which electronic records/data sets to preserve.	See answer to cell 1C this table.	Resp. 1. Embedded in our collection policy 'if collect material we consider it to be suitable for long-term preservation'.	Everything driven by regulators – then cost.	• Scholarly value • Documentation • Format.	Concerned with dynamic updates –rather than older material - more likely be an external decision – depend on owns /users.	Usefulness for various classes of users. Non-availability from other sources – if published would not need to keep it.

Table 3.3A Analysis of challenges identified by stakeholders and selection issues cont.	**1** Publisher	**2** Specialist archive (Respondent: 1 & 4)	**3** Company from science-based industry	**4** Distributed data service	**5** Advisory body concerned with networking issues	**6** A national archive
G Factors influencing selection methods used by organisation to preserve electronic records/data sets.	See answer to cell 1C this table.	See response to cell 2F Table 3.1A.	Driven by regulations. Non-regulated material = not sure what decisions individual departments would make. If had electronic archive would have rules for non-regulated data.	-Subject matter of material service providers accession. Then -judging records for scholarly value for potential for re-use.	More aware of internal issues - Aware of need to develop preservation policy	IT cost benefit factors – easier to preserve records plus particular requirements in system at the planning stage. Select records as early as possible.
H Criteria for selection of electronic records/data sets for long-term preservation.	See answer to cell 1G Table 3.1A.	See response to cell 2G Table 3.1A.	If regulated have to keep but have choice of media for legal copy – can be paper.	Specific selection criteria for each archive.	See response to cell 5H this table.	In process of creating. Philosophy of paper – records for departments admin. use select after 25 years those of administrative importance – kept to show policy and decisions – by-product information for researchers 1copy 1 person – electronic records removes this plus able to choose detailed info due to completeness.
I Particular types of electronic records/data sets which give organisation problems when selecting for long-term preservation. What problems do these cause?	See answer to cell 1G Table 3.1A.	Resp. 4. No most of data sets in social sciences field don't currently have multimedia, audio, video clips – mainly statistical format. In future stored on one system.	Living records – so inter-linked no point at which can say frozen & this is official version – answer = snapshots or long audit trails. -large size gigabytes.	Some service providers established longest = tried and tested methods. New services more difficult as no industry standard + emerging technology.	Documents mainly public – do not have selection process but of were to would be for content. Don't have complex databases.	Databases due to structural information being difficult to extract & difficult to represent changing information. Databases = technical structured information - pointers & format hidden + difficult ever changing have to decide what to represent. Cartographic images - as interlocking databases. Not in paper format, multimedia & compound documents.

Table 3.3B Analysis of challenges identified by stakeholders and selection issues	**7** Government Agency	**8** Advisory body for data preservation	**9** A national library	**10** Spokesperson from HE sector (personal view)	**11** Journal publisher	**12** Specialist archive
A Challenges organisation encounters.	Main problem = volume of material. How decide what are and not going to preserve? Resource implications if keep everything.	Resources: - people - money.	Everything. Continue ethos though nature of publishing changing. Migration. Intellectual coherence of content. Access delivery mechanisms.	Biggest challenge = management – understanding issues (e.g. who responsible within organisation) & taking right steps to preserve material over long-term	(a) Corruption (b) Vol. of storage (c) Speed of access (d) Support (e) Errata when it comes to journals.	Work on texts similar to MARC cataloguing work – learn lessons + issues unresolved not just the organisation's. Creating headers +plus SGMLplus suitably qualified staff Migration of material if a cataloguing or archival format became standard.
B Training/experience person/people who take(s) decision to select electronic records/data sets have?	Statisticians have a lot of experience but focused on their specialist area. No one formal archiving, preservation or selection experience. Plus side – no set agendas.	N/A	In future: combination of curitorial expertise + new set of information (cost of retention) then curators would make a balanced judgement – importance of material versus long-term costs, in a more proactive way than past.	In future: archival qualification - people who can take a long-term view + no direct interest in material. Work same way as historical or print archives.	N/A	In past no experience apt. from own + no criteria for selection. – couldn't prejudge what people might want. In terms of taking advice = Advisory Board's subject experience.
C Level of seniority at which electronic records/data sets selected.	How going to work this out = first or second line managers.	N/A	Future relatively senior mangers to make sure getting it right – get policy established & functioning, then assess training needs for junior staff.	Future Senior for policies & procedures = Pro-vice-chancellor level.	Senior management level.	At moment just take resources. Future: Advisory Board
D Does selection process cater adequately for 'dynamic' electronic records/data sets?	N/A	N/A	Not catered for at all – need to address. Possible increase in Web to gain access to don't have to acquire. One of strands of policy debate.	Not at the moment.	N/A	Not offered enough to have experience of how difficulties. Interested in finding ways of searching across multiple databases CD-ROMs.
E Does selection process cater adequately for 'living' electronic records/data sets?	N/A	N/A	N/A	Not at the moment.	N/A	As 'dynamic' ± present even more problems integrate into holdings? -assert identity compare this with the SuperJournal project.

Table 3.3B Analysis of challenges identified by stakeholders and selection issues cont.	**7** Government Agency	**8** Advisory body for data preservation	**9** A national library	**10** Spokesperson from HE sector (personal view)	**11** Journal publisher	**12** Specialist archive
F Factors considered important in determining which electronic records/data sets to preserve.	See response to cell 7D Table 3.1B.	N/A	See response to cell 10D Table 3.1B.	See response to cell 10D Table 3.1B.	See response to cells 11F and 11G Table 3.1B.	See response to cell 12D Table 3.1B.
G Factors influencing selection methods used by organisations to preserve electronic records/data sets.	Don't know haven't had any before.	Depends on overall mission of organisation + resources. Costs unknown. People start off retaining permanently & find not able to maintain. Advantage of electronic material can be done by consortia.	At moment service driven by -demand + best way of providing that information then decision made to purchase.	Don't know enough about overall organisation's view + new area only started to think about issues. What trends emerging? research reason to keep interaction between Students & tutors. Keep television material for historical reasons.	None as use same methods to preserve all data.	Criteria in collections policy : -scholarly -variacy of it -available at other Internet sites/archive only site -how viable to manage resource.
H Criteria for selection of electronic records/data sets for long-term preservation.	N/A	None yet – reports being produced then have better idea. Pilot project based. '98 review results - decide if something that can run with long-term.	Two areas where selection occurs -electronic catalogue records preserved forever. - important maintain corpus of literature over time.	Don't think do long-term. Archive of printed course material = keep everything. Long-term would like to say forever – with volume of material looks like impossible to do that.	N/A	N/A
I Particular types of electronic records/data sets which give organisation problems when selecting for long-term preservation. What problems do these cause?	'Dynamic' -have to take snapshots = series of freeze frames - best can do in reality. Initial problem = volume of material.	Multimedia distinct problems with sound, graphics etc to migrate without loss of links + data sets re-coding when migrate.	Multimedia access & presentation structurally bound in the information – linked to particular technology for access method = problem for organisation. XML will sort out. Proprietary influence too great. Through EU tentative approaches by national libraries to computer industry.	Digital video material – want to store electronically but due to cost impossible to store complete archive. Difficult transporting material in any great quantity over network.	N/A	Copyright material. Actual content of material itself – suitability of material offered for deposit. Multimedia may cause problems as not experts in this field. People from library community have had these problems for years.

3.4.2 DDOs defined as being of permanent importance

All stakeholders regarded some of their records to be of permanent importance, either as internal resources for potential future use within their organisation, or as a public record with limited or unlimited access restrictions. Despite identifying some records of permanent importance not all stakeholders had strategies in place to put the preservation of DDOs into practice (see Tables 3.4A and 3.4B, rows G and H), for example, a publisher whose prime concern is the creation and dissemination of DDOs rather than their retention.

The reasons why DDOs are seen by stakeholders as having permanent importance corresponds with the drivers influencing preservation, as evidenced by a comparison between the responses to questions D and G on Tables 3.1A and 3.1B. Organisational goals figure prominently within the reasons as evidenced by the following statement where an organisation claimed that it was "...responding to our duty to maintain some kind of record of UK publishing output as it moves from print to electronic..." (A national library) and "...shows the policy and actions of government and accountability..." (A government agency). The organisational goals often take account of external pressures, for example, within the pharmaceutical industry the external regulation of data is a major factor governing the type, amount of data retained, and the length of time it is kept. However, other factors such as cost also play a part. In the case of two stakeholders, one a publisher, the other an archive, the requirement to control access is also indicated as a reason for wishing to retain DDOs. That is, the ability to control the level of service (for example, speed of access) provided for data users.

The majority of stakeholders preserve dynamic DDOs, but in many cases this is a recent development and policy documents are in the initial stages of development. One stakeholder felt that there was a fuzzy boundary between dynamic and living DDOs, and this could account for the difficulty some stakeholders had distinguishing between the two terms. A spokesperson from a national archive expressed the view that, once a record has been deposited it ceases to be living. This archive requires records to be retained in a non-amendable form to ensure the integrity of the records over time. Few stakeholders create and preserve living

electronic records and the management problems these present are only just being identified. Some stakeholders provided access to living resources by hosting/mirroring resources whilst the original was retained by the creator/creating organisation. Any hosting/mirroring does have the disadvantage of being a snapshot of the original and therefore not updated in real time.

3.4.3 Legal requirement to preserve DDOs

The interview findings confirm the view expressed in Sections 1.6.2 and 1.6.3, that legislation does not exist that defines the responsibilities for the preservation of digital information; the exceptions being government agencies, national archives (see Tables 3.1A, cell 6E and 3.1B, cell 7E) and the stakeholder from science-based industry (see Table 3.1A, cell 3E) who are required by law to retain information. The need for clearer guidance on copyright for digital materials was also noted by respondents. The absence of legal deposit legislation is reflected in the answers of a stakeholder from the library community (Tables 3.1B, cell 9E)[11].

The legal requirement to preserve electronic records is shown by the findings to be concentrated on the requirements of financial and regulatory bodies. An exception to this trend is a government agency that has a legal obligation to retain certain classes of DDOs and has a secondary legal obligation to deposit DDOs with a national archive. Half the organisations contacted did not at present have any legal requirement to manage DDOs, that may be because these were within the library and archive community where the legal obligation to preserve DDOs is under review. A national library is concerned about the lack of a legal framework within its sphere of operation and is therefore interested in extending legal deposit legislation. As a result of this legislative vacuum the deposit of DDOs has developed in a piecemeal fashion on a voluntary or good will basis.

There is clearly a need to develop legislation to support those charged with the responsibility for preserving DDOs, and conversely to provide safeguards to ensure that these same institutions fulfil their obligations. However, documents that are

[11] The extension of legal deposit to cater for digital material is at the time of writing under consideration (see Section 1.6.3).

Table 3.4A Analysis of preservation policies & procedures	**1** Publisher	**2** Specialist archive (Respondents: 1, 3 & 4)	**3** Company from science-based industry	**4** Distributed data service	**5** Advisory body concerned with networking issues	**6** A national archive
A Are electronic records/data sets managed and monitored and if so how?	N/A	N/A	Managed by individual groups with few exceptions. Regulators require it.	At service level each has own arrangements as each hold different data types. Impossible for executive to impose something centrally beyond requirements of collection policy & service agreements.	Where we can but up to individual within org. to check saved internal links are working on their own pages. If Web site is redesigned then responsibility of group working on those pages.	Yes, still early days in this - only been done on small scale at moment. Already set plan for large scale for all electronic records.
B Are different methods used to preserve different types of electronic records/data sets?	A challenge have not faced up to yet.	Resp: 4 No all maintained on same system – maybe physically stored differently.	No policies on storage of dynamic and living. Well established procedures for paper records. Working on establishing a system for electronic data.	No in terms of methods that are available – emphasise migration & standards.	N/A	Only between data sets & documents.
C What procedures does organisation use to preserve information content of electronic records such as Web pages, databases, hypermedia resources?	Regular backup of internal network + external servers concretely archive each electronic entry – raw forms + routine back up. Not defined what it means to finish a product.	Resp. 4. Various methods could be used not all implemented. Do keep multiple copies - read only CDs. Shadow copy of original data. Compression. Cyclic redundancy check. Error correcting code. Access to data limited. Staff accidental damage possible. Do date stamp data – know if modified.	Suspect answer = none. May be put on a backing store with backup procedures. Suspect very little effort is made to preserve them in formal records management sense. An issue problem in making economic case, question of making senior management aware.	Varies according to material. + newly emerging technology.	Don't know physical way that is done backed up on tape, not sure where or how long keep for.	Basic view = migration and transfer records onto current software and hardware platform to maintain digital archive periodically. Everyone decided for audit reasons. Want access to look same whatever type of record looking at. + easier for user to use.
D Electronic records/ data sets kept in hard copy, electronically?	By accident have hard copy of most of published material, not deliberate strategy. Keep for 2-3 years after publish. Internal documents = no official procedure.	Resp. 1. Some paper and PDF, some paper only, some electronic only. Encourage people to send in electronic form.	Must keep documents. No coherent policy for handling. No policy for retention of information / long-term retention.	Documentation deposited with data = practice but has not always occurred affects usability.	If they are stored in both it is on an ad hoc basis. Where things have been published in both - very much a decision of the staff responsible.	No we decide the medium and if electronic records we keep them in electronic form.

Table 3.4A Analysis of preservation policies & procedures cont.	**1** Publisher	**2** Specialist archive (Respondents: 1, 3 & 4)	**3** Company from science-based industry	**4** Distributed data service	**5** Advisory body concerned with networking issues	**6** A national archive
E Estimated life spans of electronic records/data sets.	None have limited life span/underlying data does not.	N/A	N/A	N/A	Not asked due to responses to previous questions.	Not asked due to responses to previous questions.
F Preservation of electronic records/data sets which are constantly evolving?	N/A	Resp. 3. Yes. Constantly = internal database system of backups + ensure availability in future/not archived as evolving. Continuously = live database updated – additional records added incrementally.	N/A	N/A	Not asked due to responses to previous questions.	Not asked due to responses to previous questions.
G Written policy for preserving electronic records/data sets?	N/A	Resp. 4. Probably is – mission statement probably contains this.	Not applicable see earlier response.	Yes.	No policy there may be some provisos on some reports we publish on behalf of organisations. With some organisation there may be some agreement on how keep issues for the future. Host resources for other organisations.	Yes, written policy and procedures for specifics of preservation of electronic records within organisation. Generic support and guidance for departments.
H Written selection policy for long-term preservation of electronic records/data sets.	N/A	Resp. 4. Not defined but will be shortly plus will keep up-to-date mirror of holdings as backup.	Not applicable see earlier response.	N/A	At moment would not have any.	Not asked as covered in responses to earlier questions.
I Electronic records/data sets deliberately excluded.	N/A	Resp. 4. Nothing excluded as yet.	With new system would exclude Lotus Notes database plus documents.	N/A	Not asked due to responses to previous questions.	Office systems type records - have some multimedia & CD-ROM as part of contextual information not electronic record proper.

Table3.4B Analysis of preservation policies & procedures	**7** Government Agency	**8** Advisory body for data preservation	**9** A national library	**10** Spokesperson from HE sector (personal view)	**11** Journal publisher	**12** Specialist archive
A Are Electronic records/data sets managed & monitored and if so how?	N/A	N/A	N/A	[Quality of information content] All library catalogue entries checked. Course material produce electronic copy prior to printing. Course software = quality control.	N/A	N/A
B Are different methods used to preserve different types of electronic records/data sets?	Not at moment – but not from any wish not to. Probably think they would be preserved differently because have different nature and reasons for use. Preserve so much – difficulty is working out what people might want.	Inevitable. Each type if does not have its own strategy within an overall strategy needs to be recognised as a type & a strategy developed. May have to accept loss each time migrate/reformat - this may change as media improves.	Area of non-existent practice. They would be, but not done yet.	No overall policy as the archive is in its infancy. Academic archive –not considered 'living' electronic records – know there is an issue. Will be an issue in future with electronic files because of the amount & dynamic character.	Preserve CDs differently to databases.	N/A
C What procedures does organisation use to preserve information content of electronic records such as Web pages, databases, hypermedia resources?	Number of fragments in different policies, mostly out of date - pre-merger days. In process of creating a data management policy = direct bearing on what & how preserve data.	N/A	Have straight forward multimedia CD-ROMs seen as a published product not something to preserve.	Organisation wide WWW policy group. looking at guidelines Who responsible for quality & content. Have been checks in past on what there but no organisation wide policy. People maintaining servers have own procedures.	Backed up daily automatically be IT dept. Support of what have electronically -all maintained internally electronic customer support during day. Use of third parties for selling data. All other data networked.	None [no hypermedia resources] Depositors give material in many forms, archive takes a binary snapshot encode intellectual information – SGML and TEI – Encourage depositors to do this, as encoding their interpretation of data. No models to follow.

Table 3.4B Analysis of preservation policies & procedures cont.	**7** Government Agency	**8** Advisory body for data preservation	**9** A national library	**10** Spokesperson from HE sector (personal view)	**11** Journal publisher	**12** Specialist archive
D Electronic records/data sets kept in hard copy, electronically?	Awful lot of data which is released (not private or confidential). Data mostly held electronically as too great a volume to hold or publish it all in paper form.	People realising they ought to have some view on preservation they are forced rather than wanting to take it on board at start of a project. Keeping it in electronic form.	Question of available resources – since decision to develop OPAC at new library. long-term aim = convert exiting major catalogues to electronic form + smaller newer databases developed purely in electronic form. Matter of time and money.	Academic archive: trust paper more than electronic copies. Latter seen as addition to paper based archive + don't send electronic courses to students at moment archive used purely for internal use.	Commercial journals kept in electronic & hard copy as : (a) bulk of subscribers use hard copy. (b) Keep back issues to offer new subscribers. One of third parties dealing with have policy to keep material for 25 years. Developing own preservation service in 1998.	Sample print out (summary sheet) of each text so have an idea of what it contains = + paper information deposited with text. Main resource for the archive is the electronic metadata. Ideally digitize paper when time allows.
E Estimated life spans of electronic records/data sets.	Not asked due to responses to previous questions (see record life span table).	N/A	N/A	Not asked as question not appropriate see response to earlier questions.	Electronic journals kept indefinitely CD-ROM material data & programs held until updated which is likely to be much more frequently than for books. What happens to record when no longer requirement for product? Don't know yet – in practice it would be archived.	N/A
F Preservation of electronic records/data sets which are constantly evolving?	Not asked due to responses to previous questions (see record life span table).	N/A	N/A	Not asked as question not appropriate see response to earlier questions.	N/A	N/A

Table 3.4B Analysis of preservation policies & procedures cont.	**7** Government Agency	**8** Advisory body for data preservation	**9** A national library	**10** Spokesperson from HE sector (personal view)	**11** Journal publisher	**12** Specialist archive
G Written policy for preserving electronic records/data sets?	Not asked due to responses to previous questions within this table.	No not yet group looking into it. Organisation has its own digital remit.	Different parts of organisation different approaches. Some areas more interested in current usage. Organisation actively investigating issues involved in projects.	More in place in academic archive much practice & number of projects. Don't know if there is a project plan for creating a digital archive, but seen as an issue for the organisation	Written policy for backup, archiving and disaster recovery. Software programs if they are electronic records are maintained centrally by Information Technology department.	N/A
H Written selection policy of long-term preservation of electronic records/ data sets?	No selection criteria.	N/A	N/A	Not asked as question not appropriate see response to earlier questions.	Going to be a policy to preserve content. Archiving of content that make available to outside world. Third parties have an archiving policy.	N/A
I Electronic records/data sets deliberately excluded.	Exclusions by exception rather than by category - business case made on individual basis for non-inclusion.	Suspect a lot of electronic material is ephemeral, but could be of use to someone -can't retain everything - depends on mission.	Yes, multimedia games. In past maintained right to exclude traditional published material regard as ephemera. Difficult to see what will & won't be important. Past been more liberal. Future = stricter control, representative sampling.	No not at moment.	Cannot think of any. Whether we would keep it accessible to the public is a different matter.	Not unless documented elsewhere. No list that says what are classed as arts and humanities.

not within the public domain will only be subject to the preservation requirements of a future framework when they are proffered for consideration for long-term preservation by their legal owner. There would also appear to be an urgent need to develop legislation to address rights management issues (see Sections 4.11.1 to 4.11.3 and 7.3.5).

3.4.4 Administrative requirement to preserve DDOs

Administrative requirements for preserving electronic records are important to seven of the twelve stakeholders interviewed. The reasons cited for administrative requirements demonstrate the need for stakeholder organisations to re-use the material within their organisation and also to supply electronic records, in a useable form to external organisations or individuals. Two of the organisations that stated that they do not have an administrative requirement to preserve electronic records, do have systems to aid administration of DDOs within a distributed archival service of which they are both service providers. These mechanisms consist of an agreement to catalogue, document and display deposited material on an integrated catalogue.

3.5 Selection policies and practice

In the main, selection criteria appear to be very much in the development phase across the whole spectrum of organisations contacted (see Tables 3.4A and 3.4B, row H). Where organisations are concerned with materials that are either static or dynamic they generally indicate few or no types of resources which give problems, but in organisations where living resources are preserved/held, there are generally problems encountered with specific types of resources. The stakeholder from science-based industry indicated that databases such as Lotus Notes and Oracle databases cause problems because of their volatility[12].

[12]The preliminary results of the Research Libraries Group Survey (Section 3.3) suggested that member institutions also indicated that data acquired in the most volatile formats, namely, spreadsheets, Geographical information systems (GIS), video and vector graphics present the most problems (Hedstrom and Montgomery, 1998).

Stakeholders who held living DDOs felt in many cases that the selection process did not cater adequately for living DDOs, and in many cases highlighted this as an area of concern. Factors influencing which records should be preserved long-term were frequently linked to organisational collection policies or external pressures. These policies were generally founded on paper-based philosophy and as such do not cater comprehensively for DDOs.

Stakeholders were asked if they excluded any DDOs from long-term preservation in order to determine whether they had considered any parameters for selection criteria and the reasons for doing this. The findings suggest that decisions to exclude certain DDOs from long-term preservation seem to be subject to different criteria in different organisations. In some cases it rests on the DDO being published or unpublished. Other stakeholders excluded a particular type, or made a business case for the exclusion or inclusion of materials on an object by object basis. A national library has deliberately excluded the category of multimedia games. This is directly related to its policy of excluding non-digital games and a spokesperson gave the following justification for excluding this category.

> ...I think there are a number of materials which we have not traditionally preserved for example multimedia games. We have never collected ordinary games in the past so we do not see any reason to collect multimedia games, that is not our function. So we are looking at published material and to a large extent although we have, as I said earlier, maintained a lot of ephemera, it is difficult to see what in the future will be important and what will not. We have always retained the right to exclude traditionally published material which we regard as ephemeral or not central to the work of a major research library. In practice we have been a little more liberal than that... It may be that we can't be quite so liberal in the future and we will have to have a much stricter control, perhaps do more sampling rather than straight collecting, so that we maintain some representative samples. How that will be for historians of the future I do not know. If historians are honest they have to admit that even the materials they work with today, a lot of it, has survived through happenstance. So I guess the judgement of our curators is as good as anyone else's in those terms. (A national library)

This quotation raises a number of important issues relating to the role of selection in long-term preservation.

- How is selection to be achieved, particularly the cost/stringency balance?
- What is classed as digital ephemera?
- Difficulty of predicting the future needs of end-users?

The whole issue of exclusions brings into sharp focus the need to select materials even if it is only from the perspective of what should not be kept. A greater degree of stringency over the selection process is also evident from the previous quotation, underlining an awareness of costs associated with the future management of each selected resource.

3.6 Metadata

The organisations consulted all agree on the importance of metadata and eleven out of twelve are attaching this to DDOs at present (1997/98) (as seen in Tables 3.5A and 3.5B). The reason(s) each of the organisations expressed for attaching this metadata varied, reflecting not only concerns for the fragility of digital data but also organisational goals. This is perhaps reflected in the range of metadata formats that currently exist. Those archival institutions interviewed saw metadata as an extremely vital component to the functioning of their retrieval and storage systems.

> We have got such vast holdings of data here that it is the only way that internally, staff can track it and also for outside users who are using our information retrieval systems to identify the data sets that we have got. (A specialist archive)

A distributed data service expressed the view that whether they own a DDO or only provide user access to it "From our point of view we need metadata with every source to make it accessible and to manage it whether we are creating the data or specifying catalogue requirements for articles". A government agency with its large holding of data sets containing numerical data also highlighted the importance of recording the contextual information relating to the DDO when asked if they felt metadata was important they stated:

> Yes because without it, it does not mean anything. It is a heap of figures, our data particularly. Many other departments produce an awful lot of very good quality data and a lot of it will be textual or have a mixture of text and numbers, or may be they have got some fairly self explanatory numbers. Our data by and large tends to be huge amounts of data presented in a tabular form, very indigestible, called absolutely nothing in English. You need the metadata to tell you this is seasonally adjusted, this is not seasonally adjusted, this is what it is. (A government agency)

A national library sees metadata as a crucial attachment between the digital object and the highest level record. It also indicated that the use of metadata for preservation has not been thought through "Now, how we would take that into a preservation environment into the future I don't think we have a clear idea and I am pretty certain nobody else does either".[13] Although all organisations contacted acknowledged the importance of metadata some organisations, as indicated by the previous comment, had yet to define the metadata they require for long-term preservation of DDOs. This is occurring in an environment where the theoretical concept of the role of metadata varies between professions in the information community. This dichotomy of approaches to the underlying principles on which metadata is based is identified by Ng and Burnett (1997, p.337). "The two main approaches may be characterised as: (1) the library science orientated bibliographic control approach; and (2) the computer science orientated data management approach". The authors suggest the two main approaches be brought together to create an integrated approach to the concept of metadata.

[13] This is consistent with reports commissioned by the organisation and its involvement in the development of legislation for legal deposit. This comment is further substantiated by the presence of a number of initiatives aimed at creating preservation metadata (see Section 1.12).

Table3.5A Analysis of metadata formats	**1** Publisher	**2** Specialist archive (Respondent: 2)	**3** Company from science-based industry	**4** Distributed data service	**5** Advisory body concerned with networking issues	**6** A national archive
A Attachment of metadata to an electronic record/data set and reason	Beginning to look at – keeping published material not keeping records of what we have kept electronically.	Yes attached to records - part of online retrieval sys. Vast holdings – staff can track + users identify data sets that we hold.	Not created at moment <u>but</u> would be - essential for - re-finding - restore info and run application - meaning - economic decisions.	Yes. Needed with every source whether creating data or specifying cataloguing requirements for access.	Yes. Not specifically for preservation (organisation's prime concern is metadata for resource discovery). Estimate usefulness.	Yes. Part of definition of an electronic record.
B Metadata required to ensure information content of an electronic record/data set being preserved is authentic	Difficult issue – can't ensure anything keep is genuine – relying on integrity of people undertaking archiving.	Depositor sends in data + forms to describe data and Processing Section verifies data + errors corrected prior to metadata being attached.	Rely on regulations + validated system. Will be certain amount of visibility amongst data in a future electronic archive than at present – as done in separate departments.	At data capture in-house description created. Electronic record probably a product of original resource therefore can compare new resource with original.	No corporate view.	Guidelines given to departments.
C In-house versus standard format	Not possible to envisage any scheme to accommodate organisations requirements.	In-house software.	In-house.	Dublin Core at least for high-level resource discovery areas.	Several levels of adoption of Dublin Core format in other areas + made studies of other formats.	In-house but being re-done to library type EAD coding.
D Areas for improvement in adopted metadata format(s)	N/A	(a) advances in new technology (b) look for time saving ways e.g. documentation online (c) people want new things.	N/A	Dublin Core evolving Report published on Web on areas where more work needed e.g. coverage	No format perfect. Dublin Core large-scale implementation not happened, but there is some support for this.	No, designing it as think it should be. Expected to be completed for year 2001.
E Variance in the type and amount of metadata between different electronic records/data sets	N/A	Yes but not by much - system does not require completion of every section.	Yes – not significant difference between as have to supply records in very explicit way.	Yes. + metadata provided by depositor or by another agency can differ.	Yes – no particular reason for this. Higher level need more m., Web context minimal	Yes.

Table3.5A Analysis of metadata formats cont.	**1** Publisher	**2** Specialist archive (Respondent: 2)	**3** Company from science-based industry	**4** Distributed data service	**5** Advisory body concerned with networking issues	**6** A national archive
F Does metadata format cater adequately for the electronic records/data sets the organisation preserves?	N/A	Information given to users on technical aspects being developed, changed over time – old information not relevant.	N/A	Yes there are attempts. Dublin Core used for Internet catalogue developments. Catalogue depends on better management and subject specific access.	Yes – only deal with Web documents & Dublin Core is flexible & extensible, it will handle all metadata we perceive we need at the moment.	N/A
G Importance of interoperability between metadata formats.	N/A	Yes, but difficult to get those standards – all software use is in-house – difficult to find a match and if did could change in future.	Yes: encode the data - written in SGML – changes to SGML - might need to change. Translate from one language to another.	Yes – organisation objective to encourage interdisciplinary research and cross-sectoral access.	Dublin Core provides interoperability between richer, deeper formats. Interoperability avoids standardisation on 1 system. Personal view = very important.	Yes, want all our records on a common platform for ease of use.
H Stage(s) in an electronic record's/data set's life cycle when metadata attached.	N/A	N/A	N/A	N/A	Pres. m. no agreement on what consists of & difficult to know when it should be added & what included. Resource description metadata = added at creation & preservation	Creation

Table 3.5B Analysis of metadata formats	**7** Government Agency	**8** Advisory body for data preservation	**9** A national library	**10** Spokesperson from HE sector (personal view)	**11** Journal publisher	**12** Specialist archive
A Attachment of metadata to an electronic record/data set and reason	Yes. Organisation's data of no meaning without it.	Yes, in order to retrieve records in the future.	Needed. Ultimate attachment between digital object and highest level record – no clear idea how take this into preservation environment neither does anyone else.	Important to work on this. Enable organisation to keep information about content.	Full text SGML. Enables re-purposing = less costly than recreating + market requirement + international standard.	Yes
B Metadata required to ensure information content of an electronic record/data set being preserved is authentic	Don't know how going to guarantee – probably something archiving specialists will tell us & we will do.	Stick with basic level information - whatever is required by Dublin Core abided by in first instance.	N/A	If department or group within institution could validate electronic document – place electronic stamp on it would know it had gone through quality control procedure.	Parsing of SGML data at supplier and at publishers. This involves checking that all data is there and that it follows the rules of the DTD e.g. tags valid. However the content may not be accurate.	Rely on depositor + checks in the TEI header. Text in ASCII format and encoded.
C In-house versus standard format	Metadata Assistant = in – house. Consistent across data sets plus format standard and generic.	No particular format adopted because not responsible for a collection.	MARC top level record for object. How deal with information below this level? – digital library environment = similar to archive.	Would be ideal if there was one type. Already use UKMARC.	In true metadata definition the answer is none. Have SGML and DTD don't think this is strictly metadata – don't have metadata DTD.	SGML based standards and developed TEI guidelines.
D Areas for improvement in adopted metadata format(s)	Always improve + questions the system won't answer. If have monumental system too large to maintain therefore compromise.	N/A	Examination of finding aids used by archivists rather than title entry catalogue. Dialogue between professions.	Add references to electronic records. Link from library catalogue to Dublin Core description of documents. UKMARC or other to link to databases of metadata.	Relationship with HTML & Web – the hybrid being developed = XML and MML.	Requires formalisation of:- Proportions languages in texts plus dates.
E Variance in the type and amount of metadata between different electronic records/data sets	Generally no. Different areas of metadata to complete depending on type of data described + more on people inputting.	N/A	Yes.	Don't do a lot with metadata experimented with ROADS software for Web resources for courses.	N/A	Extensible so can use what is required for particular level of record.

Table 3.5B Analysis of metadata formats cont.

	7 Government Agency	8 Advisory body for data preservation	9 A national library	10 Spokesperson from HE sector (personal view)	11 Journal publisher	12 Specialist archive
F Does metadata format cater adequately for the electronic records/data sets the organisation preserves?	Hope it will – haven't come across any problems yet.	N/A	No –haven't clear structure of what is most appropriate – Dublin Core only answers certain questions – got to be more but don't know what.	Doesn't cater for WWW pages of organisation	Vast majority – exception = magazine type products.	What it makes available for us is fine – it is more a question of whether we have enough information to fill in the bits.
G Importance of interoperability between metadata formats.	Would be ideal – Is it practical ? unless take from one and put it into another, or one super set still presenting user with problems.	Technical people say it is – things that don't talk to each other defeats the whole purpose of the technology.	Yes. Envisage electronic library as single entity – everything needs to be linked for functionality.	Very nice if this was possible.	Has to be relationships between. component parts of metadata to fit within overall structure of document.	TEI designed for interchange. Parent organisation developing distributed catalogue = central database searches data of service providers.
H Stage in an electronic record's/data set's life cycle when metadata attached.	Immediately resource is thought about.	Should come with it or be created as soon as have acquired it – make sure all data is there + accessible. Can't leave until later stage.	Continuous process starting at beginning of life (acquisition) what constitutes metadata = how material needs to be handled and assessed in future.	Creation – some information can be entered at this stage, other info later when document has been finished e.g. end date of production.	As near to point of creation as possible e.g. during process of editing author's electronic file.	Follow TEI guidelines - start thinking about creating headers as soon as possible.

An organisation representing science-based industry which is in the process of constructing an in-house digital archive, has clearly defined that metadata is an integral part of the organisation's definition of a DDO and has clearly delineated what they regard as the component parts of their future metadata.

> A record in our sense would be the actual files themselves with the structures preserved and then at least four types of information. This would tell you things like who originated the data, what data was archived, when it was created, who owns it, what department was involved, how long we have got to keep it for and more technical stuff related to its status inside the industry. Then we would have to store information, very full information, on the technical side. So technical metadata which is concerned with describing each file, what version of the operating system was it created under, what piece of software created it, what was the version of the software, how big is the file, what standards does it conform to... how does it relate to other files in the same data set. Then we would have a third category of metadata which is to do with interpretation. It is no good, in our view, storing the data and then bringing it back onto another machine if you need to look at it or re-run the analysis, or do what ever you need to do with it, using just the identification information and just the files, because the people who understood the content could have left the company or might have moved on, or would have forgotten...There is a fourth category which we have been thinking about but haven't formalised and that is economic data about the information, such as the particular data set is worth x dollars to us. We need that kind of information, in my view, because the maintenance of the information is very expensive; so we need to be able to make informed decisions about its future maintenance... Unless you have all of those components you are really creating a huge electronic rubbish tip. (Spokesperson, science-based industry)

This quotation raises a number of important issues.

- The detailed specifications some stakeholders need to satisfy the aims of their organisation and the requirements of external bodies, in this example, industry regulators.

- The successful retrieval of data at a later date for future use. A process that the stakeholder identifies as requiring the attachment of metadata.

- Identification of the monetary value of data to the stakeholder and the implied cost associated with inefficient management. Cost in this example is restricted

> to monetary value. In a wider context value can also refer to the, administrative, legal, historical, scientific, scholarly or literary value of the substantive content.

Although all the organisations contacted, acknowledged the importance of metadata some had yet to establish practices for the creation of metadata for their DDOs. Metadata varies both for standard and in-house formats from a basic filename to far more sophisticated formats. As noted in Section 4.5, in-house formats are often adaptations of standard formats. In one case (see Table 3.5A, cell 5A) preservation of DDOs was not the primary concern of the organisation but rather the ability to search and find resources on a particular topic. Therefore, effort had been channelled into creating metadata for resource discovery[14]. Some stakeholders saw metadata as a mechanism for managing the DDOs by enabling digital data to be found and 'restored'[15].

The merits of placing the metadata in a separate database from the DDOs or embedding it into the resource as a header is another topic of debate. The latter was favoured by one specialist archive who stated that:

> ...thinking of conventional library cataloguing, or something like that, they have catalogue records and then they have things that are catalogued, and we don't have that distinction here, because what we rely on is electronic metadata essentially, that we glue with the item being documented. So as far as we are concerned, an item is the thing itself plus the metadata that comes with it. (A specialist archive)

The cost of replacing information versus the cost of preserving the DDO was highlighted by some stakeholders. A respondent from an electronic journal publishing house indicated that it was less costly to maintain existing DDOs than to re-create them (see Table 3.5B, cell 11A). It may be that in certain circumstances it is impossible to recreate events, not just because of the cost

---

[14] This is supported by the literature published by the organisation. Since the interview was conducted the organisation has become involved in a preservation initiative.

[15] Restoration in this context being the ability to represent the DDO in its original form (such as image quality and layout) by keeping information about data content and contextual information and maintaining the original software. This approach would appear to equate with the emulation strategy for digital preservation (see Section 1.6.8).

involved but also because it is a unique event as in the case of the data gathered from the Voyager II exploration of Mars in the 1980s. In this case the data had been recorded on seven track tapes which were deteriorating and could no longer be read without undertaking considerable error correction work and rebuilding a machine specially for this task. The corollary of this is the need for warnings to be generated within systems if we are to prevent the data becoming a victim of technological obsolescence, in other words there is a need to ensure that data migration occurs.

3.6.1 Metadata formats

One advisory organisation has not adopted a particular format, as it does not have a collection of its own. The rest of the organisations are split evenly between those utilising a standard metadata format (such as UKMARC) and those who have developed an in-house format. Within both groups there are both commercial and public organisations, however the group favouring a standard format contains both library based stakeholders. This could be the result of previous high investment in systems such as UKMARC, which is not surprising as it was designed for library use and many publishers still provide Electronic Data Interchange (EDI) services in this format. Some organisations, such as a national archive, are re-assessing their in-house format so that it more closely resembles Encoded Archival Description (EAD)[16]. As most of the stakeholders have some form of metadata format, which involved expense in its creation, it is suggested that were all organisations required to adopt a single metadata format a significant loss of investment could occur. This constitutes a case for interoperability between formats so that information can be shared. In addition it strengthens the case for the development of a higher level metadata that would contain information required to manage the resource over time, but allow for the existence of this multiplicity of formats at the metadata level (Heery, Powell and Day, 1997).

---

[16] EAD is the result of a project started by the University of California, Berkeley in 1993 to develop a platform independent encoding standard for machine-readable finding aids (see Section 1.13).

3.6.2 Areas for improvement of metadata formats

Some organisations identify areas for improvement in their chosen metadata format regardless of whether it is an in-house or a standard format. Other organisations are still in the process of constructing a new format and are therefore hoping to design a system that will rectify any previously identified weaknesses. The speed at which technology is advancing and the change in user demand has been a cause for some organisations to re-evaluate their metadata format.

> We have had 30 years to refine the system for what we want and there is new technology there, so that would be where we would be advancing. (A specialist archive)

From the analysis of the interview data it is apparent that the stakeholder organisations are at different phases in the development of metadata format(s) to cater for their needs. For example, a national archive is currently developing an in-house format with the requirements it feels it should have for the year 2001. Some stakeholders are seeking to extend standard formats, such as UKMARC records used for traditional materials, concentrating efforts on linkages such as crosswalks or mappings (described in greater detail in Section 5.6.2.2). A national library is examining the use of software based finding aids as used by the archival profession and seeking closer co-operation between librarians and archivists. A government agency stated that there are always improvements that can be made to metadata formats and questions that any designed system will not be able to answer, stating that:

> ...the only way you can accommodate that sort of thing, be perfect to all people, is to have an open monumental huge thing that we won't be able to maintain so yes it could be improved, but one has to look at the resource constraints and say is that practical? (A government agency)

This stakeholder is producing a large volume of data that it wishes to make available to the public. However, as responses through the tables in this chapter suggest (see responses for stakeholder 7), this organisation was willing to delegate responsibility to preserve most classes of DDOs to another part of the government.

The agency is developing its own in-house format to guide end-users, a development in keeping with the trend towards more open government[17].

As is to be expected, the extensibility or variance in the type and amount of metadata differed the most within organisations concerned with the greatest range of types of DDOs. The range of DDO types requiring management, together with factors such as the principles on which the format is based, were seen to influence the degree of extensibility it needs to exhibit. The principles may include the ability to leave fields incomplete or increase or decrease the detail within a field. The influence of the underpinning principles on extensibility is highlighted by a comment from a specialist archive which stated that the structure of the Text Encoded Initiative (TEI)[18] format is extensible enough to cope with any level of record one would wish to create and that the problem for them was finding sufficient information to include in the metadata.

External drivers can also influence the amount and content of metadata as evidenced by the comments of the respondent from science-based industry, who indicated that little variance would occur as they are required by the regulators to supply electronic data in a very explicit way. Similarly, within the pharmaceutical industry, there are specific regulations applicable to the United States market pertaining to electronic records, such as 21 part 11 issue relating to the use of electronic signatures (United States Food and Drug Administration, 1997). The need for metadata that enables the long-term management of these DDOs may be generated by some organisations, if this coincides with organisational goals, however if this is not the case DDOs which society may value in the future could be lost through non-existent or inadequate metadata. This points to the need for a mechanism that will monitor, create (if required) and store information necessary for the long-term management of DDOs at a level higher than metadata.

For organisations with existing metadata schemes, most cater for the majority of DDOs they wish to preserve. However a spokesperson from the higher education community identified the inadequacy of UKMARC to cater for World Wide Web

---

[17] *Your right to know: the government's proposals for a Freedom of Information Act.*
[18] See Section 1.13 for further details of this format.

data, and a national library which at present also uses UKMARC pointed to the fact that Dublin Core only solves some problems, "...by its own admission a subset" (A national library). The UKMARC format was designed to cater for the needs of tape storage and primarily described traditional print-based materials. Due to its origin, tags such as 258 have been added to allow cataloguers to express the physical characteristics of electronic media. For UKMARC to perform the role of a long-term management mechanism for DDOs, these fields would need extensive modification. The action notes field would require a change of emphasis to embrace a more proactive management of DDOs including information such as, the time interval after which an object needs to be migrated, or alternatively, records the software needed to emulate the DDO. Similarly, the provenance note would need to be revised to include information pertaining to the ownership of the resource together with copyright and access restrictions. Such a modification of UKMARC could prove to be costly in terms of financial and human resources. Therefore, it would be more advantageous to record this information in a higher level of metadata rather than incur costly changes to the metadata. Stakeholders are at present engaged in either maintaining standard format and developing new in-house extensions or creating completely new formats to cope with DDOs. Due to this prior investment by stakeholders and justified reasons for their adoption of a particular approach it would seem to be beneficial to accommodate the diversity of metadata formats, and create a higher level metadata that would supplement preservation information at the metadata level and prevent unnecessary duplication.

3.6.3 Interoperability between metadata formats

The importance of interoperability between metadata formats is widely acknowledged to be beneficial within the selected stakeholder organisations (see Tables 3.5A and 3.5B, row G). Such interoperability is seen by a national archive to be desirable and as a result have placed all their records on a common platform for ease of use, however, there was no mention of connections with outside organisations. Within the library community a national library views the electronic library as a single entity and as a result everything needs to be linked for functionality. Some of the metadata formats created by the stakeholder

organisations were designed with interchange in mind, as in the case of the TEI, and some specialist archives also interchange material on a service wide integrated catalogue. The objective of this catalogue is to provide cross-sectoral access to DDOs in order to encourage interdisciplinary research.

A government agency felt that whilst interoperability was in principle a good concept it would not be practical unless one was taking information out of one system and placing it into a second, or that one superset of information existed. In addition, they also identified that the user may still be faced with a plethora of different interfaces to negotiate.

A specialist archive although in agreement that interoperability is desirable, pointed out that it is difficult to reach agreement on standards for preservation metadata, particularly as the archive uses solely in-house software. Other archives have approached this problem in a different way by using a standard format that is extensible, such as TEI. Even if it were possible for organisations using in-house formats to find a system in which to interoperate, it may not be desirable for them to change their system to accommodate this, as these external schemes may change in the future. This concern for rapid changes in standard formats is also voiced by a spokesperson from science-based industry, who indicated that as the data is written in SGML any subsequent change to the structure of this would necessitate work in changing their DDOs to comply.

Despite these reservations on the practicality of creating interoperability between metadata formats, strong reasons were expressed by stakeholders for the imperative need for linkages between formats. This was evidenced by the view of a Director of a preservation organisation that formats that cannot interoperate "...defeat the whole purpose of the technology". This is upheld by a general comment from an advisory body concerned with networking issues.

> I think one of the good things about recent metadata thinking is the idea that a core format, like Dublin Core, could provide some form of interoperable focus between other formats which are much richer and deeper. We have a whole host of residual data which is stored in MARC format

in libraries. We have information which is in SGML based in text archives which they use, TEI, and there are specific metadata formats of quite detailed complexity for geo-spatial data etcetera. Interoperability between these systems does seem to be a means of avoiding having to completely standardize on one system, which will make everyone unhappy because you cannot standardize between different data types. It is possible that something like Dublin Core, if it is managed well, or even other formats could be used as a core for interoperability. (Advisory body concerned with networking issues)

The avoidance of standardisation on one metadata format would be highly advantageous, as it would prevent organisations incurring time and money converting data in legacy systems to a new standard, and would not mitigate against existing and future in-house formats being developed by organisations or consortia. The use of Dublin Core as a focus for interoperability may work for the creation of crosswalks within the library and archival community in which it originated, but because of its emphasis as a basic format to aid the discovery of published resources, it does not contain the essential elements required for long-term management of DDOs (which will be discussed in Chapter 7). It unnecessarily forces other communities to conform to what might for them be an inappropriate library based framework.

Creation is the favoured point in the life cycle for the attachment of metadata, but as several organisations indicated this metadata can be added throughout the resource's life. As far as a national library is concerned the constituents of this metadata depend on how the material needs to be handled and assessed in the future. The accretion of metadata is also identified by a spokesperson from the higher education sector who stated that in his personal view some information could be entered at creation whilst some could only be added when the DDO has finished being created, such as at the end of production date[19].

A range of responses exist concerning when to create metadata, a preservation organisation identified the need to actively manage the DDOs as soon as they are

[19]This comment referred to course material, where the creation and preservation processes are within the stakeholder's control.

acquired by their organisation as vital "...they must ensure that all the data exists and is accessible, something that cannot be left until a later stage".

3.6.4 Documentation and validation

The degree to which stakeholder organisations have established what constitutes 'good' documentation was seen to vary. The analysis of responses (Tables 3.6A and 3.6B) as to what constitutes valid documentation indicate that many organisations do use internal and external processes, but it is clear that within the organisations contacted little standardisation of practice exists. Depending on the aims of the organisation there can be a concern for both internal and external validity as expressed by a stakeholder from science-based industry, where any system holding the digital data would be checked by a department within the company and also externally validated by the regulator(s). For some organisations sufficient validity is gained if the organisation(s) which created the DDO is/are viewed as being reliable. Within organisations where digital archiving practice is in its infancy or at the planning stage, the measure of what is valid is that only final versions of documents are made available to the public, or are archived, and in fulfilling either criteria they are seen as valid DDOs. At the other end of the spectrum, some organisations with a mandate to preserve DDOs have established a clear definition of what for them constitutes valid documentation. For example, a specialist archive states that "...for it a data set to be fully validated we have to have sufficient documentation to be able to interpret the information within the files". As a distributed data service identified this varies according to the data type.

Validation in a second specialist archive, which has established metadata, is linked to the process of creating it, and they go on to comment that validation of this metadata does not involve a value judgement, rather it is a statement that it conforms to SGML guidelines. In the case of a national archive, to ensure that they are preserving valid DDOs they require records to be made secure and non-amendable at the point at which they are used corporately. There would appear to be a wide range of methods for validating DDOs. For example for some numerical data it is sometimes better to validate the program which created the data rather

than the data itself. A respondent from an advisory body concerned with networking issues indicated that certainly within the Web forum the subject of how to achieve validation was open to debate. In some organisations this is in the process of being established, where as in others detailed guidelines exist.

An organisation within science-based industry is in the process of developing an in-house electronic archive and has (as stated in Section 3.6) identified documentation as a key component that should be included in the development of metadata. This includes what the respondent terms technical metadata which describes the software environment in which it was created together with interpretation metadata which provides contextual information in a read-only file. Above all, the organisation requires this documentation to be timely and accurate. A specialist archive that manages statistical data considers that good documentation for a data sets provides information relating to the source and interpretation of, for example a letter, number, or series of letters and numbers within a table of data; its history; whether the variables are weighted and if it contains derived variables. A national archive also supports this view that accurate structural and contextual information are vital components of documentation as they provide an audit trail on a records creation and use. This need is evident in a specialist archive's integration of documentation into their metadata. The acid test of this integration is:

> ...that the original physical object can be recognised from the metadata that you can recognise that there is enough information in the metadata or in the header that will identify with the unique item, and that is what we constitute as well documented. (A specialist archive)

Common to many definitions is the need to establish how the document was created together with conditions for maintenance and retrieval, and also what constitutes good documentation is dependent on the organisational environment in which the document is constructed. The documentation therefore is tailored to meet the needs of a particular organisation, or is even specific to individual service providers in a distributed service who are responsible for different types of DDOs. A spokesperson from the higher education sector indicated in his personal view that during the process of compiling the documentation for the long-term

Table 3.6A Analysis of documentation & validation issues	**1** Publisher	**2** Specialist archive (Respondent: 3)	**3** Company from science based industry	**4** Distributed data service	**5** Advisory body concerned with networking issues	**6** A national archive
A What constitutes a validated electronic record/data set?	N/A	Validated data set = sufficient documentation to interpret information within files.	Any future system would be validated = tested to a certain criteria + by regulators. Checked by validation department.	Varies according to data types some use of parsing programs.	No corporate view. Area on the Web forum which is very much up in the air. One project involved in suggesting using hash values.	One that has been made secure and non-amendable at point that it was used corporately.
B What constitutes 'good' documentation?	N/A	Good:- complete picture not isolated file of nos. - Source + interpretation - Data sets = history – why collected - Note use of weighting, variables - Derived information additional variables - How derived variables Adequate documentation = some variables can't interpret 100%	No standards for documentation – have standards on preservation documentation online Factors required: Accuracy Timely.	Information disseminated to potential depositors – tailored to different services and also covers documentation common to all service providers.	No corporate view as most of information is Web based. Use guidelines for HTML – try to keep it conformative to level that could be used by links browsers.	Contextual and structural information on the record which is accurate and provides audit trail of creation and use.
C If an electronic record/data set is inadequately documented but is of legal, administrative or scholarly importance, what action would be taken?	N/A	Try to gain as much documentation as possible at deposit. Detailed checks at each stage. Where documentation insufficient contact the depositor. If very important would take without sufficient documentation.	If had system would put this info in the record. Electronic archives people/records manager. Insist something adequate is entered. Try system out in real-world.	(1)Professional judgement. (2)Depends on budget plus in some cases time. (3)Completeness of documentation that can be constructed (if electronic not adequate is there hard copy documentation). Response varies depending on importance of resource.	N/A	Occurs quite a lot – only if impossible at reasonable cost to ensue authenticity and integrity would records be destroyed. Don't keep any record can't document and access.

Table 3.6A Analysis of documentation & validation issues cont.	**1** Publisher	**2** Specialist archive (Respondent: 3)	**3** Company from science based industry	**4** Distributed data service	**5** Advisory body concerned with networking issues	**6** A national archive
D Are there written guidelines relating to accompanying electronic documentation deposited with the electronic record(s)/data set(s)?	N/A	N/A	Documentation = metadata descriptive information - depends on originating department who may have their own standards.	N/A	N/A	Yes currently partly in electronic and paper form
E What mechanisms are employed to ensure electronic records/data sets considered for preservation are authentic?	N/A	N/A	Only if falls under rules for Good Laboratory Practice, Good Clinical Practice and Good Manufacturing Practice.	Not applicable –cannot see it being an issue with historical records and text.	Do not think use any mechanism – partly as information comes from ourselves or from information providers we trust.	Request people meet principles of PD0008.

Table 3.6B Analysis of documentation & validation issues	**7** Government Agency	**8** Advisory body for data preservation	**9** A national library	**10** Spokesperson from HE sector (personal view)	**11** Journal publisher	**12** Specialist archive
A What constitutes a validated electronic record/ data set?	Has been released, it should be valid. (How guarantee validity in networked environment?)	Has all necessary bits and pieces attached to it. Has sound provenance. Know what going to do with it and why have it.	Two ways at moment: CD-ROMs plus living records – Produced by library Produced by similar organisation.	Electronic archive of a materials organisation creates in paper form – measures in place – only archive if final version.	Parses with nil error.	Validated SGML document conforms to guidelines, last stage of processing.
B What constitutes 'good' documentation?	Intelligible to somebody else outside organisation Difficult to create guidelines for staff.	Adheres to national and international standards e.g. Dublin Core and have to develop preservation metadata that allows record to be retrievable over time.	Documentation that allows: 1. to understand nature of information in record and functions of product are. 2. technical information - platform & delivery mechanism - data structure implicit in record.	Important to have information on who owns copyright on paper and electronically.	Written procedures that could be electronic that determines precisely how data created, stored and recovered.	Original physical object can be recognised from the metadata
C If an electronic record/data set is inadequately documented but is of legal, administrative or scholarly importance, what action would be taken?	If adequately documented send via national archive and they can decide whether to preserve.	N/A	Not addressed question, not sure that it has arisen.	Not at moment.	N/A	N/A
D Are there written guidelines relating to accompanying electronic documentation deposited with the electronic record(s)data sets?	No but working on development of adequate metadata to describe data as part of in-house system.	N/A	Catalogue records held separately from product. MARC plus extra information No where near the broad stream metadata required for the future.	No but in electronic of archive printed course material have documentation.	N/A	N/A
E What mechanisms are employed to ensure electronic records/data sets considered for preservation are authentic?	Looking at Internet as a delivery mechanism – period with no direct personal contact – looking at non-repudiation – need to authenticate.	N/A	Catalogue records created by the library or similar organisation, or come form a defined publisher.	Not sure if it is an issue for the organisation as know the creators are employees of organisation.	As far as author has to sign a copyright assignment form.	Recording a statement that depositor is stating this is what it is.

maintenance of DDOs it is important to resolve rights management issues for a specific resource. Where such guidelines exist in other stakeholder organisations they relate to contextual and structural information which includes details on provenance, which allows records to be re-found over time, and provides a complete picture of the DDO, rather than it being an isolated file of numbers.

With DDOs that are of importance but have insufficient documentation, stakeholders exhibited a range of strategies. Within the archival community every effort would be made, cost considerations aside, to salvage the DDO.

> Only if it was going to be impossible at reasonable cost to ensure the authenticity and integrity of the record and only if it didn't meet that, would we destroy them, but we wouldn't keep any records that we couldn't document or access...". (A national archive)

In order to reduce the number of DDOs with inadequate documentation, archival organisations, such as one specialist archive, attempt to gain as much information as possible at the time of deposit, by making detailed checks at various stages as part of the acquisitions process, which may involve contacting the depositor for further information. If a DDO were considered to be very important this specialist archive would preserve it even with insufficient documentation. In situations like this an archivist expressed the view that the decision to maintain or dispose of a DDO is a matter of professional judgement influenced by the budget and in some cases the time available, with the response being dependent on the importance of the DDO. As expected the issue of insufficient documentation has not arisen in those organisations that have, as yet, to develop a digital archive/archiving policy. Written guidelines that relate to documentation accompanying DDOs are regarded as part of the metadata.

This brings us to the whole area of metadata, what it is and what it is not and it is still an on going debate. I think, as far as the dynamic products are concerned, they are certainly catalogued so their existence is acknowledged and some information about the producer is implicit in the catalogue record, but they are held separately from the product itself and these are MARC records shoe horned to add a few bits of extra information. So they are certainly nowhere near the kind of broad stream metadata that one might want functionally in the future. (A national library)

Currently a national library has documentation in both electronic and paper form. Similarly, other organisations in the process of developing digital archives will be developing guidelines for the electronic storage of documentation.

This section of the analysis underlines that:

- Documentation is a vital part of ensuring the successful long-term management of DDOs.
- The importance attached to the quality and content of documentation created by stakeholders varies widely.

It would appear from the findings that documentation is a vital component necessary for establishing authenticity and an aid to validation processes. Therefore, it is seen to be a vital component of metadata required to manage DDOs over the long-term.

3.7 Authentication

Information required by organisations which ensure that the information content is authentic is determined by a mix of internal and external drivers (such as regulatory bodies). Other organisations which create DDOs have identified the need to manage their DDOs, but not all have implemented measures to ensure this occurs. For example, some rely on the integrity of their personnel. To an extent, some stakeholders expressed the view that they are reliant on the information provided by depositors and communicating with them if the documentation proves to be insufficient or inconsistent with the substantive content.

Some sectors use mechanisms that ensure that DDOs are genuine, such as PD0008 (Allen *et. al.*, 1996)[20]. Where DDOs are digital surrogates for paper-based resources, such as historical texts, ability to consult the original removes the

[20] A British Standard which stipulates the requirement necessary to ensure information stored in electronic document management systems is legally admissible.

need to establish the authenticity of the DDO (see Table 3.6A, cell 4E). Authentication appears to be based on trust and some organisations which have a closed community inputting and using information in this way may at present be sufficient. The remainder of stakeholders rely on a network of trust especially in the library community which is frequently generating DDOs internally (mostly in the form of cataloguing records), and accepting DDOs created by organisations of 'similar standing' or 'defined publishers'.

3.8 Integrity and access

The majority of organisations would not wish to trust another organisation with their only copy of a DDO as seen in Tables 3.7A and 3.7B, row A. Those stakeholders from the archival and library professions wished to retain their copies as they felt this would be in accordance with their goals. A national archive stated that to trust a world preservation database as proof that a DDO had been preserved and discard their copy or copies on this basis would conflict with their role "...as the National Archive it is seen as the national memory and not to be handed over to anybody else".

A specialist archive views the retention of digital resources as an issue of retaining access to the DDOs rather than responsibility for their retention; although they do recognise this policy may have to be reviewed at a future date if retention did become a problem.

> As I say at the moment we wouldn't throw things away, simply because it is cheap and it is relatively easy for us to store it and manage it. If it ever did get a problem then we might have to start throwing things away. The other thing is, that we come back to this argument of accessibility, we say even if it is available somewhere in the world, if it is available here it is easier for our academics to get at it quicker than from here than it is to go to Australia. So we may keep it just on that basis. (A specialist Archive)

An advisory body concerned with networking issues commented that under its present remit it does not strictly preserve DDOs, but if it were to do so at some future date it would be with the following proviso, "If it were preserved

somewhere else, or available somewhere else, we would not necessarily feel we would have to change that". That is they would seek to minimise duplication and provide access to the DDO through a link. Those publishers and a science-based company interviewed would similarly wish to retain their data primarily for commercial reasons and to ensure security of their data for future developments. However an electronic journal publisher did not rule out trusting a preservation database but stated that it would need more information on how DDOs were preserved, archived and the speed with which they could be accessed. The stakeholder representing science-based industry indicated that although they would wish to keep their highly sensitive data in a closed environment they may consider using a vendor, provided that this vendor was also being utilised by similar companies and had highly developed security systems such as firewalls to protect clients' data. Contrary to the rest of the stakeholders a government agency said it would trust an entry on a world preservation database if it were to exist as proof that the work has been preserved, as they do not want to be under any obligation to undertake the role of preservation themselves, that is they do not see it as part of their remit. This is because they have insufficient resources in terms of space, time and money. Stakeholders such as a national library although at present unlikely to trust such a world preservation database, would be willing to trust a co-operative venture formed of legal deposit libraries[21]. A comment from a preservation organisation highlighted the need for trust in a digital 'master', if there is to be a situation where not everyone is keeping DDOs long-term.

When a resource reaches the end of its life many stakeholders would be prepared to allow another institution or body to preserve it. The majority of stakeholders, who were collection based, felt that they could allow another body to manage a DDO but that this would be conditional on factors such as the degree of access the managing body would provide. An example of this would be the submission of a commercial electronic journal to a national library. Since it is in electronic form it would be very easy for a national library to make this journal available globally, hence requiring access restrictions to be put in place to safeguard the publisher's interests. Some form of metadata would be required to manage this.

[21] This statement is supported by initiatives such as the Networked European Deposit Library (NEDLIB) URL: http://www.konbib.nl/coop/nedlib/midden.shtml.

Table 3.7A Analysis of access & integrity issues	**1** Publisher	**2** Specialist archive (Respondents: 1, 3 & 4)	**3** Company from science-based industry	**4** Distributed data service	**5** Advisory body concerned with networking issues	**6** A national archive
A Trust of an entry on a world preservation database as proof work has been preserved and discards the organisation's copy/copies on this basis.	Would not want to. Feel strongly that it is our own problem something would not want to rely on somebody else to solve for us.	Resp. 1. No. Resp. 3. Yes, compression is a potential problem some data sets very large makes sense to compress where can. Policy = never compress anything for which archive is solely responsible. Resp. 4. Archive does not own the data – would consult depositors before deleting.	Keep it closed as highly sensitive data. Might use a vendor dealing similar companies but would want to know that the firewalls were secure.	We would recommend that we maintain our copy. However it may be an issue where we produce publications – users may have different requirements in terms of equipment.	No if organisation did preserved things (does not strictly preserve at the moment) would do so because felt it was part of our role. If resource preserved or available elsewhere would not necessarily preserve.	No as organisation is a national archive – seen as national memory and not to be handed over.
B When an electronic record/data set reaches end of its life could you allow another institution or body to preserve the record?	Yes. Key would be what access that other body was given to the material.	N/A	No, except for some classes of information which might share with others or sell.	Complex landscape number of us discussing how we would co-ordinate our collection policy.	N/A	Does not apply to the organisation, but could apply to source organisations.
C Maintenance and restriction of access to electronic records/data sets that organisation holds.	Closely maintained form of access.	Resp. 3. Section in the metadata that details access restrictions.	See responses to earlier questions.	Much of material available to users number of ways can restrict: - only available to registered users. - authentication based on individual rather than machine - material stored off-line.	No restrictions partly due to nature of organisation. Challenges presented by Web site reorganisation and frequent URLs change.	Data sets = mainframe access via local terminals & Internet. Documents = PC network system.

Table 3.7B Analysis of access & integrity issues	**7** Government Agency	**8** Advisory body for data preservation	**9** A national library	**10** Stakeholder from HE sector (personal view)	**11** Journal publisher	**12** Specialist archive
A Trust of an entry on a world preservation database as proof work has been preserved and discards organisation's copy/copies on this basis.	Yes, do not want obligation to preserve ourselves internally as have not got -space -time -means	If going towards point where not everybody has long-term copy need master copy – would have to trust it.	Current thinking + unlikely, except where legal deposit libraries are thinking of co-operating together.	Don't know - not enough into the digital age to be able to do that at the moment.	Not to begin with would need further information on how going to be preserved, archived and how quickly could access it.	Keep own resources as cheap and easy to store and manage. If ever became a problem might discard. Accessibility = easier and quicker even if available elsewhere.
B When an electronic record/data set reaches end of its life could you allow another institution or body to preserve the record?	Yes, although have certain classes of records which cannot be released to another organisation.	N/A	Yes. Already has a retention policy & certain things– expects other libraries to retain.	Yes.	Yes, depending what the record was.	Yes.
C Maintenance and restriction of access to electronic records/data sets that organisation holds.	Mostly ad hoc entry points, although some are formal. New system being developed to bring organisation's fragmented holdings together.	N/A	N/A	N/A	Journals: - online journal service using a third party -own journal service.	Three levels of access 1. free/open access 2. registered 3. restricted.

Maintenance and restrictions of access to DDOs varied and is strongly influenced by organisational goals, within some stakeholder organisations access restrictions were recorded in the metadata.

The development of a more regulated environment could increase stakeholder confidence in the quality of another stakeholder's data and the management of some of their data by a third party. The findings indicate that some stakeholders are willing to place confidence in third parties providing that:

1. Access restrictions can be specified and implemented.

2. The authenticity of another stakeholder's data can be determined.

3. Rights owners can retain access to their DDOs.

The interview findings suggest that rights management issues are complex core issues requiring resolution if a long-term management framework for DDOs is to successfully address the concerns of all stakeholders. The nature of rights management issues is explored in greater detail in Chapters 4 and 6 of this thesis to determine how they are likely to impact on long-term digital data preservation.

3.9 Record life cycle and its influence on the long-term management of DDOs

The degree of precision with which the life span of DDOs were defined would appear to be strongly influenced by the clarity with which retention policies have been developed (Tables 3.8A and 3.8B, row A and Tables 3.3A and 3.3B, rows E and F). For example, a national archive has a clearly defined span for its DDOs which go through an 'active phase' typically of between three and fifty years. During this phase updated versions of the DDO may be created and only after that does deposition occur. This entails deciding whether to discard or retain for appropriate periods, including in perpetuity. For other organisations, aims, commercial reasons and external pressures seem to contribute to the process of defining a DDO's life span. Stakeholders who had clearly defined the record life

span stated that this span is determined as close to creation as possible. However, there is a division of opinion as to when in a DDOs life cycle it should be identified as requiring long-term preservation, in some organisations it is at creation, in others this occurs when material is acquired, or in the case of a national archive, when it is deemed to have concluded its 'active' phase[22]. It should also be noted that the creator can determine the period DDOs are made publicly available and can request the removal of items from a deposit organisation.

Another important aspect of life span is the fact that a DDO may pass through more than one owner during its life span, for example if it is created or deposited with a commercial organisation, its life cycle as far as that organisation is concerned is limited to its commercially useful life span. Within this sense life span can be defined very precisely, not as x number of years but by some criteria such as commercial profitability or regulatory/legal requirements. At the end of this period however, the DDO may pass into the ownership of a national archive where its life span is reassessed in the national interests.

3.10 Balancing the needs of creators and end-users

The protection of the stakeholder's right as a creator of DDOs takes the form of in-house policies, or deposit organisation policies, that attempt to cater for the electronic environment or rely on existing copyright legislation (see Table 3.9B, cells 7A and 11A). The latter has been identified as lacking, as it has not kept pace with technological developments (see Section 1.6.2). This need for clear guidance on rights management issues is identified by a stakeholder from a government agency, who makes the point that, "Really we only use standard copyright or crown copyright... It's a bit outmoded so we are having to work at the edges and establish precedents..."[23]. The requirement to strike a balance between the needs

---

[22] Where it is not possible to establish guidance on the management of DDOs in their 'active phase', leaving the decision to preserve until this phase is complete could create problems, for example a potential loss of contextual metadata and potential physical degradation of the DDO, particularly if the 'active phase' has a lengthy duration.

[23] This opinion was expressed prior to the consultation document *Crown copyright in the information age: a consultation document* being presented to parliament in January 1998 and *The Copyright and Rights in Databases Regulations 1997* which came into force 1 January 1998.

Table 3.8A Analysis of record life cycle issues	**1** Publisher	**2** Specialist archive (Respondent: 4)	**3** Company from science-based industry	**4** Distributed data service	**5** Advisory body concerned with networking issues	**6** A national archive
A How is the duration of the life cycle of an electronic record/data set defined by your organisation?	Never dispose of material the organisation is publishing. Internal material = no particular policy.	N/A	Suspect all independently driven in different departments according to their needs.	N/A	No definition but expect final reports for a project which is non-dynamic, non-living to stay longer in an unaltered condition than conference notes.	Planning, creation, coporate use. Use could finish anywhere between 3 –50 years. Then deposition – destroy or keep for appropriate periods including permanently.
B What factors influence the duration of the life span and how?	Material publish are kept forever – no life cycle dependancy.	Don't selectively delete at the moment (have enough space). Only selectively remove data sets or study numbers incomplete - certain level useless. (even this not happening at moment).	Regulators probably influence everything.	Various things have influence depending on type of record. In a sense life span wrong [concept] - when it reaches archive it is already selected.	Seen public documents e.g. research reports as long-term.	Records held are all permanent, it is only in the departments that there is a variation.
C What are the estimated life spans of electronic records/data sets held by organisation?	Do not think of any of them having a limited life span, or at least not the underlying data. Seen as being of permanent value to the company.	Theoretically forever - so long as archive exists. No policy of culling. Whether will happen in future is a policy decision that will have to be made at some stage in future.	Make no distinction between electronic and paper version much depends on type of information	Varies because of wide range of data types and media.	No policy on time lengths.	Permanent – forever.
D With regard to 'living' electronic records/data sets that your organisation preserves is it intended to keep these alive for the foreseeable future?	Yes. Operate selection criteria. Assume material bringing in archived for as long as archive exists.	Not a frequent update, nothing that is constantly growing/changing. No true 'living' electronic records.	N/A	N/A	Yes <u>but</u> no decision on total length of time. Most information depends on depositor as holds copyright. Agreement to host items.	N/A
E When in an electronic record's/ data set's life is its life span determined?	N/A	N/A	Much of this determined before it is created – conforms to standard operating policy. During life time put some extensions on that.	N/A	Not at any stage at moment. No formal procedures – good sense to keep Web site up-to-date.	At start of electronic record's life, because users & creators have got to have responsibility under legislation.
F When are electronic records/ data sets identified as needing long-term preservation?	N/A	N/A	Defined points where have to archive information. e.g. completion of study.	When selected by people as having 'scholarly value' and included in our collections.	N/A	Appraisal. Reviewed and selected whilst in depts. = from planning system stage to 3-5 years after records created.

Table 3.8B Analysis of record life cycle issues	**7** Government Agency	**8** Advisory body for data preservation	**9** A national library	**10** Spokesperson from HE sector (personal view)	**11** Journal publisher	**12** Specialist archive
A How is the duration of the life cycle of an electronic record/data set defined by organisation?	Defined by cycle of data. Past data sets not worthless but lost currency as far as organisation is concerned.	N/A	Currently not defined. One of issues still being debated.	Defined in terms of printed documents. Electronic records run in parallel. Does not apply to WWW documents.	N/A	No policy on disposal – keep everything. Parent organisation developing policy on disposal.
B What factors influence the duration of the life span and how?	Data produced to inform central government departments. Can be cyclical – framed to answer requirements of commissioning body. External pressure.	N/A	N/A	Starting to look at issues: e.g. who responsible	N/A	N/A
C What are the estimated life spans of electronic records/data sets held by your organisation?	Don't know had to depend on someone else in past. Sit on something until – bored with it, no longer relevant, currency run out.	Not relevant.	Catalogue records by definition last as long as library. Same with other electronic materials but nobody in organisation clear what these are yet – duty to archive some material.	Life span of content of courses 7-8 years. Popularity of course affects life span. Tendency to produce shorter courses which able to update and change more easily.	N/A	N/A
D With regard to 'living' electronic records/data sets organisation preserves is it intended to keep these alive for the foreseeable future?	Some records are produced in living/dynamic environment this is how it moves forward. Releases = snapshots.	Not relevant.	N/A	Information stored on electronic conferences people who maintain servers control length of time kept before deleted.	If 'living' includes electronic journals – yes.	N/A
E When in an electronic record's/data set's life is its life span determined?	Determined by data grouping. Style of data production factor making it vastly different from most data sets. Not much ad hoc material.	Reliant on acquisition, selection & retention policy of institution Why acquiring Why retaining plus strategies from beginning to maintain.	N/A	N/A	At moment on day one it is forever – all records treated the same.	N/A
F When are electronic records/data sets identified as needing long-term preservation?	Not identified at moment In future system they would be flagged on release as requiring preservation and it will be part of metadata.	At point of creation and acquisition.	Acquisition stage – not currently being done.	Do not know – one of things to address within a digital archiving strategy.	N/A	Parent organisation's view encouraging archive to actively encourage people to think about preservation issues.

Table 3.9A Analysis of stakeholder opinions to creators' & users' needs	**1** Publisher	**2** Specialist archive (Respondent: 1)	**3** Company from science-based industry	**4** Distributed data service	**5** Advisory body concerned with networking issues	**6** A national archive
A Policy to protect organisation's rights as creators of electronic records/data sets from external users.	N/A	N/A	Yes.	Divided hypothetically into three: (1) Mechanical ways (2) Legal contractual access agreement (3) Type of material provided.	No policy – not been an issue to date.	N/A
B Does organisation balance needs of creators with users, and if so how?	Rarely deal with living authors. Situations with licensing rights where third parties have worries what customer will be able to do with material Very difficult in electronic age to give third parties absolute reassurances. All products paying products – good security at server end. Can' t check if something has been altered = major issue. Theoretically possible that it could happen.	Difficulty is balancing the needs and rights of the depositors against providing as free access as possible to users. Constant tension between the two.	Factors affecting balance: Small community designated people putting information in and restore it back again. Not imagine a wide community of users being able to look at the metadata. Customers given opportunity to collaborate with us.	Yes. Varies depending on creators. Majority creators = academics interested in collecting authors rights, credits plus if re-used/re-interpreted that, then service not held responsible for any re-interpretation. If have commercial depositors in the future need different criteria. Try to negotiate as wide access as possible plus make access process as simple as can for users whilst balancing needs of creators.	Some ways not much contact or knowledge of users – fairly common in Web area. People see organisation as a place to store reports and are referred from elsewhere = response to fact organisation had well developed Web site before some other organisations that they deal with.	Weighted towards users in the sense of the absolute need for authenticity and completeness and we believe this is best.
C Who are main users?	N/A	Academic use across no. of disciplines + use from outside academia.	Regulators, business users, legal challenge = lawyers. Need to go back to old data.	UK lecturers and students within higher education + many of services working within wider field.	Records preserve or hold of interest to: (1) Web spiders (2) People involved in research and development in library and information world – students and related groups.	Largest class = genealogists, then social and economic researchers, journalists, government departments, members of public with general interest queries.
D Who are main creators/originators?	N/A	Vast majority academic research, a lot material funded by funding bodies.	Scientists and laboratory workers.	Main category for acquisitions = funding councils and research bodies. Preservation plus improving access	Most information generated by organisation. Store reports of other organisations plus host mirrors.	Government organisations particularly major government departments.

Table 3.9B Analysis of stakeholder opinions to creators' & users' needs	**7** Government Agency	**8** Advisory body for data preservation	**9** A national library	**10** Spokesperson from HE sector (personal view)	**11** Journal publisher	**12** Specialist archive
A Policy to protect organisation's rights as creators of electronic records/data sets from external users.	Standard copyright or crown copyright – bit outmoded try to establish precedents.	Not appropriate as not collection based.	Information currently on Web site - truncated version of full MARC record which users apply and pay for.	Not formal policy – in practice not accessible for external use. If owned by organisation restricted access unless sold to somebody else.	Written copyright laws [cover] all articles produced in journal = copyright assignment forms – originator assigns to organisation the rights to reproduce article in print or electronic form.	N/A
B Does organisation balance needs of creators with users, and if so how?	Yes, users always want to know more than able to tell them, want more detail quicker than can get it to them.	Do not think we have to balance the needs. We have to say what is best practice to creators, institutions charged with preserving long-term and in terms way users are allowed to use materials. Have to work within capacity of individual institutions – help it take incremental steps towards best practice.	N/A	Comments refer to electronic archive of printed course materials. At moment set up for creators not users. People who want to create documents may re-use their own materials, but not users such as students and people external to organisation	Support copyright. Concerned about user breaking copyright whether through photocopying or more seriously when material available electronically.	N/A
C Who are main users?	Us, mixture between internal researchers and academic researchers plus some commercial usage.	[Comments which follow relate to the organisation's Web site] Libraries, archives nationally and internationally, anyone concerned with preserving records electronic or traditional.	Researchers accessing catalogues.	Academic archive = academic and production staff. Students marginal.	Academics at institutions, libraries, colleges, individual academics, similar companies libraries.	Five years ago = small core community of specialist users. Recently under graduates and postgraduates. Subscription or registration will help to profile users.
D Who are main creators/originators?	Statisticians – ask them to take responsibility for data they say they can release.	Originators from organisation or parent organisation.	Ourselves plus one or two institutions creating catalogue records.	Academic and production staff.	Research academics, research authors, PhDs.	Past almost exclusively academic Now Web site = casual users.

of the creators and those of users is greatly influenced by the nature of the stakeholder's enterprise, as this is a main factor influencing the degree of public and in-house access to their records.

According to the stakeholder from science-based industry it is clear that they do have measures in place to protect their rights as creators of data that they deem to be commercially sensitive and have a high financial value. Access is restricted to a closed community that may include customers who are given the opportunity to collaborate. A future framework for the management of DDOs would need the capacity to accommodate a range of access restrictions, from open through to the most stringently restricted access, to encourage private sector companies to participate in such a scheme.

A distributed data service has indicated how protection of their rights could be achieved. The first is through the use of mechanical mechanisms such as electronic management systems, copyright and electronic watermarking, the second is to establish legal contractual access agreements and the third by providing different levels of quality of material or images, such as a thumbnail image of a work of art. A stakeholder from a specialist archive remarked that there is constant tension between balancing the rights of creators and giving as free access as possible to users. As far as a national archive is concerned the relationship is weighted in favour of the users.

Over the whole sample of stakeholders there is a wide range of users, which serves to emphasise the importance of the long-term management of digital data to society. These users include: academics, scientists, students, research and development departments within a specific community, journalists, government departments, historians, genealogists and casual users.

The above users can also be creators of information within their own organisation or within different information dependent organisations, and unless an organisation restricts its user base on a geographical basis in accordance with its aims, these users can be international. Although this data demonstrates current usage it is impossible to predict with a degree of certainty, future end-user

requirements. This inability to predict future needs, some would argue, is the justification for retaining all DDOs especially as storage costs are falling all the time and the selection process itself could be a costly exercise in terms of staff resources. However, the scales are tipped in favour of positively selecting DDOs for long-term retention because if everything which was created digitally were retained it would become a greater task to retain, monitor and retrieve the data. There are identified disadvantages of zero selection (see Section 8.6.4). As with paper-based systems, as the number of objects stored increases the need to have some means of identifying each object becomes increasingly imperative, that is the need for metadata increase. In the digital environment, this is true to an even greater degree due to the intangible qualities of DDOs and the challenges this presents (see Section 1.6 to 1.6.11). Therefore, it would seem that metadata also becomes essential for the long-term management of DDOs. What cannot be identified, cannot be managed over the long-term, accessibility and long-term management are one and the same (see Section 1.6.1). The levels of access to DDOs differ according to the remit of stakeholder organisations (a discussion explored in Section 4.11.1).

It is seen that the balance between creators and current users is linked to a stakeholders' remit. To maintain future access to DDOs this research considers that it will be necessary to develop guidelines to balance the needs of different stakeholder groups. Establishing such guidance appears to be a grey area that merits further investigation (see Chapter 6).

3.11 Future developments

As the interviews focused on the stakeholders organisation, interviewees detailed specific initiatives that were being developed, but few mentioned cross-sectoral co-operation, or the need for an integrated approach to digital preservation challenges. The need to collaborate with organisations within the same field would it is felt allow similar experiences/challenges to be tackled, but also, as this analysis hopes to demonstrate, it is beneficial to involve a wide range of organisations who see their information as a valuable resource in these discussions. This would

encourage a co-ordinated approach to the long-term management of DDOs and reduce duplication of effort.

3.12 Summary

The conclusions of this analysis indicate that a variety of approaches to digital data management exist and that by combining best practice from a range of stakeholders, with an examination of the issues involved leads to a possible way forward for the management of digital data. In summary, the tabulated findings from the interviews indicate that a range of challenges are affecting stakeholders, many of which are interrelated. There is a need to develop legislation to:

- Ensure the deposit of publicly available DDOs.
- Provide safeguards to ensure that deposit institutions fulfil their obligations.
- Resolve rights management issues.
- Increase stakeholder confidence in the quality of another stakeholder's data.
- Increase stakeholder confidence in the management of their data by other stakeholders.
- There is a need for training staff to acquire additional skills to meet the new challenges presented by DDOs.
- There is a need to encourage organisations within all sectors of the information community to actively consider developing written policies and selection criteria for the preservation of DDOs where applicable.

Chapter 4 utilises SSM to develop a rich picture (as detailed in Section 2.5.2) which enables representation of stakeholders' activities to be shown pictorially and to indicate the type of interaction that was seen to occur between them.

Chapter 4

Analysing stakeholder views - a Soft Systems approach

4.1 Analysis of the relationships between stakeholders

The analysis in Chapter 3 has looked at both broad ranging and specific issues for the stakeholders interviewed, determining similarities and differences in what practices are being undertaken, and the reasons for and appreciation of why they are being undertaken. Chapter 4 builds on this analysis using Soft Systems Methodology (SSM) approach to focus on the relationship between stakeholders, drawn from the opinions they expressed, in response to questions tailored specifically to their organisation.

Through involving the various stakeholders, this methodology hopes to produce an outcome that is in tune with both the composite of stakeholders' views and the setting of these within the context of prominent literature in the field from various countries. The construction of a rich picture aids a researcher to build up an holistic view of the 'problem situation', which takes the stakeholder views into consideration. As with systems diagrams, symbols are used to convey concepts that are difficult to express succinctly in words and to aid the reader in following the picture.

4.2 Description of the rich picture

The extraction of critical trends from the tabulated interview data

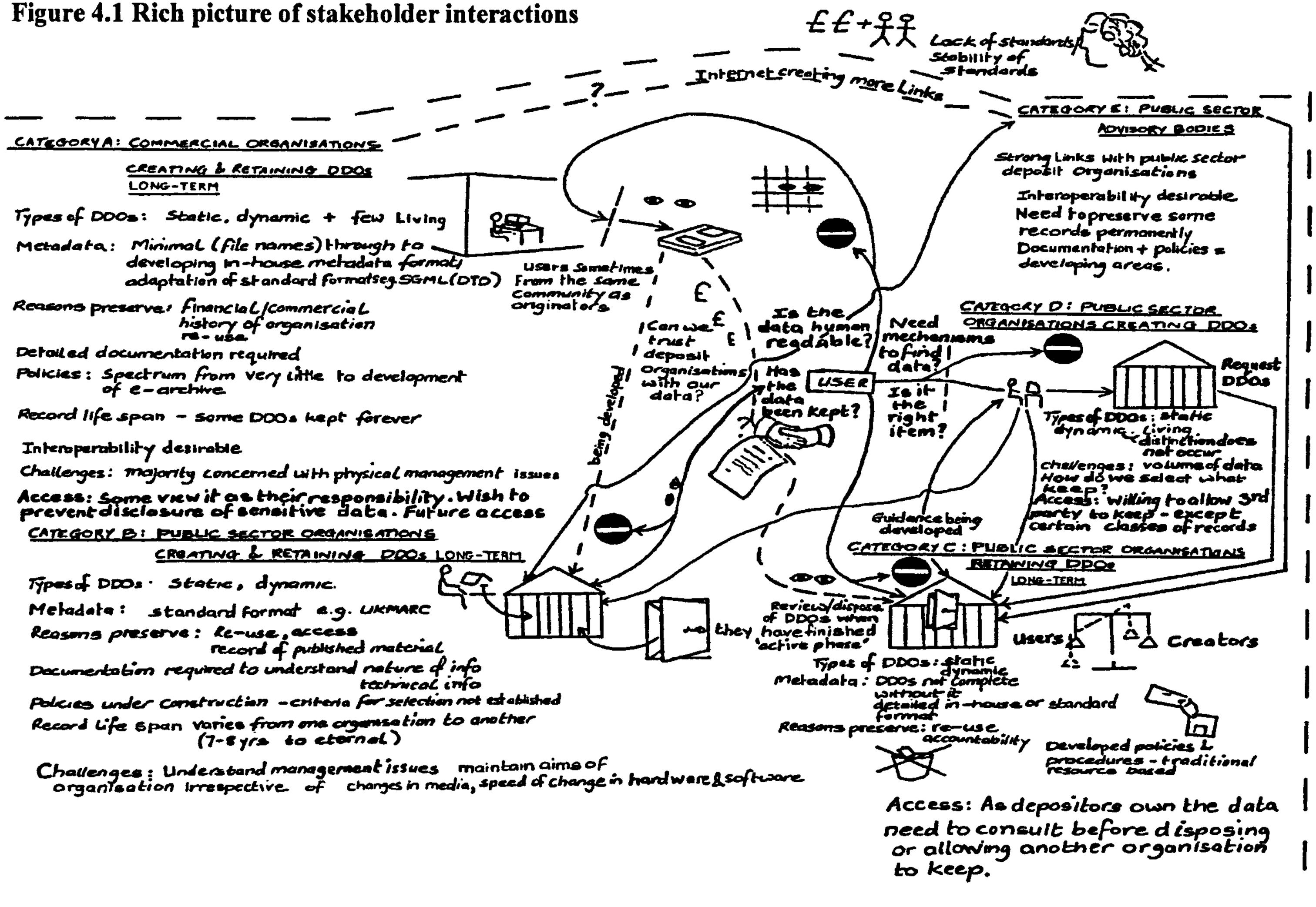
together with information unique to individual stakeholders, was combined and placed in a succinct form in the rich picture (Figure 4.1). This picture enabled the formation of a composite impression of the research problem, as expressed by the stakeholders at a particular point in time.

The constructed rich picture showed areas of consensus and areas where divergence of practice existed, and critically for this research, it also identifies areas where there was a lack of co-ordination. The broken lines and question marks in Figure 4.1 represent the lack of co-ordination existing between different sectors of the information community, some of whom had not forged links in the traditional paper-based environment. The principal external pressures common to stakeholders include, finance, human resources, the lack of standards/stability of standards and legislation (as shown in the top right of Figure 4.1). Through analysing the rich picture, it was possible to gain information on the personal and corporate views of stakeholders that quantitative methods, such as statistical analysis, would probably have failed to uncover. It was therefore possible to define the research problem with greater clarity by developing a root definition (as discussed in Section 2.5.2).

4.3 Categorisation of stakeholders

The use of SSM to create a rich picture enables the categorisation of stakeholder organisations into relatively homogenous groups, based on their approach to the long-term management of DDOs. This categorisation aided the understanding of the interaction, or lack of it, between the various groups. The analysis of the rich picture shows that it is possible to sub-divide the stakeholders into one of the following five groups, on the basis of whether the primary concern of the stakeholder is creation, retention, or a mixture of both. There is also a fifth group that is purely advisory.

Figure 4.1 Rich picture of stakeholder interactions



A. Private/commercial organisations creating and retaining DDOs long-term
B. Public sector organisations creating and retaining DDOs long-term
C. Public sector organisations retaining DDOs long-term
D. Public sector organisations creating DDOs
E. Public sector advisory bodies

Although all organisations retain some DDOs, this research is focused on long-term preservation, and in this context refers to a minimum of seven to eight years to perpetuity, depending on the DDOs concerned and the stakeholder's remit. To construct a detailed view of how these categories intersect, it is necessary to examine the characteristics of each (as shown in Figure 4.1). This process prepares the ground for the clarification of the problem, and the ability to express it as a root definition.

Category A includes representatives from publishing and from science-based industry, where the overriding concern is the commercial value of the information they create and the need to retain control of this information for commercial reasons. These organisations recognise the value in developing a long-term management policy for re-use of these resources. In some cases, work is already in progress to achieve this end.

Category B contains public sector organisations that both create and retain DDOs long-term. These organisations include one educational institution that is producing course material and depositing this within an archive, whilst another stakeholder creates catalogues of their digital data holdings which are disseminated widely, although at the present time acquisition is not supported by legal deposit legislation.

Category C, public organisations retaining DDOs long-term, comprises organisations in the archival community that have traditionally been concerned with maintaining a record of specialist subject fields, or a more broad based national record of unique objects. These public sector organisations which are retaining DDOs can be sub-divided into those holding the original DDO, and those maintaining access to them through mirrors and pointers. Within the public sector, there are also category D

organisations, which create DDOs, but do not themselves retain these objects. This category applies to government agencies that create large numbers of DDOs but pass on the responsibility of preserving the majority of DDOs to national repositories.

Advisory bodies within the public sector constitute category E. These organisations may create DDOs, even if it is only their own Web pages and advisory documents. However, both of the organisations studied were aware of the need to manage digital data, and were actively seeking opportunities to create collaboration between different stakeholders.

4.4 Questions arising from the rich picture

Analysis of the rich picture established areas of consensus and also areas conflict that require resolution. The questions listed below arose from the analysis of the rich picture. Each of these is discussed in the sections that follow.

- How can data in standard, in-house and legacy metadata formats be managed over the long-term?

- How can the balance between issues of data access and rights of access be achieved?

- How different are the systems for the creation and long-term management of data in the electronic and traditional environments?

- Who can data creators have confidence in to manage their data?

- Is the process of retaining DDOs driven by technological or organisational factors?

- Does a boundary exist between technological and organisational factors and if so where does it lie?

- Do stakeholders work with other sectors within the information community to solve challenges?

- How can the issues of disclosure, copyright and rights management be resolved?

- How do stakeholders select materials for long-term retention?

- How can costs be determined?

- What form should a future framework take for the deposit of DDOs?

4.5 Standard, in-house and legacy metadata formats

The rich picture illustrates that there is a division between stakeholders who use standard systems and those who have, or are in the process of creating, in-house schemes. This division does not appear to correspond consistently with the type of organisation, as can be seen in Figure 4.1, where most categories, apart from libraries, have presented examples of both standard and in-house schemes. At this point it is worth noting that some in-house formats may have been adapted from standard formats, therefore the distinction between the two is not as marked.

The ability to control the management of digital data includes a consideration of the metadata requirements. A national archive identifies the development of a records management system as a challenge which entails organising and setting up enough metadata on the records initially, to enable them to be managed and used to determine what an archive holds. In this stakeholder's view, once this has been overcome, it is

then a matter of preserving the record. Tied to the development of schemes is the perceived need for interoperability.

There is also a division between organisations, those with data in legacy systems and those without, which needs to be taken into consideration in any future long-term management framework for DDOs.

One stakeholder in the commercial sector (category A) has an in-house metadata scheme and another, a journal publisher, uses the international standard SGML with a customised DTD. This generation of in-house formats could indicate that, at the time their systems were created there was a lack of standard formats flexible enough to accommodate individual organisation requirements. As these stakeholders are primarily concerned with the creation of DDOs for commercial reasons, the retention of DDOs would appear to be driven, primarily, by the potential to generate extra income through re-use. Exceptions to this apply where regulatory bodies require information. Although, the stakeholder from science-based industry commented that metadata was not currently created for DDOs, there is nevertheless a developed set of requirements that any future system would need to fulfil (see Section 3.6).

The public sector organisations (categories B, C and D) also exhibit a range of different options, but, unlike category A there are more sub-groups, possibly because they comprise a larger group in numerical terms. The first sub-group are libraries that utilise the UKMARC format, however, an Advisory body concerned with networking issues has opted to develop Dublin Core (to demonstrate the formats feasibility). The second sub-group comprised entirely of stakeholders from the archival community, where there is an even split between organisations using an in-house format and those with a standard format. This division occurs irrespective of whether the archive is part of a distributed service or a centralised archive. The third sub-group, represented by a government agency (category C), has developed an in-house format to cater for its needs. Developing interoperability between metadata formats, although seen to be beneficial by the majority of respondents, raises questions of achieving this at the

metadata level (see Tables 3.5A and 3.5B, row G and also Section 5.6.2 which discusses this issue in more detail).

The composite view emerging from stakeholder opinions suggests that metadata is crucial to managing DDOs, but that its role and composition appear to be influenced by organisational aims, evidenced by the number of different formats tailored to suit the needs of stakeholder organisations.

4.6 Issues of data access and rights of access

The issues of data access and rights of access are seen as central issues to the long-term management of DDOs. The in-depth interviews with stakeholders and the collation of material obtained from these which is presented as a rich picture, shows the access restrictions which organisations will tolerate. These restrictions are shown in the rich picture as no-entry signs, and eyes behind bars (Figure 4.1). Access to DDOs has been a primary focus in the development of metadata formats as evidenced by the current concerns surrounding the development of formats such as Dublin Core. This focus on resource discovery needs to be followed through into the long-term management of DDOs. Long-term access is seen to be the primary goal of managing DDOs. Policies and procedures to facilitate this goal would appear to exist within stakeholder institutions and are at varying stages of development. Strategies on a community wide level however, would appear to be absent although this situation is changing.

A co-ordinated approach appears to be absent in the following crucial areas:

- The management of a diversity of DDOs over the long-term, which addresses the crucial concerns expressed by stakeholders, namely the need for general guidance, guidance on specific data types and the resolution of rights management issues (see Tables 3.3A and 3.3B, cells 4A, 7A and 10A).

- Mechanisms to support future end-user access.

Through analysing the rich picture, it is evident that commercial publishers in category A would be willing to allow their data to be preserved by another body, once it has ceased to be of commercial value to them. However, this acceptance is conditional on the degree of accessibility, and rights, such a body would have to their data. To satisfy the requirements of a stakeholder from science-based industry they would demand that a third party, such as a vendor, could provide assurances of the security of data within their systems. Some organisations would not relinquish any DDOs deposited with them (public organisations in categories B and C), as it is within their remit to retain them for the national interest.

Many organisations expressed the view that they would not necessarily have confidence in the records of a future world preservation database as sole proof that a work exists discarding their own copy or copies on this basis. This does seem to suggest that these organisations require a degree of control over the DDOs they have created, or are expected to maintain, as part of their organisational remit. This points to the need for a universal framework which would allow control of the actual DDOs to remain with the stakeholders, whilst providing co-ordination of intellectual preservation standards and the production of guidelines. This framework would need to accommodate the long-term management of both scholarly and non-scholarly DDOs.

A review of such stakeholder control appears to require a thorough examination of rights management issues, to find strategies to overcome areas of difficulty. These issues include responsibility for retaining information, legal deposit, copyright in the electronic environment, disclosure, authenticity and integrity.

The remit of stakeholder organisations is seen to influence opinions on access to DDOs (see Table 3.7A and 3.7B). Such opinions raised the following issues.

- Who may view the data
- The purpose of access
- Duration of access.

The nature of the stakeholder's organisation affect who the creators and current end-users are (see Tables 3.9A and 3.9B, rows C and D). The stakeholders *raison d'êtr*e therefore influences the amount of control they 'expect' to exert over access. The duration over which a stakeholder expects to provide access to a DDO is again linked to the organisational aims. For example, a national archive or library expects to retain DDOs in perpetuity, whilst a government agency wishes to relinquish this duty for all but the DDOs it must legally maintain, which has an impact on the length of time over which records are maintained. However, a number of other factors are seen to influence the duration and degree of access to DDOs, these include organisational resources and external pressures, such as regulators. In addition, access is linked to disclosure, copyright and rights management (issues discussed in Sections 4.11 to 4.11.3).

4.7 The effect of differences in the systems for the creation and long-term management of data in the electronic and traditional environments

The publication and long-term management of data in the traditional print-based environment is characterised by formalised processes, supported by legal structures that safeguard interested parties' rights, and reinforce the deposit of data. In order to gain public credibility for their work creators submit this for examination through processes such as peer review. Informal processes are however open to creators, such as commissioning a private printing of their work.

Systems for the creation and management of digital data are seen to vary from the publishing practices for traditional materials. The principal distinction is the lack of formality of the processes, with virtually no standards governing both the creation and

retention of records; this is noted as one of the constraints shown outside the broken line containing the problem situation in Figure 4.1. The processes that presently exist cope with static DDOs that are relatively similar to paper-based resources, but it does not take into account the monitoring and documentation required to ensure continued access to dynamic and living DDOs.

The interviews suggest that the majority of stakeholders are dealing mainly with static resources, many of which are the products of digitization projects, and few are meeting the challenges posed by dynamic and living DDOs, this is illustrated in the rich picture by broken lines and question marks. This may be because the quantities of these resources that are being generated or retained at present, are low, and a conceptual framework to deal with these resources has yet to be developed.

There is seen to be a need for a legal structure and a conceptual framework to facilitate a co-ordinated approach to managing objects in the digital environment. Support for a new framework is evidenced by a spokesperson from an advisory body concerned with networking issues who stated that, "...some people are trying to impose a traditional publication model on the Web, it breaks down at many points, it is not the same". The term published is defined as making material available to the public. In the electronic environment it is not essential to use the services of a publishing house so managing access is more difficult to achieve. Even the term 'published' in the electronic environment can mean anything along a continuum, from formal processes, such as the peer review of papers for referred electronic journals, to very informal means, as in the case of creators producing Web pages. The construction of the rich picture enables the researcher to see the stakeholders' views not only in relation to each other, but also in relation to the influences of the wider environment. Some of the issues concerning the selection of material for long-term retention are broadly similar for both traditional and electronic resources, however, others are not, and in this respect require the development of a more proactive management structure and procedures for retaining DDOs (see Section 4.12). Differences do occur in the type of information required to manage the electronic material long-term, and also in the

mechanisms to implement this, namely the type of metadata, where it is held, and the role of stakeholders and their collaboration/co-operation with each other (see initial stages of Figure 7.1).

4.8 Stakeholder confidence

4.8.1 Stakeholder confidence in third parties to manage their data

The rich picture highlights the fact that creators have reservations about relinquishing their DDOs to third parties such as deposit organisations, or commercial vendors (see Section 3.7). The rich picture shows that some classes of DDOs would not be released to third parties, or into the public domain (demonstrated by the no entry signs in Figure 4.1), while the access to other DDOs may be restricted (shown by the eyes behind bars in Figure 4.1).

The reasons stakeholders give for not allowing others to manage their data vary. Some saw it as part of their remit to retain their own records (for example public organisations categories B and C), whilst others saw it as a matter of maintaining the security of records that have commercial value (for example the stakeholder from science-based industry, category A). Where stakeholders will allow their DDOs to be preserved by another body, acceptance is dependent on the rights of access which such a body would control. These results show that the issue of ownership of DDOs at different stages in their life cycle is identified as a crucial factor affecting long-term preservation. By adopting an SSM approach, it is possible to identify the views of stakeholders on the life span of DDOs, both between certain categories and within the same category. To resolve ownership issues requires an examination of areas such as legal deposit, disclosure, authenticity and integrity.

It would seem essential to construct mechanisms that provide protection for creators against the misuse of their data by such parties. These rights management

mechanisms would also be required to provide a framework for deposit organisations to administer the requirements of creators. There are developments taking place within the archival community such as the Arts and Humanities Data Service's *Digital collections: a strategic policy framework for creating and preserving digital resources* (Beagrie and Greenstein, 1998) and the CURL Exemplars in Digital ARchiveS (CEDARS) project in the research library community[1].

Many of the factors being discussed in this analysis of the rich picture are inter-linked, exemplified by the influence of disclosure issues over access and rights management. The mechanisms that exist at present, such as those which can be embedded into Web pages to prevent information being misused, are seen as being in need of further development. Some stakeholders are developing mechanisms, such as restrictions in the metadata, to control the resources that can be viewed. This relates to the copyright for digital material which stakeholders also identify as an area requiring reassessment and clarification[2].

4.8.2 Stakeholder confidence in technological developments

The speed of technological change is causing concern amongst some stakeholders, they are reluctant to commit the organisation's resources to a particular system, only to find that it is superseded in five years time by another system which has become the industry standard. The development of co-ordination mechanisms at a higher level than metadata would seem to be necessary; this would allow formats that contain residual data to be accommodated. This shows the benefits of SSM in helping to discover if the long-term management of digital data is driven by either organisational aims, or technological developments, or a combination of both.

It would seem to be advantageous to develop non-proprietary systems, or standard

---

[1] URL: http://www.leeds.ac.uk/cedars/

[2] Since the majority of these interviews were conducted, the Statutory Instrument, *The Copyright and Rights in Database Regulations 1997* has come into force (1 January 1998) and the Green Paper, *Crown copyright in the information age*, published.

systems, which exhibit sufficient flexibility to be customised. Long-term management of DDOs within systems is paramount to their longevity; to facilitate this, it is important to develop higher level metadata that transcends changes to the software and hardware systems available at any one point in time. This higher level information would link to the descriptive metadata held about a DDO and provide the administrative metadata, or point to it where it exists, to a sufficient standard at the metadata level. This would also facilitate deposit organisations in monitoring living DDOs earmarked for retention prior to their deposit, and encourage a spectrum of stakeholders creating such resources to deposit them at a predetermined point in the future.

4.9 Relationship between technological and organisational factors in the process of retaining DDOs

From an analysis of the rich picture, it would appear that stakeholders were concerned with physical preservation challenges, such as technological obsolescence. Concerns for technological obsolescence centred primarily on how stakeholders were to maintain a suitable medium to support each DDO and the cost implications which this involved. The possibility of separating the message from the medium enables the emphasis to change to focus primarily on the preservation of the message and a secondary need to consider the medium supporting a DDO. It would appear that stakeholders continue to use traditional approaches in the absence of new strategies more sympathetic to the management of digital resources. The rich picture shows that the boundary between organisational and technological factors vary from one stakeholder to another, as their technical expertise and financial resources vary.

Organisational challenges were also identified by some stakeholders as areas that affect the success of initiatives in the field of long-term preservation. These

challenges can be classified into two levels, those affecting the information-dependent organisations in general, and those specific to particular organisations.

Addressing the first of these categories, the rich picture demonstrates that limited co-ordination between stakeholders exists at present (Figure 4.1). As stated in Section 4.7, this occurs mostly through existing structures developed for traditional resources, as no proven model exists for the management of DDOs. Newly formed organisations are, however, attempting to rectify this situation by developing new structures that more accurately reflect the requirements of managing digital data. The rich picture also demonstrates that organisations are concerned about insufficient solutions to technical issues although they are attempting to resolve them. The technological factors could become less important, as these are less intractable, and the organisational factors could increasingly drive the process.

The following view from an advisory body concerned with networking issues underlines that both technological and management issues play an influential role.

> If you think of conventional cataloguing you have agreements about technical stuff. You have agreements about particular services, some commercial, some collaborative. Whatever, and you have agreement about particular organisational patterns in the same sense that you have some understanding of co-operative cataloguing, some understanding of straight forward commercial relationships and a market with particular business models in operation. So it seems to me in this new environment you have none of those things. You do not have technical agreements, you do not have services that people are coming to expect, and you don't have particular business or organisational models for dealing with this type of thing.
>
> It seems to be the issue of how you begin to manage large collections of electronic information objects, whatever they are, whether they are maps or engineering models or documents, is something about which we don't really yet have a lot of experience in an open network environment. (Advisory body concerned with networking issues)

The stakeholders comments relating to organisational challenges within their own organisation, revealed a number of areas of concern. The definition of roles and

responsibilities within the organisation, was evidenced by a spokesperson from the higher education sector, who in his personal view identified the major challenges to developing a digital archive as:

> ...who is responsible for it within the organisation, especially because there is a history of distributed responsibility for the physical archive? Who is going to take on the role of keeping an electronic archive?
>
> I think the biggest challenge is the management, to be honest, to understand the issues and to take the right steps in order to be able to preserve materials in the longer term. (Personal view of a spokesperson from Higher Education sector).

A government agency saw the need to change the culture within the organisation, so that DDOs are seen as the definitive record. This concurs with the government view to ensure electronic records become the definitive version. Even where policy and practice have been developed to manage DDOs within specific organisations they have still to become routine, and issues of cultural change within the organisation are still viewed as a challenge.

4.10 Approaches to the resolution of challenges within the information community

Some of the challenges presented in managing DDOs would seem to require a collaborative effort to find solutions. What seemed to be in its infancy, at the time the interviews were conducted, was a collaboration between stakeholders from diverse backgrounds (such as archivists, librarians and records managers from commercial enterprises) working together to create a co-ordinated framework. As a result, there were individual organisations attempting to achieve 'best practice' for their enterprises, which might have benefited from collaborative interchanges of experience and challenges faced by other types of organisations. This may be changing, with initiatives emanating from the Joint Research Libraries Group and National

Preservation Office conference *Guidelines for digital imaging* (National Preservation Office and Research Libraries Group, 1998), which did highlight preservation metadata and digital archiving issues.

Stakeholders also highlighted that there is a danger that resources, which are not seen as scholarly, or having commercial interest, may not be preserved and therefore lost to future generations.

> ...people will create metadata for data that is of commercial interest and somebody is paid, a librarian, archive, an academic, or where the stuff is of commercial interest and somebody reckons that they can sell it. That part of the selling would be to create a product description, or to make sure it has some value or packaging. There are a lot of the things you might want to preserve that don't come into either of those categories. (Advisory body concerned with networking issues)

A co-ordinated approach may ensure that more of these DDOs are managed over the long-term. The same stakeholder identified that the volume of resources in digital form is increasing, and as this occurs, it is becoming increasingly difficult to know what to preserve and what to discard. This would seem to underline the need for selection guidance, and collaborative efforts would be helpful in developing a conceptual framework, where responsibilities for implementing selection criteria could be established, and difficulties in maintaining DDOs resolved. Some stakeholders are all too aware that these issues are not easy to resolve, "...even if you look at the British Library or other national libraries, thinking about their legal deposit obligations, it seems to me, even if you narrow it down to that, it is still quite open and complicated" (Advisory body concerned with networking issues).

4.11 Rights management: issues of disclosure, copyright and legal deposit

The ownership of DDOs, and the protection of owners' rights would appear to be crucial areas requiring resolution in the management of DDOs. The rich picture illustrates the manifestation of this: a concern over disclosure of information and lack

of legal deposit mechanisms (shown by the broken lines between creators (A) and repositories (B and C)). This is combined with the comments from responses to questions on copyright. These specific issues will be discussed in Sections 4.11.1 to 4.11.3.

4.11.1 Access

Determining access restrictions is seen to involve negotiation between stakeholders, particularly where organisations are concerned with maintaining public access to DDOs, as the following quotation shows:

> A lot of the time we are brokers between the creators and the users, and all we can do really is act as middlemen between the two. (A specialist archive)

If organisations have not developed co-ordinated deposit agreements in the past, but rather have entered into one-off agreements, this can cause problems when access restrictions change due to technological advances, such as increasing access to DDOs by providing Internet access.

4.11.2 Disclosure

The issue of disclosure is inter-linked with issues of access, copyright and rights management. At present the mechanisms which enable controlled disclosure are at an early stage in their development as an advisory body concerned with networking issues noted:

> From a technical point of view, you are going to have to begin to have agreement about what types of things people will want to disclose about these particular collections they have, and how they will disclose them. I think, increasingly, people will want to disclose information about their own metadata in a more organised way. (Advisory body concerned with networking issues)

According to this stakeholder organisation, disclosure mechanisms such as embedding pages with Dublin Core metadata were said to be, informal and unregulated. Some specialist archives place embargoes on particular data sets at the request of depositors and also anonymize the data. The disclosure of some DDOs requires it to be completely restricted for a defined period of time, as defined by copyright laws, although these are ill defined for the digital environment, as discussed later in this section. One distributed data service has developed three categories of access to its services, namely, a free or open category, a category where material is available for registered users and finally, a restricted category. Changes in access restrictions such as these can entail archives in consultation with depositors to redefine the access restriction(s) that apply to their resource(s). It is also noted by one specialist archive that their practices, had for historical reasons, not always adopted the latest technologies:

> ...the legacy of the history of the organisation has not exactly prevented new technologies being used, but things have not progressed, you usually have to work backwards to tackle individual issues within the data sets. (A specialist archive)

There is seen to be a need for the reappraisal of the existing legal tools of copyright and legal deposit, as witnessed from the findings detailed in Chapter 3. The comments of the majority of stakeholders pre-date the publication in the United Kingdom of the consultation document, *Crown copyright in the information age*. In the absence of universal guidance, individual stakeholders have attempted to develop and implement measures to negotiate copyright clearance. In one specialist archive this has been an evolutionary process, each stage being more rigorous than the last.

In general, the absence of parameters governing copyright in the electronic environment has lead to confusion and a plethora of different approaches to overcome situations, such as negotiating copyright clearance in a situation where multiple copyright claims exist, and creating access agreement documents. For example, a distributed data service is working towards a common deposit agreement as part of

its rights management framework, which at present is on a voluntary basis, except in cases where funding bodies recommend fund holders to deposit material. This agreement requires users of the service to sign a copyright contract, and makes it clear that the service is not in a position to pursue infringements of copyright, though this does not prevent individual depositors taking legal action. Where data sets are purchased by the archive service then the process is more stringent. In addition, the archive service is establishing written guidance for good practice on the copyright of digital materials. Similarly the journal publisher contacted has adopted a number of strategies to protect their data, such as requesting creators to sign copyright assignment forms, displaying copyright notices on electronic publications and actively participating in copyright committees. A specialist archive noted how some learned societies actively encourage the deposit of their material, and the ability for it to be viewed on the archive's Web site, in the belief that wider circulation of the electronic version boosts their paper-based sales. This same archive has expressed the view that managing some material which was of commercial origin was less of a problem when their metadata was not as widely available, but commercial depositors became concerned when mirrors of the site were developed.

Some specialist archives are less affected by copyright than others by virtue of the material they preserve, for example, some are concerned with historical material that is outside the 100 year copyright ruling. Some respondents expressed a personal view, feeling that copyright insufficiently supports their organisation's needs. A government agency noted that copyright had not kept pace with technology, especially with regard to the Internet.

A spokesperson for a specialist archive felt that existing copyright legislation was not very workable in their context as each deposit 'is subtly different'. The spokespersons from this same organisation also voiced some of their concerns and insecurities relating to current copyright legislation. This is shown in the following quotation.

> It is difficult to know what the existing copyright law is at the moment, that is part of the problem, that is why people have got absolutely confused about all

> this. They need to have ten very basic simple sentences on what copyright means, then people may start to really understand this sort of thing... There are too many grey areas and that is why people are sceptical about even using it or using it properly. (A specialist archive)

An added dimension to the intellectual property right issue occurs where stakeholders both retain DDOs acquired from depositors, and create DDOs in their own right. For example a national library may acquire DDOs through voluntary deposit and also publish DDOs such as research reports in both electronic and paper form. Therefore, reform of current copyright legislation would need to address both of these roles, as evidenced by the following corporate view:

> Certainly, as far as we are concerned, we would want to see the intellectual, the IPR of producers, publishers and others safeguarded, in order to assure them of our good intentions; and of course in certain areas, where we are ourselves a producer and a publisher, we would want to maintain our slice of the action. (A national library)

4.11.3 Legal deposit

The absence of legal deposit legislation, specifically for electronic publications, is of concern to stakeholders with a remit to maintain a record of material published in the United Kingdom. Establishing such legislation would appear to require the resolution of specific collection policy issues, such as, defining what for them is a United Kingdom publication, and developing mechanisms to acquire living DDOs.

> ...if one is to clarify exactly for us, what exactly is meant by a UK publication. What kind of process is going to be involved in maintaining that? For example, if they are in the living electronic record area, what kind of snapshot are we going to be taking of this material and what kind of additional information in terms of data structure etcetera, which are very often of course, commercially in confidence information. (A national library)

The extension of legal deposit legislation, as a regulated process for digital objects, is also deemed to be beneficial by a preservation organisation, stating that in their view,

"It can't be left as a voluntary deposit" (A national library). This research advocates that such an extension to legal deposit cover all publicly available DDOs, including those which do not have paper-based equivalents, such as online databases. Development of precise boundaries are a matter for further research.

The complexity of issues surrounding this extension is documented in the literature (see Section 1.6.3). The principal driver behind the developments are the public good arguments for the transmission of culture, and the potential to increase societal productivity. It would appear that there are a number of grey areas that are contributing to the absence of a formalised process for legal deposit, these include; insufficient definition of the roles of stakeholders, the stated parameters of such legislation, and the mechanisms for achieving the deposit of different types of DDOs.

4.12 Selection criteria for long-term retention

The need for selection processes would appear to be an integral part of the management of DDOs over the long-term, for organisations with a remit to retain published or unique material for current or future public access. However, transferring this ethos into the digital environment is presenting challenges, as the corporate view of a national library demonstrates:

> There is a constant process of intervention and renewal, that means that our view of the selection process for preservation may have to be far more rigorous and sparing than perhaps it has been in the past with paper material. Now nobody, we understand this is the case, has really yet thought through what the real implications of that are, both for us as an archive and of course for the future researchers who want to use the archive. That, clearly, is going to be an issue we have got to think about over the next couple of years. (A national library)

This view indicates important questions that remain unresolved, a view supported by a spokesperson from an advisory body concerned with networking issues who stated

that fundamental questions remain unresolved, "The question is who decides what to preserve and who preserves it, which is a fundamental problem which no one has really answered" (Advisory body concerned with networking issues).

Although as Section 3.5 shows, selection policies and practices for long-term management of DDOs is in its infancy, some stakeholders are considering issues such as when to select DDOs, even if this is only theoretical at the moment. The advisory body concerned with networking issues quoted above also noted that selection was, from an organisational point of view, best undertaken at creation as this was the time when most information about the object was likely to be available. For example, the decisions taken at the creation of Web documents are not in the respondent's view irreversible, "...you could say that every time you change a document, a consideration of its longer term preservation should be made. In practice this is not going to happen, it is the nature of the Web" (Advisory body concerned with networking issues). This stakeholder organisation indicated some of the factors felt to be important in determining which resources to select.

> A national library would have a mission to want to collect for preservation, certain things, and that mission would translate into certain criteria. It might be possible to make decisions against those criteria, based on metadata, or to be informed by the metadata in some cases and in some cases, not. Then there might be other criteria which would not allow that type of thing, and then also if you are in this open and uncontrolled environment, there are a lot of things you might be interested in but they will not have anybody there to create metadata for them. (Advisory body concerned with networking issues)

The diversity of activities that contain data is growing, and the ability to sift through, to determine which of these DDOs have long-term value, becomes an increasingly difficult task. This point is noted by the following quotation from one stakeholder,

> ...there is an informational component to more and more of what we do, what you might want to select and keep becomes bigger and bigger, and it becomes much more difficult, it seems to me, to know what to keep...you are not just looking at formal literature or particular archives or collections, you

> are actually looking at the informational component of a whole range of activities. (Advisory body concerned with networking issues)

The same advisory body identified the lack of snapshots of global networks as a recent example of how information might already have been lost which could potentially have been interesting. It was suggested that these snapshots would have given the ability to look back at the early history of the World Wide Web and Internet discussion groups.

When a national archive was questioned on the possibility of selecting evolving DDOs, if particular government departments developed these in the future, a spokesperson commented that, because DDOs took up so little storage space, it might be cost effective to undertake minimal selection.

> Yes, if the departments decide to use them for major business uses then we will need to preserve them, because we take 25% of the most important records [of some departments] currently, and we will continue to take at least that many. Now, given the compactness of electronic records, we might decide it is not worth selecting, or hardly selecting. We might take 80% and leave out the ones that are obviously not interesting. (A national archive)

This minimal selection approach to the management of electronic records in automated office systems might seem to circumvent selection issues, but is based on the Archive having legal responsibility to maintain a substantial proportion of government records to demonstrate their actions for the Nation. In the longer term the cost of managing DDOs, particularly those which have been living in their 'active phase', could outweigh the initial cost of selecting these resource in the first place. This is because such resources require more close monitoring of their content than their paper-based counterparts, such as the need for periodic migration.

The decision to deposit material with the Archive would largely fall on government departments and in particular the 'user in the business area'. It is seen as a department's responsibility to develop a deposition schedule defining which are classed as ephemeral records and which are of long-term importance. This comment

echoes the views of one government agency who felt that preservation of DDOs was a specialist area and best left to the professionals to undertake. This is illustrated in the rich picture (Figure 4.1) by the lines which link categories C and D.

One specialist archive (category C) has adopted a proactive approach and identified that it is easier to select materials with this structured approach. It enables the archive to target particular areas of interest which correspond to identified gaps in their collection. The exclusion of certain DDOs, either, on an item by item basis, or, by excluding a category is an essential part of the selection process, and as indicated in a national library's comment (Section 3.5) it can be linked to parallel categories in print-based/analogue resources and organisational remits. The decision to dispose of, rather than manage, DDOs within the archival community involves a consideration of the amount of work needed to bring the resource into the archive, and process it.

There are however, differing views regarding the re-appraisal of DDOs during their life span. From stakeholders who felt that there was a need to re-appraise selection decisions over time, to a spokesperson from a national archive who stated that.

> Records having been taken into the archive, essentially, would be kept permanently, the decision having been taken once. There could be a review of an area of records and it could be decided that they are not kept, but in a way, it is saying the original decision was wrong. (A national archive)

Whether a DDO is owned by stakeholders, or only hosted by them, can influence the ability to dispose of it and also affect copyright issues. However according to one stakeholder, views do vary. The advisory body concerned with networking issues cited the examples of Brewster Kahle's Internet Archive (see Section 1.8.7) and the 'Alexa concept', which raises copyright issues as a blind Webcrawler collects and stores pages.

> ...partly for preservation purposes, which could be seen as acceptable use, but also to make it available to future users who may be looking for a URL that is updated. This has caused some controversy, and this is partly because the control of the data is no longer in the hands of the organisation.

> Something maybe where national libraries do have a role, but there is no official position for 'dynamic' data at the moment. (Advisory body concerned with networking issues)

Although there are many unanswered questions surrounding the selection process, many of which have been identified in this section; specific stakeholders identified a need to take on the role of selectors in the electronic environment. In addition, there is a recognition by some stakeholders who create DDOs of the need for the continuance of public deposit organisations in the electronic environment.

4.13 Determining the costs of long-term digital data preservation

The costs associated with the management of DDOs would appear to be a concern to some stakeholders, as evidenced by the following comments. None of the stakeholders were asked questions relating to costs, rather the concern was expressed indirectly in relation to specific issues or general challenges. When answering a question concerned with the relationship between access and preservation in the digital environment, a publisher noted that the costs associated with preservation may not always be self evident.

> In summary, the act of preserving the electronic material does not mean you will ever be able to access it. I suppose it carries with it the implication of expenditure. It may be very easy to 'keep electronic material' – there are an awful lot of hidden costs. (A publisher)

The expense of maintaining DDOs was supported by other respondents, and in one case metadata was viewed as crucial to managing DDOs cost effectively.

> We need that kind of information in my view because the maintenance of the information is very expensive, so we need to be able to make informed decisions about its future maintenance, whether we migrate it from one format to another for example... (A spokesperson, science-based industry)

Concerns over costs are not limited to commercial organisations (category A); a national library suggests the overheads for maintaining DDOs, viewed as 'a constant process of intervention and renewal', could be greater than for traditional materials.

> What thinking there has been to date, and really we are talking here about policy discussion rather than set policy, is in fact that we are aware that the long-term overheads for maintaining this material are likely to be far higher than the long-term overheads for traditional material, particularly paper-based material. (A national library)

4.14 Developing the root definition

Having analysed the rich picture it has been possible to construct a root definition which states the challenges to stakeholders that require resolution, as detailed below.

4.14.1 Root definition

Root definition - It is suggested that a flexible framework is created that preserves DDOs over the long-term to ensure future access, using selection criteria and metadata to support this aim. The framework hopes to benefit end-users of information and meet the conditions, and/or concerns of creators, owners and funding bodies. This should be undertaken by organisations who would have the responsibility for preserving digital materials and this may be achieved by stakeholder organisations working in consultation with advisory bodies. This takes into consideration the constraints of cost, human resources, legislation and the lack of and instability of standards.

4.15 Summary

Stakeholders identify the need to manage DDOs over the long-term, for a range of organisation dependent reasons. In some cases they also expressed a number of unresolved challenges which included a need for guidance. The corporate views of stakeholders have allowed the construction of a root definition, given above, from which two further hypotheses were generated. This analysis concludes that the following parameters would appear to be crucial to developing a way forward for the long-term management of DDOs.

- The need to define the roles and responsibilities of stakeholders.
- Consensus over the need to achieve long-term access.
- An identification of the need for selection, but an absence of proven ways to achieve this for a range of DDO types.
- The need for an extension to legislation to cater for DDOs[3].
- The importance of rights management to long-term preservation.

In addition, the analysis of the rich picture identified the grey areas that require more detailed analysis, using conceptual modelling to determine the overlap between these areas and the importance of each.

This chapter has demonstrated the value of SSM in drawing out the deeper trends from the semi-structured interviews conducted with selected stakeholders, and the subsequent construction of a rich picture on which the text in this chapter is based. Control of digital data, speed of technological advance, legislative guidance, interoperability, cost, access, confidence and authenticity are seen to be major issues of concern that stakeholders are having to contend with in the evolving electronic environment. Having identified these issues, they are by no means clearly delineated, that is 'grey areas' exist. Exploration of the nature of such 'grey areas' would appear

[3] Note that some of the recent developments in copyright legislation had not been published when the majority of stakeholder interviews were undertaken.

to be critical to the development of a future framework for long-term management (see Chapter 6). It is the purpose of Chapter 5 to demonstrate the support for the hypotheses generated from pertinent literature from various countries.

Chapter 5

Testing the hypotheses

5.1 The initial hypotheses

The following section reiterates the hypotheses developed in Chapter 2 as a result of researching the literature, before progressing to test and support these hypotheses.

A. The selection of 'dynamic' DDOs for long-term preservation requires the use of similar criteria to those used for 'static' resources.

B. The selection of 'living' DDOs for long-term preservation requires different selection criteria from those used for 'static' resources.

C. It is possible to define universal threshold standards for a framework of metadata for long-term digital data preservation.

5.1.1 Background to support the initial hypotheses generated from the literature review

To establish the support for the hypotheses it is necessary to evaluate the current practices of specific stakeholders, discussed in the previous two chapters, against developments in the literature both in the United Kingdom and in various other countries which were identified as important to the resolution of digital data management. In particular the schema to categorise types of DDOs, selection criteria for long-term retention and the possibility of the development of a minimum threshold of metadata.

From the literature review it became apparent that selection criteria do exist which are designed to cope with hybrid collections, that is, ones containing both traditional print-based resources that have been digitized and DDOs that have only ever existed in digital form. Such selection criteria would appear to be based on data content but also express concerns for managing the physical attributes of the medium which supports the data content (see Section 1.10). Through analysing the interview data some answers to why stakeholders wish to preserve DDOs long-term and how vital selection is to this process began to emerge.

The findings from the literature review suggest that selection criteria are often tailored to meet the requirements of specific organisations or sectors of the information community. Within these broad trends there are differences in the purpose, degree of access, dissemination and period over which DDOs are retained (see Sections 3.4 to 3.4.4 and 3.5). Additionally, specific stakeholder organisations have legal and or administrative duties to retain DDOs as a result of their remit, as in the case of a national archive and a national library (Table 3.1A, cells 6E and 6F and Table 3.1B cells 9E and 9F) or due to their accountability to outside agencies such as regulatory bodies (Table 3.1A, cell 3E). These reasons for preservation are seen to be paralleled with those for retaining print-based material, a fact underlined by responses such as that of a national library (Table 3.1B, cell 9G).

The 'book' has been seen as a powerful symbol of preservation and transmission of 'reliable' knowledge (Hunter, 1998). The credibility of the information contained in books and the amount of confidence readers place in this are issues which still present challenges with digital data. What has changed, is the greater degree of influence creators exert over the dissemination of their work, for example by publishing on the World Wide Web. As a consequence the dissemination role of the publisher has been challenged. However, given that an increasing number of unevaluated resources are being mounted on digital systems, some form of evaluation mechanism is considered necessary for end-users and deposit organisations to judge the quality of resources for their use. Although the reasons for preserving DDOs are seen to be compatible with

traditional objects, DDOs do possess different characteristics (see Sections 1.6.1 to 1.6.11). Therefore, it was also necessary to examine stakeholders' views on selection criteria and mechanisms currently utilised to manage living and dynamic DDOs. The evidence gained from the stakeholder interviews were analysed and used within this chapter to test the hypotheses.

5.2 Testing hypothesis A: the selection of 'dynamic' DDOs for long-term preservation requires the use of similar criteria to those used for 'static' resources.

This research attests that there is no fundamental difference between static paper-based objects and read-only digital data objects (as defined in Table 1.3), as the underlying data in both is fixed. Customisation of DDOs by the end-user is confined to selecting whether they wish to view the entire object or just sections of it. At the operational level parallels can be drawn between selection criteria for the two, such as the need to assess their substantive content for re-use, assurance of authenticity[1] and physical extent. There are however differences at this operational level. For example, a 100 page paper-based book has 100 pages no more no less. Users can access the substantive content of the static paper or digital object by using indexes and contents pages to locate the information. Increased speed of access and choice of access points is achieved in digital form by use of keyword searching.

The above paragraph serves to define the static category and illustrates that in most respects the category is fundamentally the same as paper-based objects and indeed contains digitized versions of traditional resources in addition to resources which have been 'born' digital. As extensive work is being undertaken on the digitization of resources (Kenney and Chapman, 1996; Prescott, 1998) it was therefore not felt necessary within the scope of this research to explore this category in-depth. The

[1] The work of Duranti (1997) and Bearman (1994) as described in Sections 1.8.1 and 1.8.2, have developed frameworks to provide mechanisms for establishing and maintaining the authenticity of DDOs.

review of the literature suggests that selection criteria do exist for hybrid collections based on:

- Substantive data content

- Managing the physical attributes of the medium that supports the substantive data content.

However, selection criteria for static DDOs need to be integrated into a framework that manages aspects such as rights management, preservation procedures, and also support long-term access to dynamic and living DDOs.

Progressing from the static to the dynamic category there are fundamental differences between the two types. These differences are based on dynamic resources allowing a greater degree of interaction between the user and the DDO than static DDOs, through the temporary manipulation/customisation of the data they contain, without affecting the data in the permanent ROM files. Dynamic DDOs can possess interactive capabilities that often rely on multimedia linkages, for example they can rely on sound and graphics files to function. The interactive facilities may enable the user to present this underlying data in may ways, for example a CD-ROM that aims to teach photographic techniques could enable a user to experiment and view the results of many exposure settings. Effectively, the 100 page book used as the example in the static category could increase dramatically in volume to include thousands of permutations. The following bullet points summarise the similarities and differences between static and dynamic DDOs:

Similarities between static and dynamic DDOs:

- Substantive content cannot undergo change.

Differences between static and dynamic DDOs:

- Dynamic DDOs enable choices in the way data is presented, that is the information can be reconfigured.

- Dynamic DDOs are often dependent on multimedia linkages.

The fundamental characteristics of dynamic DDOs are examined in greater depth in the following paragraphs and selection criteria developed accordingly.

There is a need for more consideration of how dynamic resources are migrated from one generation of technology to another because of their dependence on application software, which is seen to change over a relatively short time span (see Section 5.10). Stakeholders who are charged with the responsibility to select and retain dynamic DDOs need to be able to cope with the financial and human demands of undertaking this additional periodic process as part of the long-term management of dynamic resources. The dynamic category can cover a range of different DDOs from those that most closely resemble static data objects to interactive objects dependent on software to execute multimedia functions. Some dynamic DDOs do have links to living information as in the case of the multimedia encyclopaedia Microsoft Encarta 97, which has the majority of its information stored on a CD-ROM, with links to World Wide Web pages. The information stored on CD-ROM can be supplemented using an online link to the World Wide Web. Encarta 97 can be used solely as a dynamic resource by using only the information stored on the CD-ROM or as a resource that is part dynamic and part living through using the WWW link. This shows that different copies of the same DDO can exist in different states and that the boundaries between different categories of DDOs are not rigidly defined.

Where dynamic DDOs most closely resemble static DDOs a strategy of removing the dependency on standard systems could be used to reduce the need to migrate as frequently. This also gives the added benefit of interoperability between data within a

collection. However, it should be stated that transferring data in this manner might be possible for numerical data but could lead to loss of functionality for DDOs reliant on multimedia[2].

Within the dynamic category of DDOs this research has identified a sub-division between resources designed from their creation to be dynamic, and living DDOs converted to a dynamic state to allow their management to be achieved within repositories (see Table 3.2)[3]. The concept of keeping DDOs as totally living entities would at present seem to be unfeasible in terms of maintaining the authenticity of the object, as the continuous or constant change that such objects undergo mean that the information content of the original object may have been changed by updated versions.

When stakeholders were asked if the current selection mechanisms were adequate for the dynamic DDOs three out of the six organisations which held this category of DDOs stated that they were not catered for (Tables 3.3A and 3.3B, row D), the remaining three felt that these resources were catered for within their organisation. The first, a publisher, was satisfied that the organisation's selection mechanisms were adequate, however, it is worth emphasising that from the interview findings long-term management strategies did not appear to be well advanced (Table 3.3A, cell 1C). The second was a specialist archive which remarked that the dynamic category was easier to manage than those DDOs in the living category as the former are finished products (Table 3.3A, cell 2D) and the third was from science-based industry.

As static DDOs are composed of read only, time independent files, the need for complex mechanisms to manage them over time is not as great as for dynamic DDOs. This is because dynamic DDOs are interactive, which often involves maintaining

---

[2] The need to maintain the 'look and feel' of the original has been put forward as an argument in favour of emulation strategies for digital data preservation (see Section 1.6.8).
[3] Stakeholders within the archival community justified this need to convert living DDOs to a dynamic form by stating that without an object which has a fixed intellectual content it is impossible to establish its authenticity over time (see Section 3.2.2).

multimedia links together with the possibility of maintaining the links between several versions or accumulations of the resource, some of which may still exist as living parts of the DDO. In addition, the degree of change occurring within the hardware and software that support dynamic DDOs will result in a greater dependency on standard systems, requiring a greater amount of metadata to ensure future accessibility.

The findings from the stakeholder interviews suggest that current theory is in advance of best practice and is insufficiently based on empirical research. There are several reasons why formal policies for managing dynamic DDOs are in their infancy, as the following organisations demonstrate. Some specialist archives which deal mainly with static resources only retain low numbers of dynamic DDOs; an advisory body concerned with networking issues disseminating information mainly produced as living DDOs does not have a remit to preserve these; and there are publishers who create dynamic DDOs, with archiving policies which are more concerned with shorter term commercial re-use (although some wish to retain an outline history of their company). A national library has also seen many of the CD-ROM products it holds as immediate resources and not something to be managed over the long-term.

As the testing of hypothesis A shows, dynamic DDOs can be assessed using the same criteria as static DDOs, but with the addition of criteria that take account of their interactive qualities. The points listed below show the generic selection criteria felt to be applicable:

- Has the repository sufficient hardware/software to support the potential extra cost such preservation could entail over and above that required for static DDOs?
- Has the repository sufficient resources to support dynamic DDOs over the long-term?

- Assessment of the quality and benefits of the interactive capabilities. For example, with educational software, is it approved by a recognised authority in the particular field for its interactive features?

In summary hypothesis A is supported, although there are differences between the characteristics of static and dynamic DDOs, data content is fixed in both types of DDO and sufficient similarities exist to adopt broadly similar selection criteria.

5.3 Testing hypothesis B: the selection of 'living' DDOs for long-term preservation requires different selection criteria from those used for 'static' resources.

When dealing with living DDOs the analogy that they are sufficiently similar to static DDOs to have similar selection criteria breaks down because of the ability to change almost their entire substantive content. Unlike static and dynamic DDOs, archived snapshots of living DDOs have a relationship to the living form of the object, that is, the object continues to live.

- Living resources continue to change over time, unlike static and dynamic resources, which in turn could require a greater amount of metadata to record changes to their substantive content. Living DDOs may also contain multimedia and therefore require metadata to record the supporting hardware and software required to access these objects in the future. It will be important to develop criteria that will assess the cost and time involved in deciding the frequency of archiving 'snapshots' of the living object and in some cases maintaining the 'look and feel' of the resources (through maintaining access to documentation and appropriate hardware and software).

- The malleability of living DDOs could lead to problems such as a loss of data if a version of the same DDO is overwritten, or changes are made without being

recorded. An assessment of a resource needs to ensure that necessary mechanisms exist to support the recording of authorised changes or whether a system to achieve this needs to be developed (namely metadata).

- The origin of living DDOs is likely to be more diverse due to the ease with which data can be disclosed on global information systems.

- Living DDOs can exist in one of two forms as detailed below.

    - The first form is composed of DDOs whose past forms have been archived (for example captured as snapshots that become static or dynamic DDOs), but the present form exists as a living DDO, for example monthly government statistics (see Section 3.2.2, Type 1).

    - The second form is composed of DDOs that exist solely as living resources in their active phase, for example Web pages (see Section 3.2.2, Type 2).

As the results from the interviews with stakeholders indicate living resources are seen as being the most difficult to manage due to their volatile nature, time dependency, and interactivity. The results of the interview data suggests that the number of stakeholders who posses both dynamic and living DDOs is small. The whole area of the long-term management of living DDOs is new to the majority of interviewed stakeholders and, as a result, the distinctions between the way categories are selected has not been drawn for a range of reasons (see Section 3.5 and Table 3.3A cells 1E to 3E and Table 3.3B cell 12E).

The living nature of these objects means that selection needs to take the form of a consultative process with the rights owner, to ensure that objects of potential future long-term 'value' are evaluated; that is they go through a selection process before they cease to exist. Furthermore, as objects are developing in their 'active phase', capture of their information content at frequent intervals is required to prevent information

being lost during subsequent revisions. The frequency of these captures is influenced by factors such as the importance of the resource, the quantity of information changing between revisions and the frequency with which the information content changes. As well as the evanescent nature of living resources the advent of global information systems, such as the World Wide Web, has altered the relationship between rights owners and repositories. Authors are seen to have increasing power to circumvent publishers in the market, publishing directly onto the Web. Where stakeholders fit on this value chain affects their perspective of the process (Feldman, 1997). The altered relationship between authors, publishers and the market affects the selection processes needed to enable long-term selection to occur, because in the current climate information can be placed on the Internet without going through a peer review process. The economic challenges presented by the altered market have been identified by Getz (1999).

In the digital environment it is possible for authors to communicate directly with customers without needing to use publishers as an intermediary, although through taking this route the control mechanisms which help to ensure a degree of quality control have effectively been removed. In this situation, how do users evaluate DDOs? Publishers are evaluating their role in this digital environment and this has included the development of electronic initiatives such as Elesevier Science's ScienceDirect[4] service or the production of the CD-ROM version of Encyclopedia Britannica[5]. The relationship of ownership to access is seen as crucial in establishing how DDOs are managed. Ownership of DDOs can determine access rights as highlighted by the development of OCLC's Digital Archive[6], where access to back runs of journals are guaranteed to participating libraries as long as they subscribe to the OCLC service, providing access to back runs if the publishers cease trading[7].

---

[4] URL: http://www.sciencedirect.com/science/page/static/scidir/
[5] URL: http://www.eb.com/about/index.html
[6] URL: http://www.oclc.org/oclc/eco/archive.htm
[7] In such a situation concern has been raised about a repositories lack of access to back runs of journal, and promoted the development of initiatives such as JSTOR (Finholt and Brooks, 1999). Similarly, access to back runs can be placed in doubt if a library ceases its subscription to a journal title.

The characteristics of time dependency and interactivity exhibited by living DDOs render existing mechanisms for their selection and future management inadequate. Therefore, because of the fundamental differences living DDOs exhibit, new selection criteria need to be formulated which address these issues. The research suggests the following generic criteria be considered when selecting living DDOs.

Storage capacity

1. Has the repository sufficient storage capacity to devote to a living DDO, bearing in mind the probable snapshot frequency and the amount of data generated by each snapshot?

Nature of change

2. How does the DDO change over time? For example, are the changes to the substantive data content cumulative or is the content periodically overwritten?

Frequency of change

3. How frequently should a snapshot be taken: daily, weekly, monthly quarterly, annually?

Life span

4. Has the living DDO an end point expressed by the creator?

5. When was the living DDO first created?

In summary, support for hypothesis B rests on the fact that living DDOs require additional selection criteria, as listed above, to the criteria needed for static and dynamic DDOs. Hypotheses A and B were supported by a combination of observation

of the findings from the initial interviews held with stakeholders and logical deduction. The evidence from the interviews suggests that both static and dynamic DDOs are, from a long-term preservation point of view, treated almost the same, but living DDOs are treated differently.

5.4 Testing hypothesis C: it is possible to define universal threshold standards for a framework of metadata for long-term digital data preservation.

The minimum requirements for this metadata include access rights, a declaration of provenance, creation date and preservation review period (see Section 7.3.8 for a more detailed explanation). Whilst it would appear that there are no practical reasons why this information should not be included in the metadata, in practice the inclusion of this data varies between formats and organisation. Stakeholders were in agreement that metadata provides a means of managing DDOs, but the requirements necessary to ensure that the information content was managed varied widely between organisations due to a number of factors. To tease out the reasons why metadata was not universally implemented, a series of questions were put to stakeholders.

- Stakeholders were asked if their metadata format was in-house, standard or modified from standard. The findings showed that out of the ten respondents responsible for a collection, five stakeholders had created in-house formats, whilst five had standard formats. The comments associated with these figures showed that some stakeholders had found it essential to create customised metadata formats to fulfil their requirements (Tables 3.5A and 3.5B, row C).

- To assess how satisfied stakeholders were with their existing metadata format they were asked if there were any areas where improvement could be made. This identified problems with current standard formats such as the lack of extensibility of UKMARC. Also expressed was the hope that mark-up languages such as XML and MML would improve the relationship of SGML

to HTML and the WWW. Similarly, in-house formats were at varying stages of construction or refinement (Tables 3.5A and 3.5B, row D).

- The findings of the previous point led to a further question being asked to determine whether these improvements were linked to an inability of the format to cope with certain types of DDOs. This provided a more specific indication of which DDOs are not catered for by a stakeholder's particular metadata format(s). Those stakeholders using the UKMARC format emphasised its inability to cater for WWW based information, whilst the reverse is true of newer standard formats. In-house formats also needed to be modified to cater for new developments, such as the updating of documentation to correspond to changes in hardware developments. (Tables 3.5A and 3.5B, row F)

- Variations in the type of metadata applied to different DDOs within a stakeholder's organisation showed the extent to which the use of metadata has been adapted to cater for different resources. This linked in to the degree of input into the metadata from external sources, the range of materials being dealt with and how the metadata format was used by each stakeholder within their organisations (Tables 3.5A and 3.5B, row E). It would also appear that the greater the degree of extensibility the format is capable of, the more potential variation could exist in the way metadata is applied to individual resources.

- Stakeholders were asked what metadata is required to ensure that the information contents of DDOs being preserved are authentic. This indicated that stakeholders were aware of the need for checks to ensure that the metadata faithfully represents the data yet few mechanisms exist within currently used metadata formats. Stakeholders who are developing new digital archiving systems stated this was being considered.

- Views were sought from stakeholders on how important interoperability between metadata formats was to them. This tested the need stakeholders have to assimilate data from outside sources and the dissemination and disclosure of data to other institutions or bodies. The pattern exhibited largely reflected the remits of the stakeholders. Stakeholders did highlight difficulties in trying to achieve interoperability such as obtaining the standards and having to amend data within the organisation to match changes in standard formats. There is a need for a structure to support the selection of DDOs that utilises the fact that global connectivity (achieved through global information systems such as the World Wide Web) can help to eliminate duplicate DDOs. However, this assumes that a repository can check the authenticity of both the metadata and the DDO.

The evidence from both the interviews and literature survey suggests that diverse approaches to the long-term management of DDOs exist, even though there is consensus on the importance of metadata in this process. Therefore although the case for hypothesis C is supported it is not universally implemented. The outcome of this analysis suggests that a minimum threshold metadata, although theoretically feasible, would not be practically feasible on the following grounds:

1. It would seem unlikely to gain agreement between a diversity of stakeholders, given the costs of metadata conversion.

2. Such a universal metadata may have to be expanded over the course of time in order to incorporate new developments. If these developments continue to occur the metadata format could become unwieldy.

3. Retrospective conversion of metadata in legacy systems could prove to be uneconomical.

The result for hypothesis C prompted further analysis of the data obtained from the stakeholder interviews, using SSM to re-evaluate the problem situation, resulting in the generation of the rich picture described in Chapter 4. Two new hypotheses (as detailed in Section 5.5) resulted from this analysis. These hypotheses are tested in Sections 5.6 to 5.10.

5.5 Justification of the new hypotheses emerging from the root definition

As shown in Chapter 4 the construction of a composite view, a *rich picture*, aided the development of a more precisely defined expression of the research problem, the *root definition* and, stemming from this, the development of new hypotheses D and E which are detailed below:

D. A universal conceptual framework for the long-term access to DDOs at a level higher than metadata is possible.

E. Criteria for the selection of DDOs for long-term access are a fundamental component of such a universal system.

5.6 Testing hypothesis D: a universal conceptual framework for the long-term access to DDOs at a level higher than metadata is possible.

To test hypothesis D, it is necessary to examine how existing frameworks that cater for more than one metadata format operate, together with the strengths and weaknesses of each of these. This allows an assessment of the difficulties already encountered by existing initiatives that cater for more than one metadata format to be made (see Table 5.1). In addition, existing mechanisms that facilitate data interchange are evaluated to determine their suitability for accommodating multiple metadata format management over the long-term (see Sections 5.6.2 to 5.6.2.5). Some of these mechanisms have

been designed to co-ordinate formats for specific purposes, such as, interoperability for resource discovery and/or data interchange.

Table 5.1 Review of a range of current initiatives

Name of initiative	Purpose for which developed	Comments
Pittsburgh Project – Functional Requirements for Evidence in Recordkeeping	Electronic record system using metadata to support capture, maintenance and re-use.	A comprehensive specification of metadata requirements. Requires records to be created within a predefined system to maintain evidentiality.
University of British Columbia Project Preservation of the Integrity of Electronic Records	Identify and define requirements for creating, handling and preserving reliable and authentic records. Physical custody and description by archival institutions.	Addresses a range of important issues. Considers the automatic capture of metadata to provide insufficient historical context compared to the use of archival description methods.
Digital Rosetta Stone (DRS)	Provides long-term access by preserving information necessary to recover and reconstruct digital object.	Metadata archive that enables data objects with different metadata formats to be accessed and emulated in its original format.
National Library of Australia Preserving and Accessing Network DOcumentary Resources of Australia (PANDORA) Logical Data Model	To develop policies and procedures for preservation & access, set up an archive of selected publications, quantify resources necessary and develop a national approach.	Attaches preservation metadata to digital objects to enable long-term preservation and to give online user access to the descriptive metadata. Links to other projects developing selection procedures for online Selection Committee for Online Australian Publications (SCOAP).
Reference Model for an Open Archival Information System (OAIS)	Stems from ISO development of standards for long-term preservation of terrestrial and space information.	Aims to produce international standards for many of the processes involved in digital data preservation.
Cedars Project (CURL) exemplars in digital archives	Examine strategies to aid digital libraries to preserve both digital and non-digital information resources over the long-term.	Combines metadata elements from a number of initiatives compliant with OAIS model, but is specifically focused on the United Kingdom higher education community.

5.6.1 Limitations of existing frameworks

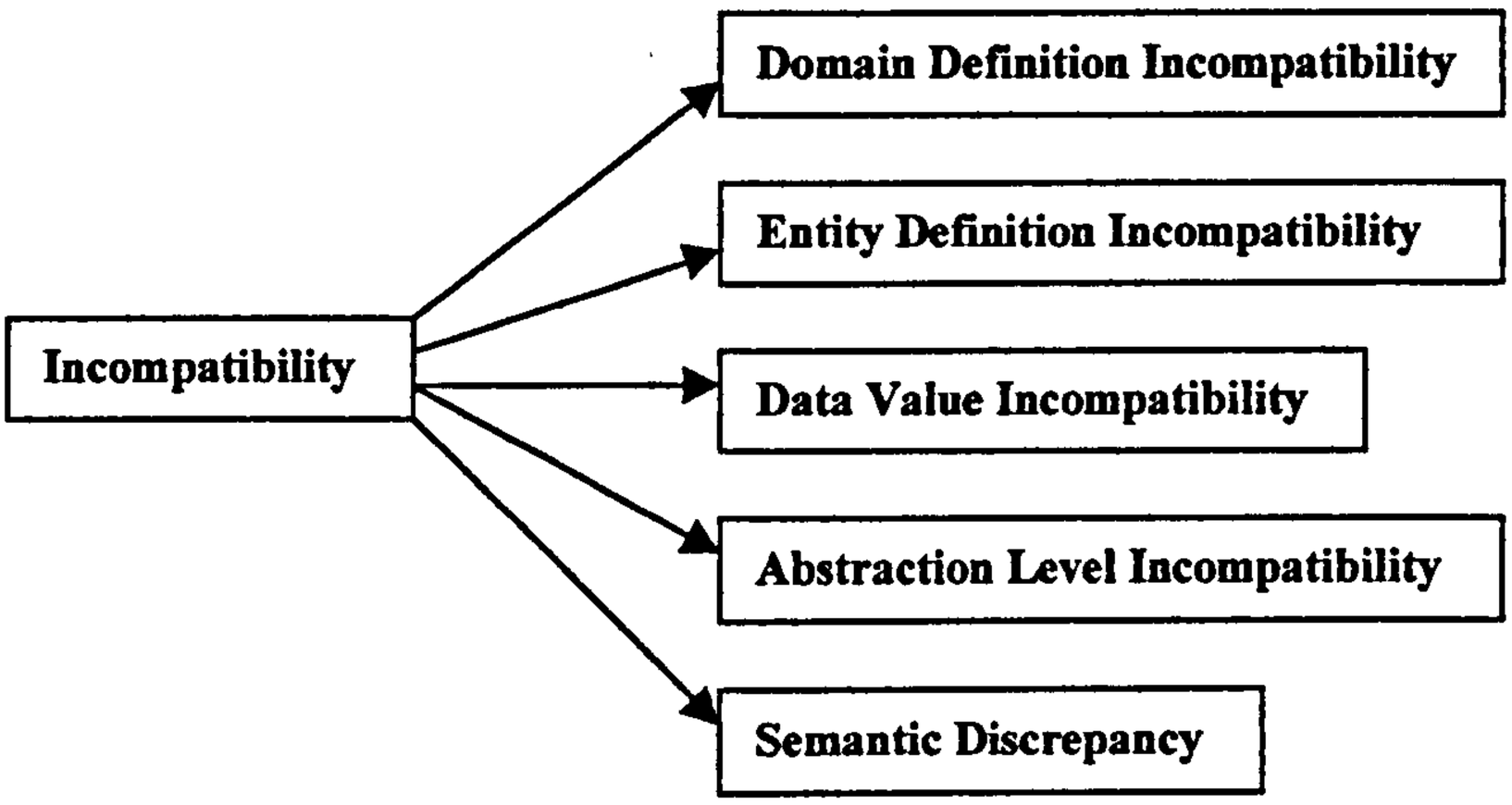
Many of the frameworks that have been developed exhibit some features that could be beneficial to the long-term management of DDOs, such as establishing authenticity, development of metadata specifically for preservation, and the development of models based on individual organisational needs (see Sections 1.8.1 to 1.8.9). There would appear to be an absence of a conceptual framework that co-ordinates long-term digital data preservation. What appears to be required is an intra-organisational structure which co-ordinates these approaches.

Existing frameworks for the management of DDOs are rooted in distinct sectors within the information community, and are focused on specific aims such as resource discovery and interoperability, and/or the long-term management of digital data within specific parameters. Table 5.1 reviews a range of current initiatives and describes their purpose and other comments on their advantages, how they achieve their aim and their potential uses.

The development of different metadata formats has led to semantic discrepancies as shown in Figure 5.1. It is one of many areas where incompatibilities can occur due to heterogeneity (Sheth and Kashyap, 1993). The figure demonstrates the different levels where incompatibility can occur in a heterogeneous digital environment; from the subject field (domain definition), the individual DDO (entity definition), down to an individual item in the data (data value). Similarly, abstracted data in the metadata (abstracted level) could contain inconsistencies as could the way it is structured (semantic discrepancies). Sheth and Kashyap (1993) suggest that data which is schematically different (described by different metadata schemes) can be semantically very similar (consists of very similar subject information). This difference between formats, together with the other incompatibilities they highlighted, demonstrate the difficulties of achieving interoperability at the metadata level. Therefore, for a successful future long-term management framework to be developed integration must

occur at a higher level than metadata, at the super-metadata level, as detailed in Section 1.2.6.

Figure 5.1 Structural incompatibilities due to heterogeneity



Source: (Sheth & Kashyap, 1993, p. 284)

The integration of formats at the metadata level would reduce the structural incompatibilities but creates problems associated with standardisation (see Section 3.6.3). As we have seen in the selection of specific frameworks mooted to date, detailed in Sections 1.8.1 to 1.8.9, they have focused on developing structures which support the combining of metadata formats rather than their co-existence and are focused on particular organisations or domains. This research proposes to co-ordinate the strengths of existing frameworks and to add to them the concept of super-metadata as a management tool; integrating selection, authentication, physical preservation strategies, rights management and interoperability. At present, there are a number of systems designed to enable interoperability to occur across the range of existing metadata formats as shown in Section 5.6.2. In addition to this there are frameworks that hope to address interoperability across a range of existing metadata formats by a variety of mechanisms shown in the sections below. These have been designed to co-ordinate formats for specific purposes, such as, interoperability for resource discovery and/or data interchange. As shown in Sections 5.6.2 to 5.6.2.5 these different

mechanisms could be incorporated into an over-arching framework to support the long-term management of DDOs.

5.6.2 Mechanisms to cater for multiple metadata formats

5.6.2.1 The need for interoperability

The findings of the stakeholder interviews suggest that organisational drivers play a crucial part in the corporate view of an organisation to the importance of interoperability. Stakeholders with organisational remits involving the interchange of digital information viewed interoperability as advantageous, especially one stakeholder operating a distributed service as this concurred with organisational goals to encourage interdisciplinary research and cross-sectoral access (Table 3.5A, cell 4G). Stakeholders who held data in closed systems, possibly because of its commercial sensitivity, were not necessarily concerned with the need for interoperability. The stakeholder from science-based industry highlighted that if a standard like SGML changed then the organisations' data may need to be changed to reflect this, requiring additional work. However this same stakeholder did see interoperability as advantageous in enabling translation between languages. One specialist archive recognised the advantages interoperability would bring but questioned the feasibility of actually achieving 'common' standards, as the metadata format which the archive uses was developed in-house and the chances of finding a format with sufficient similarity were seen as being remote (Table 3.5A, cell 2G). A reason for encouraging interoperability put forward by an advisory body concerned with networking issues was that it prevents the need for everyone to standardise on one metadata format (Table 3.5A, cell G5). However, there are difficulties in integrating different metadata formats that can lead to data loss, as detailed in Section 5.6.2.2.

Similarly, the practicality of developing interoperability between metadata formats was challenged by a government agency which felt that, unless a superset was formed,

users could be confronted with a plethora of different interfaces to negotiate. However, as a national library envisages the digital library functioning as a single seamless entity, the interoperability issue seems to be one which needs to be urgently resolved. There are a number of different approaches to achieving the interchange of information between metadata formats in multiple databases. Semantic heterogeneity is seen as one of the crucial challenges to achieving interoperability and co-operation between DDOs, and research has been undertaken focusing on databases (Kashyap and Sheth, 1998). 'The problem of semantic heterogeneity is the identification of semantically related objects in different databases and the resolution of schematic differences among them.' (Sheth and Kashyap, 1993) Semantic heterogeneity is seen as a characteristic of the current networked environment and at present there are a selection of approaches, as detailed in Sections 5.6.2.2 to 5.6.2.5, to address the need to interchange information between metadata formats. The creation of Global Information Systems, such as the World Wide Web, has made this task more difficult to achieve by increasing the volume of diverse material whilst at the same time increasing the need for semantic interoperability. Metadata is seen as a vehicle for achieving interoperability by providing reference points to link different formats. It is also felt that metadata would reduce 'information overload' by using it to determine the relevance of data without the need to return to the raw data (Kashyap and Sheth, 1998).

5.6.2.2 Crosswalks

The mapping of information between metadata formats has been seen as a potential way of creating linkages between them. The DESIRE-RE 1004 project (Development of a European Service for Information on Research and Education) (1997), under the *European Union's Telematics for Research area of the Fourth Framework Programme*, has examined 22 metadata formats used for resource description. These formats ranged from generic to domain-specific. This project recognised the increasing importance of metadata in the networked environment, that it is not only important for resource discovery but for the management of found objects. This is

achieved through the establishment of frameworks and the promotion of interoperability across a variety of metadata formats. Drawbacks to mapping across metadata formats have been identified (Day, 1996) and are listed as follows.

- Inevitable data loss, as information deemed necessary in some metadata formats will not be seen as relevant in another.

- "The ethos underpinning a metadata format, their 'cultural differences'" (Day, 1996), affects factors such as; revision procedures, the degree of change permissible, that is their extensibility (see Section 5.6.2.4) and the longevity of the format.

- The information content of a particular field can vary between formats.

- Some fields within existing metadata formats cannot be searched, for example the note field in MARC.

- The syntax used to subdivide fields in some instances is not translatable across formats.

(Abstracted from a ROADS workshop presented by Day, 1996)

5.6.2.3 Container Architecture

Container Architecture, known also as the Warwick Framework, is a modular based framework, which was developed from a OCLC/UKOLN Metadata Workshop (DC-2) held at Warwick University, United Kingdom. The framework allows for the 'packaging' of metadata, that is aggregating multiple independent metadata sets for networked resources into linked units. This framework aims to support interoperability between formats within a domain specific area and also across domains, based on the

premise that it would be futile to create one metadata scheme to cater for all domains. Consideration is given to the absorption of metadata formats such as MARC, by allowing interchange between such systems and new developments, and also the adaptability of the framework to encompass new formats. As identified in Chapter 1 access is a key component to the management of DDOs and this is a strong feature of this framework. This emphasis on access includes consideration of rights management issues. This would appear to provide a suitable framework for the future management of DDOs, however there are a number of fundamental issues which are not addressed, namely the selection of resources and the lack of a structure for managing DDOs differently, according to the various types which exist. The framework provides an interoperability mechanism between multi-metadata formats, but due to its purpose of aiding resource discovery, does not provide a long-term management strategy. However, some of the factors this framework considered to be important for resource discovery and storage are also important for long-term retention, however for long-term retention additional structures need to be built. It would appear from Section 1.6.5 that an essential part of any future framework for the long-term management of DDOs requires consideration of the types of resources based on their function. The selection mechanism is informed by these DDO types, which is fed directly into the future management strategy. This strategy, as will be shown in Chapter 7, is dependent on a framework which, unlike container architecture, considers that there are distinct operational levels, namely metadata deposited with the DDO and management metadata, super-metadata (see Section 1.2.6), each with specific functions.

Extensions to the Warwick Framework have been developed, such as, the Distributed Active Relationships (DARs) proposed by Daniel and Lagoze (1997). These DARs express relationships between networked resources. The concept of container architecture has been adapted, and to a large extent been superseded, by the Resource Description Format (RDF). This format is specifically designed to describe Web based resources, using a data model to develop a common syntax. However, the

development of a model for Web based resources could make it difficult to accommodate metadata formats based on legacy systems.

5.6.2.4 Extensibility

Metadata formats vary with respect to the degree to which their structures can be customised. As a general rule the more rigid the rules on the structure of the format the less extensibility is permissible. To an extent the greater the degree of flexibility in interpreting a metadata format's syntax, the more difficult the task of generating interoperability between metadata formats becomes. The Text Encoded Initiative (TEI) is an example of a highly extensible format developed with interchange of data in mind. However the high degree of extensibility permissible within the format can create problems of inconsistency in the application of the format (as noted in Table 3.5B, cell 12D), according to a specialist archive more formalisation of the format is required. Metadata created to specifically cater for networked resources tends to exhibit higher levels of extensibility than legacy systems.

5.6.2.5 Development of metadata registries

To overcome the problems caused by semantic discrepancies between metadata formats, one solution has been the development of metadata registries. Such registries are seen to be advantageous in creating semantic agreement on element meanings and 'enumerated lists', publishing of qualifiers and publishing existing metadata schemes so that others can utilise them, preventing duplication of effort according to Powell and Dempsey (1997).

5.7 Integrity and authenticity issues

The high dependency of society on digital data coupled with the malleability of DDOs, has brought with it concern for ensuring its integrity and authenticity. With

global connectivity the opportunities to integrate systems are increasing and conversely so are the diversity of approaches to managing digital data. The ability to decipher authentic metadata is becoming increasingly important to producers of digital data, repositories and end-users. The trust which stakeholder organisations exhibit towards other institutions' data is largely determined by being certain of its provenance and the regard in which a creator/producer is held. At present no uniform standards apply to registries of approved metadata formats.

5.8 Stakeholder confidence

A combination of factors, as seen in the following discussion, affect the development of a framework for long-term digital data preservation. The confidence (or lack of it) which organisations show towards data produced outside their own organisation is seen to be a major issue requiring resolution if co-operation is to be achieved. Some national institutions would consider interchanging data with organisations of a similar standing to themselves, as in the case of a national library (Table 3.7B, cell 9A). Stakeholder confidence in DDOs produced by other organisations which are offered for deposit, and confidence that the details of DDOs held by other organisations are what they purport to be, will enable disposal of duplicate DDOs as a result. As far as commercial organisations such as publishers are concerned, deposition of material with a third party would be dependent on the access this third party had to their data (Table 3.7A, cell 1B). A stakeholder in science-based industry would be willing to contract out digital data preservation to a third party or vendor dealing with similar companies, providing assurances could be given that their data would remain secure, in addition to any potential financial return from further re-use (Table 3.7A, cells 3A and 3B). Confidence that the holdings of an outside organisation are what they say they are, is a vital part of developing a framework for the long-term preservation of DDOs. This trust rests on establishing the integrity and authenticity of individual DDOs, something which it is hoped super-metadata will aid.

The degree of confidence some stakeholders have in a potential global framework was tested by asking them whether they would discard their own resources if they were to find an identical record on a world preservation database. The findings (as stated in Section 4.8) show that the majority of stakeholders would have confidence in another organisation to manage some of their DDOs over the long-term, however this would not be without assurances that issues such as security, access, and procedures for archiving and preservation met their requirements. As seen in Tables 3.7A and 3.7B, row A, requirements varied between stakeholders, but responsibility for long-term management was consistently linked to ownership of DDOs. Some stakeholders were unwilling to allow others to manage their resources as they felt access to resources was inextricably linked to ownership. The responsibility for the 'digital master' according to an advisory body for data preservation is the responsibility of the owner. The uncertainty of access to resources and means of restricting access are seen to affect stakeholder confidence. To develop a co-ordinated framework for long-term management it is essential to build into it mechanisms to protect creators, end-users and to increase stakeholder confidence. Questions of how to evaluate the integrity of resources, and even of the integrity of organisations that proffer resources, are highlighted by this research. The framework proposed by this research (see Chapter 7), aims to embrace developments in achieving integrity and authenticity of DDOs to develop trust between depositors, repositories and future end-users. There is a need to develop an infrastructure in which there are established criteria for evaluating whether the handle given to a digital object (its metadata) is an authentic representation of the digital data.

5.9 Ownership, origin and use

The development of global interconnectivity raises the expectations that 'collections without walls' are possible. Increasing access to DDOs may be dependent on technological progress, but as Gorman (1997) has identified, ownership rather than access to digital objects will influence the shape of future structures. "The time has

come, it seems to me, to redefine the idea of 'collection' and to rethink the ways in which users gain access to documents in the redefined 'collection'". (Gorman, 1997, p.498) Gorman's use of the word collection is confined to the library field, however he recognises that interpretations of what constitutes a collection vary depending on the stakeholder's point of view. Within this collection are four main components of resource;

1) Tangible objects (books, sound recordings, videos, etc.) owned by the library and housed in the library.
2) Intangible electronic documents (on CD-ROMs, mounted on the library's computers, etc.) owned by the library and available without a fee at terminals connected to the library's systems.
3) Tangible objects owned by other libraries and available through interlibrary loan and other resource-sharing systems.
4) Intangible electronic documents owned and maintained by other agencies and available electronically for a fee or without payment.

(Gorman, 1997, p. 498-499)

In Gorman's view, it is preservation and bibliographic control which support the new definition of collection and are crucial to making these resources available. In tandem with ownership are issues concerning the origin of DDOs, that is their provenance. The origin of DDOs links to three further issues that affect the long-term management of DDOs, namely stakeholder confidence, authenticity and integrity. At present very little research exists which ties usage of DDOs to ownership and origin which to some extent contribute to the 'grey areas' surrounding the long-term management of digital objects. In order to develop a framework to achieve the long-term management of DDOs the development of procedures to resolve issues such as access controls; verification of provenance; integrity and authenticity will be required. The framework will need to be set in a legal and administrative environment that will support it. This will include the need for legal deposit and rights management issues to be resolved; training of those charged with the responsibility to preserve DDOs over the long-term and a thorough investigation and construction of a cost model (see Chapter 8), to consider the possible cost implications of the long-term management of digital objects for a variety of stakeholders.

5.10 Testing hypothesis E: Criteria for the selection of DDOs for long-term access is a fundamental component of a universal framework

Hypothesis E proposes that criteria for the selection of DDOs for long-term preservation are vital for the super-metadata framework, as without such selection by 'informed professionals' the ability to re-find material would become proportionately more difficult to achieve and the costs of storing vast quantities of data become increasingly expensive. As the numbers of DDOs in existence increases it becomes important to employ selection criteria in order to maximise the use of human and financial resources. "With high interconnectivity and access to many information sources, the primary issue in the future will not be how to effectively process the data that is known to be relevant, but which data is relevant" (Sheth and Kashyap, 1993). If a selection process sensitive to digital data requirements is not adopted, than as the quantity of material being managed increases the cost and difficulty in achieving long-term access would seem to be prohibitive. It is argued that the selection of material is best undertaken through evaluation by institutions holding specialist collections of purely digital or hybrid collections. Such institutions would employ experts to critically appraise DDOs on an informed basis. This decentralisation of selection decisions would be supported by super-metadata, which would act as a universal standard, removing the need for centralised selection and the attendant danger that this dependence could bring. Such a centralised data repository could be vulnerable to manipulation and data corruption intentionally or unintentionally. As geographic distance is ameliorated by global interconnectivity, it would seem to be more desirable to adopt a decentralised structure with a centralised co-ordination structure, such as super-metadata, at a level higher than metadata.

Existing selection criteria have limitations for living and dynamic DDOs, as has been demonstrated in the previous sections within this chapter. This is supported by respondents' needs for guidance in the development of strategies to cater for the long-term preservation of digital objects, and as an outcome of this, the need for training. Through stakeholders comments, it is evident that this training may take many forms,

from guidance and tutoring in electronic records management for archivists with a grounding in managing traditional materials, through to the need for knowledge of the financial side of managing DDOs (see Tables 3.3A and 3.3B, row B).

Whereas the selection of traditional paper-based objects can be deferred, the nature of DDOs requires that selection takes place at their creation or as soon as possible afterwards, otherwise the object may be lost (Table 3.1B, cell 9H). Shepherd (1996) has identified the need to bridge the gap between the design and construction of systems (traditionally regarded as the domain of information technologists) and the care of records after they have been created (the domain of record managers), to ensure that the effective management of DDOs is achieved. However, as noted earlier in this thesis, it is not always possible or desirable for records managers to intervene at the system design stage. If a proactive management stance is not adopted by records managers, Shepherd (1996) fears that electronic records will be neglected as these managers may be excluded from important decisions taken before the resources are created. The lack of co-ordination in the management of DDOs remains a 'grey area' between stakeholders. The evanescent nature of DDOs requires that the process of selection be more closely integrated with the creation stage. This is twinned with the fact that hardware and software obsolescence is also occurring at a rapid rate. From a software vendor's perspective it has been suggested that applications require major revision after 18 to 24 months and become fundamentally obsolete after five years (Myrhvold, 1999)[8].

> Perhaps it is very difficult because you can't always see what is going to happen. I mean, SGML is perhaps the standard now, but will it still be in so many years time? Will Microsoft still be around, it probably will but will it still be around in ten years time? (Personal view of a spokesperson from higher education sector)

---

[8] A comment expressed in an interview conducted by Peter Day, 11 January 1999. *In Business: Cyberbrain.* BBC Radio 4.

Factors that stakeholders identified as influencing selection of DDOs for long-term management include:

- adherence to the collection policy;
- influence of outside agencies;
- measurement against criteria such as scholarly value;
- documentation and format;
- usefulness to various classes of users;

(see also Tables 3.3A and 3.3B, row F).

These factors influence what is selected and also determine how selection is undertaken. The selection process and subsequent careful monitoring of DDOs are characteristics of what this research defines as management and an absence suggests DDOs are merely being stored.

In summary, the testing of hypotheses A, B and E supports the need for selection. The conceptual feasibility of the selection process is tested by the construction of a conceptual framework (Chapter 7).

5.11 The need for the super-metadata framework

The need for a framework which may act as future guidance for information-dependent organisations to follow is supported by the findings of a survey commissioned by the Research Libraries Group (RLG), conducted in April 1998 referred to in Section 3.3. This survey showed member institutions both in the United States and in the United Kingdom looking for guidance in the area of DDO management.

> Member institutions are seeking leadership in the development of standards and best practices, guidance on model policies and practices, and various types of training from consortia, such as the RLG. (Hedstrom and Montgomery, 1998, p. 15)

The advantages of developing a co-ordinated structure based on super-metadata are outlined in the following points.

- As many information systems are expected to operate in a networked environment it would be unrealistic to expect any one ontology or classification to be able to cater for the diverse range of digital data (Kashyap and Sheth, 1998). If one ontology was to be used it would need to cater for the whole diversity of recorded knowledge which, if it were possible to achieve, would probably possess an unwieldy structure.

- Metadata in legacy systems would co-exist on an equal footing with more recent formats (designed for use solely with digital resources) within the flexible structure of the super-metadata framework. This prevents organisations having to expend resources on retrospective conversion of data to a standard metadata format. In addition, it could mitigate against the errors and problems that may occur during the retrospective conversion process.

- The fact that different metadata formats exist, regardless of whether they are content-dependent or content-independent would not in itself appear to promote interoperability, despite the existence of crosswalks. There is seen to be a need for a co-ordinating mechanism that facilitates interoperability by accommodating the diversity of metadata formats.

- Within global information systems which exist today there are seen to be differences in language and culture the harmonising of which may not be possible or desirable at the metadata level. Super-metadata would enable these differences to exist whilst at the same time allow unification for the purposes of long-term management at a higher level.

- It is anticipated that the flexible nature of the super-metadata framework will allow future metadata formats and as yet unknown types of DDOs to be catered for.

- The super-metadata framework allows for the distributed selection of DDOs by subject specialists, the ability for such specialists to construct contextual information within domain-specific metadata and an overarching mechanism for co-ordinating long-term management, to prevent loss of data and duplication of effort.

- The framework would support the existence of hybrid collections of material, that is a mix of both traditional paper-based resources and digital data within one collection. This allows the future support of metadata formats to support these hybrid collections as they themselves become legacy systems.

- The approach brings together a number of initiatives (detailed in this chapter), being worked on in different sectors of the information community, into a cohesive framework.

- Aims to accommodate the diversity of organisational aims which different stakeholders bring to the issue of the long-term management of DDOs through its flexibility.

- In the current situation some stakeholders are looking for guidance in a changing field, which is still in its infancy, while others are developing their own solutions to these problems, therefore pointing to a need for a conceptual framework.

The development of a framework for the long-term management of DDOs needs to resolve a number of complex interrelated issues. It is hoped that this may be achieved by drawing on the strengths of a diversity of initiatives developing in different sectors of the information community.

5.12 Summary

A symbiotic relationship could be achieved between a distributed selection mechanism for DDOs and the standardisation of preservation metadata through the common link of super-metadata. For example, the selection of statistical data sets by a specialist archive could be carried out on site but with adherence to the guidance enshrined within the super-metadata framework, this in turn would ensure continued access to the archive's resources over time and thus promote use of its resources and the archive's continued existence. Decentralisation of selection would therefore be achieved on a co-ordinated basis, hence allowing specialist subject fields to be catered for. As far as purely digital collections are concerned geographic location is less crucial, subject specialists aside, than ownership of the data. As noted earlier ownership in the digital environment goes hand in hand with the responsibility to manage the DDO over the long-term. The distribution of DDOs could well be based in particular institutions for a variety of reasons which could include:

- A need to develop a base of subject specialists who can evaluate specialist materials and with the knowledge and interest to maintain these same resources over time.

- A need to concentrate specialist knowledge and resources within particular institutions and the ability to link digital resources with traditional paper-based resources contained within hybrid collections.

Distributed archiving is facilitated by the co-ordinated access mechanism of super-metadata[9]. This super-metadata does require a legal and administrative structure in which to place the framework (Chapter 7). It is anticipated that the responsibility for the smooth running of super-metadata is best placed in the hands of a consortia of

---

[9] Hofman (1998) in a paper entitled *Towards a United but Distributed Archives of Europe?* identifies some initiatives that are occurring within Europe and the differences in approaches to handling/managing electronic records due to differing legal contexts but suggests that the "... arrival of 'global archives' is just a matter of time...". (Hofman, 1998, p. 336).

public organisations, the arrangement of which would not be unlike a present day intranet. It is anticipated that such an arrangement would include an agency creating unique identifiers. As well as resolving the structure of selection mechanisms and super-metadata content, questions of how the future development and maintenance of super-metadata is to be achieved requires further research (see Chapter 9).

Through testing the hypotheses generated by this research, it has been demonstrated that the current approach to categorising DDOs needs to move away from its current concern for the medium on which it exists, to a concentration on the function of the digital data itself. The research has identified three broad, though not mutually exclusive, categories under which DDOs can be classified and used to determine the strategy needed to ensure their long-term preservation. These strategies in turn need to be an integral part of the selection process. The findings from the initial stakeholder interviews, supported by the literature, state that metadata is central to the development of a cohesive framework. In addition, the problems of creating a single metadata format for preservation purposes outweigh the benefits. This observation required the research to explore the possibility of developing a co-ordinating mechanism at a level higher than metadata and examining the importance of selection at this level. The tangible outcome of testing these hypotheses has been the exploration of potential solutions through conceptual modelling (Chapter 6) and the development of a framework for long-term management of DDOs (examined in Chapter 7).

Chapter 6

Conceptual modelling

6.1 From root definitions to conceptual models

The findings from the data analysis (Chapters 3 and 4) indicate a need for guidance concerning the management of DDOs. The previous chapter has demonstrated that a range of current initiatives are focusing on addressing some of the identified issues, often treating them solely as technological issues. This may be appropriate for some issues but, as this chapter aims to emphasise, there are a number of human centred issues crucial to an investigation of the 'research problem' (identified as 'grey areas' in Chapter 4). It is the purpose of this chapter to explore the nature of human centred issues using conceptual models and use this process to inform the development of a conceptual framework that considers both the human centred and technological issues.

Continuing with the SSM approach, the root definition (see Section 4.14.1) and alternative root definition (see Section 6.1.2) are used to aid the generation of conceptual models of the 'research problem' by examining the possibilities that are contained in the root definitions. In keeping with SSM philosophy, the conceptual modelling describes the activities necessary to achieve the 'transformation process' detailed in the root definition and displays the logical dependency of the parts that form the system. The transformation process being defined as a "...process in which some defined input is changed, or transformed, into some defined output" (Checkland and Tsouvalis, 1997, p. 156). Conceptual modelling expresses the systems defined in the root definition according to the essential activities required for the system to function. "The relationship between a root definition and a conceptual model in SSM is therefore a relationship between a 'being' and a 'doing' or, between a 'schema' and the 'performance' of that schema" (Checkland and Tsouvalis, 1997, p. 158). The

conceptual modelling stage of SSM promotes the generation of alternative 'systems', rather than promoting the selection of the 'most relevant' system from the beginning: the selection process occurring after comparing the possible solutions with a 'real-world' action. However, these systems can only be validated in terms of their conceptual feasibility, as they are only theoretical constructs of reality.

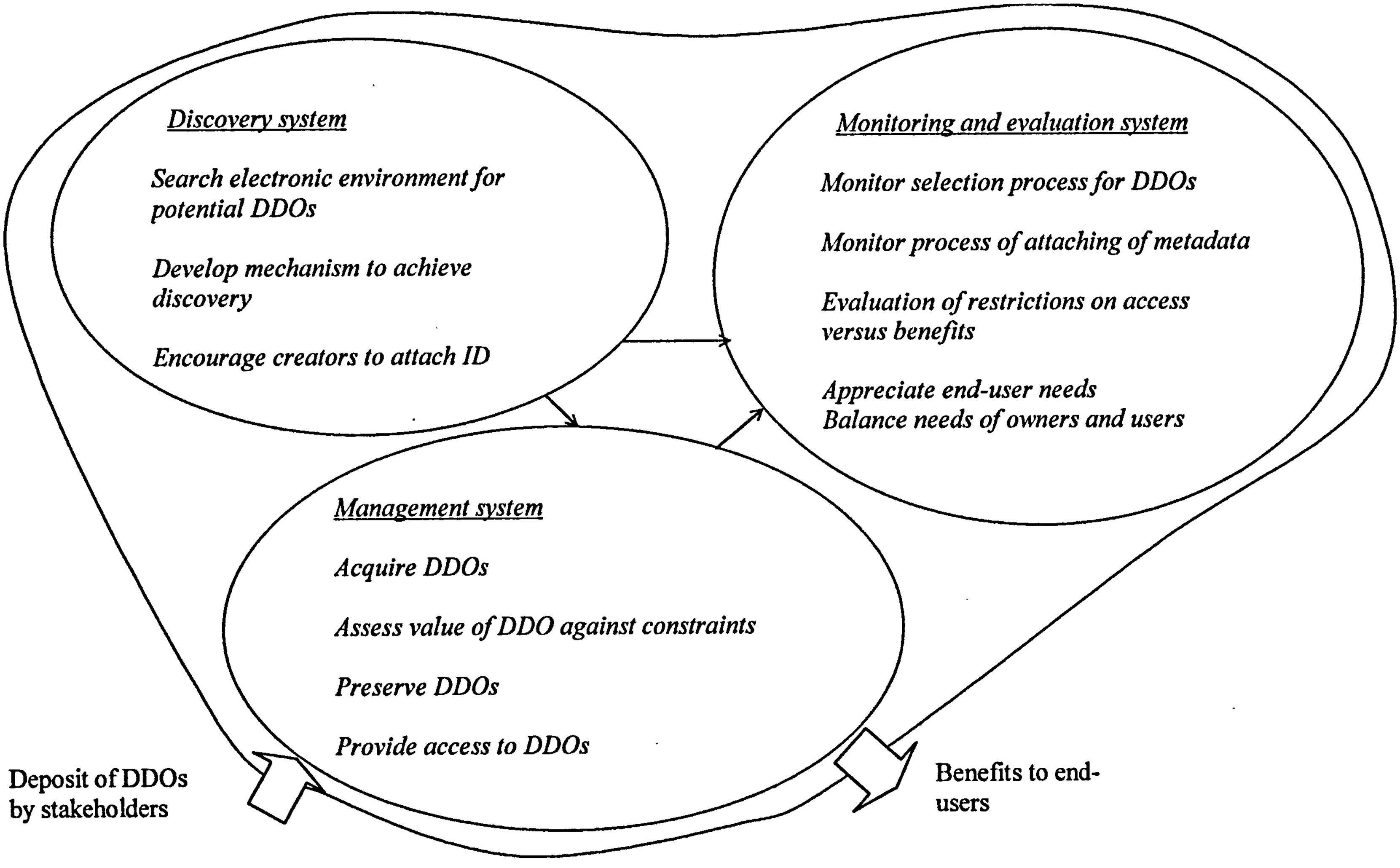
Taking the root definition, or alternative root definition as a baseline it is possible to generate conceptual models of the systems implied in the definition. The conceptual models aim to express the 'activity systems' implied in the root definition.

6.1.1 Initial conceptual models developed from the root definition

The root definition (see Section 4.14.1) suggests the need for 'appropriate' structures to support long-term preservation and from some stakeholders the view that guidance is required.

Figure 6.1 illustrates the first stage of this process, where a general structure of the systems described in the root definition is modelled. The modelling focuses attention on the main components, their relationship to each other, together with the input and output of the system needed to carry out the activities named in the root definition. To manage future access to DDOs, the conceptual modelling suggested the existence of three inter-linking systems concerned with discovery, evaluation and monitoring, and maintenance. Additionally, the main stakeholder responsible for preservation was assumed to be the repository, the 'problem-owner', as determined by their central aim of preserving DDOs over the long-term. As the 'research problem' was characterised by the lack of an existing 'all embracing' framework, a more structured approach was pursued in the conceptual modelling process. However, an alternative root definition was considered that did not specify the means of achieving the aim of future access, such as the development of selection criteria.

Figure 6.1 The general structure of the systems described in the root definition



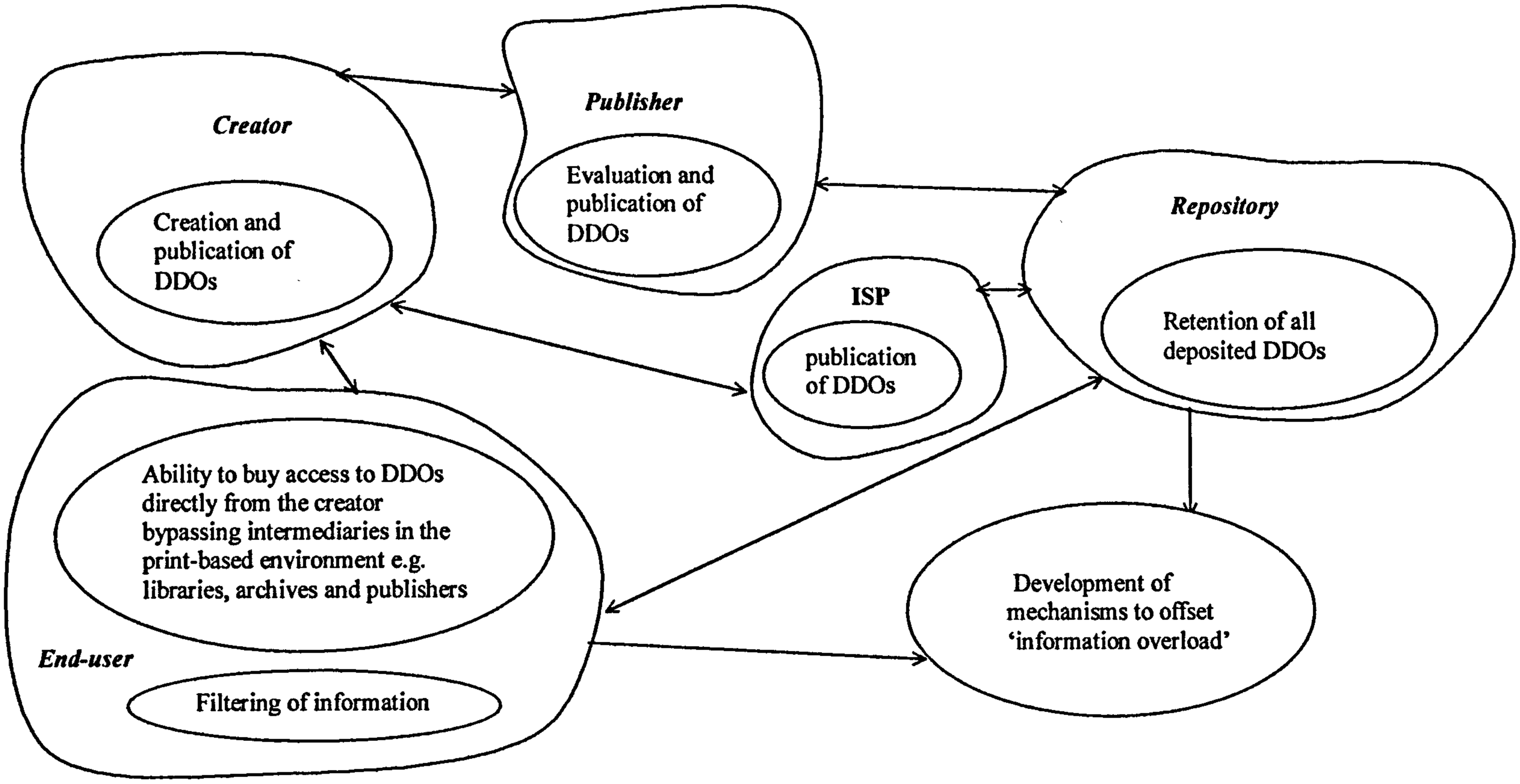
6.1.2 Alternative root definition

Alternative root definition - It is suggested that a flexible framework is created that preserves DDOs over the long-term to ensure future access. The framework hopes to benefit end-users of information and meet the conditions, and/or concerns of creators, owners and funding bodies. This should be undertaken by organisations who would have the responsibility for preserving digital materials and this may be achieved by stakeholder organisations working in consultation with advisory bodies. This takes into consideration the constraints of cost, human resources, legislation and the lack of and instability of standards.

From this alternative root definition a second conceptual model was developed (see Figure 6.2), which concentrates on developing a system that exerts minimum structure on the management of DDOs. This conceptual model centres on the roles of stakeholders and allows for the possibility of more than one stakeholder performing several functions and the duplication of functions between stakeholders, for example, the American Physical Society's involvement in the Physical Review Online Archives, PROLA (Thomas, 1998). This second conceptual model examines the question of how long-term preservation could be achieved by assuming that responsibility for preservation is shared by creators, disseminators and repositories. However, it also highlights the fact that too much fluidity was detrimental to the functioning of a future framework. Therefore, it was felt that amalgamating the specificity of the first conceptual model with the concentration on stakeholder roles and the flexibility of the second, gave the most desirable outcome.

The development of this alternative conceptual model (Figure 6.2), showed that the introduction of a measure of fluidity in a proposed framework could be beneficial to its implementation, as it accommodates variations in the way stakeholders view their

Figure 6.2 Conceptual model providing a minimum structure to the long-term management of DDOs



roles. However, there are identified disadvantages of this second model marked by the absence of proposed structures to resolve 'grey areas'. This could lead to the creation of a framework with minimal structure and perpetuation of the present absence of guidance. The need for selection is not yet supported, but is included in the framework for further consideration.

6.1.3 Summary of the initial conceptual modelling stage

It would appear from the above sections that the conceptual modelling developed from the root definition is more structured than the model created from the alternative root definition. However, the latter is useful in demonstrating that flexibility is an important element in a future framework.

From the general structure shown in Figure 6.1, increasing detail was built into the model of stakeholder roles considered in Figure 6.3. This aided the assessment of the 'grey areas' requiring resolution for a proposed future framework to be successful. As increasing levels of specificity are built into the conceptual model, the relationship between the levels becomes relative to where it is viewed. The system that emerge from the root definition (Figure 6.1) need to be assessed against the stakeholder roles defined in Figure 6.3. This process allowed an analysis of 'grey areas' between stakeholders to be matched with system processes, and aided the development of mechanisms to resolve these areas. As the interviewed stakeholders belong to different organisations the question of who is to be responsible for particular operations is a vital element in the successful functioning of a proposed framework. Therefore, it was felt to be important to use both the conceptual model of stakeholder responsibilities and that of the chosen system in order to create a proposed framework (see Chapter 7). For example, the creation of a rights management system is viewed as one of the 'hows' of achieving a framework for long-term preservation, but in turn it could be viewed as a 'what', which needs to be developed by discovering further 'hows' to implement it. Therefore, the conceptual modelling process needs to take into consideration not only what systems can be generated and how these could be

implemented, but also the roles and responsibilities of the stakeholders. SSM encourages the analysis of the 'research problem' in conjunction with its social context, therefore the analysis would not be comprehensive if an examination of stakeholder roles and responsibilities were not considered.

Figure 6.3 identifies the repositories as the major stakeholder illustrated by the fact that the whole of their role is within the core area of the proposed system. The roles for each stakeholder are delimited more than in Figure 6.2, as some issues in the 'grey areas' have been identified. The super-metadata framework, detailed in Chapter 7, aims to take these issues into account, by creating structures to aid their resolution. The definition of stakeholder roles and responsibilities has been identified by other research into the preservation of electronic records, as an area requiring further investigation (Kandur, 1992; Haynes *et. al.*, 1997). As this research has created a conceptual model of the area of responsibility for each stakeholder, it is possible to define where overlap of stakeholder roles is likely to occur. These overlaps need to be taken into consideration when developing future conceptual frameworks. The boundary of the proposed framework can be super-imposed on the model of stakeholder roles (as shown by the broken line in Figure 6.3). The ability to define the parameters of a conceptual system that overlaps individual stakeholder organisation's boundaries is seen as a strength of this methodology, "...part of the strength of systems thinking lies in the analyst's power to select a viewpoint he considers relevant and to denote systems, whose boundaries do not coincide with organizational boundaries..." (Checkland, 1984, p. 257). As the researcher has placed the primary focus for long-term preservation on repositories, the whole of their role is encompassed by the new framework, whilst only part of the role of most stakeholders is relevant to the system. The sub-sections that follow examine the interaction between stakeholders and the nature of 'grey areas' identified through the creation of the rich picture (see Chapter 4). This exploration of the issues that each area contains aims to aid the development of strategies to cater for these within the super-metadata framework (created in Chapter 7).

Figure 6.3 Conceptual model of the stakeholder responsibilities and concerns based on the root definition

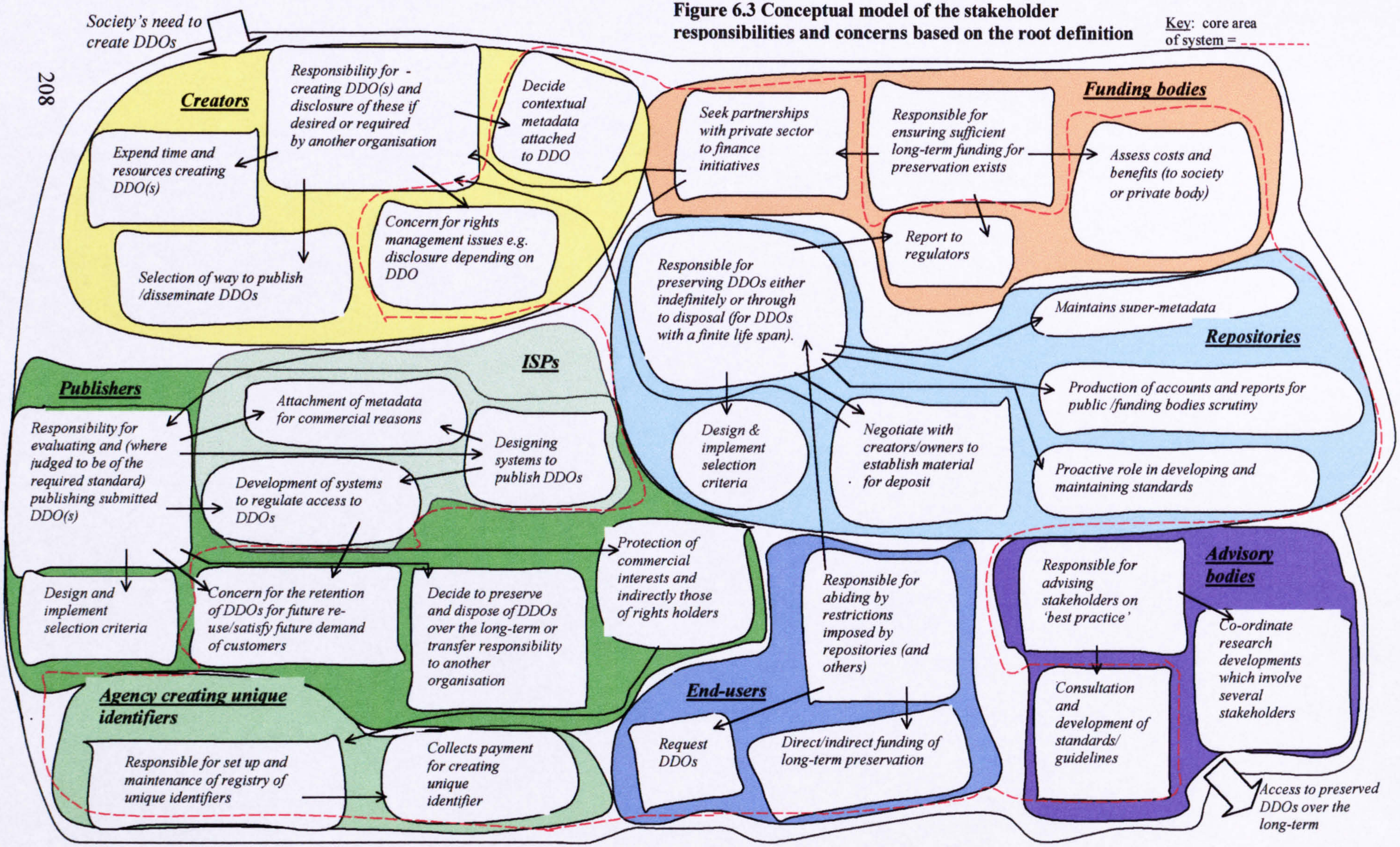
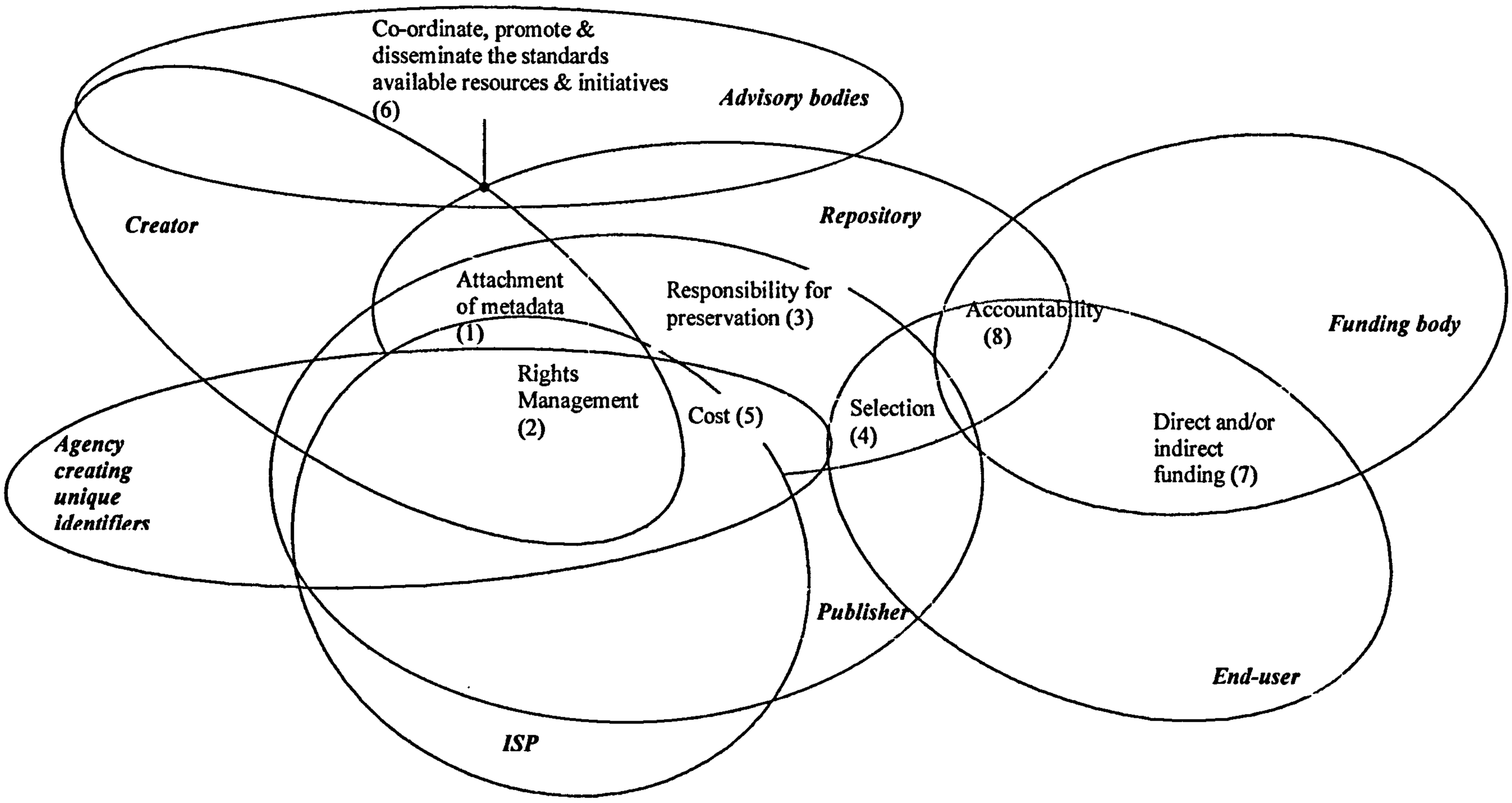


Figure 6.4 Intersection of stakeholders' roles in relation to the issues contained in the identified 'grey areas'



6.2 The benefits of employing conceptual modelling

Conceptual modelling encourages the exploration of alternative solutions to the 'research problem'. Therefore, this process allows for the analysis of the models against alternatives before a system is chosen.

The exploration of alternative models is beneficial to this research, as viewing the challenges and potential system workings from different stakeholder perspectives encourages flexibility in the proposed framework and examines the way technical mechanisms would need to function to fulfil the aim(s) which the framework hopes to achieve. In addition, the conceptual modelling process allows objects being preserved to be viewed, not only in terms of their intellectual content and physical attributes, but also to assess their significance to society.

> ...the focus on the content and format of information to the exclusion of consideration of the social contexts and functions of knowledge is a distortion of the nature and dynamics of scholarly communication and the academic community. (Lyman, 1999, p. 377-378)

6.3 Conceptual model of stakeholders' responsibilities and concerns based on the root definition

At the time stakeholder interviews were conducted (Autumn 1997 to Summer 1998) challenges were presented by the lack of a structure (evidenced by the opinions of stakeholders together with concerns voiced in the literature) resulting in the research focusing on the modelling of a framework. The exploratory nature of the conceptual modelling process, aided the analysis of 'grey areas' identified in the rich picture. Through this process, it became apparent that a proposed solution to long-term digital data management needs to embrace an understanding of the issues involved in such areas, and develop strategies to resolve potential challenges. Figure 6.4 demonstrates the issues contained in the 'grey areas' identified by this research.

6.3.1 Creators

The nature of DDOs (Sections 1.6. to 1.6.11), and the increasing reliance of society on digital data combined with the concerns of stakeholders, strongly point to the need for society to maintain its 'societal memory'. It is the creator, in the form of an individual, group or organisation that controls the intellectual content of a DDO and selects what is to be published or disseminated within a closed environment. If creators wish the intellectual content of their DDOs to be preserved over the long-term, it would appear to be advantageous for them to attach sufficient metadata to the objects. This metadata enables the identification of objects, explains their use and significance of the content, and allows creators to assert their ownership of the information. The creation process represents time and effort expended by creators to create DDOs and therefore this stakeholder group is concerned with maintaining ownership of their intellectual property (see Figure 6.4 (1)). Rights management issues bring the creators into conflict, or, consultation with other stakeholders (see Figure 6.4 (2)). How creators wish to influence the disclosure of their information is one facet to this identified 'grey area'. This creates an area of potential conflict for which co-operation for mutual benefit appears to be the solution. At present informal disclosure mechanisms exist in the networked environment but if confidence is to be increased these will need to become more sophisticated.

The interaction of creators with other stakeholders within a networked environment, unlike the traditional print environment, has yet to be regulated. In the latter the relationship between author and publisher is regulated by copyright custom and practice. The relationship between the creator and other stakeholders is clearly understood, with regulations between the creator and the publisher governing how the product created between them can be used. This lack of regulation in the networked environment is a concern to creators who do not wish the DDOs that they create to be misused, but wish to receive due recognition for their work. There may also be monetary disincentives for creators to deposit certain material with public repositories. Some DDOs will be of greater monetary value (have a higher market value) than

others, or represent a greater investment in terms of time and, or, money. The present system of compensating creators for lost revenue created by public library access to traditional print-based materials, namely the Public Lending Right may, if modified to include DDOs, obviate the disincentives. However, depositing with an archive may be beneficial to the revenue of commercial publishers, by providing another means of advertising products.

In the electronic environment the nature of the relationships between stakeholders changes. This is evident from the significant overlap of stakeholder roles in the area of rights management. There is therefore a need for repositories to become much more proactive and become actively involved in the process of rights management. This role involves resolving issue such as: how to attach ownership to a DDO, assessing access rights, and communicating any restrictions where necessary to other stakeholders. Repositories need to establish rights management processes that fulfil two purposes:

- Instil confidence in the depositor that appropriate processes are in place to prevent inappropriate disclosure of their information taking place.
- Determine the repository's role in relation to the management of individual DDOs.

The negotiation of access rights ultimately influences the end-user, as restrictions on access limits the amount and also which materials are available for consultation in the future. The roles of the publisher, creator, and the repository overlap to a certain extent, as the following paragraphs demonstrate.

6.3.2 Publishers

The conceptual modelling of the root definition identifies that the publisher's role is concerned with the evaluation and publishing of DDOs submitted by creators. The role as with those of other stakeholders shown in Figure 6.4, is primarily defined by

their aims, and is then influenced by the way other stakeholders interpret their roles and responsibilities at any given time.

There is a potential for overlap of roles between creators and publishers with respect to the attachment of metadata (see Figure 6.4 (1)). If the creator has attached insufficient metadata then the publisher may need to supplement the information to aid their commercial activities. Similarly, if publishers and/or creators attach insufficient metadata to DDOs then repositories need to attach more information to their records to facilitate management and ensure future access.

The regulation of access to DDOs is also a challenge within the publisher's domain (see Figure 6.4 (2)). This restriction of access links to a concern for the protection of the publisher's commercial interests. The wide availability of DDOs and ability to create paper copies of them could undercut the publisher's ability to sell copies. Indirectly restricting access protects the rights holder's interests. For these reasons, it would appear to be in the publisher's interest to develop systems to regulate access to the DDOs that they publish to protect both commercial interests and the interests of creators.

Publishers are concerned with the retention of published objects as they feel customers may wish to acquire, or, pay for access to the same DDO at a future date, or, they themselves may wish to re-work or re-package a DDO to encourage further sales. Nevertheless, the publisher has to balance these gains against the costs associated with maintaining an archive or of delegating responsibility to a third party, be this a commercial organisation or a public repository (see Figure 6.4 (3)). As developments in electronic archiving suggest (Harnad, 1999; Thomas, 1998), there is potential for an academic society that publishes material to also undertake an archiving role. The transfer to a third party is seen, from the stakeholder interviews to be conditional on factors such as security and speed of access to the data for the depositing organisation.

6.3.3 Internet Service Providers

The present role adopted by some ISPs suggests that the basis of their revenue streams are different from publishers and long-term retention of hosted information is generally not a priority, rather as clients' subscriptions lapse their disk space is re-used (Thompson, 1998). If in the future ISPs adopt some of the characteristics of publishers then the distinction between the two could become blurred.

6.3.4 Agency creating unique identifiers

Stakeholder concern for securing the intellectual property, as demonstrated in the interview data (Table 3.7A and 3.7B, 1A to 12A), points to the need for a system of attaching ownership of intellectual property to the DDO at an early stage in the life cycle. An agency responsible for creating a mechanism to attach ownership may be responsible for maintaining the mechanism, and collecting a fee from stakeholders who wish to make use of such a service (Figure 6.4 (2)). It is debatable who would be responsible for maintaining such a service: a public body, such as the Copyright Office and the Library of Congress who together have instigated the Copyright Office Electronic Registration, Recordation and Deposit System (CORDS) project[1], or a commercial agency as with the present ISBN system in the United Kingdom. Several stakeholders could be interested in attaching something that denotes ownership, although who actually attaches it could vary in practice. Creators who are initially not intending to publish a DDO, or are not concerned with establishing ownership for a particular DDO may not be interested in participating in such as scheme. If the same object is published, or is identified by a repository as having long-term importance at a later stage, then the publisher, or the repository, will need to establish ownership of individual DDOs in such cases.

---

[1] The project aims to create and test a system of registering and recording copyright applications, copies of works and related documents, transmitted over public electronic networks. The intention is that the Library of Congress will be able to use digitized copyrighted information where the conditions of access have been stipulated by the owner (URL: http://lcweb.loc.gov/copyright/cords/cords.html).

6.3.5 End-users

Future access for end-users is the principal reason for the preservation of DDOs over the long-term. However, the demands of future end-users are unknown, and therefore this makes selection decisions difficult to undertake (see Figure 6.4 (4)). As a stakeholder group, they are the primary drivers for long-term preservation, whilst at the same time presenting challenges as their future needs are difficult to predict. End-users act as drivers, not only in terms of their future needs but also as a source of either direct funding through payment at the point of access, or indirect funding through taxation (see Figure 6.4 (7)), which can raise accountability issues (see Figure 6.4 (8)).

At present there would appear to be an absence of mechanisms that co-ordinate concerns for disclosure, expressed by those creating and, or, retaining DDOs, with end-user access to these objects (see Figure 6.4 (2)).

The role of the end-user is changed in the electronic environment as they have the chance to perform a more active role, they can search for the information. Creators and rights managers have to take on a different role. They need to establish how the documents are to be used at the point at which they are created. The documents are published at the point at which they are placed on the network. The creator can publish in a networked environment without significant intervention from others. There is no distinction between formal and informal publishing in the electronic environment and this is where it differs from the traditional paper-based environment.

The user has to do his/her own selecting to check the validity of DDOs when using the World Wide Web, whereas DDOs that are formally published or selected for a repository, do not require this check by the user.

6.3.6 Repositories

The primary role of the repository, whether situated in the commercial or public sectors, is the preservation of DDOs, either for an indefinite period, or where DDOs have a finite life span through to their disposal or deposit with another organisation. Such preservation involves a number of stages that are generally recognised to be the responsibility of a repository (see Figure 6.4 (3)). However, it is evident that responsibility for preservation could be shared between stakeholders, as evidenced by the interaction or overlap between stakeholder roles exhibited in Figure 6.4. As repositories re-define their role in the digital environment their functions may embrace those of other stakeholders, such as the British Library's decision to launch its own Internet service (Copps, 1999). This section demonstrates that many of these stages involve negotiation with other stakeholders. Many of the communication channels shown in Figure 6.3 are formalised channels in the traditional print-based environment, but have yet to be formally established in the digital environment.

The interviews with stakeholders who are attempting to preserve DDOs, demonstrate that there is an absence of guidance on how to select DDOs and implement administrative structures that ensure the management of DDOs over the long-term. The embryonic nature of structures to co-ordinate rights management issues between stakeholders, means that 'grey areas' exit where creators are unsure of the protection of their information from inappropriate disclosure, and repositories are trying to balance creators' rights with the maintenance of access to as many DDOs as possible (see Figure 6.4 (2)). The absence of a framework based on finding solutions to the human centred challenges that long-term digital data preservation presents, and the interrelation of these with technological issues, has prevented the assessment of costs. Costs form one strand of the 'grey areas' between funding bodies and repositories (see Chapter 8). Part of establishing the conceptual feasibility of a proposed framework would be aided by the identification and apportioning of costs, together with the potential range of scenarios for different stakeholders. The benefit to society of preserving DDOs over the long-term needs to be assessed and balanced against the

costs of managing DDOs (see Figure 6.4 (5)). It is apparent that, for a future framework to be successful, structures that provide sufficient incentives or benefits to all stakeholders need to be developed.

6.3.7 Advisory bodies

It is apparent from the root definition that advisory bodies could play an essential role in co-ordinating cross-sectoral developments. Such developments are seen to be advantageous on a number of levels and areas. On a conceptual level there is a need to co-ordinate the different philosophical approaches to long-term preservation that stem from a number of factors. These include different experiences and concerns for different types of digital objects; and/or the overriding concerns of a discipline. This is demonstrated by the difference in approach between computer science and its concern for access to resources versus information science approach to focusing on the importance of institutions and services.

It would appear that advisory bodies could perform a vital future role in co-ordinating and disseminating research developments, for example, by bringing together a diversity of stakeholder groups through conferences and workshops (see Figure 6.4 (6)). Similarly, advisory bodies could play an important role in the development of standards and guidelines for preservation, which at present are a concern to stakeholders who wish to preserve DDOs over the long-term. A plethora of metadata standards already exist, but there is an absence of an examination of the conceptual issues that underpin long-term preservation, and the nature of DDOs. The re-evaluation of existing practices incorporating the opinions of stakeholders is seen, as evidenced by the empirical data from this research, to be an essential part of this process.

6.3.8 Funding bodies

The funding of digital data preservation is an emerging area. In the higher education sector in the United Kingdom funding has been given to develop a distributed data service and private finance initiatives are being developed to facilitate partnerships between private and public sectors. How the funding for digital data preservation is to be achieved is uncertain, because the roles of the different stakeholders are still being formulated. Therefore, who will be responsible for funding, and how the costs are to be apportioned, are at present undefined (see Figure 6.4 (7)).

6.4 Summary

The development of a framework would appear to be a logical progression from the soft systems modelling. The 'grey areas' are concerned with stakeholders' areas of responsibility and their concerns are inter-linked with the development of the framework. As the previous sub-sections demonstrate, these 'grey areas' occur where overlap between roles exists, creating the uncertainty of where the responsibilities of one stakeholder end and another begin, and where the boundaries between technological and human issues lie. The results from the analysis of these 'grey areas' using soft systems modelling informs the development of a framework for long-term digital preservation (see Chapter 7). The conceptual underpinning for the framework draws on the results from the empirical research supported, where appropriate, by the literature.

The conceptual models that this research has used, derived from the root definitions, examine the roles and responsibilities of stakeholders. The first model allows greater flexibility in the interpretation of roles to occur, whilst the second provides more details of how roles may be interpreted. This greater level of detail exposes the 'grey areas' between stakeholders and allows an examination of the issues each of these areas contains. It was decided to combine the detailed examination of the 'grey areas'

with the flexible interpretation of the stakeholders' roles demonstrated in the first conceptual model to inform the creation of a new framework. The framework is repository focused, as the assumption is that the remit of such stakeholders would make it highly likely that responsibility for long-term preservation would rest with this stakeholder group. Therefore, the repository was labelled as the primary 'problem-owner'. In addition, the 'public good' argument outlined in Sections 1.6.1 and discussed in more depth in Sections 8.5.1 to 8.5.2, supports the continuation of public sector repositories to maintain future access to DDOs. The framework does allow for the overlap of stakeholder roles and for differences in the way individual stakeholders interpret their role.

The fundamental aim of long-term preservation, to provide future access, remains unchanged as one moves from a print-based to a digital environment. However, many of the elements surrounding this aim are in a state of flux on a number of levels. On a management level the roles and responsibilities of stakeholders are evolving, leaving questions, such as, who will be responsible for which tasks, what costs will be incurred by different stakeholder groups, unanswered for the present. Similarly, the issues of a lack of standards, the stability of these standards and the inadequacy of the existing legal deposit framework to support advances being made by stakeholders, remain unsolved. The development of a framework needs to be in tune with the 'research problem' and take account of the effects of a large number of factors that affect its resolution.

Chapter 7

Development of a new framework for the long-term management of DDOs

7.1 The need for the creation of a new framework

In order to develop a management system that supports the long-term access to DDOs originating from varied sources, a common metadata is required. As Sections 5.4 and 5.6 to 5.6.2 demonstrate, there are practical difficulties that result from attempting to integrate formats at the metadata level and thus the development of a higher level metadata for long-term management, super-metadata, is suggested. To support the functioning of super-metadata this research advocates the construction of a new framework, which seeks to define stakeholder roles, and suggests mechanisms to resolve the issues raised by stakeholders and identified in the rich picture.

Identification of the need for new structures in the digital environment is expressed by the Task Force on Archiving of Digital Information, "It is a problem of building -- almost from scratch -- the various systematic supports, or deep infrastructure, that will enable us to tame anxieties and move our cultural records naturally and confidently into the future" (Task Force on Archiving of Digital Information, 1996, p.7). The underlying principles of the super-metadata framework are detailed below.

The conceptual modelling process has pinpointed crucial human-centred issues that need to be addressed, and the solutions to technological issues this research considers to be essential to aid the development of a framework, have been considered. Many of the facets detailed in this framework are dependent on the co-operation of many stakeholders and often involve sharing responsibility.

7.2 Principles underpinning the framework

This research attests that a future framework needs to be underpinned with core principles to provide a conceptual structure that could achieve global applicability on a technical level and capable of being adapted to take into account the differences in administrative and legal practices in different countries.

7.2.1 Recognition of the functional difference between metadata and super-metadata

Within the considered framework, the function of metadata is to provide the information which describes the content, context and structure of the data, whilst the super-metadata contains the management information essential to ensure DDOs receive appropriate long-term preservation. The framework proposes a conceptual approach which repositories can use to create appropriate super-metadata, that it is hoped will have universal applicability.

7.2.2 Life cycle[1]

The first point to consider in the life cycle of a DDO is, if it is to be preserved, it is advisable to 'capture'[2] it at the earliest date, to avoid possible loss. From the point of view of the repository, the life cycle commences when the selection of the DDOs for preservation takes place[3]. Part of this selection process involves specifying the required life span or review period (period of time after which the DDOs requirement is re-evaluated). Where the DDO has a review period, the manager is alerted when

---

[1] The term life cycle within this research refers to the stages through which a digital data object passes from its creation through to its disposal.
[2] Capturing refers to the identification of a DDO for preservation.
[3] Occasionally the repository has influence over the development of a system for the creation of DDOs and therefore has an input from an earlier stage in the life cycle of a DDO. However, as noted in Chapter 6, the electronic environment is changing the nature of the relationship between stakeholders. Therefore, the research suggests that repositories adopt a more proactive role in such an environment in particular through the provision of guidance and management of rights management issues.

action is necessary to re-evaluate the DDO. It is possible that more than one repository will select a given DDO for preservation. In such instances, the DDO will have more than one life cycle. The life cycle criteria may differ between different repositories, for example one repository may keep the DDO indefinitely whilst another may set a review date. To prevent unnecessary duplication of effort and resources and to aid the synchronisation of management between repositories this framework encourages intercommunication between repositories. It is anticipated that this could be achieved by collaborative development and maintenance of the super-metadata and the design of intelligent software agents which work collaboratively. This research suggests that a certain amount of duplication could be advantageous in physically protecting DDOs from man-made or natural disasters. Universal bibliographic control mechanisms are considered necessary to co-ordinate the super-metadata structure.

When a DDO reaches its review date the decision to continue preserving the DDO can be re-assessed along with an examination of physical preservation issues, such as, the need to migrate or to maintain the facilities to emulate the DDO. To aid the decision to continue to preserve the DDO or to dispose of it, experts could refer to a mechanism in the super-metadata that records the amount of use a DDO has had since its last review. Selection for further retention could be based on a number of criteria, including the statistics obtained from such counters. However, DDOs that have not been accessed in the early years of their retention may still have long-term value. Each time a resource is reviewed, the decision taken is recorded in the super-metadata, this enables DDOs which have been selected to be preserved for a further time period to be automatically assigned for future review. Similarly, where the decision has been taken to dispose of a DDO at such a review, the information needs to be recorded and retained in the super-metadata whilst the DDO and its metadata are deleted. It is important to record this information because although the DDO no longer exists, users who have previously accessed it may enquire as to its whereabouts in the future.

7.2.3 Selection based on use to society

Selection is an essential element of the framework, based on the DDO's usefulness to society, which can include administrative, legal, historical, scientific, literary or scholarly value. Long-term access to digital information needs to concentrate resources on preserving selected DDOs. Selection, therefore, evaluates the DDOs for its 'worth' in preservation terms and ensures that in the process the significance of the intellectual content is recorded in the metadata for future end-users. Public accountability of the selection process needs to be developed to protect the public interest in maintaining long-term access to a wide range of DDOs.

A good selection process avoids the expense of retaining DDOs unnecessarily. Although this selection itself has a cost associated with it, this cost is only small when spread over many years (see Section 8.2.4). When establishing parameters for selection, the balance between social and economic benefits needs to be considered, as demonstrated in Section 8.5.4. If a certain DDO is disposed of, the economic benefits of not having to preserve the DDO may be outweighed by the loss of that information to society. From a corporate records management perspective, Megill (1997) identifies the cost implications of information loss, "The loss of information from corporate memory can be very costly, no matter how it is stored. Estimates of the cost of lost or misplaced information are difficult to make, but some attempts to quantify these costs can be made" (Megill, 1997, p. 8).

Selection costs can be minimised by utilising automated processes. These include the selection of DDOs from 'reputable sources' determined by experts in their fields, and by the use of intelligent software agents (see Section 7.3.2.1). These software agents could be programmed to recognise sites adhering to codes, such as the self governance initiative Health On the Net Code of Conduct (HONcode), designed to address the varying quality of health information available on the Internet[4]. This research

[4] URL: http://www.hon.ch/HONcode/background.html

suggests that subject specific selection criteria developed by experts within the relevant fields will supplement the generic criteria detailed in Section 7.3.3 and 7.3.4.

The present absence of selection criteria, that are part of an all embracing structure, as identified in Section 1.10, and the results of testing the hypotheses relating to different types of DDO (see Sections 5.2 to 5.3), suggest the need to create additional selection criteria for dynamic and living DDOs.

The value attached to a DDO, in the selection process, is measured according to the criteria of the repository managing the super-metadata. Valuation of a DDO is therefore dependent on who is selecting and the criteria by which it is being measured. Within a commercial organisation the monetary 'value' is often paramount, however, within the context of this thesis, 'value' also embraces the concept of 'value' as a public good (see Section 8.5.1). This includes DDOs deemed to be of administrative, cultural, literary, scholarly, scientific, legal or historical importance.

Value is also influential in determining the life span of the DDO. As stated earlier in this thesis, the owner of the DDO may regard a particular object as having 'value' for a defined period of time, whilst other stakeholders, assuming they are aware of the existence of a resource, may view its life span in different terms. The value of a DDO does not necessarily remain constant over time, and can be influenced by its connection with other DDOs. Therefore, the philosophical underpinning of selection criteria is seen to be important to the evaluation of an object and the assessment of its dependence and relative worth to other DDOs. This could potentially link an enormous number of DDOs. This research considers this to be impractical and therefore advocates the selection of DDOs based on their individual merits.

7.2.4 Legal deposit and the development of rights management and disclosure restrictions

If the proposed extension to legal deposit occurs, then computer-assisted mechanisms will need to be developed to ensure designated legal deposit repositories receive newly published DDOs. However, this arrangement will not cover the growing amount of materials published by individuals or organisations other than 'traditional publishers', and therefore, stakeholders wishing to preserve publicly available DDOs will need to proactively search and evaluate this material (using the methods described in Section 7.3.2.1).

The deposit of DDOs is a 'grey area' (identified in Figure 6.4) requiring the balance of a number of stakeholder's interests, some of which conflict. Owners' concerns for the loss of revenue need to be safeguarded, and a compensation scheme developed which repositories can use when it is not possible to compensate owners by other means, such as restricting access through limiting distribution. A repository's right to enforce the deposit of published DDOs, judged to be of long-term importance in the interests of the public good, requires legislative support. Future end-users' requirements of access to as much information as possible also require balancing against the owner's right to restrict access. Thus, the super-metadata framework suggests the extension of legal deposit from a voluntary to a mandatory structure that includes the development of rights management practices, to ensure the balance between owners and users is maintained and regulated.

Ownership is seen as the primary driver in the preservation process (Feeney, 1999), determining the form and content of the object in the creation phase, its disclosure and the degree of access to it, through to its long-term preservation or disposal. The digital environment presents new challenges to existing mechanisms for establishing ownership. If global production and distribution are merged, as occurs on the Internet, it could present problems for repositories retrieving information. The collection of foreign material that has not been copyrighted highlights some retrieval problems. If

this material has been preserved by a repository it could later present problems if this same information has been included in a copyrighted work. These difficulties could include negotiating and assigning different access restrictions for the same material and challenges such as, how to restrict access and use of DDOs, if required, to national users in a situation where global access is possible. It may be that mutually acceptable international agreements are required to establish procedures to harmonise intellectual property issues of resources on global networks. This research considers that for a framework to function effectively it is vital to link ownership with a particular DDO as early as possible in its life cycle. Ownership of DDOs controls a number of crucial factors that determine future preservation procedures such as how it is kept; by whom; for how long; and who has access to it. Encouraging owners to register the DDOs ownership at an early stage helps to avoid additional administrative expenses at a later stage, such as negotiating rights with creators.

Various information-dependent organisations have been developing mechanisms to manage aspects of digital data particular to their needs, for example, record managers in commercial organisations are concerned with ensuring security of their information. It is essential to include rights management information in the super-metadata as this determines who can read the information, and establishes whether a licence has to be paid or authorised. The rights management component of the super-metadata framework seeks to establish the necessary information, where possible from the metadata. The use of identifiers is presently gaining ground in the information community as the following example shows. The Interoperability of Data in E-Commerce Systems and the Digital Object Identifier (INDECS/DOI)[5] initiatives have combined with the aim of finding common semantics to support resource description metadata. This they aim to achieve by focusing on the shared requirements of the two initiatives to create a common conceptual model (Bearman *et al.*, 1999). The

[5] INDECS like Dublin Core has the same goal of resource discovery, but the main thrust is establishing intellectual property rights of objects. The DOI is a sponsor of the INDECS project, and is concerned with developing persistent unique identifiers and standardised metadata for individual digital objects (and consists of three basic elements; people, objects and deals). The IFLA Functional Requirements for the Bibliographic Record and concepts detailed in the Resource Description Format underpin these developments (URL: http://www.indecs.org/).

development of the two initiatives shows the interdependence of resource discovery and rights management metadata.

Collaborative ventures show the importance of intellectual property rights management to both the user and the creator. Proponents of such initiatives stress the importance of finding synergies to prevent unnecessary duplication of standards in the future. However, for the reasons expressed earlier in this thesis (see Section 4.10), super-metadata is required to manage the range of metadata formats necessary to support the range of DDOs maintained. The combining of metadata formats in this manner may lead to an escalation of additional elements, ultimately creating a format that could prove to be unwieldy to implement. The common base for the DC/INDECS/DOI consists of published works, although further work on formulating fundamental categories for information resources based on the IFLA FRBR model is underway. As differences can occur in the interpretation of standard formats, mapping of all fields may not be possible to achieve. However, it may be possible to achieve standardisation on a particular field, for example, rights management. The super-metadata framework allows for the possibility of having a regulated environment for some areas, and a flexible system that also allows individuals and networked environments to be catered for.

The estimated costs involved in linking legal deposit responsibilities with the formulation of super-metadata through the attachment of unique identifiers at creation, may prove to be acceptable to stakeholders (see Appendix C). The set up and running costs that are associated with either modifying the functions of an existing organisation, or the formation of a new agency, require factoring into a cost model which is considered in Chapter 8.

As demonstrated above it is beneficial to link information concerning ownership with the information content of the DDO. To achieve such linkages it is important to develop the necessary infrastructure to support these processes. If unique identifiers are attached to the DDO, such as the developments discussed above, then a re-

examination of the existing structures that support the generation and attachment of this information to traditional paper-based objects is considered necessary. If, as the previous chapter suggests, an agency is to be responsible for the creation of unique identifiers, its implementation will require new legislation to give it enforcement powers, provide training for publishers and advise repositories on its implementation, together with the development and dissemination of procedures for creators. It is anticipated that within the proposed super-metadata framework each country would have an agency responsible for creating unique identifiers. To prevent duplication of identifiers these national agencies could be part of a co-ordinated global network of agencies.

7.2.5 Authentication processes

As the rich picture illustrates, (Figure 4.1), the reticence of a variety of stakeholders wishing to disclose or grant unrestricted access to some of the DDOs they create, appears to be linked to the level of confidence they have in current authentication mechanisms. The reasons stakeholders require assurances range from a creator's wish for DDOs not to be misused, to repositories being confident that deposited DDOs are authentic. As the level of need for authentication and how this is manifested varies between creators, it is important for those preserving DDOs over the long-term to develop a standard to ensure the authentication of deposits at the point of acquisition. Some stakeholders in the archival community share a concern for ensuring authenticity of the data content of electronic records, and advocate recording this in the metadata (see the work of Bearman and Sochats, 1999). However, for assurances of authenticity to be met, such a system needs to be put in place prior to the creation of records. As not all DDOs will be created in managed environments, the super-metadata framework needs to cope with a range of scenarios. Therefore, repositories can choose to authenticate DDOs at the point of deposit by processes such as time-stamping (see Section 7.3.6.1), which is used to indicate the point when this authentication took place. Within the super-metadata framework, the substantive data content and its metadata need to be authenticated. The former relies on the evaluation

skills of subject specialists in creating appropriate criteria for computer-assisted selection, and applying them where manual selection occurs. Authenticating the structure of the metadata format involves checking it against the information held in the metadata registry (Section 7.3.2.1). If the metadata does not pass the authentication process then deposition of the DDO cannot take place. Since, at this stage, the suitability of the DDO for deposit has been determined, consultations with the depositor will be necessary in order to correct the metadata for successful authentication.

It should be noted that time-stamping is only one suggested mechanism and is not advocated as the only means of authenticating DDOs. The researcher is aware of the potential danger for unauthorised alteration of the generated hashes. Research into the development of authentication systems would appear to be a research priority.

7.2.6 Management of DDOs with links to other DDOs

The dependency of a DDO on another digital object, which hypertext and hypermedia encourage, introduces a new element to the selection of some DDOs. These DDOs include CD-ROM encyclopedias that contain Internet links to the latest information on a given topic, such as, the Grolier Multimedia Encyclopedia[6]. These links essentially introduce a living component to an otherwise dynamic DDO. Repositories need to evaluate each DDO as an independent entity, since the continued existence of links in the present Internet environment is difficult to guarantee. Therefore, it is necessary to assess at the selection stage the value of a DDO without its links, and in some cases this may be a reason for its rejection. These links could also exist where a repository is preserving living DDOs at given time intervals, and in this case the super-metadata can reference earlier versions, in addition to linking to the current living version (see Section 7.3.4). An example is the monthly snapshots of the Retail Price Index

[6] URL: http://www.gi.grolier.com/gi/products/reference/97gme/docs/gmexplor.html

produced by the Office for National Statistics and held at the Data Archive at Essex University[7].

7.3 Structure of the super-metadata framework

The following sections define the characteristics of the super-metadata framework that will create a supportive structure to allow for past, present and future metadata developments within the information community. The new framework does not attempt to be prescriptive, but to provide a flexible structure that aims to cater for as wide a range of metadata formats as possible. Section 7.4 describes the actual components of super-metadata.

7.3.1 Description of the framework using entity-relationship diagrams

In constructing the framework, there is an attempt to balance the need for long-term preservation guidance, evident from the findings of the interview data (Chapters 3 and 4), with the experience of other initiatives that have occurred in the digital preservation field (Feeney, 1999). The process of constructing the framework uses the findings from analysing the data using SSM, to inform its structure. This enabled the examination of different scenarios, and an assessment of the inter-acting processes, before the framework was constructed.

The series of entity-relationship diagrams, which follow, show the complex linkages between and within the component parts of the super-metadata framework. These diagrams are sequential, and follow the stages in the data management processes which gradually build up a description of the DDO, by aiding the creation of the super-metadata. These diagrams, see Figures 7.1 to 7.10, are as follows:

---

[7] The Data Archive site lists the monthly snapshots of data sets (URL: http://dawww.essex.ac.uk/online/ons/monthly/index.html).

- Figure 7.1 Proposed super-metadata framework
- Figure 7.2 Construction of the super-metadata
- Figure 7.3 Subroutine 1: Selection and rights management
- Figure 7.4 Subroutine 1.1: Selection of DDOs with links to other DDOs
- Figure 7.5 Subroutine 3: Computer-aided checks of the metadata
- Figure 7.6 Subroutine 4: Determining the review periods of a DDO
- Figure 7.7 Subroutine 5: Processing an end-user request to access a DDO held in a digital data repository
- Figure 7.8 Linkage of DDO snapshots using version codes
- Figure 7.9 Subroutine 2: Digital time-stamping
- Figure 7.10 Division of responsibility for attaching metadata to a DDO

Legend for Figures 7.1 to 7.10

——— = Information flow

——— = Positive decision flow

——— = Negative decision flow

= Process that could be achieved using manual or computer-assisted means, or a combination of the two

= Manual process

——— = Stakeholder boundary

7.3.2 The origin of DDOs and initial capture mechanisms

7.3.2.1 Computer-assisted capture of DDOs from public networks

Public networks, such as the Internet, provide a vast quantity of data which is constantly being added to and changed. This research suggests that monitoring potential data sources such as public networks will be a necessary part of some repository's work. The remit and the collection policy of a repository governing

whether such monitoring is necessary and how it is to be conducted. However, the researcher is mindful of the cost implications of manual selection of data for preservation on a large-scale, and therefore, recommends repositories employ computer-assisted techniques wherever possible. This research advocates that the formulation of subject specific selection criteria is best achieved by involving experts in the particular field. This approach aims to take into account the differences in interpretation between disciplines when considering DDOs for long-term preservation. Mechanisms to support devolved selection, require a balance between human and computer-assisted elements, that is, "A judicious blend of automatic and human methods of analysis of meaning should be maintained in order to exploit information resources on the Internet" (Burke, 1997, p. 76).

The creators of the DDOs have a number of options open to them if they wish to release them into the public domain, or at least acknowledge their existence to other organisations. These options being: formal submission to a publisher[8], submitting it to a repository, publishing it on a non-networked media, or on the Internet through an Internet Service Provider (ISP). Within the super-metadata framework, all four of these routes would lead directly or indirectly to the repository (see Figure 7.1). To ensure material going through formal publishing channels or appearing on the Internet is viewed by the appropriate repository, it is necessary to extend existing legal deposit and develop new capture mechanisms. The current informal nature of publishing on the Internet demands a different solution to ensure the relevant repository can identify potential DDOs. This involves the repository developing intelligent software agents to search out DDOs matching their criteria. For resources to be selected by computer-aided means, these agents rely on Internet resources possessing metadata which conforms to a known format (such as DC or TEI) to aid resource discovery. The use of

[8] The term 'publisher' includes electronic publications created by specific academic communities such as that for biology (URL: http://www.soton.ac.uk/~library/elec/biosis.html) and academic societies such as the American Physical Society's development of an electronic peer-review journal. In addition, this society has developed an image archive for the Journal of Physical Review, namely the Physical Review Online Archives (PROLA) (URL: http://prola.aps.org/). Such developments are seen to pave the way for new avenues for collaborative research (Lyman, 1999). The current informality within the electronic publishing environment is taken into consideration in the super-metadata framework by allowing, through its flexible structure, for the emergence of new stakeholders and working practices.

metadata, by these intelligent software agents, being for resource discovery purposes. Intelligent software agents would enable repositories to discover, and select DDOs, by computer-assisted means, and return copies of these objects to the repository with the super-metadata already attached.

Where these intelligent software agents are unable to identify the provenance of documents (to help to establish their authenticity), or where the intellectual content does not match the repository's criteria, then details of these marginal cases would be sent to the repository. These cases would be subject to manual review undertaken by subject specialists. If the DDO is worthy of long-term preservation, then super-metadata is attached and the copy is preserved within the repository. If the DDO is living and subject to frequent change, the repository needs to set criteria of how often to capture this object. This instruction could be placed in the super-metadata to enable the repository to remotely monitor the DDO, in order to record any changes made to it whilst it is still active, (see Section 7.3.4).

It is desirable to use computer-assisted mechanisms to undertake routine tasks such as; searching for potential resources, monitoring changes and disseminating information, thus reducing costs (see Section 8.2.2). The use of computer-assisted components is replicated throughout the framework. It is suggested that as the proposed framework continues to evolve further computer-assisted developments could be assimilated, not only in the resource discovery phases, to cater for the amount of digital material repositories will have to deal with. Once the intelligent software agent selects the DDO it automatically submits it to the repository and it is digitally authenticated by time-stamping (see Section 7.3.6.1). Therefore, the super-metadata framework envisages that repositories will use intelligent software agents to discover and evaluate DDOs for their functional value to society. Figure 7.3 (Subroutine 1) shows the generic criteria for this selection process to which individual repositories would attach subject specific criteria. The effectiveness of these intelligent software agents in identifying, and selecting, potential DDOs is linked to the number and variety of known metadata formats known to the controlling repository. It is anticipated that

Figure 7.1 Proposed super-metadata framework

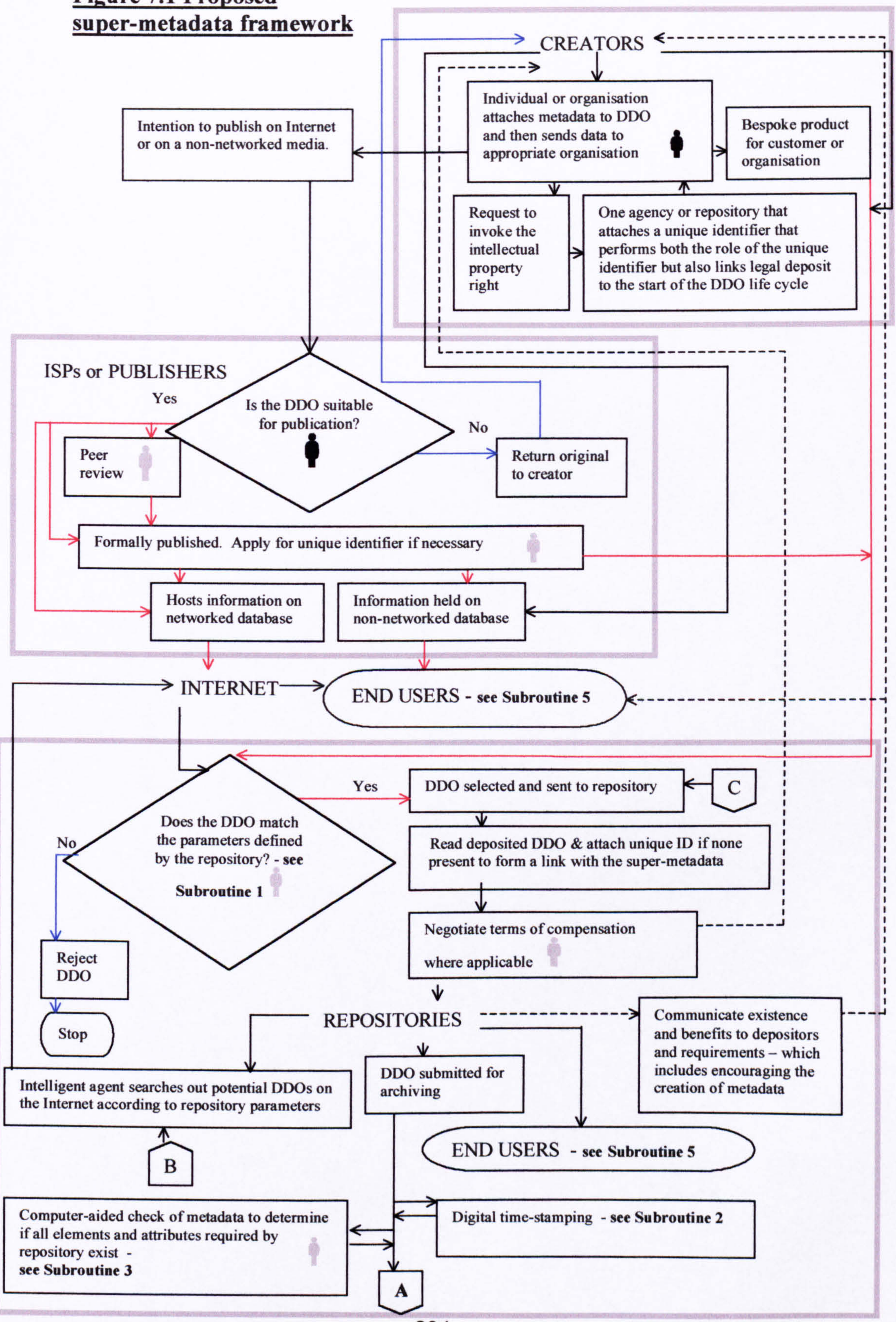
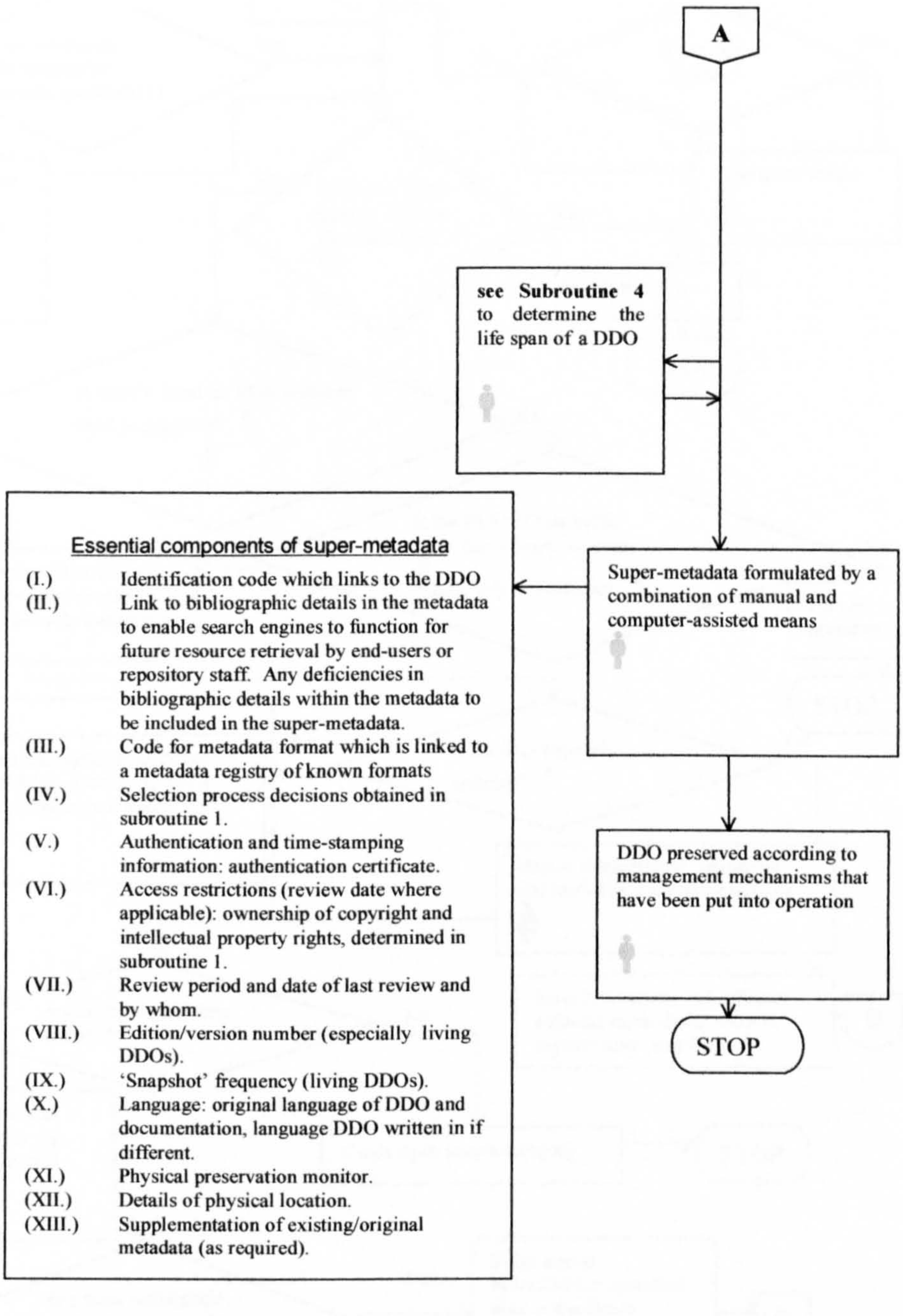
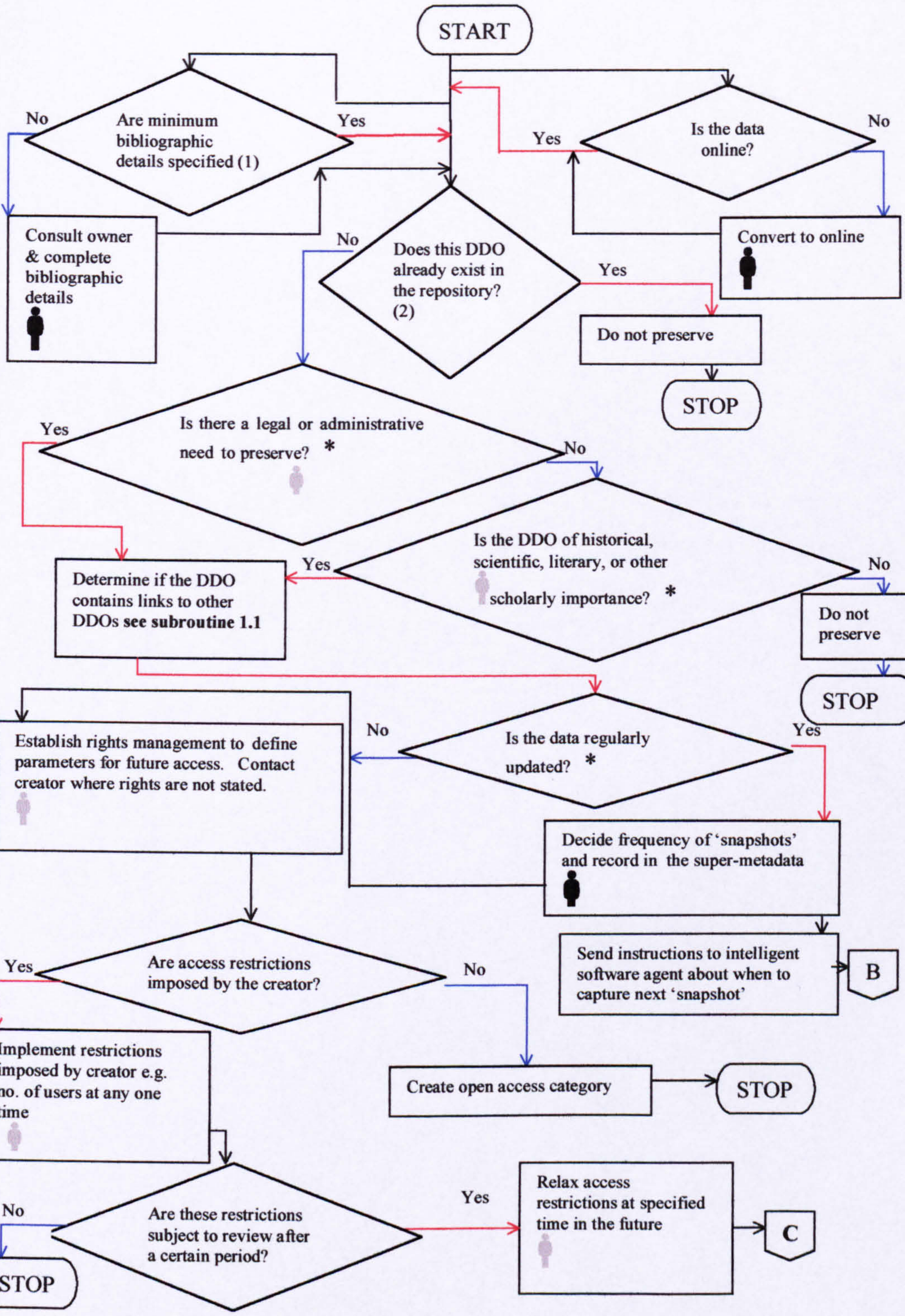


Figure 7.2 Construction of super-metadata



Note: super-metadata parameters may differ between the originator and the archive e.g. life span and access.

Figure 7.3 Subroutine 1: Selection and rights management



NOTES

All selection and access restriction decisions to be recorded in the super- metadata

(1) Specify author, title and creation date and unique identifier.

(2) Search engine interrogates super-metadata and metadata.

* Feather (1996).

Figure 7.4 Subroutine 1.1 Selection of DDOs with links to other DDOs

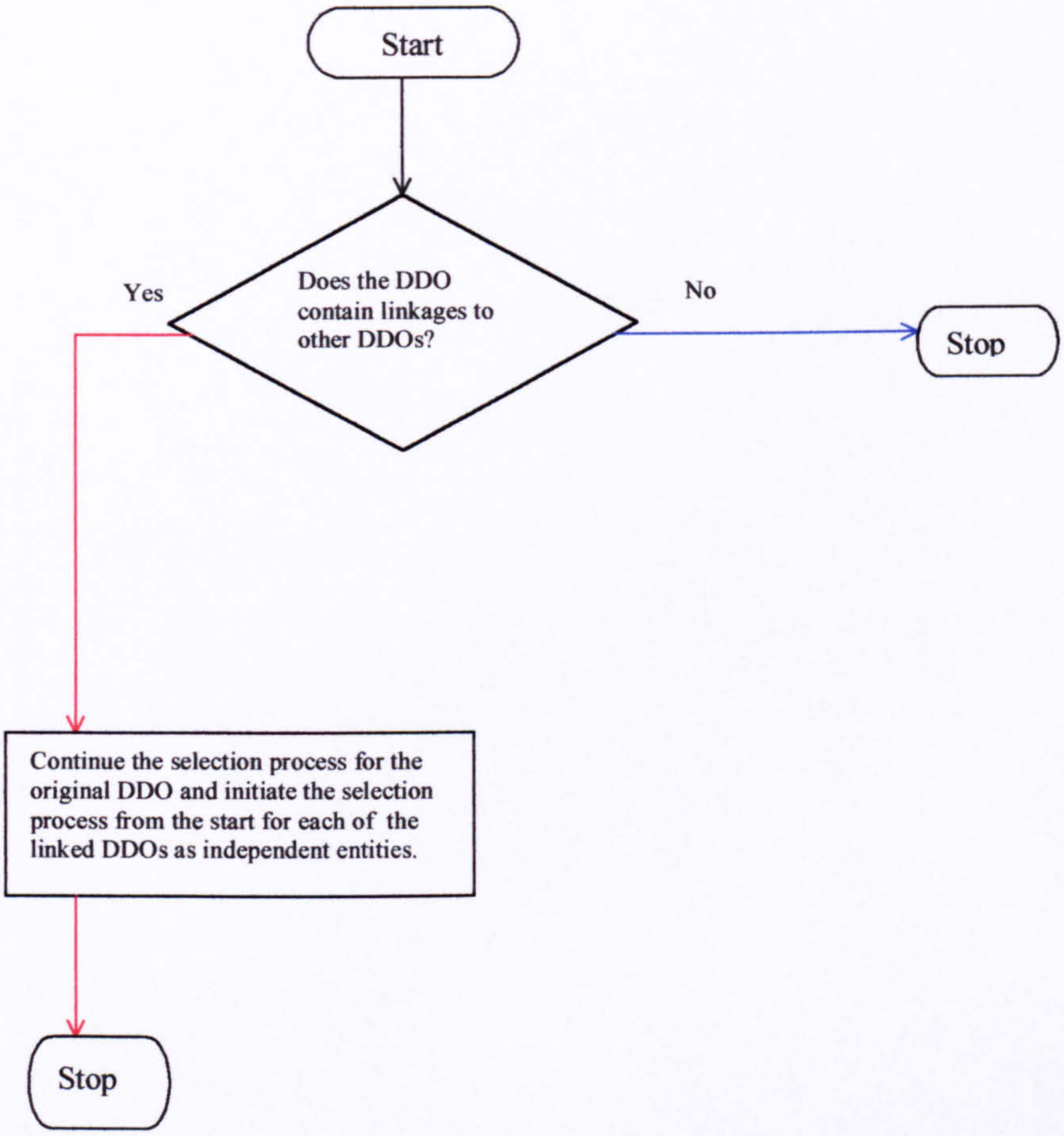


Figure 7.5 Subroutine 3: Computer-aided check of the metadata

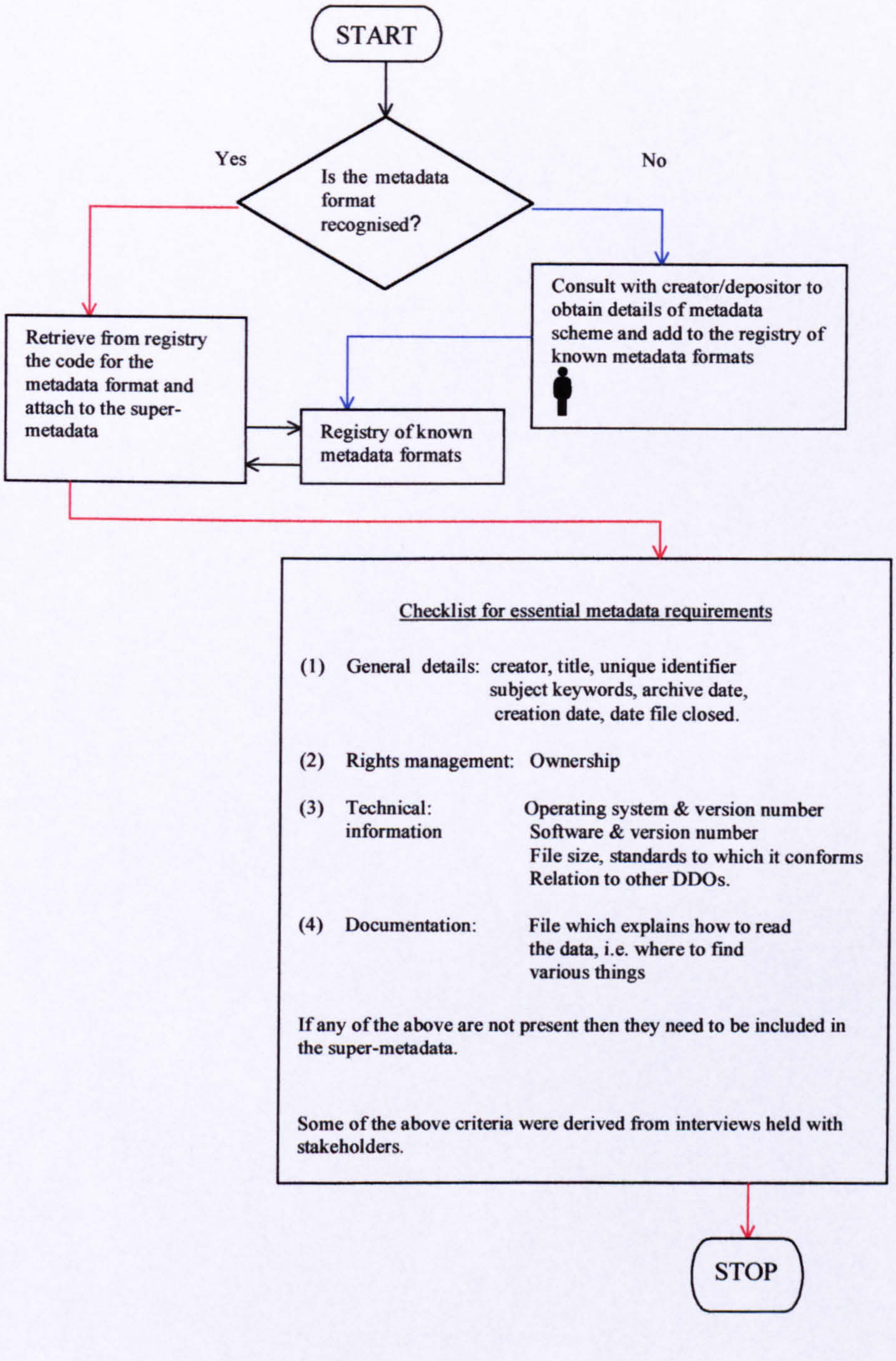


Figure 7.6 Subroutine 4: Determining the review periods of a DDO

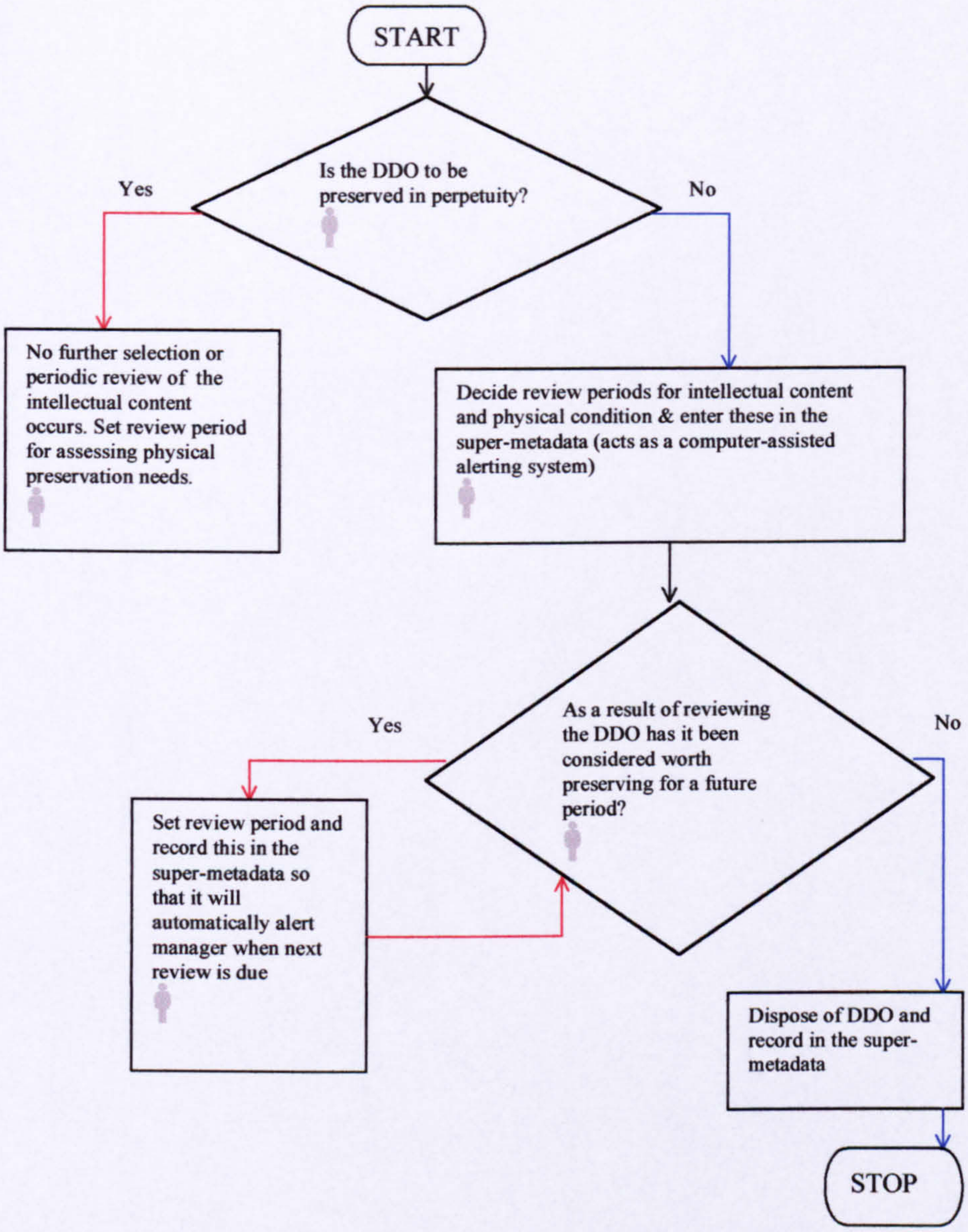
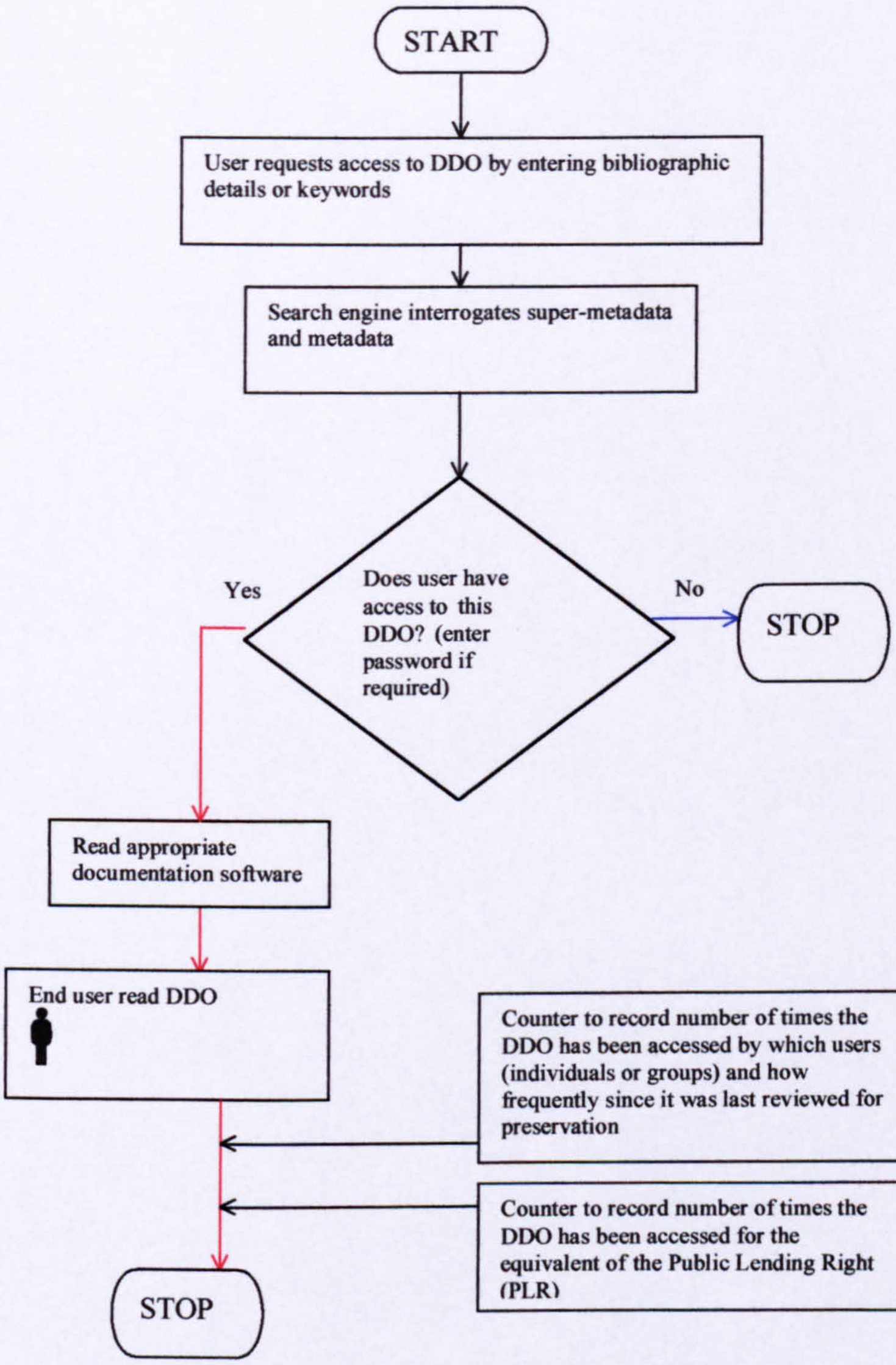


Figure 7.7 Subroutine 5: Processing an end-user request to access a DDO held in a digital data repository



these agents would also have regularly updated links to a central registry of metadata formats, which could be managed by an organisation which could either be self supporting, or, be supported by a consortia of organisations with an active concern for long-term preservation of DDOs. Each time a new format is encountered, it is vetted and if required added to the registry. New DDOs are, therefore, tested against a growing number of metadata format details stored in the registry.

Mirroring the contents to several sites could alleviate the vulnerability of a central metadata registry to unauthorised change. For this strategy to be effective, regular updating of these mirrors would be necessary. In a situation where large parts of the metadata for individual DDOs are identical, it would appear to be more efficient to store this information centrally, rather than duplicate the information by embedding it with each object. The greater the number of known formats within this registry, the greater the number of formats the intelligent software agents will be able to identify. As resource discovery is dependent on intelligent software agents recognising the format, there is the potential danger that DDOs with potential long-term importance may remain undiscovered. As DDOs in the Internet environment can be short lived (Koehler, 1999), there is a danger that there will be a time-lag between the appearance of new formats on public networks and software agents learning to identifying these new formats. This may be resolved by professionals, at individual repositories, being aware of the emergence of new metadata formats and the ability to re-program their intelligent software agents accordingly. Another drawback to using intelligent software agents, concerns selection criteria that the repositories develop; if the devised criteria are too specific, some DDOs that are of long-term importance may be overlooked during the selection process.

A variety of approaches are being adopted to monitor objects on the Internet, that although not intended to aid the long-term management of information, could possibly aid repositories. For example, one such service is the SearchEngineWatch site (1999) that provides a rough guide to the comprehensiveness and freshness of the information

provided by specific search engines[9]. This is necessary because the search engines' database of Web pages is composed of samples from sites, and once these source documents cease to exist the links are lost. Internet Archive is another initiative that aims to capture the entire Web as it is evolving. However, the exponential growth of the Web is highlighting the increasing difficulty of maintaining this approach, and in turn highlights the need to base a future framework on the measurement of DDOs against rigorous selection criteria.

7.3.3 Selection

The range of characteristics of DDOs and the challenges experienced by stakeholders, which are stated in the findings (Sections 3.3 and 4.4 to 4.12), shows the need to re-appraise existing selection methodology. Where possible the super-metadata framework encourages stakeholder organisations, wishing to deposit DDOs with repositories at a future date, to develop a records management system that will provide the necessary information for long-term preservation. Although this is possible to establish where repositories are dealing with defined organisations, such as government departments with statutory obligations to deposit records, the system breaks down with the present informality of the Internet environment. The development of evidential information aids long-term management, but to impose such a system on individual creators could be unworkable, as this stakeholder group may not see the benefit of undertaking this task, unless they identify the advantage of asserting ownership and conditions of access.

The processes of selection and rights management are integral parts of the super-metadata framework (detailed in Figure 7.3; Subroutine 1). All but the final stages of this subroutine form the generic structure on which manual and computer-assisted selection are based. The initial stages are concerned with ensuring data is converted to

---

[9] 'Comprehensiveness' is measured in terms of the number of pages gathered from each site. The higher the figure the more comprehensive the site. 'Freshness' refers to the maintenance of links between documents in the search engine directories' databases which is measured by assessing how frequently the data is refreshed and the sites are revisited (URL: http://www.searchenginewatch.com).

online format (this relates to manual selection only and it is anticipated that in the majority of cases data will already be online) and ensuring minimum bibliographic details are specified, the latter being to establish the provenance, content and context of the DDO. Where details are incomplete, the framework advises consultation with the owner. The DDO is then checked to ensure it has not already been deposited.

Once submitted for consideration, DDOs go through a selection process to assess the potential long-term value of its substantive content. To ensure that a 'representative sample' of resources is maintained for future end-users, it is essential that each DDO is measured against objective selection criteria. It is at this point that any metadata already constructed by creators may be useful in determining the importance of a DDO. These checks are followed by the selection of the DDO against the generic criteria, developed by Feather (1996) to assess the long-term importance of the information content (see Section 1.10). The generic selection issues are qualitative decisions concerned with 'societal importance'. This importance is divided into two main areas, legal or administrative; and historical, scientific, literary and scholarly. The development of selection criteria aims to provide generic guidance to aid creating organisations, publishers and repositories develop selection policies (identified as a 'grey area' in Figure 6.4 (4)). The sections which follow discuss each of these generic criteria.

7.3.3.1 The legal and administrative importance of DDOs

Legal importance applies to DDOs created and held to comply with legislation, whilst administrative importance applies to DDOs that have been developed within an organisation to aid its management. The legal and administrative value of a paper-based object to society is often enshrined in the existing legislation or formal procedures that have developed. In the digital environment, this is often not the case, because legislation and administrative practice is lagging behind technological developments. If on the other hand the DDO is not of legal or administrative

importance, then it is evaluated for its potential historical, scientific, literary or scholarly significance to society.

7.3.3.2 The historical, scientific, literary or scholarly importance of DDOs

The importance of an object in history, or science, or its merit as an example of literary style requires consideration when assessing long-term value. Determining the 'intellectual importance' within this framework is felt to be best achieved by ensuring resources are selected by experts in relevant fields. This increases the potential to select DDOs on an informed basis. However, the cost of achieving this is considered prohibitive without the extensive use of computer-assisted mechanisms (see Section 8.2.4).

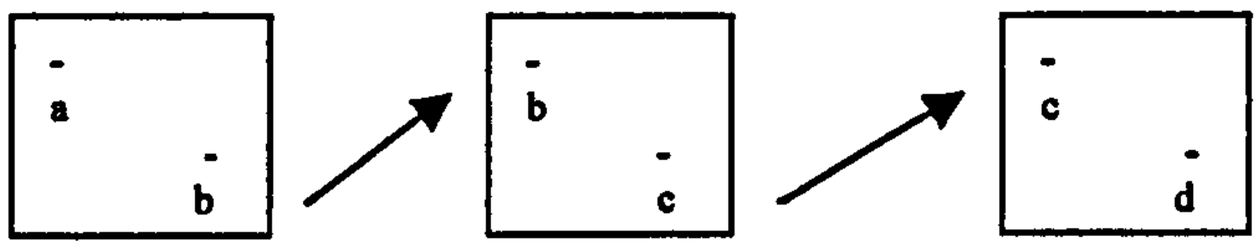
7.3.4 Additional metadata for living DDOs

With living DDOs the process of archiving involves removing the ability to change the information content permanently. Living DDOs then have the same characteristics as dynamic DDOs. This loss of interactivity is balanced by the ability to maintain authenticity in a repository setting. Where the DDO is part of an ongoing resource, the most recent versions of which are still living, it is important to ensure that the super-metadata supports this continuum (see Figures 7.3 and 7.10).

For the documentation and interpretation of living data to be meaningful, it needs to be in sympathy with the frequency and type of change that occurs. Where change is occurring at intervals, there is a need to take snapshots of the DDO to record these changes. Attached to each of these snapshots would be a version code indicating the position of each snapshot in the sequence. Figure 7.8 shows how the follow on code could operate to identify individual snapshots of a living DDO as separate entities, and also identify its place in a sequence of snapshots. The first snapshot in the sequence is given a machine-readable code at the start of the information content, to signify that it is the first in the series, and a second code placed at the end of the information content.

This end code appears at the start of the second snapshot (signified by the letter b in Figure 7.8). This sequence of codes continues until the last snapshot is archived. The last snapshot would have a code to signify this (marked by the letter d in Figure 7.8). Such a system of version control could aid future access where end-users require access to several versions of a DDO to examine trends over time.

Figure 7.8 Linkage of DDO snapshots using version codes



Legend: Bottom right hand corner of a given square shows its current version code and the top left hand corner gives the previous version code.

Where there is continual change occurring these snapshots may be insufficient and need to be used in conjunction with audit trails[10]. The potential frequency and extent of change a living DDO is expected to undergo is a criterion which needs to be considered in the selection process, not least because of the financial implications of managing these DDOs (see Chapter 8).

In the case of continuous monitoring, this would need to be negotiated with the creator(s). Some organisations which have adopted rigorous internal management mechanisms which comply with the super-metadata requirements, will be actively encouraged to continue this, and to supply repositories with details at previously agreed intervals. Where this does not occur it is anticipated that the repository would set up the above mechanisms in consultation with the creator(s).

[10] An audit trail is defined as, "ISO data, in the form of a logical path linking a sequence of events, used for tracing the transactions that have affected the contents of a record" (Nader, 1998, p.42).

7.3.5 Mechanism for establishing rights management of DDOs

Once the need to retain a DDO is established, it is important to establish its ownership, as this influences the nature of future access. It is essential to establish access rights at this stage as these influence a number of facets of the preservation process. These include restricting the way repositories administer access, who is restricted, over what period these restrictions apply, and how these restrictions are recorded in the super-metadata. The first stage in this process is to ask what restrictions the creator has imposed on the DDO. In some cases, the creator may not be the owner, for example where a creator is an employee of an organisation. These details may already be provided by the individual or depositing organisation, or be recorded in the metadata of DDOs where these are captured from a publicly accessible network. If the creator states that no restrictions are to be imposed then the repository can assign an 'open access' category to the DDO.

Where the metadata does not contain this information, repositories need to consult creators to establish these. In some cases, the access restrictions put in place at deposit are for a finite amount of time and can be reviewed when the initial restrictions lapse. In such cases the access restrictions may be relaxed and replaced with an open access category. The next phase is to establish what implications these restrictions will have on the way users can access DDOs, for example, where creators perceive a loss of revenue by allowing widespread access to a DDO. In such cases, owners may request that repositories restrict the number of users able to view a networked DDO at any one time, or specify its use on a stand-alone machine in a specific repository. Details of the negotiated or stated access restrictions, the length of time over which they are applicable, together with the their implications for systems architecture, are recorded in the super-metadata. The fields within the super-metadata governing access, link to a computer-assisted system using this information to control future end-user access.

7.3.6 Establishing and recording the authenticity of DDOs

Repositories would need to establish the authenticity of the metadata attached to DDOs by the creators or depositors. A computer program checks the metadata against the super-metadata registry of known formats. Mirroring it to other sites may offset the vulnerability of a centralised registry to unauthorised change. If a match is found the appropriate metadata format code is attached to the DDO. The computer program then matches the content of each field with the specified description of the element detailed in the metadata format. This ensure that the intellectual content and the metadata of the DDO is what it purports to be, a process known as authentication. Where these programs fail to prove conclusively the authenticity of the metadata, a manual check is made by the repository which may involve dialogue with the depositor to augment or amend the metadata. This process may entail negotiating access to privately held DDOs, together with a fee where applicable. To aid this process it would be advantageous to be able to determine when, and by whom, the metadata was created.

7.3.6.1 Digital time-stamping

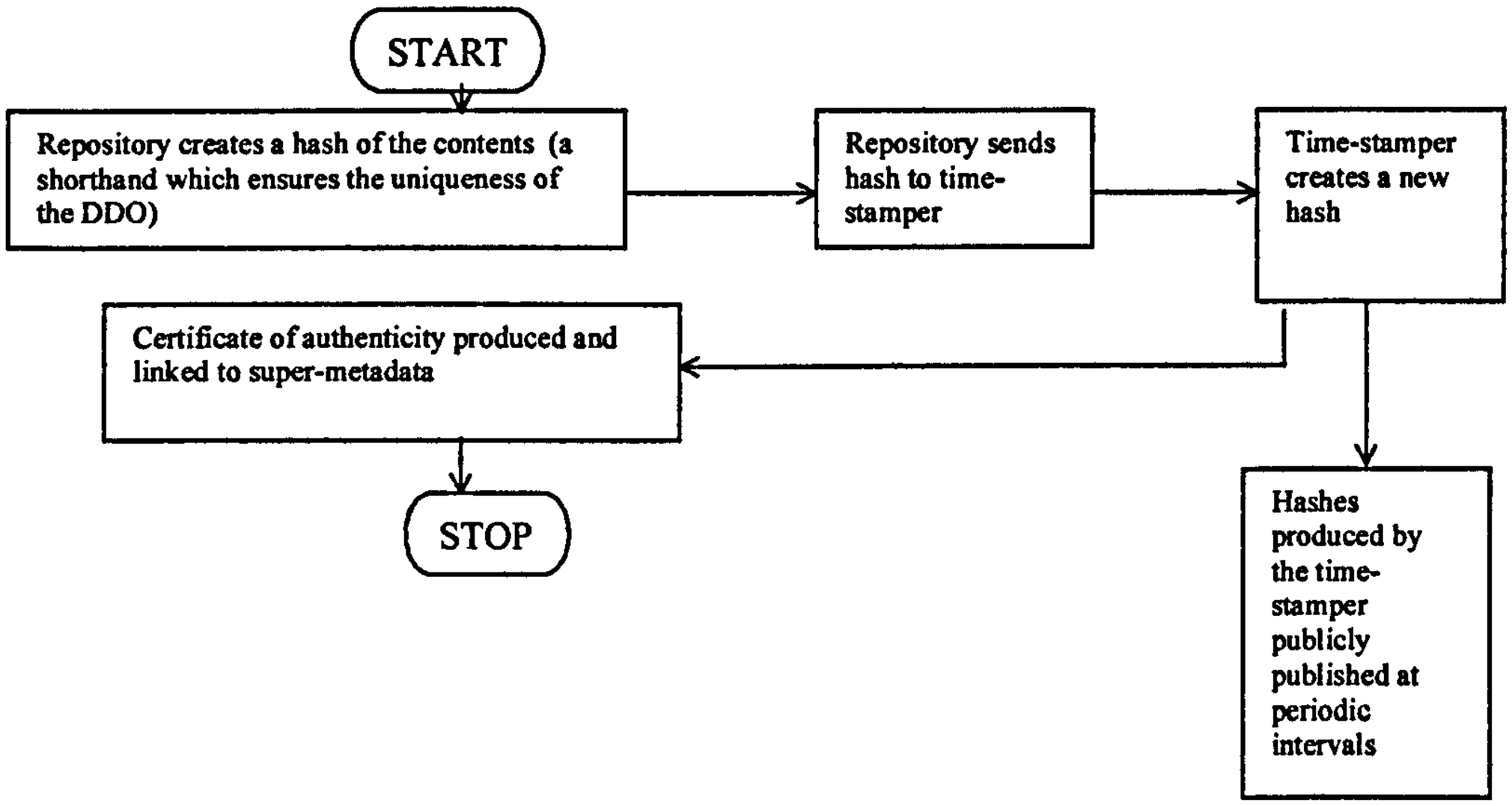
One process of securing the information content and uniqueness of DDOs can be achieved by time-stamping. Once a DDO has undergone this authentication process, it cannot be amended. Any amendments would be recorded as a separate DDO. This follows the philosophy of the Reference Model for Business Acceptable Communications (Bearman, 1997)[11].

[11] Workshop hand-out *Strategies for the management of organisational electronic records*: Workshop on Electronic Records, held at the Public Record Office, Kew, 12-13 September 1997.

Time-stamping[12] is seen as a potential mechanism to ensure that selected DDOs are authenticated. "It is typical for a document to be mundane at the time of its creation; it is only later that a document becomes important. Therefore an authentication mechanism is needed that is so cheap and easy that documents can be authenticated as a matter of routine" (Graham, 1993, p.90). One means of achieving authenticity is to time-stamp all DDOs which enter a repository to ensure that the metadata, the substantive data content, together with the date and time it first arrived in the system are recorded. Figure 7.9 (Subroutine 2) outlines the procedure for creating the time-stamp envisaged in this framework. Some organisations are going a stage further and exploring the potential for time-stamping to ensure that both the metadata and the intellectual content of DDOs can be used as legal evidence. As 'evidentiality' is not necessarily a requirement sought by all creators, universal adoption of such schemes may not occur. As DDOs selected for deposit may not all have been managed with evidential requirements in mind, the repository needs to ensure that each DDO is authentic as it is acquired. The super-metadata framework therefore accommodates both DDOs whose intellectual content have undergone stringent authentication and those which have not gone through this procedure.

[12] Time-stamping, developed by Haber and Stornetta (1991), relies on the technique of hashing. Hashing is the assignment of values to each portion of an electronic document. The results of computations give the 'hash totals' or hashes, which are more compact to send over a network than the original. This also maintains the privacy of the uniquely described object. To enable this process to work, computational methods which people have confidence must be developed, be in a useable form and be relatively ubiquitous. The whole content of the document is involved in this process and is entirely public (including the content of the document unless there is a wish for it to remain private). Each time a DDO is created or saved, a hash is generated; similarly, if it is published the hash also needs to be published. The hash comprises of two parts, the hash developed by the organisation authenticating the DDO (the public key) which is released into the public domain, for example published in a newspaper. The second part (a private key) is the code that is capable of breaking the electronic seal.

Figure 7.9 Subroutine 2: Digital time-stamping



Source: After Graham (1993)

7.3.7 Computer-aided check of the metadata

A computer-assisted program reads the DDO on submission to identify the metadata format. If the format is not recognised by the program, then the repository contacts the depositor to encourage them to deposit details of the metadata format (see Figure 7.5; Subroutine 3). These details go into a metadata registry that contains all the known formats, and an identification code is assigned to each format. Where the metadata format is identified by the program, but does not contain the essential metadata components given in Section 7.3.8, the repository requests additional information which is then added to the super-metadata. In cases where the metadata format is recognised by the framework, construction of the super-metadata template occurs by computer-assisted means. If the information contained in a DDO's metadata is complete, the super-metadata may consist solely of management information required to preserve the DDO long-term. Where insufficient information required at the super-metadata level has not been provided by the depositor, the repository needs

to engage in a consultation process to obtain it. Once manually selected DDOs have reached this stage, they are time-stamped. Ideally, records management systems should be developed within deposit organisations before any DDOs are created. However, the super-metadata framework is flexible enough to cope with DDOs regardless of how they have previously been managed and it advocates mechanisms to establish the authenticity of the metadata and the substantive data content at the time of deposit.

7.3.8 Essential components of metadata

(1) General information

The general information of the metadata is concerned with recording the duration of the life cycle and ownership information. The life span of the DDO may be defined in a number of ways, whether this is permanent; of a shorter defined span; or subject to periodic review. It is important to record who created the DDO; the date it was created; date it was archived and who owns it if this is different from the creator. This information is necessary to aid rights management of the DDO over time, by establishing who are the stakeholders for any one object and the time over which it is to be held. Similarly, it is important to know the date the DDO was deposited. This is particularly important for living DDOs as it indicates the date on which a particular snapshot was archived.

(2) Technical information

This information determines the status of the DDO, whether it is static, dynamic or living. The assigned status of the DDO determines the type and amount of documentation obtained to ensure the maintenance of functionality and access. To ensure technological preservation concerns are addressed, the following physical characteristics of the data and its information carrier are required:

- The operating system and version number
- The software and version number

- File size
- Standards to which it conforms
- Relationship to other files, for example within the same data set.

(3) Interpretational information

This area of the metadata explains the various component parts that are required to access the DDO. This involves indicating the links to the content, context and structure of the DDO, for example, the raw data file, documentation that explains how to read the data and files which describe the relationship to other DDOs.

(4) Search aids

These may take the form of keywords that aid future users to select useful DDOs. Construction of the contextual metadata, such as an abstract to a paper, is an important component of the DDO that the creator is felt to be best placed to write. It is also for this reason that these processes need to be an integral part of digital data management. This could include detailing the class numbers assigned to DDOs which may exist in the metadata.

(5) Value of the data

This could include the commercial value of a DDO to a particular stakeholder in the medium term, in addition to its long-term value to society. It is particularly important to determine the value of living DDOs, as the management of these resources can be costly over time, a factor that needs weighing against its intellectual value.

7.3.9 Responsibility for metadata formulation

The previous sections detail the elements of metadata required by repositories to preserve DDOs over the long-term. To achieve this level of detail it is anticipated that the creator will attach information relating to provenance, documentation and interpretational information (see Figure 7.10). The creators with their in-depth knowledge of the DDO are said to be best placed to attach contextual information to

the object. It would appear that there is a need to promote the benefits to creators of attaching metadata to establish their ownership of the DDO.

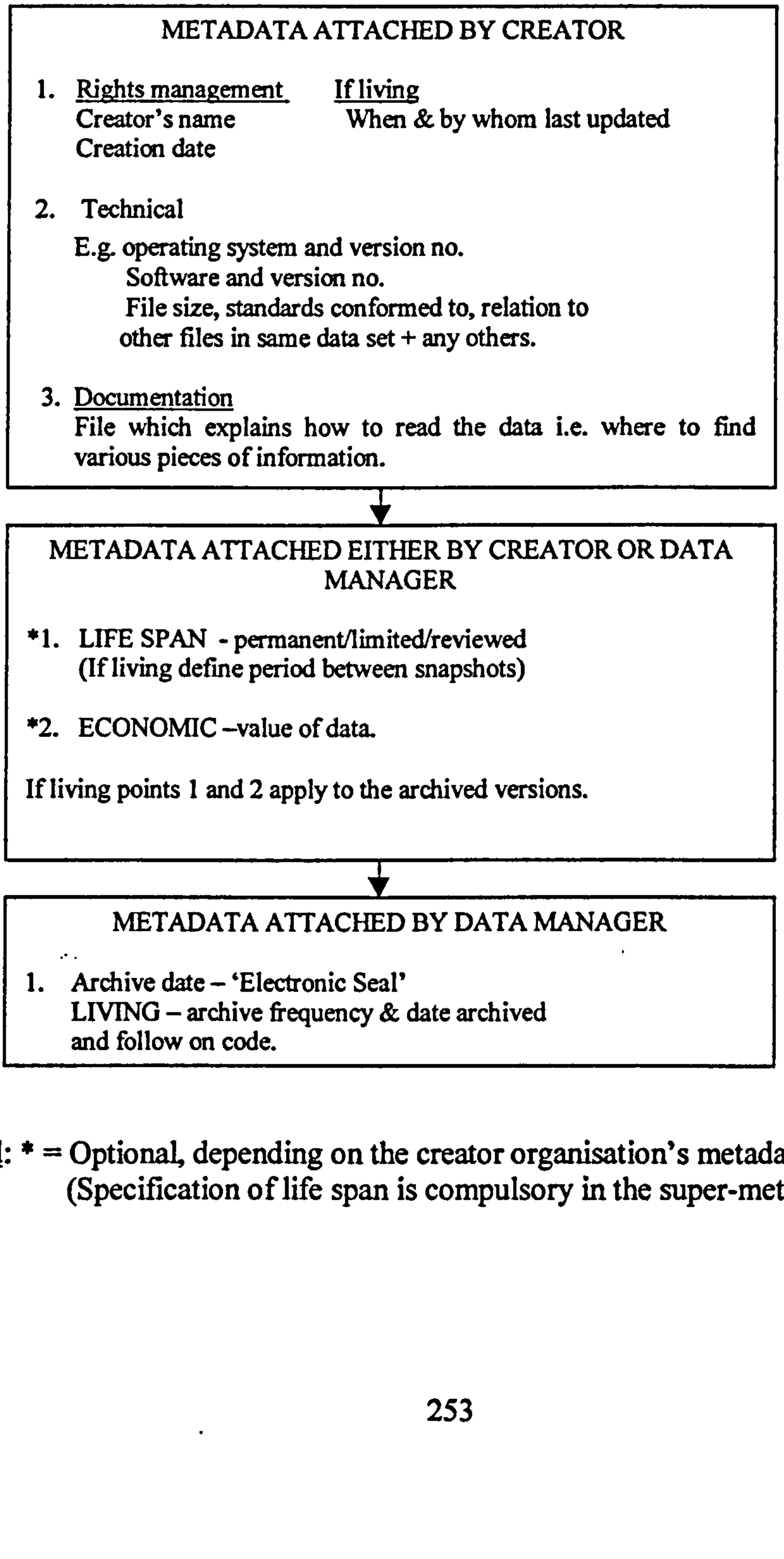
If the attachment of recognised metadata does not occur, then it is less likely that the discovery of DDOs for potential preservation will be successful. Placing the responsibility for the attachment of metadata on creators rather than other stakeholders, for example publishers and repositories who are concerned with maintaining and preserving the DDO later in its life, may maximise the amount and the detail of the contextual information which can be recorded early in the DDOs life cycle. Although as stated earlier, the creator's familiarity with the purpose and content of their DDOs makes them best placed to attach contextual metadata, this does not necessarily mean that they can accurately assess the DDOs long-term significance. Placing the responsibility on creators to attach metadata to DDOs could also be problematic, as creators are not necessarily a cohesive group that can be addressed. There is a need for repositories to disseminate information to creators that details the metadata needed for deposit. To facilitate the development of these linkages, it is important that the repositories produce information to inform potential depositors of the subject coverage, documentation and metadata requirements, and the location, (virtual or physical) of suitable repositories (see Figure 7.1). Some public and specialist archives in the United Kingdom are starting to addresses this issue by posting information for depositors on their Web sites. For example, the Visual Arts Data Service (VADS)[13] guidelines for depositors' page stipulates the Web page formats which this service provider can accommodate, although it will consider deposits in other formats.

The attachment, as a minimum, of a unique identifier, a statement of ownership, and the addition of contextual information by creators and publishers, would increase the efficiency of the super-metadata framework by reducing the need for repositories to re-establish the ownership of a DDO and negotiate access rights. If the creation of a DDO is occurring within a company, information on the life span and economic value

[13] URL: http://vads.ahds.ac.uk/depositing/depositor-pack.html

is provided either by the individual creator or by their data manager. Finally, the addition of some information needs to be added by the data manager, or archivist at a repository (where an organisation does not posses its own digital archive), this would include the date a DDO is received by the archive or repository (see Figure 7.10).

Figure 7. 10 Division of responsibility for attaching metadata to a DDO



Legend: * = Optional, depending on the creator organisation's metadata requirements. (Specification of life span is compulsory in the super-metadata).

If all or only part of the metadata which the creator supplies is not in electronic form the data manager needs to digitize it on receipt of the data. The creator may not choose to attach any metadata whilst the DDO is within his/her work area. However, when the DDO is submitted to the organisation's data manager, then metadata must be attached, if people other than the creator are to be able to read or interpret the data. If the organisation wishes to ensure the authenticity of the DDOs it creates, a system needs to be developed that retains DDOs in a non-amendable form (one possible approach is detailed in Section 1.8.2). The record manager, in the depositing organisation or repository, adds information that determines a DDOs life span and its economic value (if applicable), if this has not already been added to the metadata by the creator/depositor. The final stage is to check, and change if necessary, the metadata parameters to meet the requirements of the archive. In reality the division of responsibility between stakeholders may not be as clearly defined, especially where repositories are dealing mainly with individual creators rather than organisations.

If any elements are not included in the original metadata then they are attached to the DDO as part of the super-metadata. In effect a two-tier system operates where 'inadequacies' in the metadata are identified by experts in a particular field and are included in the higher level super-metadata, a process which is described in greater detail in the following section.

7.4 The components of super-metadata

Once the selection process described in Section 7.3.3 has established the long-term value of a DDO, then super-metadata is created for each digital object (see Figure 7.2). The essential components are listed below.

7.4.1 Unique DDO identification code

Each DDO requires a unique code to ensure that the data, metadata and super-metadata are synchronised. This is the same as the unique identifier attached to each DDO at creation. In the same way links exist between the data in a relational database. The same code needs to be embedded in the data and metadata (where this is in a physically separate database) and the super-metadata to ensure the maintenance of links.

Creators would be encouraged to apply for a unique identifier to attach to DDOs that they intend to publish. The generation and allocation of unique identifiers could be administered by an agency, as in the case of the ISBN in the United Kingdom, or a consortia of public bodies, who may or may not also be responsible for overseeing the development of the super-metadata framework.

7.4.2 Metadata code

This identifies the metadata format which is stored in the super-metadata registry. If the metadata format has not been listed before in the super-metadata registry of metadata formats then full details of the format, are requested from the creator for storage in this registry, and a new code generated.

7.4.3 Time-stamping and establishing evidentiality

This consists of the authentication date, that is when the DDOs metadata was time-stamped, a reference to the location of the public key and the location of the authentication certificate. The mechanism for generating this certificate is described in Section 7.3.6.1).

7.4.4 Access restrictions

This details the restrictions imposed by the creator. The protection of creator's rights involve repositories building restrictions into the super-metadata, which may include the following:

1. Placing restrictions on the amount of material that is viewable.
2. Restricting access to certain groups of users.
3. Stipulating the length of time over which restrictions apply. For example setting an embargo to expire on a predetermined date.

These restrictions cater for three stakeholder groups namely, owners, users and repositories. This framework envisages that owners (creators or publishers) have a controlling stake in the management of access to their DDOs.

7.4.5 Physical and intellectual content review periods

Having selected the DDO for long-term deposit, it is necessary to set a review period for the future assessment of whether a particular DDO continues to have long-term intellectual value. The stated review date could be linked to a computer-assisted system which brings DDOs, requiring review, to the attention of managers or archivists when it reaches this date. At this point, the repository re-evaluates the DDO to establish if it still has long-term value. This review period is not necessarily set for the same duration each time a DDO is re-evaluated. Where DDOs are initially selected to be preserved in perpetuity, they are exempt from periodic re-evaluation of their intellectual content.

In addition, all DDOs require assessing periodically to determine the need for physical preservation, such as migrating data to another software platform, or checking that it is still possible to emulate a DDO. It is possible that this physical preservation cycle

could be of a shorter duration than the cycle assessing the intellectual content of DDOs.

7.4.6 Edition/version number

Where the preserved DDO forms part of a continuum of snapshots, it is necessary to attach a follow on code to all but the final snapshot, to ensure that the position of each in the sequence is recorded. These snapshots may be augmented by audit trails, depending on the nature of the change (see Section 7.3.4).

7.4.7 Language

The language in which the DDO and its documentation were written needs to be recorded, to ensure that future users are guided to documents which are appropriate to their needs. It may be that the same DDO is subsequently published in other languages and a reference to the original language recorded in the super-metadata.

7.4.8 Details of physical location

Although the electronic network environment negates barriers to access produced by physical location, recording the geographic location of the master copy or copies is important. This ensures that an adequate number of copies are preserved in different physical locations, thereby ensuring the security of the data should disaster such as a fire occur at one site.

7.4.9 Supplementation of original metadata

Where the metadata, attached by the creator/depositing organisation, is regarded as deficient in some respect by experts at the repository, then these areas need to be add to at the super-metadata level. These include the essential components detailed in Section 7.3.8.

7.5 Data retrieval mechanisms

As end-user access over the long-term is the primary reason for establishing the super-metadata framework, information retrieval mechanisms are an important component. Figure 7.7 details the subroutine for processing an end-user request to access a DDO. The process is initiated by the end-user entering bibliographic details or keyword(s) into a repository's online search form.

To aid future users to search for DDOs, certain contextual data is included in the metadata. This may include creators of DDOs inputting keywords. Some DDOs may not have any contextual data within their metadata, and this would make the DDO difficult to find by way of keyword searches. It is therefore desirable to include within the super-metadata framework the ability to add contextual data, where the metadata is lacking in this respect. It is anticipated that classification of the DDOs held in a repository could aid browsing and retrieval, as it would appear to be aiding the management of Web based resources. The end-user request would be processed by a search engine that searches the repository's catalogue. A catalogue which would be formed by copying from the super-metadata and metadata created during the active and accessioning phases of the DDO's life cycle. If a DDO meets the user's requirements, he/she submits a request to the deposit organisation. At this stage, the super-metadata system can check whether the user's access code or password entitles him/her to view the requested material.

Assuming the user has the right to access the DDO, then the object is sent to the user as a 'read only' file, together with the appropriate documentation and software. The structure and original software format of the DDO is likely to influence the costs involved in delivering objects to end-users, and therefore, need to be considered when selecting objects for long-term preservation. The conversion to 'standard' formats needs to be considered when assessing access costs.

The framework aims to use counters, attached to the retrieval process, to provide information to aid the future management of DDOs. These counters could aid the review process by providing information on which groups of users access which DDOs, and the frequency with which this occurs. The access counter, documenting the number of times a DDO is accessed, could also be used for establishing compensation payments for depositors, where public access to DDOs affects the owner's commercial return.

7.6 The potential benefits to stakeholders of adopting the framework

7.6.1 Benefits to depositors

- Delegating responsibility to manage some of the DDOs they create to a repository can spread the cost of preserving DDOs.

- Ability to develop metadata formats now and in the future tailored to the individual or organisation needs, with the knowledge that integration into a digital preservation framework is possible at some point in the future.

- Ability to determine access restrictions for DDOs in order to maintain intellectual property rights. The degree of influence stakeholders have over the DDOs they own is seen to have a major impact on their confidence in third parties managing their data (see Section 4.8.1).

- Ability of depositors to access their own resources for future re-use.

- Allows preservation of DDOs which are recognised to have permanent importance to society and which have actual or potential re-use value for the creator and other end-users.

- Acknowledgement of the investment of the creator/owner producing a DDO that a repository subsequently identifies as having long-term importance, through the protection of intellectual property rights.

7.6.2 Benefits to the repositories

- Ability to systematically manage DDOs for long-term preservation within a repository.

- Ability to co-ordinate deposit strategies with other repositories.

- Opportunity to develop links with groups of potential depositors.

- Integration of preservation with selection, resource discovery and rights management.

- Stability gained from a system based on repository requirements rather than one controlled by commercial vendors.

- Addresses both human and technology centred issues.

7.6.3 Benefits to end-users

- Existence of a 'representative sample' of DDOs that might otherwise have been discarded.

- The knowledge that DDOs preserved using the framework have been evaluated by subject specialists.

- Existence of mechanisms that aid retrieval.

- Use of repositories as gateways to access collections.

7.6.4 Benefits to society

- The range of DDOs available to users from a single access point could be greatly increased, since there are no limits on metadata formats.

- Deposited DDOs could bring benefits to a wider community. For example the cross fertilization of ideas could result in works developed in one academic discipline giving rise at some point in the future to a new discipline or invention. Administrative benefits could, for example include increased effectiveness in solving crimes.

- Evaluation of data through the selection process.

- The long-term preservation of data will be assured and achieved in an as efficient a way as possible.

7.7 Summary

This chapter supports the view that by using higher level metadata it is possible to create a universal framework; that it is possible to include in this framework mechanisms to cater for a diversity of issues relating to the management of DDOs including authentication, selection and rights management. However, to test the conceptual feasibility of the framework and thus support hypotheses D and E (see Section 5.5) it is important to return to stakeholder organisations to seek their evaluation of the framework and cost model assumptions and to set the framework in a wider context. Chapter 8 therefore evaluates the framework from a number of perspectives, the anticipated costs, the critical evaluation of it by stakeholder

organisations, the management issues arising from its creation and the political issues at both national and international levels.

Chapter 8

Evaluation of the super-metadata framework

8.1 Introduction

The previous two chapters detail how the super-metadata framework has progressively evolved using Soft Systems Methodology. This chapter evaluates the framework by returning to selected stakeholders, setting it in a wider context in order to demonstrate the potential value of its adoption for digital data management and outlines the supporting environment that it requires. Sections 8.2 and 8.3 demonstrate the cost issues requiring consideration in developing such a framework and Section 8.4 critically appraises the stakeholders' evaluation of the recommendations of this research. Section 8.5 examines the impact of such a framework on society and Section 8.6 demonstrates the ways in which the framework advances the management of digital data. In the absence of an existing cost model that caters for the whole range of DDOs identified by this research (see Section 1.6.5), it was felt to be beneficial to generate such a model to conceptually test the super-metadata framework, and to provide guidance to aid stakeholders implementing such a framework at a future date. The lack of a framework supporting the preservation of a wide range of digital data has prevented the construction of basic measures to estimate costs. It is recognised that the cost model detailed in this chapter has limitations, which include: the inability to produce precise costs, determining the precise percentages of DDOs originating from specific sources and the exact division of costs between stakeholders. These factors can only be established when such a framework is implemented.

The analysis of the super-metadata framework has enabled the researcher to identify the contributing cost factors, but not necessarily who will incur these costs. In some cases, these cannot be determined within the scope of the present research but could present avenues for future research. For example, which stakeholder(s) will incur the

costs involved in developing and maintaining a system to assign unique identifiers to the DDOs at creation?

8.2 Cost factors requiring consideration in the proposed super-metadata framework

The development of a costing structure for the proposed super-metadata framework aims to aid repositories managing digital data to make informed decisions on preservation. An explanation of the assumptions and variables, and brief results are given in Appendix C. Within the parameters of this research, it is not possible to fully develop a costing model, but it is possible to factor in the essential elements required for such a model, and these are discussed in the following sub-sections.

8.2.1 Linkage of intellectual property rights with super-metadata

As demonstrated in Section 6.3.1, it is beneficial to link information concerning ownership with the information content of the DDO. To achieve such linkages it is important to develop the necessary infrastructure to support these processes. If as Chapter 7 suggests, unique identifiers are attached to the DDO which link ownership with the information content, then it is considered necessary to examine the cost implications. For example, if an independent agency is to be responsible for the creation of unique identifiers, its implementation will require new legislation. This legislation would appear to be necessary to provide ground rules for publishers and repositories on the functioning and maintenance of a unique identifier scheme and the provision of guidance for creators.

8.2.2 Resource discovery

The DDOs which the super-metadata framework aims to support are varied and a range of mechanisms to aid resource discovery may need to be employed by repositories (those felt to be most appropriate being determined by the individual

repository). Costs could include: staff time proactively seeking DDOs that are in the process of being created/refined, development and refinement of intelligent software agents, through to identifying objects covered by legal deposit that have not been received (subject to the development of new legislation). As many DDOs as possible would be identified by computer-assisted systems such as intelligent software agents. As the efficiency of these increases it is anticipated that manual costs would reduce proportionately. However, development costs will need to be considered.

8.2.3 Authenticating the intellectual content of DDOs and their metadata

The initial cost of developing an authentication system and the continued functioning of this process needs to be taken into consideration. This could raise issues of who will pay for the authentication, will external monitoring of metadata take place and if so, who will fund this monitoring?

This research suggests that the processes of ensuring metadata elements required by a repository for long-term preservation becomes as computer-assisted as possible. However, there may be DDOs without the requisite metadata attached at deposit and the cost of its addition needs to be taken into consideration at the selection stage. Routine costs can also be attributed to selection such as checking each DDO against a registry of known metadata formats and the addition of unknown metadata formats to such a registry (see Section 7.3.2.1).

8.2.4 Selection

At this point issues such as resource type (static, dynamic or living) and the economic value of a resource can determine whether the level of monitoring required for its effective management is justified. In addition, the distribution of selection costs can vary depending on the origin of a DDO, for example, the content of articles in a referred journal has been peer reviewed before it reaches a repository, whilst this is not necessarily the case with many Internet resources. Therefore, origin affects the amount of work a repository has to undertake once it acquires a DDO. Where stakeholders,

such as publishers, evaluate the intellectual content this spreads some of the selection costs between the publisher and the repository, however, whatever the origin of the DDO, the repository still has to stand the cost of assessing the object's long-term value. Content and the physical condition of the medium supporting the DDO could also affect how much time and resources are required for its management.

As can be seen, the cost of long-term management of digital data involves a number of issues that are not only important from a budgeting point of view, but also influence selection criteria. The case for selection being a vital component of a future framework still exists, as the repository is not the only stakeholder incurring costs. If the total cost to society of storing and accessing digital data are considered, it is apparent that creators, publishers, ISPs and end-users could all incur a range of costs (see the itemised lists of likely costs given in Appendix C, Section C.3.7). These costs show the maximum and minimum cost each group may incur.

Cost benefit calculations of the selection process need to take into consideration the financial implications of preserving digital resources measured against their value in monetary, scholarly or societal benefit terms.

If a scenario exists where no selection is undertaken and everything is kept, then stored DDOs could include a substantial number of objects which may not have long-term value, either because they have no intrinsic value in the first place, or are judged to have become out of date. Consequently, many users searching for data will incur the cost of weeding out many data objects which are irrelevant to their particular needs. When multiplied by a large number of users this could amount to a substantial cost to society, incurred by users. The cost of no selection, in terms of increased time expended by end-users trying to locate 'quality' information, can outweigh the purely financial argument (see Section 8.6.4). Once information has been retrieved by an end-user, the next challenge they face is assessing whether the DDO has specific intellectual merit, a task which can be aided by the 'value added' information attached by professional staff or intelligent software agents. If end-users are to be 'protected'

from spurious information, then the selection process is vital to maintaining 'quality' DDOs. Provenance becomes an important determinant in evaluating a DDOs intellectual value, therefore, the attachment of metadata is a vital component in promoting stakeholder confidence in using DDOs. If on the other hand, one uses simple selection procedures aided by intelligent software agents that select the majority of DDOs, this weeds out both duplicate DDOs and those which do not meet an organisations selection criteria. The more thorough the selection, the greater the potential benefit to the end-user in terms of successful discovery of material, as the DDOs that remain have been evaluated by experts in the particular field.

8.2.5 Access and rights management

Costs are associated with negotiating access to DDOs with creators and protecting their intellectual property rights need to be considered. Such mechanisms involve staff time and resources setting access conditions and ensuring these are applied. The implementation of these restrictions is a cost associated with end-user access to repositories and needs to be considered in the general maintenance costs (see Section 8.2.8).

8.2.6 Construction of the super-metadata

Although the framework determines the crucial elements required at the super-metadata level the construction costs of this could vary between DDOs depending on the amount and accuracy of the information supplied by the depositor. Costs could therefore vary considerably, however, the increasing development of computer-assisted systems may in time ameliorate some of these costs.

8.2.7 Review cycles

Costs are associated with the review cycles, (1) monitoring the physical media which support a DDO, and (2) those assessing the intellectual value of retaining a DDO for a further specified period. The latter is applicable where objects are not deemed to be

worthy of permanent retention at creation. Software dependency, the type of medium, and size of resource, could be significant in assessing the physical review cycle costs. Similarly, a number of factors could affect the costs of reviewing the intellectual content such as, the cost to society of not having access to the information, or the monetary value of a DDO (see Section 8.6.4).

8.2.8 User access costs

Access costs vary depending on the type of material the user wishes to consult, for example, access to dynamic DDOs will need to ensure that the necessary software is available to support the interactive elements of the data. Previously living DDOs may require more storage space on a repository's server, although the use of compression techniques, in cases where this does not lead to loss of functionality or presentation, may ameliorate this situation.

The development of online guidance for end-users and the periodic revision of this information need to be considered. This could involve the development of classification schemes to aid end-user browsing of repository holdings. It is expected that this would work in conjunction with the ability to use keyword searching. The choice of scheme would reflect the nature of the DDO, the subject coverage, and aims of the repository. In addition, the use of electronic mail for correspondence between the repository and its depositors and end-users could contribute to access costs.

8.2.9 Potential future users of a super-metadata framework

There are likely to be a range of end-user communities that may benefit from using super-metadata as a vehicle for digital data management. In general, any organisation whose data is not restricted for commercial or other reasons and which has a remit to retain DDOs long-term, may be interested in pooling its information in order to gain the benefits of a wider knowledge base. For example, institutions such as health authorities and higher education institutions, which for historical reasons may have

data in different formats, may be interested in using super-metadata to pool their information. The contents of super-metadata that end-user communities require will inevitably vary and be influenced by a number of factors such as stakeholder goals, nature of the digital data and resources at the stakeholders' disposal. The ability of the super-metadata framework to cater for a wide range of digital objects is essential as the diversity includes DDOs created for administrative[1], legal[2] and scholarly uses.

8.2.10 Staff re-skilling and training costs

It is anticipated that implementing the super-metadata framework will involve staffing costs associated with training and re-training to keep staff abreast of current developments. Additionally, senior managers will need to develop the skills necessary to plan and maintain a super-metadata framework and to develop staff training programmes appropriate to their organisation. This concurs with some of the stakeholder responses gained from the semi-structured interviews, which indicated a need to re-equip staff with the necessary additional skills digital preservation demands. These skill requirements cover a number of important areas, from selection procedures, intellectual property rights issues, to intelligent software agent programming skills. In addition, management skills need to be developed to resolve issues of responsibility in 'grey areas' between stakeholders.

8.2.11 Disposal costs

DDOs that are not to be preserved in perpetuity, could at some point in their life span, be re-evaluated and disposed of by a repository. These actions incur costs that have to be included in a future cost model. Building into a framework mechanisms to ensure the accountability of repositories to end-users would act as a driver ensuring disposal decisions are recorded. If disposal is to be contracted out to a third party, then the cost

---

[1] The Federal Statistical Office of Germany generates statistical information from censuses it undertakes (http://www.statistik-bund.de/erg_e.htm).

[2] Singapore has developed a national electronic court system with the following features, a filing system, information services, extracts and service of documents (http://www.lawsoc.org.sg/links.html).

of monitoring such disposal to ensure it is undertaken in the manner prescribed by the repository, also needs to be considered.

8.2.12 Findings from the costing model

From the general identification of costs described in the above sections, the analysis was taken a step further using realistic figures to gain an impression of the costs likely to be incurred using the super-metadata framework to preserve DDOs over the long-term (see Appendix C). It is not intended to explore the costs associated with every possible scenario, but to show that the framework could be managed in such a way as to keep the costs within reasonable limits. The production of a cost model enables key cost factors in the super-metadata framework to be identified and its sensitivity to various cost parameters explored.

It is assumed, for the purposes of this cost model, that the majority of costs are incurred by a repository whether in the public or private sectors. As repositories are anticipated to continue to be the main organisation or lead partner in long-term digital preservation this section focuses on analysing the costs from a repository perspective.

As well as itemising the costs associated with individual processes occurring at different stages of the framework, it is possible to determine which costs link to general inflation and which to hardware/software inflation. Unlike the model produced for Yale's Project Open Book (Task Force on Archiving of Digital Information, 1996, Appendix 2), a distinction is made between set up and annual costs. This is necessary to determine the proportion of costs attributed to selection and set up in year one and how many years it takes to reduce these costs to an annual cost of minimal amount.

For this model, the repository costs are divided into four main components; selection, set up, storage and access. All these costs are per DDO held by the repository. Hence, the access costs are factored by 0.2 when the access rate is 20%. Figure C.1 (see Appendix C) shows these costs over a ten year period as a bar chart. By plotting

the costs for the items it is possible to show that by applying the standard assumptions (see Table C.1), there is a steep decline in the cost of preserving a DDO, from £13.77 in year one, to £1.92 at the end of year ten (see Table C.10). Selection costs account for the largest proportion of costs over the ten year projection, decreasing as the number of years increases, due to the cost being spread over the number of years the DDOs is preserved. Set up costs are the second largest single cost a repository incurs and like selection, this cost decreases over the ten year period. The costs of selecting DDOs for long-term retention and attaching super-metadata, although quite costly initially, if kept for more than ten years becomes small in comparison with the on-going costs of storage and access.

This analysis suggests that where DDOs are of ephemeral interest (that is to be kept for less than ten years), minimal selection could be undertaken at deposit since they would not be kept long enough to justify the selection costs. In cases where these ephemeral resources are found to be of longer term importance at a later stage, more thorough selection criteria could be applied. This two-stage selection process would negate the need to commit the same amount of expenditure to all resources acquired by the repository. Balanced against the benefits of not selecting ephemera, due to its potential shorter term retention, has to be a consideration of the possible disadvantages to the end-user of retrieval problems (see Sections 8.2.4 and 8.6.4). If a policy of minimal selection is adopted then the end-user is faced with the problem of searching through a greater number of DDOs.

Unlike selection, costs associated with storage and access decrease due to software/hardware inflation (-13% per annum). The costing model shows that labour costs and computerisation are important factors in many parts of the framework. As labour rates are high and are inflation-linked, it is important to limit human intervention to the most critical stages and maximise computerisation where possible, the costs of which are anticipated to fall with time. The financial value of a DDO is not the only indicator of its long-term importance, rather it is one of many facets to take into consideration when selecting objects. Determining the financial value of

DDOs is seen to be a less complex process than judging its legal, administrative or scholarly value. It is possible that a DDO may be of cultural significance to society but not a high financial value. Therefore, it is also important to view the conceptual feasibility of the super-metadata framework from the societal benefits derived from managing the preservation of DDOs in order to achieve future access (see Sections 8.5.1 to 8.5.2).

Until the development and implementation of a conceptual framework occurs the costs cannot be predicted with accuracy. There is therefore, a need to develop costing models based on pilot tests of such conceptual frameworks, since those that exist do not explore the range of resources and the cost implications of a wide network of digital repositories. It may be possible to calculate the cost of static DDOs and multiply these by a factor to determine the costs of managing dynamic, or living DDOs, however, these extrapolations only give a rough indication of factors such as the greater degree of software dependency, the need for more detailed metadata and closer monitoring for migration purposes.

As part of this process, it is was necessary to create a cost model and to define the parameters to assess overall costs. This is inevitably a mechanistic approach, necessary to provide realistic costs. However, this inter-links where appropriate with the results of the conceptual modelling (see Sections 6.3 to 6.3.7). Such modelling helped the researcher focus on developing mechanisms that aimed to alleviate a range of stakeholder concerns.

Appendix D[3] provides users with the opportunity to vary the costing assumptions and to view the effect on cost. This uses macros to enable the user to easily view a ten year projection of costs and pie charts showing cost breakdowns, similar to Figures C.1 to C.9, displayed in Appendix C.

---

[3] The pocket attached to the inside of the back cover of this volume, contains one diskette holding an interactive spreadsheet (file name: smcosts.xls), written in Excel 97, together with user instructions (file name: readme.txt).

8.3 Relating the cost model findings to the results of the conceptual modelling

As Section 8.2.12 shows, the cost model provides an indication of the likely costs of operating the super-metadata framework. The major components of the framework can be assessed as a proportion of the overall costs and the effect of varying individual variables on the total costs assessed. Although it has been assumed for the purposes of the cost model that the repository assumes the majority of the preservation costs, the conceptual modelling revealed this to be a simplification of reality. Therefore, the cost model considers the costs from different stakeholder's perspectives, based on their existing roles, as evidenced from the semi-structured interview data (see the breakdown given in Appendix C.3.7). Although it is possible to establish cost scenarios for different stakeholder groups, the conceptual modelling demonstrates that roles and responsibilities in the digital data preservation environment are continuing to evolve. The cost model considers realistic costs that stakeholders may encounter, but only for situations where stakeholders' interpretation of their role coincides with the parameters detailed in the cost model. During the analysis of stakeholder costs it became apparent that a number of questions remained unanswered from the results of the cost model. Therefore, the researcher combined the findings from the conceptual modelling stage of the SSM, with the costs assigned to stakeholders in the cost model. The conceptual modelling demonstrates the difference in the way roles may be interpreted and the potential effect of this on the roles of other stakeholders (see Sections 6.3 to 6.3.7). The allowance for variations in the way stakeholder roles are interpreted is translated into the design of the cost model. However, the majority of the costs are ultimately either directly or indirectly assigned to the repository (the problem owner).

The SSM aided the development of future modelling and reinforced the need for flexibility as new stakeholder roles emerge, existing roles merge, and new communication channels, technology and management structures change. However, the fundamental functions of access and preservation remain unchanged, but the nature of roles, who performs them and the costs incurred by stakeholders, are subject

to change. In summary, this research supports the view that flexibility is an essential strength of a successful framework, required for its survival in an environment where many facets are in a state of flux. Returning to stakeholders for their evaluation of the super-metadata framework and the assumptions on which the cost model is based supported the case for the conceptual feasibility of the framework, as the following sections demonstrate.

8.4 Evaluation of the recommendations and assumptions by selected stakeholder organisations

In order to test the conceptual feasibility of the proposed super-metadata framework the researcher returned to three stakeholder organisations to interview the same people interviewed initially, in accordance with the methodology described in Section 2.5.6.

8.4.1 Evaluation of the rich picture and the root definition

The stakeholder organisations' comments suggested that the rich picture was an accurate reflection of their position at the time the interviews were conducted and that the category to which their organisation had been assigned was also an accurate reflection of their role.

As stakeholders adopt new roles which have been traditionally associated with other stakeholders in the paper-based environment, the categories detailed in the rich picture become blurred. A national library commented that although the major part of their role as the holder of a national deposit collection was contained within category B, they could also be classed in category A, a commercial organisation, with respect to their role as licensees of electronically published material.

> ...as both a local or remote document supplier in which we obtain rights to use either local or remote electronic material. In those cases we do compensate because we are charging the end-user and a proportion of that

> money is then passed back to the publisher, but that is us acting as an agent of dissemination rather than our role as the national deposit collection. (A national library)

This blurring of roles is also occurring between creators and publishers as noted in Section 8.4.2.

Minor points were raised by stakeholder organisations on the applicability of certain characteristics detailed in the rich picture under the category applicable to their organisation. In the opinion of a national library the category B heading needs the addition of the word 'receiving' to reflect more accurately their activities. Similarly the specialist archive suggested that the word 'retain' in the category D heading implies that they already own the material, but did not suggest a substitute term. To take account of this comment perhaps the word 'retaining' could be replaced by the word 'manage'.

The root definition (see Section 4.14.1) was viewed by the journal publisher and the national library as an accurate reflection of the problem situation from their perspective, with the latter commenting that in their view it was thorough and succinct with no obvious omissions. The specialist archive suggested that hardware changes need to be specifically mentioned in the definition, however the phrase 'lack of and instability of standards' aims to include hardware issues.

8.4.2 Extent to which the proposed super-metadata framework meets the needs of the selected stakeholder organisations

All three stakeholders stated that from their organisation's perspective the framework was conceptually feasible, although the specialist archive had reservations as there was no specific mention of the migration process or updating of information.

The degree to which stakeholders saw the proposed framework meeting their organisation's needs varied. The journal publisher and the national library indicated

that generally the proposed framework caters for their needs. The national library also stressed the importance of an independently developed framework.

> ...by the nature of where we have got to we are having to develop our own business model of this activity anyway, but I guess what is encouraging is that someone working independently has come up with something that verifies the way we think we should go... It is useful to have some external frame of reference which people can map onto and as a verification that this is the best way of approaching the situation. (A national library)

Similarly the specialist archive saw the framework as a useful aid to demonstrating how the interests of a variety of stakeholders might be accommodated.

> To have a conceptual framework that you can demonstrate, that this is how it can work, this is where it protects your interests, these are the steps that can be taken both from a data producer's point of view and from the repositories' point of view is very worthwhile.
>
> I think one of the great values of this [the framework] is that you have approached it from the perspective of lots of different types of data producers rather than just social scientists or just scientists... (A specialist archive)

Stakeholders do however express some concerns which focus on specific issues pertaining to their organisation. For example a journal publisher raised the issue of the time frame within which a central repository would require the receipt files from publishers. The electronic environment is shortening the time from acceptance of corrected manuscripts to publication. For a journal publisher, if a repository were to expect deposit of electronic files prior to publication, or within a very short time period after publication, this would present difficulties. The specialist archive agreed with the fundamental principles on which the super-metadata framework is based, but suggested that in order to meet their requirements a mechanism for returning to the creator at the point where resources are selected by the repository would be necessary for the reasons given below.

> ...there is no allowance for working on the resource to improve it or to make it suitable for users. In a traditional archive that is the case. We would not

> change it. We would not do anything to it. So the model that you propose is going to work for a great number of archives that work on this basis... we are not an archive in that traditional sense, we are more a data centre and data service in that we will work to transform some of the material that we get so that the DDO will change shape, will be reformatted, provided in a more user friendly format ... (A specialist archive)

The suggested adaptation was to insert an iterative loop between the creator and the repository in Figure 7.1, which enables the repository to consider accepting DDOs which require further work to ensure that they are suitable for re-use, rather than rejecting them without consideration at an early stage in the selection process. This has implications for the cost model as indicated in Section 8.4.3.

Similarly, when stakeholders were asked if the framework was conceptually desirable they responded positively, indicating that the framework was desirable from their organisation's perspective. This was supported by comments that it concurs in some cases with their business models or is not in conflict with their organisation's aims.

In addition, a national library stated that a proposed framework would need to recognise the blurring of creator and publisher roles, which they observe through their experience of operating the voluntary legal deposit system. This concern is addressed in the super-metadata framework by allowing for DDOs created by individuals to be sent direct to the repository rather than utilising the services of a publisher, or publishing it on the World Wide Web through an Internet Service Provider.

Stakeholders received general agreement with the recommendations to aid the implementation of the proposed framework (see Appendix E, background information Section 1). The specialist archive underlined the importance of developing staff training and re-training programmes although potential problems were noted, such as having sufficient funding to take the time out to train staff and then to implement the changes. However, if the benefits of such training/re-training can be demonstrated it may be possible to gain the necessary funding. The discussion of these recommendations for implementation also raised issues relating to the scope of the

framework with respect to; (a) its geographic coverage, that is whether it is applicable solely in the United Kingdom or has wider applicability, and (b) the framework that the super-metadata framework would operate within.

8.4.3 Costs and standard assumptions considered to be realistic by selected stakeholders

Prior to the interviews the three stakeholders were sent copies of the costs and assumptions on which the framework is based (see Appendix E). At the interviews stakeholders raised a number of issues which they felt ought to be considered when calculating costs and formulating the assumptions. A journal publisher felt that the overall costs were reasonable, but that costs specifically attributed to peer review, stated in Table C.6, were thought to be double the figure given. As far as costs involved in complying with legal deposit for DDOs, the journal publisher considered the costs involved to be negligible apart from the overheads to cover sending electronic files in batches to a central repository. As a result of further discussion with this stakeholder, it was identified that costs would be incurred by both publishers and repositories to establish the parameters for the management of DDOs which include future access as the following quotation suggests.

> So if we win a journal tomorrow the first time we start publishing it there could be some sort of set up cost with a central repository, even if it is purely advising them that it has changed from publisher A to publisher B. If it was a new journal then there is obviously a greater set up cost of things like the publishing management, the digital rights management information and so on, and I suppose there is probably an even greater cost the first time you would get the first journal from us because we would set up a number of rules that would apply to every journal. I think that is the major set up cost, there is then a minor set up cost per journal and then from thereafter it is negligible and I wouldn't even say it is worth costing. (A journal publisher)

The national library was concerned about the cost of establishing such a framework and if it had been considered in the calculations. These costs were seen by this stakeholder as their responsibility for the following reason.

> ...the likelihood is that most organisations will not have a substantial infrastructure to manage DDOs in place at the moment. If we don't the likelihood is that most other places don't and clearly that is an initial cost of quite substantial proportions. (A national library)

The set up costs may add to the overall costs but by assuming that the repository would be a large scale operation acquiring 200,000 DDOs per annum, and also by spreading the costs over a minimum of ten years, these costs are minimised. The need to increase the cost attributed to the initial set up would require verification by establishing pilot projects. If the capabilities of the super-metadata are increased, that is the functions it is expected to perform increase, the initial set up costs could increase. In addition, the national library was concerned that the initial set up costs of hardware, software and technical infrastructure required to support the proposed framework did not appear to be included in the costs. If the Yale Project Open Book figures for the access and storage costs are examined, a portion of these costs provide for the initial acquisition of the hardware and software. Also included in the cost model is allowance for the replacement of equipment and a time over which hardware is written off and replaced, but not the resources necessary to design the system architecture.

The specialist archive saw limitations in assigning a cost to a DDO because this approach assumes there is such a thing as a 'standard' DDO.

> It is difficult to fit everybody within an equation. It would be pretty complicated if that would work because you need to include the type of data and the variety of things that happen to it.
>
> If you are dealing with complex statistical data then you have a whole range of issues that you need to deal with, if you are dealing with visual images then you have another set of issues that you might want to deal with and the costs, you may have legal issues that are more relevant in one sector than in another sector. (A specialist archive)

Firstly for them the costs were 'massively underestimated'. The cost of processing, creating metadata, creating user guides and long-term preservation for a data set is

approximately £1,200. Secondly, they viewed the cost breakdown as being too specific as differences in exchange rate, inflation and the passing of time influence costs preventing the cost model being extended into the future. Thirdly, processes are likely to be expensive during the interim period before computer-assisted elements of the framework become operational. The specialist archive therefore suggests that stating the wages that one is costing for a data set of a given size, at each stage in the framework, although these figures may change over the years, equivalent costs can be calculated. However, this approach would appear to exclude computer based costs.

Although it is appreciated that the amount of work the specialist archive undertakes varies and that the DDOs they handle are complex data sets, the super-metadata cost model is capable of being adapted to consider larger scale DDOs such as data sets. Therefore, this research suggests that there are benefits to be gained in calculating both human and computer based costs as these can be used as a baseline from which work on costs can progress to aid future planning and development and do not have to be applied rigidly.

8.4.4 Stakeholder evaluation of the inclusiveness of the costs

As far as the journal publisher was concerned the costs were seen to be inclusive, whilst a national library suggested that the costs of the physical activities associated with long-term preservation, such as writing emulation software or migrating a DDO from one software platform to another, are at present unknown. It would appear from evaluating the views of stakeholder organisations that evidence of likely costs are required from pilot projects such as those instigated by the *JISC/NSF International Digital Libraries Programme: emulation for preservation*[4]. It may be that future developments in technology may make the physical activity of preserving DDOs less expensive.

---

[4] Through the development of emulation tools the project aims to assess the cost/benefit of using emulation as a long-term preservation strategy and to develop a model for collection management, in addition to guidelines for other strategies such as migration (http://www.leeds.ac.uk/cedars/JISCNSF/proposal.htm).

8.4.5 Summary of stakeholder evaluation

The results of interviewing selected stakeholders to obtain their evaluation of the super-metadata framework and the assumptions on which the cost model is based, detailed in Sections 8.4 to 8.4.4 above, indicate that the framework generally meets stakeholder requirements. The in-depth discussions with the three selected stakeholders did not identify major unresolved challenges which the framework fails to address, rather it was viewed as a positive contribution and they provided constructive criticism of minor points which in their view could be included to improve the framework. The researcher is aware that the framework would need to be evaluated in the light of findings from any pilot project which may be set up. These pilot projects would appear to be essential to verify that the identified costs are inclusive and are realistic for the majority of stakeholder categories (see Figure 4.1).

On balance, the framework received a positive reception, but there are seen to be differences of opinion with regard to the cost model. However, as the framework is flexible the cost model can be adapted to accommodate some of these criticisms such as the inclusion of issues pertinent to particular data types, or the generation of less specific costs based on man hours.

This research advocates setting up pilot projects to assess the practical feasibility and desirability of developing the super-metadata framework on a large scale. The sections which follow concentrate on assessing the public good case for the super-metadata framework and the management benefits such a framework is likely to produce.

8.5 Political issues

8.5.1 Public good argument for the provision of digital information

The need to maintain the identity and history of human societies are cited as reasons for maintaining access to data (Higgs, 1998). The recognition of international co-

operation to preserve documentary heritage is evident from the existence of initiatives such as the UNESCO programme, *Memory of the world*, with its aims to preserve, democratize access, distribute and increase awareness of the significance of retaining such material (Abid, 1997). If one links this to society's increasing reliance on data solely in digital form, then it is becoming evident that traditional forms of paper based preservation are no longer adequate for the range of created objects. Evidence of this reliance has been expressed in the trend for government records in countries such as the United States of America and the United Kingdom to adopt the electronic version as the definitive record and to set up initiatives for access, such as, the Electronic Access Project (National Archives & Records Administration, 1999) and AD2001 (Public Record Office, 1999). Similarly, the academic research community and science-based industry are increasingly moving to solely digital working practices, because of the new opportunities such digital media provides. As identified in Chapter 1, the evanescent nature of digital media requires proactive strategies to prevent its demise. Combining the nature of DDOs with the reliance which society is placing on digital data, it is becoming imperative to manage such resources effectively in an attempt to prevent "...placing the chronicle of our entire period in jeopardy" (Rothenberg, 1995, p. 24). The composition of this digital cultural heritage is likely to consist of a range of data types and a rich diversity of intellectual content, if present trends in the digital environment are any indication of the future. The diversity of digital data refers to the range of data types (visual, audio, text)[5], the range of operating systems, together with a range of purposes for which the objects have been created. For example, data of historic interest can originate from a diversity of sources ranging from, the diary of a deceased famous person on a Personal Digital Assistant (PDA), to a register of contaminated land.

This section aims to identify the benefits society gains by maintaining access to digital information. Some of these benefits are identical to those for retaining print-based objects and can be supported by and translated into the digital environment, such as

[5] The European Visual Archive project (EVA) with partners that include the European Commission on Preservation and Access, aims to generate use of national and regional multimedia resources held by public archives for public and multimedia industry use (URL: http://www.eva-eu.org/proposal11.html).

the ability to make informed decisions for the future aided by recourse to existing digital objects. There is however, a counter argument to this, which suggests that many of the documents that exist today do so through happenstance. The need to record past events, or results for posterity also translates from the traditional to the digital environment.

Digital data however, does create new challenges for the preservation of cultural heritage, for example, at an operational level there is a need to ensure that any copy of a DDO made by an archive has the date when it was copied and the original creation date attached. This ensures the protection of the original from accidental or intentional change. Possibilities exist with digital data that do not exist with paper based data, such as the option to copy the data to a new DDO where it can be reworked. For example, an engineering student 100 years hence may be able to take a personal copy of the 1990s space shuttle design and use it for a design exercise to improve the characteristics. The digital environment is providing new opportunities to improve accessibility for the general public to physical objects considered to be of cultural value, through digitization projects such as the 24 hour museum[6] (Kennedy, 1999), funded by a grant from the government of the United Kingdom. This drive to embrace new technology, to provide access and minimise social exclusion has been highlighted by the museum sector through the publication of *A net full of jewels: new museums in the learning age* (Keene, 1999).

With the advent of the Internet the general population is gaining a growing awareness of the wide range of data available through the World Wide Web and is appreciating the benefits to be gained from this. Since this data is continually evolving and expanding it remains relatively up-to-date, however, this is at the expense of previous data being superseded. A public need may arise for preserving some of this information, since some of this data could be lost by being overwritten or replaced. Unlike traditional print-based material, versions of DDOs are characteristically fluid. Editions of traditional materials are for all intents and purposes complete at the point

[6] URL: http://www.24hourmuseum.org.uk

of public release, which is not necessarily the case with digital publications. There is therefore a need for a system that manages version control. Not all information retains value once its currency has been lost, and hence the need for a selection process for such a repository or repositories. As the costing model in Appendix C shows, the need for a highly selective retention policy can significantly impact on preservation costs.

The socio-economic value of having an effective framework for the long-term management of DDOs is becoming evident, through the generation of reports by Public Bodies emphasising the need to develop strategies. This involves addressing the social issues of achieving and maintaining access for end-users. Provision of material for the public good has been voiced in plans to develop the public library network in the United Kingdom to provide access to digital resources for the whole of society (Department for Culture, Media and Sport, 1997). As the 'Information Society' (Bangemann Report, 1994) evolves, knowledge management becomes an increasingly important commodity to commercial activity. This concern for the functioning of the economy is evidenced by the *Green paper on public sector information in the information society*, produced by the European Union and in the United States, the report of the President's Information Technology Advisory Committee (PITAC, 1998)[7] with research objectives that include measuring socio-economic impacts. Similarly, the European Union's Info2000 programme aims to encourage the awareness and promotion of the multimedia content industry[8]. In a modern advanced society education is extremely important and as an increasing amount of data is created for educational purposes, solely in digital form, it is important that means are found to provide long-term access to the full diversity of data (Joint Funding Councils' Libraries Review Group, 1993).

The Arts and Humanities Research Board is funding the Arts and Humanities Data Service for a further three year term, because it is seen as a benefit to the academic community. Such publicly funded ventures are considered by the researcher to be very

---

[7] URL: http://hpccpublic1.hpcc.gov/ac/report/exec_summary.html
[8] URL: http://www2.echo.lu/info2000

important, since a lack of publicly supported access to DDOs over the long-term could create marked inequalities between those with the income and skills to access DDOs and those without.

8.5.2 Public good argument for the legal deposit of digital materials

Legal deposit facilitates public provision to a society's recorded cultural heritage by providing the legal support, but at present this does not exist for digital objects in the United Kingdom. Currently, legal deposit libraries within this country hold DDOs that are donated through a voluntary deposit scheme, a situation which is unlikely to encourage comprehensive coverage to exist within these libraries. Where legislation exists, as in the specific case of government records in the United Kingdom (Macfarlane *et al.*, 1998), a more managed approach has been adopted, as evidenced by the production of guidelines to support the preservation process.

For legal deposit to function as an effective mechanism in providing public access to DDOs, the focus of attention needs to be on resolving 'grey areas' where responsibilities and provision are indistinct. Conceptual modelling using the SSM approach indicates a number of potential human centred 'grey areas' (see Chapter 6), many of which involve ownership, public provision of DDOs and funding.

If legal deposit legislation is extended to include digital materials, it will need to include an official selection process, with the legal deposit archives reserving the right to dispose of DDOs deemed to be unsuitable for long-term preservation. As the costing model (Appendix C) demonstrates, selection for the majority of DDOs could be undertaken by computer-assisted means and where this proves impossible manual selection could be used. To reduce the costs of manual selection, where this is undertaken for a high percentage of deposited DDOs, it may be necessary to introduce restrictions on the amount of time devoted to the selection of each DDO. The Kulturarw^3 Heritage Project (Arvidson and Lettenström, 1998) instigated by the

National Library of Sweden in 1996[9], aims to investigate the long-term preservation of published electronic documents. The project has adopted a strategy of collecting as many objects as possible as a first stage and then disposing of unwanted material at a later stage. According to (Arvidson and Lettenström, 1998) the Kulturarw[3] project has influenced a number of other projects, namely the Central American Web Archive (CAW3A) and the Nordic Web Archive (NWA).

In the digital environment, the absence of a physical object to deposit requires systems to verify that the correct items are deposited and that the actions taken throughout the DDOs life span are documented, in order to ensure long-term access. As identified earlier in this thesis (see Section 1.9) the sources of origin of digital data are varied and this needs to be taken into consideration when formulating legal deposit legislation. It has been stated that there are distinctions in the way publications are deposited, depending on whether they are '...complete or finished, live or dynamic...' (Ratcliffe, 1999, p. 90-91).

8.6 Management issues

Having examined the cost and political issues that surround the long-term preservation of DDOs, attention now focuses on how the super-metadata framework advances this aim. The development of the super-metadata framework takes the long-term preservation of DDOs forward in a number of areas.

8.6.1 Personnel costs

It is evident through the examination of costs that the total cost of preserving DDOs is sensitive to increases in both the labour rate and the amount of staff time associated with the super-metadata framework (see Appendix C, Table C.4). Management policies will therefore need to focus on minimising staff time. Using automation as

[9] URL: http://kulturarw3.kb.se/html/projectdescription.html

effectively as possible could aid the reduction of manual costs, as Appendix C demonstrates, by increasing the efficiency of intelligent software agents. Some stages of the framework are concerned with making evaluative decisions and may remain human centred activities, even if the capabilities of intelligent software agents increase to a point where complex decision-making between agents is possible. Confidence in delegating responsibility for decisions to such intermediaries, and establishing the accountability of stakeholders for their agent's actions, may prove to be inhibiting factors to advances in agents (Brenner, Zarnekow and Wittig, 1998). Management policies that actively seek ways to improve the operation of the framework's components, possibly through automation, will be required to keep staff costs to a level which will keep the costs of managing a DDO over the long-term within reasonable limits.

8.6.2 Management reporting structure

The monitoring and recording structures within the super-metadata framework record both the actions taken by repositories and the content of DDOs, information that may be abstracted using intelligent software agents and counters to provide management information. The framework could have reporting structures built into it, such as placing counters on decision processes. For example, how many DDOs have been deposited, sent for manual selection, or disposed? Such counters would provide some of the management information required by repositories to monitor the performance of the framework. The assessment of costs for each stage of the super-metadata framework is seen to aid management by giving an overall impression of the effectiveness of the process.

8.6.3 Development of links between stakeholders

Arising from the costing model are a number of 'grey areas' that raise questions such as, who is likely to incur particular costs. The division of some costs may not be equal between stakeholders, or divided in the same way for all resources. Legal deposit

provides an example of an area where the digital environment is challenging existing concepts. This includes the concept of when a DDO should be deposited, and whether it is essential for the object to physically reside in the repository, or whether it is sufficient to maintain an electronic link to its source of origin. In turn this raises issues such as: how does one determine when a DDO falls within the scope of a 'national collection' when it is part of a collaborative research project involving people in a number of countries? Some argue that there is a case for making the digital objects 'smart' and the archives that play a minimal role in the preservation process 'dumb' (Nelson *et al.*, 1999). If digital objects do not physically reside in a repository, it raises issues of how the processes of refreshing data and reviewing DDOs for further retention, where applicable, are to be co-ordinated.

The rich picture (see Figure 4.1) demonstrates that at the time the initial interviews with stakeholders were conducted many of the interconnections between stakeholders either did not exist, used existing channels formed in the traditional resource environment, or were informal or unregulated. The super-metadata framework aims to provide a starting point for the co-ordination of long-term access to digital data by providing a management structure for the preservation of DDOs. A vital part of this process is to engender co-operation between stakeholders who contribute to the creation and management of DDOs. The possibility of hybrid organisations that perform several roles, for example, publishers who also maintain a digital archive of their own publications, make it difficult to predict who is likely to incur what costs. Collaboration and consultation may occur between existing stakeholders, roles may continue to be fluid, and new stakeholders emerge. Repositories that are just starting to develop digital archive services, may be classed as new players in the current digital preservation environment.

8.6.4 The value of selection to society

As the ultimate aim of the super-metadata framework is maintaining long-term access to DDOs, part of determining its feasibility is finding a cost effective way to work

towards this aim. Balancing the costs of retaining all DDOs against the case for selection, is a necessary financial consideration. This gives an indication of the overall costs, whether it is realistically sustainable over time, and balanced against this are the societal benefits, namely end-user benefits. Even where the case for a selection process is not justified on cost grounds, there may be other critical reasons why retention is necessary. Adopting a strategy where DDOs are retained for long-term preservation without any selection taking place has a number of advantages on solely cost grounds. Firstly, there would be a saving in staff costs normally incurred by the manual selection of DDOs and secondly, the relatively low cost of storage allows large numbers of DDOs to be preserved. However, this would appear to be a minor consideration compared to the problems created by not providing some system of evaluation, through selection, which engenders end-user confidence in a system and aids retrieval (Burke, 1997). In addition zero selection has cost implications for the end-user in terms of increasing the time taken to execute searches and to evaluate the results. If costs for use of equipment to access DDOs are incurred by the end-user then the longer the search takes the greater the costs could become for the time spent on-line. The vital link between appraisal and selection of DDOs and the preservation process has been noted.

> The problem is not so much that it is impossible to retain all information created in electronic form as a by-product of contemporary activity, but that it is not feasible to document it suitably. Because of the limited value of undocumented digital information, appraisal and selection will become increasingly important in the process of preserving information. (Ross, 1998, p. 17)

The paragraphs that follow explore the perceived benefits of selection and the implications for society.

Various permutations of selection procedures are likely to arise depending on factors such as institutional policy. Such a situation may occur where it is essential to record important decisions taken by government in the form of authenticated digital objects, to ensure accountability to society. If the word 'value' is taken to include the tangible

benefits to different groups of stakeholders that together constitute the wider society, a number of issues emerge that present challenges to this approach.

Although a policy of no selection gives end-users the ability to select from a wider range of materials, there are a number of significant disadvantages to society in not undertaking any selection. As the number of preserved DDOs increases, the end-user will face increasing amounts of information to search through to find required DDOs. Similarly, as the number of preserved DDOs increases the time taken to access the results of searches could also be increased. It may be possible to increase the processing power of hardware and the use of intelligent software agents to filter some of this information. However, it may be that the task of evaluating the quality of objects may become more difficult for end-users as the number of stored DDOs increases.

Where a stringent selection policy is applied to all acquired DDOs then, as the cost model (see Appendix C) shows, it incurs high financial cost if a large number of these DDOs are subsequently retained for less than ten years. If repositories with aims that support stringent selection policies envisage offsetting storage costs, they would need to keep the majority of selected DDOs for longer than ten years.

An alternative to the above is to adopt a two stage selection process. As DDOs are deposited, minimal selection criteria are used to evaluate whether the material is suitable for retention, subject to review at some date in the future, or to be disposed of immediately. 'Minimal' selection in this context refers to the use of the generic criteria, cited in the super-metadata framework, to separate 'ephemera' from DDOs with long-term value. The expectation is that by filtering ephemera at an early stage, it will result in a decrease in the time, money and effort expended by end-users on filtering DDOs considered 'valueless' over the long-term. However, this brings into question the issues of what constitutes ephemera and how to determine if some has long-term value. To develop answers to these questions at the organisational, national and international levels, it is important to develop selection policies and management

structures that will support long-term access. This discussion demonstrates that selection is an important component of a future framework and the cost model qualifies the financial parameters selection may need to operate within. The second stage of this selection process is to enter all the DDOs that passed the 'minimal' selection into a more rigorous selection after a period of perhaps five years. This process reduces the number of DDOs that have to be subjected to the full selection process.

This supports the case for hypothesis E, that the criteria for the selection of DDOs for long-term access are a fundamental component of such a universal system.

8.6.5 Development of policies and standards

The importance to society of long-term access to DDOs requires the formulation of a co-ordinated approach to achieve this aim. It is suggested that such co-ordination involves a range of stakeholder groups to develop policies. There are a number of critical areas that need to be addressed if co-ordinated long-term management is to be achieved.

- It is evident from this research that there is a need to determine how the selection process would operate in practice. Issues such as who would be qualified to select DDOs, how many people would be required, which institutions would be responsible and who would fund such ventures need to be resolved.

- The development of legislation to support the long-term management of DDOs is seen as a priority, however in its absence there is a need for closer co-operation between stakeholders to develop voluntary agreements. For example, whilst the extension of legal deposit legislation to cover digital materials does not exist in the United Kingdom it is important for national institutions to strengthen links with publishers to encourage voluntary agreements.

- At the individual repository level, the research indicates that a broad range of staff with different strengths will be required to develop and operate the super-metadata framework. The number of staff required and the staffing structure are expected to be determined at the individual repository level. However, it would appear that staff with the following skills/knowledge are key to the creation and maintenance of a DDO repository. Staff with a detailed knowledge of intellectual property rights are required at the early stages of set up, whilst others with technical expertise and detailed knowledge of metadata systems are required at a later stage.

In addition to requiring the combined skills of specialist personnel to set up and maintain the super-metadata framework, it is important that such personnel be part of a committee, or committees, concerned with national and international issues, such as, legal deposit and protection of intellectual property. It would seem that without international agreement these issues are difficult to resolve.

The global connectivity possible with electronic resources is encouraging the development of standards of international applicability, as evidenced by the strong tradition in the library community to unify holdings of geographically disparate libraries holdings, for example the development of UNIMARC and the Functional Requirements for Bibliographic Records (Electronic Records) (FRBR(ER)). Similarly, the development of the OAIS model for the functioning of archives. In keeping with this trend the super-metadata framework subscribes to the philosophy of a global network of repositories that create super-metadata entries nationally, but provide international end-user access to DDOs over time. The development of super-metadata attempts to cater for version control and integrate rights management and selection components into an holistic data management framework. A framework that, through the use of metadata, provides an approach to addressing the issues which information-dependent organisations face as they seek to preserve data in a changing technological environment.

8.7 Summary

This chapter has evaluated the super-metadata framework firstly by identifying its potential costs and secondly by returning to selected stakeholders for their evaluation of the framework and the cost assumptions on which it is based in order to demonstrate its conceptual feasibility. The conceptual feasibility is a balance between the defined financial constraints and the societal need for such a system. Similarly, the feasibility of selection is not seen to be justified on cost grounds alone, but when set against the needs of the wider society a strong case for the inclusion of selection emerges. Relating the findings of the conceptual and cost modelling provided support for hypothesis D, that a universal conceptual framework for the long-term access to DDOs at a level higher than metadata is possible. By setting the framework in a wider context it has been possible to show the political issues which it addresses and the challenges which at this stage remain unresolved. It has also been possible to demonstrate that the proposed super-metadata framework raises a number of management issues which require resolution.

Chapter 9

Conclusions and recommendations

9.1 The main findings

The continuation of the philosophy to maintain access to a society's 'memory' remains unchanged by our increasing reliance on digital data. However, there are fundamental differences in the nature of DDOs and traditional print-based objects that necessitate a shift in conceptual thinking about their management if long-term access is to be achieved. This research has developed a series of hypotheses with the aim of using metadata, as an approach to addressing the issues which information-dependent organisations face, as they seek to preserve data in a changing technological environment. The empirical research, supported by the literature in the field, points to the urgent need to re-evaluate current management practices to develop a consistent approach to the long-term preservation of digital data. The positive evaluation of the super-metadata framework by selected stakeholders suggests that this framework could be a possible way forward.

The world is changing increasingly to using data in electronic form, leading to the development of the information society across the globe. The need for universal access to data is continuing to grow and, with the advent of the Internet, users are coming to expect access to a diversity of subjects, making users aware of the potential for access to any publicly available resources. The provision of up-to-date information on the Internet can be viewed as a mixed blessing, since the ability to access, for example, the previous version of particular data may have been lost by the updating process. As people use the Internet as a delivery mechanism for receiving information and potentially become less dependent on paper-based resources, there becomes the need for long-term preservation that applies to paper to be extended to the digital

form. In this digital environment, managing digital data in a global systematic way is crucial to establishing how long DDOs are to be retained and under what circumstances.

9.1.1 Assessing current theory and 'best practice'

The range of existing approaches to the preservation of DDOs over the long-term requires a re-evaluation of existing 'best practice' and an examination of existing initiatives to determine their suitability to the challenges presented by the digital environment. The data obtained from interviews with stakeholders demonstrates that a range of approaches were being adopted and tailored to suit defined parameters, be this within an organisation or a specific community. Examination of the reasons for retaining DDOs, challenges identified by stakeholders and an examination of existing practices, do not appear to be central to the majority of current initiatives. The development of metadata formats has followed the same pattern, with many of the formats designed for electronic objects being focused on the resource discovery phase of a DDOs life cycle. Formats which have been specifically developed for long-term preservation tend to be linked to particular data models, or be concerned primarily with different aims such as establishing ownership for control of rights management rather than maintaining long-term end-user access. This research shows that current practices over a range of information-dependent organisations do not always meeting the challenges presented by DDOs. The development of a conceptual framework, that provides a philosophy for managing the DDO throughout the whole of its life and which can be tailored to suit a particular repository's requirements, was seen to be a necessary development.

As Chapter 1 attests, a number of professions and disciplines are active in the field of digital data management, each with different aims and approaches. This is emphasised by the range of existing frameworks and at a lexicographical level in the development of parallel terms or in some cases different meanings for the same term, for example the term archive. The findings prompted the formulation of a generic term (DDO) to

describe digital objects and the need to re-evaluate current practice. The nature of the research area demands the analysis of both societal issues and the nature of the objects considered for preservation. Examining the issues from the stakeholder perspective highlighted the difference between theory and practice and the range of challenges that a potential solution would need to address. The examination of current theory on long-term management and a re-assessment of their relevance to the management of digital data, has led to the need to develop a conceptual framework in sympathy with the developing digital preservation environment.

9.1.2 Development of a universal systematic management approach

The diverse nature of DDOs requires a framework that can manage all of these objects for long-term preservation. This research through stakeholder interviews and using Soft Systems Methodology has identified the challenging management issues that this framework must accommodate. As the researcher is independent of any stakeholder groups, an objective overview of current data management, informed by SSM, has been carried out and this is used as a basis for the construction of the super-metadata framework. The case for a universal framework for the management of DDOs is supported by the hypotheses of this research and considered achievable by the development of management metadata at super-metadata level (see Sections 5.6 to 5.11). The support for the hypotheses also shows that there is a need for a selection process in such a framework.

The current environment is characterised by frameworks focusing on specific types of resource. In contrast, the super-metadata framework aims to reverse this trend and cater for the wide range of digital data covered by the definition of a DDO (see Section 1.2.3). The overriding concern is long-term access to DDOs and a secondary consideration for the physical properties where these influence long-term preservation. The super-metadata framework therefore seeks to be inclusive and to concentrate on managing all aspects of the DDO during its life cycle. The component parts of the framework provide guidance on the management of the resource within the repository.

These parts include, determining the life cycle, the creation of the super-metadata and components that aim to aid the resolution of 'grey areas' identified by this research (see Sections 4.4 to 4.13). The development of a universal systematic management framework catering for a wide spectrum of DDOs encompassed by the definition of a DDO is vital for their long-term preservation and ensuring future access. The sub-sections that follow detail how the super-metadata framework takes the long-term management of digital data forward in a number of areas:

- Formulation of generic selection criteria to assess the long-term value of the intellectual content of DDOs.

- The management of current and potential metadata formats together with metadata formats in legacy systems.

- Mechanisms for authenticating DDOs.

- Consideration of rights management issues.

- Development of a model to estimate costs.

- Linkage of ownership to the management of DDOs.

- Resolution of stakeholder needs.

9.1.2.1 Re-evaluating the selection process

This research has examined the issues of selection and concludes that it is a vital element of a future framework. An examination of the nature of DDOs has established that new selection criteria need to be created to accommodate the characteristics of living and dynamic DDOs (see Sections 5.2 and 5.3). The

challenges that these DDOs present require new structures to enable their preservation to be managed in order to ensure future access.

Selection mechanisms currently employed cater for resources most closely matching paper-based objects, but require adaptations if they are to cater successfully for dynamic and living DDOs. The research findings indicate the need for additional criteria for dynamic DDOs to take account of the greater dependency on multi-media linkages. Living DDOs continue this process of revision a stage further. Due to the differences 'living' DDOs exhibit, additional selection criteria are required. This supports the need for the formulation of new generic selection criteria to assess the long-term value of the intellectual content of DDOs (see Sections 5.2 to 5.4). The research also attests that selection is a vital component of a long-term preservation strategy. The super-metadata framework paves the way for a structured approach to the generation of selection criteria, by providing generic criteria for stakeholders to utilise, which individual repositories may then extend by adding specific criteria pertinent to their requirements.

9.1.2.2 Managing diverse metadata formats

The importance of metadata as a tool for managing digital data has been recognised by many stakeholders, but from different standpoints, as the interviews with stakeholders and the literature review confirms. The philosophy underpinning many of the existing metadata formats is focused on processes other than preservation, such as resource discovery, or the requirements of a particular organisation or community. Formats specifically designed for preservation are emerging, however, as these are designed to operate at the metadata level, they can create problems when used with some existing metadata formats (see Section 5.6.1). Similarly, improving the way data is managed through metadata varies from one organisation to the next, for example new systems are being developed in some organisations, whilst in others existing systems are being refined.

The inability of one metadata format to cater for the wide range of DDOs currently being created, is demonstrated by the existence of the wide range of metadata formats. Similarly, in the past some information-dependent organisations have expended significant resources creating metadata formats. In addition, some stakeholder's rate highly the ability to use data maintained in legacy systems. Therefore, the perpetuation of the existing range of metadata formats is seen to be beneficial and requires provision in a future framework.

The creators of current metadata formats are recognising the need to accommodate interoperability. Strategies to achieve this are being developed (see Sections 5.6.2 to 5.6.2.5). However, methods of achieving this interoperability tend to rely on standards designed for particular DDO types, or the requirement to map existing metadata into a new format at the metadata level. Super-metadata requires the recording of administrative metadata at a higher level than metadata, but not the transfer of the original metadata. Where information at the metadata level is sufficient, links in the super-metadata point to the information, rather than requiring its transfer or replication.

This research supports the hypothesis that states that it is theoretically possible to define universal threshold standards for a framework of metadata for digital data preservation at the metadata level, but concludes that it would not appear to be a practical solution. This is because adopting threshold standards at the metadata level could lead to errors in the metadata arising from conversion, which could lead to loss of access to DDOs. Conversion could be expensive and is not always possible due to metadata incompatibilities.

The super-metadata framework allows for a wide range of metadata formats to exist within a managed repository, but also allows the standardisation of procedures for long-term preservation at a level higher than metadata. There is therefore a division of metadata into two levels. The developing field of digital data management is presently characterised by a wealth of initiatives taking place across a wide spectrum of human

activity. It is also characterised by the absence of a philosophical examination of the nature and management of objects currently viewed as digital entities, as Chapter 1 shows. The super-metadata framework is one possible conceptual approach to the long-term preservation of DDOs achieved by using bi-layered metadata, which allows metadata formats in legacy systems, current and potential formats to be accommodated.

9.1.2.3 Mechanisms for authenticating DDOs

As the malleable nature of DDOs makes them vulnerable to deliberate or accidental change, it becomes important to install a mechanism to protect the authenticity of the data. The empirical data of this research demonstrates the need to increase the confidence of stakeholders to invest time and resources in developing 'best practice' for data management. Super-metadata aims to provide stakeholders with an approach to managing DDOs that monitors authenticity, and aims to ensure the integrity of the data. The development of society's confidence in the ability to authenticate digital objects and construct environments that will prevent unwanted disclosure and misuse, may resolve some of the 'pressure points' that exist between stakeholders. What constitutes authentication varies from one stakeholder to the next. Similarly, the degree to which authentication systems are required is influenced by the degree of control owners have over information flows.

9.1.2.4 Resolving rights management issues

Ownership would appear to be the major determinant of life span and access. This and the purpose of the DDO often influences how long an organisation retains a DDO, whether it is deposited with a third party, and the ability to set access restrictions.

Major concerns for owners of DDOs in the digital environment are the protection of their intellectual property rights and confidence in controlling the disclosure of DDOs. Mechanisms to protect ownership by preventing unauthorised or accidental change to

DDOs, vary between organisations and the degree to which their resources are available to the public. How the maintenance of access to DDOs occurs varies, that is whether a repository maintains a link to a resource or the actual resource. Repositories are equally concerned in maintaining stakeholder confidence in depositing DDOs and of providing current and future access for end-users.

It is apparent that a re-examination of mechanisms to support the protection of intellectual property rights is necessary. This research advocates the development of mechanisms to link ownership with the DDO at the creation phase to encourage creators to define and register the access parameters that are acceptable to them. If one accepts the justification of the public good argument for maintaining access to cultural heritage, and acknowledge that legal deposit provides a channel for ensuring this occurs, then an extension of legal deposit to cater for digital objects would appear to be essential. This research advocates the need for a management structure, such as the super-metadata framework, to support long-term access. It is advisable that any future legislative framework that supports the long-term preservation process, considers the trade-offs between owners and end-user needs implicit in the 'grey areas' identified by this research. Section 9.2.6 puts forward recommendations for developing supporting legal structures. Ownership is seen to be a major determinant in shaping future access to DDOs, therefore to ensure that access is maintained, it is crucial that a future framework addresses stakeholder concerns for the protection of their rights, by establishing a strong link between the DDO and the owner.

9.1.2.5 Development of a cost model

The framework is used to estimate the costs of managing the long-term preservation of DDOs and concludes that these costs can be quite small if the selection and set up costs are spread over many years.

A challenge identified by some stakeholders was assessing the costs involved with the future management of DDOs (see Section 4.13). This has been considered whilst

creating the super-metadata framework and attempting to find possible solutions to reduce what were perceived to be the highest costs (see costing model, Appendix C). The ability to generate a cost model for the proposed framework, based on 'realistic figures', is seen as a first step towards aiding stakeholders in identifying likely costs. The researcher recognises that many issues relating to the distribution of these costs may best be resolved through the development of collaborative pilot projects.

As well as identifying possible costs attributed to stakeholders, the effects of varying individual parameters on the overall costs were examined. The analysis drew a number of conclusions that aid stakeholders in calculating their costs. The interaction between stakeholders appears to be advantageous in reducing overall costs. The findings from the evaluation of the cost model by selected stakeholders (see Section 8.4.3) suggest that the organisational aims are likely to be major determinants affecting how stakeholders define what, to them, are acceptable preservation costs. This suggests that to make a viable financial case for selection, it is advisable to stringently control the manual costs associated with selection. On balance, the societal case of preserving DDOs for the public good, overrides the financial cost of preservation. That is the value, not just in monetary terms which is attached to the digital preservation of the 'societal memory'. To establish accurate costs it is advisable to run pilot projects, as detailed in Section 9.2.4. The ability to generate a cost model is therefore beneficial for stakeholders considering implementing the super-metadata framework and helps to demonstrate its conceptual feasibility.

9.1.2.6 Resolution of stakeholder needs

The development of a framework that identifies the components felt to be necessary for long-term preservation also provides a starting point from which to assess the division of responsibility. The cost model attempts to define the costs that are likely to be incurred, but due to the 'fuzzy boundaries' that exist in the current environment, it is difficult to attribute set costs to particular stakeholders. However, the framework aims to provide a basis on which to start to define stakeholder roles. However,

establishing the division of responsibility is likely to be a complex process, as many interrelated factors contribute to the position of each stakeholder.

By ensuring access to both end-users and repository personnel, the super-metadata framework fulfils its primary function of facilitating future access to DDOs. The framework takes into consideration the nature of DDOs and builds on the increased potential for access that global networking provides. In addition, it also tries to anticipate some of the potential hazards, such as respecting the owner's intellectual property rights. The proposed framework takes the management of DDOs forward in a number of areas as shown above. The research advocates a series of recommendations detailed in the following section.

9.2 Recommendations to aid implementation of the super-metadata framework

This research suggests that the implementation of the proposed super-metadata framework could be aided by the recommendations listed below.

9.2.1 Promoting the spirit of co-operation and co-ordination of revisions

This research advocates a spirit of co-operation and co-ordination towards the development of solutions to long-term digital data preservation. Co-operation could be encouraged by arranging conferences, workshops and working groups consisting of stakeholders with a commitment to achieving long-term access to DDOs. Hand in hand with co-operation, is the need to encourage discussion on taking the super-metadata framework forward, by developing an organisation to co-ordinate revisions and provide advice to stakeholders.

9.2.2 Staff training and re-training

Developments in the preservation of information in digital form challenges staff who have the task of retaining DDOs. To counter-balance these challenges, this research emphasises the need to provide training for new staff and re-training for existing staff, as evidenced by some stakeholder's comments. Until the roles and responsibilities of stakeholders are established, it is difficult to determine exactly what training would be required and the costs this would incur.

9.2.3 Development of selection mechanisms beyond the generic level

Chapter 7 details the generic criteria that this research advocates for all DDOs. However, to be workable in practice these need to be extended along subject lines by qualified panels of 'acknowledged experts', that is individuals, who are recognised by their peers as experts within their field.

9.2.4 Pilot projects to test the resilience of the super-metadata framework to practical challenges

There is a need to develop small-scale projects that will identify whether the framework is resilient to practical challenges and to explore ways to solve problems that may emerge. The expectation is that developing such projects could hasten the resolution of some of the 'grey areas' by bringing them into sharper focus. For example, the funding of a unique identifier scheme would have to be resolved, and the definition of selection boundaries established (that is who is responsible for a particular subject area).

9.2.5 Development of standards

The division of responsibility for preserving digital data over the long-term, together with the development, maintenance and enforcement of standards are 'grey areas'

identified by this research and supported by the literature in the field. Consultation through workshops such as the JISC/British Library and JISC/NPO/UKOLN

conferences[1] are suggested as a means of developing consensus on the division of responsibility and the degree of centralised control over the super-metadata framework; for example whether it is formed from a consortia, what the composition of stakeholders would be within such consortia and the degree of support shown by government.

9.2.6 Development of a supportive legal structure

This thesis emphasises the importance of developing a legal structure that is conducive to the aim of providing long-term access to DDOs the absence of which could restrict the diversity of available objects (Warwick, 2000). This in both the narrow sense of establishing a functional basis for legal deposit and copyright in the digital environment, and in a wider sense of developing confidence in digital technology, such as social policy aimed at inclusion of the whole society.

9.2.7 Emulation and migration strategies

This research recommends further research to establish a basis for information-dependent organisations to judge which strategies are most suitable for which DDOs and the potential challenges each entails.

---

[1]A JISC/British Library workshop organised by UKOLN as part of the Electronic Libraries Programme (eLib) concerned with the *Long term preservation of electronic materials,* (URL: http://www.ukoln.ac.uk/services/elib/papers/other/preservation/intro.html) and the JISC/NPO/UKOLN Digital preservation workshop, March 1999
(URL: http://www.leeds.ac.uk/cedars/OTHER/warwick2.htm).

9.3 Recommendations for further research

Having identified potential solutions to some of the 'grey areas', further research will be required to define these and to address further challenges that have arisen. For example, one grey area that influences the framework, but which is outside the parameters of this present research, is the involvement of the international information community in legislation relating to preservation of digital data. Increasing globalisation and integration of networked resources is making such efforts increasingly necessary. Super-metadata goes some way to providing a structure to support such research developments. The exploration of the research problem has inevitably given rise to more questions to which further research may provide some answers (detailed in the sections that follow).

9.3.1 Confidence in computer intermediaries

There is current interest in the use of intermediaries to aid end-users in the networked environment, be they individuals or organisations. Although developmental work is progressing on the technical aspects of constructing intelligent software agents to work in a range of specific environments and perform designated tasks, there is potential to explore the human centred aspects of utilising computer intermediaries. Once the framework is operational, future research may involve examining whether information-dependent organisations identify boundaries to the confidence placed in computer intermediaries, and if so, where these lie. How does the end-user know whether the selection of DDOs returned by the intelligent software agent is representative, comprehensive and the 'most relevant' to their needs?

9.3.2 Collection boundaries

The distributed nature of digital resources is challenging the nature of a 'collection', in terms of its physical existence and intellectual content. The 'virtual' as opposed to the 'physical' repository has implications for access, by creating technology 'info-rich'

and 'info-poor' (Haywood, 1995). On a second level, the intellectual content of the collections may be questioned by assessing its composition, from the point of view of geographic coverage and the degree of overlap with similar repositories. Are collection boundaries unnecessarily artificial in a world of networked distributed repositories? Are the influences on these boundaries determined more by the level of confidence of stakeholders and/or their level of funding, rather than being an objective measurement of the use of resources? Similarly, rights management or technological advances may strengthen or weaken the influence of a collection based approach to preservation, for example the effect of mirrors and electronic licences on the concept of a collection.

In the light of the current trend towards the globalisation of commercial organisations supported by the growth of e-commerce, there may be a need for international agreements on the preservation of DDOs relating to multinational companies. Does it follow that countries, where companies are registered, will wish to hold all the material deemed to be of long-term importance that a company creates? Will other countries only maintain links to relevant Web pages? The use of links to Web sites may resolve duplication of resources on the Internet but will countries wish to rely on another country's Web pages or databases?

9.3.3 Functioning of the super-metadata framework under different funding and political scenarios

Chapter 8 shows some possible scenarios for the way costs may be apportioned between stakeholders, however as the discussion of cost model notes, these can vary if different stakeholders shoulder more or less of the costs. The empirical research has concentrated on the United Kingdom due to the time allowed and the financial constraints, further research would be necessary to evaluate the framework's applicability under different legal frameworks and variations in administrative practices. The political support, both in terms of public funding and legislative support, may significantly influence the structure of the framework. Future research

may therefore wish to consider the impact of different funding and political scenarios on the functioning of the super-metadata framework.

9.3.4 How to increase co-operation between stakeholders

Will the management of preserved information encourage co-operation between stakeholders? The success with which preservation initiatives such as the super-metadata framework can promote confidence in stakeholders, could be an influential factor that determines the degree of co-operation between the different groups involved at various points in a DDOs life. Research into this area may bring benefits for the management of the super-metadata framework, particularly in the 'grey areas' identified by this research.

9.3.5 Importance of the 'traditional repository' in future long-term preservation of digital data

In a rapidly changing information environment, how will the role of the repository be viewed? As the digital environment is characterised by distributed networked resources and new possibilities to change working practices become available, could traditional repositories assume a greater role as islands of permanence in a sea of change, or will they be viewed as obstacles to change? This challenges the role of existing organisations and communication channels as the most appropriate means of providing long-term access to DDOs.

Bibliography

Abid, A., 1997. *Memory of the World: preserving our documentary heritage.* Paris: Unesco, Information and Informatics Division.
(URL: http://www.unesco.org/webworld/memory/Abid.htm)

Allen, R., *et. al.,* 1996. *A code of practice for legal admissibility of information stored on electronic document management systems.* (PD0008) London: BSI

Andersson, U., 1997. Identification and control of business records and metadata at information and processing modelling: first draft. *In:* D. Bearman & K. Barata, eds. *Electronic Records Research 1997: resource materials.* [CD-ROM] [Pittsburgh]: Archives & Museum Informatics.

Arvidson, A. & F. Lettenström, 1998. The Kulturarw^3 project – the Swedish Royal Web Archive, *Electronic Library,* 16(2), 105-108.

Bainbridge, D., 1999. *Intellectual property.* 4th ed. London: Financial Times [with] Pitman.

Barker, P., 1997. Electronic libraries of the future. *In*: A. Kent ed. *Encyclopedia of Library and Information Science.* vol. 59, New York: Dekker, pp. 119-153.

Beagrie, N., 1998. Digital preservation. *Multimedia Information & Technology,* 24,(3), 173-174.

Beagrie, N. & D. Greenstein, 1998. *Digital collections: a strategic policy framework for creating and preserving digital resources.*
(URL: http://www.ahds.ac.uk/manage/framework.htm)

Beard, K. & T. R. Smith, 1998. A framework for meta-information in digital libraries. *In*: S. Amit & W. Klas, eds. *Multimedia data management: using metadata to integrate and apply digital media.* New York: McGraw-Hill, pp. 341-365.

Bearman, D., 1994. *Electronic evidence: strategies for managing records in contemporary organisations.* Pittsburgh: Archives & Museum Informatics.

Bearman, D., 1996a. *Virtual Archives.* ICA Meeting, Beijing, September 1996.
(URL: http://www.lis.pitt.edu/~nhprc/prog6.html)

Bearman, D., 1996b. *Functional requirements for evidence in recordkeeping.*
(URL: http://www.sis.pitt.edu/~nhprc/prog1.html)

Bearman, D., *et al.,* 1999. A common model to support interoperable metadata: progress report on reconciling metadata requirements from the Dublin Core and INDECS/DOI communities, *D-Lib Magazine*, **5**,(1). (URL: http://www.dlib.org/dlib/january99/bearman/01bearman.html)

Bearman, D. & K. Sochats, N.d. *Metadata requirements for evidence.* (URL: http://www.sis.pitt.edu/~nhprc/BACartic.html) 15 March 1999.

Bell, L., 1979. The archival implications of machine-readable records. *Archivum*, **26**(1), 85-92.

Brake, D., 1997. Lost in cyberspace. *New Scientist*, **154**(2088), 12-13.

Brenner, W., R. Zarnekow & H. Wittig, 1998. *Intelligent software agents: foundations and applications.* Berlin: Springer.

Burke, M. A., 1997. Meaning, multimedia and the Internet: subject retrieval challenges and solutions. *In*: B. Micheline, E. Davenport & N. O. Pors, eds. *Library and information studies: research and professional practice*, proceedings of the 2nd British-Nordic Conference on Library and Information Studies, Queen Margaret College, Edinburgh, 1997. London: Taylor Graham, pp. 61-78.

Checkland, P., 1984. *Systems thinking, systems practice.* Repr. crr. Chichester: Wiley.

Checkland, P. & J. Scholes, 1990. *Soft systems methodology in action.* Chichester: Wiley.

Checkland, P. & C. Tsouvalis, 1997. Reflecting on SSM: the link between root definitions and conceptual models. *Systems Research and Behavioural Science*, **14**(3), 153-168.

Chilvers, A. & J. Feather, 1998. The management of digital data: a metadata approach, *The Electronic Library*, **16**(6), 365-372.

Consortium of University Research Libraries, 1998. *CURL Exemplars for Digital ARchiveS.* (URL: http://www.curl.ac.uk/)

Consultative Committee for Space Data Systems, 1999. *Reference model for an Open Archive Information System (OAIS): draft recommendations for space data system standards.* CCSDS 650. 0-R-1 Red Book, Issue 1, May, 1999. (URL: http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html)

Conway, P., 1996. Selecting microfilm for digital preservation: a case study from Project Open Book. *Library Resources & Technical Services*, **40**(1), 67-77.

Copps, A., 1999. British Library opens up to free Web access. *Times Interface,* 8 September 1999, p. 2.

The Copyright and Rights in Databases Regulations 1997. SI 1997:3032. London: Stationery Office.

Copyright, Design and Patents Act 1998. London: Stationery Office.

CORDS: *Copyright Office Electronic Registration, Recordation & Deposit System.* 1996. (URL: http://lcweb.loc.gov/copyright/cords/cords.html)

Cox, R. J., 1994. Re-discovering the archival mission: the recordkeeping functional requirements project at the University of Pittsburgh, a progress report, *Archives & Museum Informatics*, 8(4), 279-299.

Crown copyright in the information age: a consultation document on access to public sector information. 1998. London: Stationery Office. Green paper. (Cm3819)

Daniel, R. & C. Lagoze, 1997. Extending the Warwick Framework: from metadata containers to Active Digital Objects. *D-Lib Magazine*, Nov. 1997. (URL: http://www.dlib.org/dlib/november97/daniel/11daniel.html)

Day, M., 1996. *Interoperability between metadata formats.* ROADS Workshop, Wellcome Institute, London, 8 August 1996. (URL: http://www.ukoln.ac.uk/metadata/presentations/roads-august1996/)

Dempsey, L. & R. Heery, 1998. Metadata: a current view of practice and issues. *Journal of Documentation*, 54,(2), 145-172.

Department for Culture, Media and Sport, 1997. *New library: the people's network.* London: Library and Information Commission. (URL: http://www.ukoln.ac.uk/services/lic/newlibrary/)

Department of National Heritage, the Scottish Office, the Welsh Office and the Department of Education Northern Ireland, 1997. Legal deposit of publications: a consultation paper, [London]: Department of National Heritage, 1997.

DESIRE-RE 1004: *selection criteria for quality controlled information gateways.* Work package 3 of Telematics for Research project. 1997. (URL: http://www.ukoln.ac.uk/metadata/desire/quality)

Dollar, C. M., 1978. Appraising machine-readable records. *American Archivist*, 41(4), 423-430.

Duranti, L., 1997. The preservation of the integrity of electronic records. *In*: *Proceedings of the DLM-Forum on electronic records, Brussels, 18-20 December 1996.* Luxembourg: Office for Official Publications of European Communities, pp. 60-65. (European Archives News. Supplement; 2)

Duranti, L., T. Eastwood & H. McNeil, N.d. *The preservation of the integrity of electronic records.* (URL: http://www.slais.ubc.ca/users/duranti/)
7 October 1997.

Emmott, S., 1998. At the event: SGML, XML, and databases. *Ariadne*, **18.** (URL: http://www.aridadne.ac.uk/issue18/emmott-sgml/)

Erlandsson, A., 1997. *Electronic records management: a literature review.* Paris: International Council on Archives. (Studies; 10)

Europe and the global information society *- Recommendations of the high-level Group on the Information Society to the Corfu European Council, 26 May 1994.*
[Bangemann Report]
(URL:http://www.ibspan.waw.pl/html/wise/english/rd/prog/general/report.html)

Exon, M., 1995. *Long-term management issues in preservation of electronic information.* National Library of Australia.
(URL: http://www.nla.gov.au/3/npo/conf/npo/95me.html)

Feather, J., 1996. *Preservation and the management of library collections.* 2nd ed. London: Library Association.

Feather, J., G. Matthews, & P. Eden, 1996. *Preservation management: policies and practices in British libraries.* Aldershot: Gower.

Feeney, M., ed. 1999. *Digital culture: maximising the nation's investment: a synthesis of JISC/NPO studies on the preservation of electronic materials.* London: National Preservation Office, British Library.

Feldman, T., 1997. *An introduction to digital media.* London: Routledge.

Finholt, T. A. & J. Brooks, 1999. Analysis of JSTOR: the impact on scholarly practice of access to on-line journal archives. *In*: R. Ekman & R. E. Quandt, eds. *Technology and scholarly communication.* Berkley: University of California Press, pp. 177-194.

Flood, R, L. & M. C. Jackson, 1991. *Creative problem solving: total systems intervention.* Chichester: Wiley.

Gardner, S. R., 1997. The quest to standardize metadata, *Byte*, **22**(11), 47-48.

Getz, M., 1999. Electronic publishing in academia: an economic perspective. *In*: R. Ekman & R. E. Quandt, eds. *Technology and scholarly communication.* Berkley: University of California Press, pp. 102-128.

Glazier, J. D., 1992. Qualitative and nonqualitative research methodologies: thesis, antithesis, or synthesis? *In:* J. D.Glazier, & R. R. Powell, eds. *Qualitative research in information management.* Englewood: Libraries Unlimited, pp. 201-214.

Gorman, M., 1997. Ownership and access: a new idea of "collection". *College & Research Library News*, **58**(7), 498-499.

Graham, P. S., 1993. Intellectual preservation and the electronic environment. *In*: R. S. Martin, ed. *Scholarly communication in an electronic environment: issues for research libraries.* Chicago: Association of College and Research Libraries, 1993, pp. 71-101.

Graham, P. S., 1996. Preserving the digital library. *In*: *Long term preservation of electronic materials*: a JISC/British Library Workshop as part of the Electronic Libraries Programme (eLib), 27th-28th November 1995, at the University of Warwick, organised by UKOLN, report prepared by Marc Fresko Consultancy. London: British Library, British Library Research & Development Department. 1996, pp. 5-12. (BLRD Report 6238)

Gould, C. C. & K. Pearce, 1991. *Information needs in the sciences: an assessment.* Mountain View: Research Libraries Group.

Government Policy on Archives. 1999. (Cm 4516) London: Stationery Office.

Haber, S. & W. S. Stornetta, 1991. How to time-stamp a digital document, *Journal of Cryptography*, 3(2), 99-111.

Harnad, S., 1999. "E-Biomed: very important NIH Proposal." 8 May 1999. lis-elib@mailbase.ac.uk

Harvey, D. R., 1993a. *Preservation in libraries: a reader.* London: Bowker-Saur.

Harvey, D. R., 1993b. *Preservation in libraries: principles, strategies and practices for libraries.* London: Bowker Saur.

Haynes, D., *et al.*, 1997. *Responsibility for digital archiving and long term access to digital data*: a JISC/NPO study with the Electronic Libraries (eLib) Programme on the Preservation of Electronic Materials. London: Library Information Technology Centre.

Haywood, T., 1995. *Info-rich – info-poor: access and exchange in the global information society.* London: Bowker-Saur.

Hedstrom, M., 1995. Preserving the intellectual record: a view from the archives. *In*: L. Dempsey *et al.*, ed., *Networking and the future of libraries 2: managing the intellectual record*, an international conference held at the University of Bath, 19-21 April 1995. London: Library Association & University of Bath, pp. 179-191.

Hedstrom, M., 1997. *Research issues in migration and long-term preservation: Electronic Records Meeting, Pittsburgh, PA, May 29, 1997: session V: maintaining and using records.* (URL: http://www.sis.pit.edu/~cerar/s5-mh.html)

Hedstrom, M., 1998. The role of national initiatives in digital preservation, *DigiNews*, **2**(5).
(URL:http://www.rlg.org/preserv/diginews/diginews2-5html#features2)

Hedstrom, M. & S. Montgomery, 1998. Digital preservation needs and requirements in RLG member institutions. *RLG DigiNews*, **2**(6).
(URL: http://lyra.rlg.org/preserv/diginews/diginews2-6.html)

Heery, R., 1996. Review of metadata formats, *Program*, **30**(4), 345-373.

Heery, R., A. Powell & M. Day, 1997. *Metadata.* London: Library & Information Technology Centre. (Library & Information Briefings: 75)

Hendley, T., 1996. *The preservation of digital materials.* London: British Library, Research and Development Department. (BLRD Report 6242.)

Higgs, E., ed. 1998. *History and electronic artefacts.* Oxford: Clarendon Press.

Hildreth, C. R., 1996. Preserving what we really want to access, the message not the medium: challenges and opportunities in the digital age. *In*: A. H. Helal & J. W. Weiss, eds. *Electronic documents and information: from preservation to access.* Germany: Publications of Essen University Library, 20, pp. 78-95.

Hofman, H., 1998. Towards a United but Distributed Archives of Europe? *In:* E. Higgs, ed. *History and electronic artefacts.* Oxford: Clarendon Press, pp. 331-337.

Hunter, M., 1998. The nature of the book: print and knowledge in the making. *Nature*, **369**, 534-536.

Intellectual Property and Innovation. 1986. (Cm 9712) London: Stationery Office.

Interoperability of data in e-commerce systems: overview. 1999.
(URL: http://www.indecs.org/project/project.htm)

Jenkins, G. M., 1969. The systems approach, *Journal of systems engineering*, **1**(1), 3-49.

Johnson, P., 1996. Selecting electronic resources: developing a local decision-making matrix, *Cataloging & Classification Quarterly*. **22**(3/4), 9-24.

Joint Funding Councils' Libraries Review Group, 1993. *Report for HEFCE, SHEFC, HEFCW and DENI,* Chairman Sir Brian Follett. Bristol: Higher Education Funding Council for England.

Kahle, B., 1997. Preserving the Internet, *Scientific American*, **276**,(3), 72-73.

Kandur, H., 1992. *The management of electronic records*. PhD thesis, University College London.

Kashyap, V. & A. Sheth, 1998. Semantic heterogeneity in global information systems: the role of metadata, context and ontologies. *In*: M. P. Papazoglou & G. Shalgeter, eds. *Cooperative information systems trends and directions*. London: Academic Press.

Keene, S., ed., 1999. A net full of jewels: new museums in the learning age: a report from the National Museum Directors' Conference. London: National Museum Directors' Conference.
(URL: http://www.vam.ac.uk/about/natmusa4.pdf)

Kenney, A. R. & S. Chapman, 1996. *Digital imaging for libraries and archives*. New York: Cornell University Library, Department of Preservation and Conservation.

Kennedy, M., 1999. Online images allow 24 hour entry to cultural gems in leading national galleries. *Guardian*, 29April 1999, p.13.

Kenny, A., 1996. Increased deposits on the next century. *Times Higher Education Supplement*, 8 March 1996, p. vii.

Koehler, W., 1999. An analysis of Web pages and Web site constancy and permanence, *Journal of the American Society for Information Science*, **50**(2), 162-180.

Kranch, D. A., 1998. Beyond migration: preserving electronic documents with digital tablets, *Information Technology & Libraries*, **17**(3), 138-148.

Lawrence, K., 1997. Document Management Systems Workshop. *In*: D. Bearman & K. Barata, eds. *Electronic Records Research 1997: resource materials*. [CD-ROM] [Pittsburgh]: Archives & Museum Informatics.

Lesk, M., 1997. *Preserving digital objects: recurrent needs and challenges*. (URL: http://www.lesk.com/mlesk/auspres/aus.html)

Long term preservation of electronic materials: a JISC/British Library Workshop as part of the Electronic Libraries Programme (eLib), 27th-28th November 1995, at the University of Warwick, organised by UKOLN, report prepared by Marc Fresko Consultancy. London: British Library, British Library Research & Development Department. 1996. (BLRD Report 6238)

Lyman, P., 1999. Digital documents and the future of the academic community. *In*: R. Ekman & R. E. Quandt, eds. *Technology and scholarly communication*. Berkley: University of California Press, pp. 366-379.

Lynch, C. A., 1997. Searching the Internet. *Scientific American*, **276**(3), 44-48.

Macfarlane, I., *et al.,* 1998. *Guidelines on the management and appraisal of electronic records*. Kew: Public Record Office.

MacNeil, H., 1995. Metadata strategies and archival description: comparing apples to oranges. *Archivaria*, **39**(Spring), 22-32.

Mandel, C. A., 1996. Enduring access to digital information: understanding the challenge, *European Research Libraries Cooperation*, **6**(4), 453-464.

Mangan, E. U., 1995. The making of a standard, *Information Technology and Libraries*, **14**(2), 99-110.

Meadows, J., 1996. Preserving the digital imprint, *Learned Publishing*, **9**(4), 215-218.

Megill, K. A., 1997. *The corporate memory: information management in the Electronic Age*. London: Bowker Saur.

Miles, R. K., 1988. Combining 'soft' and 'hard' systems practice: grafting or embedding, *Journal of Applied Systems Analysis*, **15** (April), 55-60.

Moody, G., 1998. A new dawn, *New Scientist*, **158**(2136), 34-37.

Murdock, A., 1997. Roles and responsibilities in managing an electronic archive. *In: Proceedings of the DLM-Forum on electronic records*, Brussels, 18-20 December, 1996. Luxembourg: Office for Official Publications of European Communities, pp. 118-122. (European Archives News. Supplement; 2)

Nader, J. C., 1998. *Prentice Hall's illustrated dictionary of computing*. 3rd ed. Sidney: Prentice Hall.

National Archives and Records Administration, 1999. *The Electronic Access Project, 1996-1999.* (URL: http://www.nara.gov/nara/vision/eapover.html)

National Library of Sweden, 1999. *The Kulturar³ Heritage Project: long-term preservation of published electronic documents.*
(URL: http://kulturarw3.kb.se/html/projectdescription.html)

National Preservation Office and Research Libraries Group, 1998. Guidelines for digital imaging: papers given at the Joint National Preservation Office and Research Libraries Group preservation conference in Warwick, 28th – 30th September 1998.

Neavill, G. B., 1984. Electronic publishing, libraries and the survival of information. *Library Resources & Technical Services,* **28**(1), 76-89.

Nelson, M. L., *et. al.,* 1999. *Metadata and buckets in the Smart Objects, Dumb Archives (SODA) model.*
(URL: http://computer.org/proceedings/meta/1999/papers/53/mnelson.html)

Ng, K. B. & K. Burnett, 1997. Control or management: a comparison of the two approaches for establishing metadata schemes in the digital environment. *In:* C. Schwartz & M. Rorvig, eds. *Digital collections: implications for users, funders, developers and maintainers*: ASIS '97: proceedings of the 60th ASIS annual meeting, Washington, DC, November 1-6, 1997. New Jersey: Information Today, pp. 337-346.

Owen, J. S. M. & J. V. D. Walle, 1996. *Deposit collections of electronic publications.* Luxembourg: Office for Publications of the European Communities. (Libraries in the Information Society; EUR 16910).

Pasquinelli, A., ed. 1997. *Information technology directions in libraries: a Sun Microsystems white paper: August 1997.*
(URL: http://www.sun.com/products-n-solutions/edu/libraries/libtechdirection.html)

Phillips, J. T., 1995. Metadata - information about electronic records, *Records Management Quarterly,* **October,** 52-55 & 72-73.

Powell, A., 1997. Collecting numbers: identifiers for digital objects, *Library Association Record: Library Technology,* **2**(5), 103-104.

Powell, A. & L. Dempsey, 1997. *Dublin Core metadata: a tutorial.* Metadata Workshop, Luxembourgh, 15 December 1997.
(URL:http://hosted.ukoln.ac.uk/ec/metadata-1997/tuto...ion/sld053.htm)

Prescott, A., 1998. Constructing electronic Beowulf. *In*: Carpenter, Leona *et al.*, eds. *Towards the digital library: the British Library's Initiatives for Access programme.* London: British Library, pp. 30-49

President's Information Technology Advisory Committee interim report to the President: August 1998. (URL: www.ccic.gov/ac/interim/)

Proposal for a European Parliament and Council Directive on the harmonization of certain aspects of copyright and related rights in the Information Society. (COM(97) 628 final) 10 December 1997.
(URL: http://www.europa.eu.int/comm/dg15/en/intprop/intprop/copyen.pdf)

Public Record Office. 1999. *What is the AD2001 programme?: creating an online public record office.*
(URL: http://www.pro.gov.uk/ad2001/programme.htm)

Ratcliffe, F. W., 1999. Legal deposit. *In:* C. J. Armstrong ed. *Staying legal: a guide to issues and practice for users and publishers of electronic resources.* London: Library Association, pp. 75-96.

Ratcliffe, F. W. & D. Patterson, 1984. *Preservation policies and conservation in British libraries.* Boston Spa: British Library Board. (Library & Information Research Report, 25)

Ross, S., 1997. Consensus, communication, and collaboration: fostering multidisciplinary cooperation in electronic records. *In: Proceedings of the DLM-Forum on electronic records, Brussels, 18-20 December 1996.* Luxembourg: Office for Official Publications of European Communities, pp. 330-336. (European Archives News. Supplement; 2)

Ross, S., 1998. The expanding world of electronic information and the past's future. *In:* E. Higgs, ed. *History and electronic artefacts.* Oxford: Clarendon Press, pp. 5-28.

Rothenberg, J., 1995. Ensuring the longevity of digital documents. *Scientific American,* **272**(1), 24-29.

Rothenberg, J., 1999. *Avoiding technological quicksand: finding a viable technical foundation for digital preservation.* Washington D.C.: Council on Library and Information Resources; European Commission on Preservation and Access.

Saunders, T., 1998. Into the future: on the preservation of knowledge in the electronic age. [Video recording] Washington D.C.: Commission on Preservation and Access.

Shepherd, E., 1996. *The management of electronic records.* London: Library & Information Technology Centre. (Library & Information Briefings: 69)

Sheth, A. & V. Kashyap, 1993. So far (schematically), yet so near (semantically). *In: D. K. Hsiao, E. J. Neuhold & R. Sacks-Davis, eds. Proceedings of the IFIP WG2.6 database semantics conference on inoperable database systems (DS-5), Lorne, Victoria Australia, 16-20 November, 1992.* Amsterdam: North Holland, pp. 283-312.

Solbakk, S. A., 1995. Long-term preservation of electronic materials - preliminary experiences from the Norwegian National Library. *In: Legal deposit with special reference to the archiving of electronic materials*, proceedings of a seminar organised by NORDINFO and the British Library, (Research and Development Department), held at the Castle Hotel, Windsor, England 27-29 October 1994. London: British Library, 1995. (BLRD 6197)

Steering Committee for the Study on the Long-term Retention of Selected Scientific and Technical Records of the Federal Government, 1995. *Preserving scientific data on our physical universe: a new strategy for archiving the Nation's scientific information resources.* Washington. D.C.: National Academy Press.

Sturges, P., 1987. Policies and criteria for the archiving of electronic publishing. *Journal of Librarianship*, **19**(3), 152-172.

Swartzburg, S. G., 1995. *Preserving library materials: a manual.* 2nd ed. London: Scarecrow Press.

Task Force on Archiving of Digital Information. 1996. *Preserving digital information. Report of the Task Force on Archiving of Digital Information commissioned by the Commission on Preservation and Access and the Research Libraries Group.* Washington, D.C.: Commission on Preservation and Access and the Research Libraries Group. (URL: http://www.rlg.org/ArcTF/).

Taylor, A. G., 1999. *The organization of information.* Englewood, Colo.: Libraries Unlimited.

Thomas, T., 1998. Archives in a new paradigm of scientific publishing: Physical Review Online Archives (PROLA). *D-lib Magazine*, **May 1998.** (URL:http://mirrored.ukoln.ac.uk/lis-journals/dlib/dlib/dlib/may98/05thomas.html)

Thompson, B., 1998. The mother of all filing jobs. *Internet Magazine*, **48**, November, 62-68.

United States Food and Drug Administration, 1997. *Code of federal regulations: 21 Food and drugs: Part 11: Electronic records; electronic signature.* (21CFR11). (URL: http://frwebgate2.access.gpo.gov/)

Visual Arts Data Service. 1998. *Guidelines for depositors.* (URL: http://vads.ahds.ac.uk/deposit-pack.html)

Wallace, D. A., 1995. Managing the present: metadata as archival description. *Archivaria*, **39**(Spring), 11-21.

Warwick, C., 2000. The lowest canonical denominator: electronic literary texts, and the role of the information professional.
(URL: http://www.shef.ac.uk/~is/publications/infres/paper71.html)

Webb, C., 1999. Preservation of electronic information: what we should be thinking about now. *International Preservation News*, **18**, March, 8-13.

Weibel, S., *et al.*, 1995. *OCLC/NCSA Metadata Workshop Report.*
(URL:http://www.oclc.org:5046/conferences/metadata/dublin_core_report.html)

Working Party on Legal Deposit. 1999. [Report]. (URL:
http://www.bl.uk/information/legal-depositgrp.html)

Your right to know: the Government's proposals for a Freedom of Information Act. 1997. London: Stationery Office. White paper. (Cm3818)
(URL: http://www.official-documents.co.uk/document/caboff/foi/foi.htm)

Zarnekow, R., 1998. Fundamental concepts of intelligent software agents. *In*: W. Brenner, R. Zarnekow & H. Wittig, *Intelligent software agents: foundations and applicatons.* Berlin: Springer, pp.19-34.

Appendix A

PILOT QUESTIONNAIRE

Purpose of Research

The research aims to examine the issues relating to the long-term preservation of Digital Data Objects within libraries and archives.

Definition of Terms

Digital Data Objects (DDOs) for the purposes of this research are defined as, machine-readable, computer-based information resources (for example CD-ROMs, databases), which may or may not be formally published and their information content can be stored and accessed independently of the format from which it was originally created.

Preservation of digital data in the context of this research is concerned with the financial, managerial and technical aspects involved in the preservation of Digital Data Objects (DDOs) in libraries and archives with the aim of prolonging its useful life.

All answers will be treated in strict confidence.

To enable follow up research to be undertaken please complete the following information.

Name of organisation:.....

Address: .....

..... Telephone number:

Contact name:.....

Use of Digital Data Objects (DDOs)

1. Does your library stock any DDOs?

Please circle appropriate number.

Yes	1
No	2

If Yes please circle appropriate number(s).
If No please go to question 17.

CD-ROMs	1
Electronic books	2
Electronic journals	3
Computer software	4
Other	5
None	6

If other please specify: .....

2. Does your library own any DDOs?

Please circle appropriate number.

Yes	1
No	2

If Yes please circle appropriate number(s).
If No please go to question 3.

CD-ROMs	1
Electronic books	2
Electronic journals	3
Computer software	4
Other	5
None	6

If other please specify: .....

3. Do you give your users unrestricted access to any of the following DDOs?

Please circle appropriate number(s).

Electronic journals	1
Data archives	2
Internet	3
Other	4

If other please specify: .....

4. What use of DDOs is made in your organisation?

Please circle appropriate number(s).

Teaching	1
Research	2
Reference work	3
Income generation activities	4
Issue to user	5
Other	6

If other please specify: .....

5. What type of DDOs do you encounter in your work?

Please circle appropriate number(s).

CD-ROMs	1
Electronic books	2
Electronic journals	3
Computer software	4
Databases	5
Other	6

If other please specify: .....

Preservation of DDOs

6. Does your organisation preserve DDOs?

Please circle appropriate number(s).

Yes	1
No	2

If Yes please go to question 7.
If No please go to question 13.

7. How many DDOs did you acquire last year?

Please circle appropriate number on the right.

0-10	1
11-30	2
31-100	3
101-300	4
301-1,000	5
More than 1,000	6

What percentage of these do you expect to preserve for each of the following periods?

Please state percentage next to each option.

Less than one year	..... %
One to 5 years	..... %
6 to 10 years	..... %
More than 10 years	..... %

8. What subject area(s) do these preserved DDOs cover?

Please circle appropriate number(s).

Science	1
Humanities	2
General reference	3
Social science	4
Arts	5

9. Why does your organisation preserve DDOs?

Please circle appropriate number(s).

To meet user demands	1
For legal deposit	2
For development of a special collection	3
For development of the whole collection	4
Other	5

If other please specify: .....

10. Have you experienced any difficulties in preserving DDOs?

Please circle appropriate number(s).

Yes	1
No	2

If Yes please specify.....

Collection Management

11. Does your organisation have a written collection management policy for DDOs?

Please circle appropriate number.

Yes	1
No	2
In process of being developed	3

12. What criteria do you apply to the selection of DDOs?

Please circle appropriate number(s).

Choice of subject matter	1
Quality of content	2
Legal deposit requirements	3
To fulfil a co-operative agreement with other organisations	4
Other	5

If other please specify: .....

13. What sources do you use to obtain DDOs?

Please circle appropriate number(s).

Commercial producer	1
In-house	2
Consortium	3
Internet	4
Donation	5
Other	6

If other please specify: .....

14. Do you create DDOs?

Please circle appropriate number.

Yes	1
No	2

If No please go to question 16 .

15. What type of DDOs do you create?

Please circle appropriate number(s).

CD-ROMs	1
Electronic journals	2
Databases	3
Web sites	4
Other	5

If other please specify: .....

16. Do you belong to a consortium that shares DDOs?

Please circle appropriate number.

Yes	1
No	2

17. Should there be legal deposit for DDOs?

Please circle appropriate number.

Yes	1
No	2

Please use this page for any additional comments.

Thank you for your time and co-operation.

If you have any queries please contact Alison Chilvers on 01509 223079, or E-mail a.h.chilvers@lboro.ac.uk

Please return the completed questionnaire in the enclosed stamped addressed envelope.

Appendix B

Stakeholder interview questions

Interview held on:
At:
With: Tel:

<u>Brief explanation of the research</u>

This research is concerned with the preservation of the intellectual content of data. By preservation I am referring to the financial, managerial and technical aspects involved in the retention of digital data. The research is not concerned with the physical preservation of data.

The aim of this research is to examine the management of records that have only ever existed in electronic form. It is clear that there is a large volume of electronic data being created daily and that our ability to physically preserve greater numbers of electronic records/data sets is becoming increasingly feasible both in terms of computer storage capacity and cost. However, given that it is increasingly easy to preserve the data, does it cause more management problems than it solves and is it therefore desirable to preserve more data?

This chart (see Table B.1) shows the three categories of data types that occur at present. Static, which covers digital data, paper-based and non-digital, audio and audio-visual materials. 'Dynamic' resources which consist of electronic data sometimes with none electronic data (polymedia), in which information can be manipulated but not changed totally, for example CD-ROMs. 'Living' electronic records/data sets contain information that can be manipulated and changed, for example constantly updated data sets.

<u>Nature of electronic records/data sets and reasons for preservation</u>

- In your corporate view, what is an electronic record?
- Does your organisation preserve electronic records/data sets?

  <u>If the respondent answers **no:**</u> Why does your organisation not preserve electronic records/data sets?

- Are 'dynamic' electronic records/data sets preserved by your organisation? (For example information held on CD-ROMs.)

- Are 'living' electronic records/data sets preserved by your organisation? (For example evolving databases.)

  - Are any of these living electronic records/data sets constantly evolving?

Table B.1 Grouping of electronic books based on Barker's (1997) Typology

Static[1]			Dynamic Electronic Record	Living Electronic Record
Paper	Cassette	Video		
Textbooks Picture-books	Talking Books	Moving picture books	Multimedia Polymedia books Hypermedia books Intelligent electronic books	Telemedia books Cyberbooks

[1] This category includes static electronic records such as the products of digitization projects.

- In your corporate view does your organisation make a distinction between 'dynamic' and 'living' electronic records/data sets?

Preservation policy and procedures

- For what reasons does your organisation preserve electronic records/data sets over the long-term?

- Is there a legal requirement for your organisation to preserve electronic records/data sets?

  If respondent answers **yes**: To which types of electronic records/data sets does this apply?

  If respondent answers **no**: Why is this the case?

- Is there an administrative requirement for your organisation to preserve electronic records/data sets?

  If respondent answers **yes**: To which types of record(s) does this apply?

  If respondent answers **no**: Why is this the case?

- Are the electronic records/data sets which your organisation preserves of permanent importance, and if so, why and to whom?

- What procedures does your organisation use to preserve the information content of electronic records/data sets, such as web pages, databases and hypermedia resources?

  Hypermedia: This is very similar to multimedia, except hypermedia has embedded non-linear information structures interlinked in different ways to create complex web structures.

- Does your organisation have a written policy for preserving the electronic records/data sets it creates?

  If respondent answers **yes**: What is the policy of your organisation concerning electronic records/data sets?

  If respondent answers **no**: In your corporate view is this an issue for your organisation?

- Does your organisation have a written criteria which it uses to select electroinic records/data sets?

- Are your organisation's records stored solely electronically, or also in hard copy form, and if so, why?

- In your corporate view, do you think that electronic records/data sets are short-term in nature, and if so, why?

- In your corporate view, do you think the Internet is short-term in nature, and if so, why?

- What are the estimated life spans of the electronic records/data sets held by your organisation?

  For example; Do you clear the archive periodically? Has thought been given to the preservation of electronic records/data sets created by individuals, in the same way your archivist/librarian has given thought to the preservation of paper records?

- What happens when an electronic record gets to the end of its life?

- With regard to the 'living' electronic records/data sets that your organisation stores, is it intended to 'keep alive' these for the foreseeable future?

<u>Challenges/selection</u>

- What challenges does your organisation encounter with regard to the long-term preservation of electronic records/data sets?

- What systems does your organisation employ to maintain and restrict access to the electronic records/data sets it holds?

- For what reason(s) have these system(s) been adopted?

- Are the organisation's electronic records/data sets managed and monitored, and if so how?

- Are different methods used to preserve different types of electronic records/data sets? (For example; 'living' and 'dynamic' data.)

- In your corporate view, what factors relating to the data content, influence the selection methods used by your organisation? For example: significance (scholarly, legal or administrative importance)?

- What training/experience does the person have who takes the decisions to select these electronic records/data sets for long-term preservation?

- At what level of seniority within your organisation are electronic records/data sets selected for long-term preservation?

- What criteria does your organisation use to select electronic records/data sets for long-term preservation?

- In your corporate view, are there any particular types of electronic records/data sets that give your organisation problems, when selecting them for long-term preservation, and if so, what are these, and what problems do they cause?

- Are there any types of electronic records/data sets that are deliberately excluded from your organisation's long-term preservation strategy?

- In your corporate view, does this selection process adequately cater for 'dynamic' electronic records/data sets?

- In your corporate view, does this selection process adequately cater for 'living' electronic records/data sets?

<u>Documentation and authentication</u>

- In your corporate view, what constitutes a validated electronic record/data set?

- In your corporate view, what constitutes 'good' documentation?

- Are there written guidelines relating to accompanying electronic documentation?

- If the record has inadequate documentation but is of commercial, legal, administrative or historical importance, what action would be taken?

- What mechanisms does your organisation employ, to ensure that electronic records/data sets, which are being considered for preservation, are authentic?

<u>Record life cycle</u>

- How is the duration of the life cycle of a document defined by your organisation?

  Life cycle refers to the stages a record goes through from its creation through to its disposal

- When in an electronic record's/data set's life is its life span determined? (When it is first created?, At a periodic review stage?)

- What factors influence the duration of the life span of the records held by your organisation and how? (For example: current user demand, future scholarly demand, currency of the data.)

- When are electronic records/data sets identified as needing long-term preservation? (For example; when they are created, when they are revised.)

Metadata formats

- In your corporate view, do you think metadata (data about data) should be attached to the electronic records/data sets your organisation intends to preserve and why?

  What factors influence this view?
  For example; financial resources, human resources, development of national/international standards

- In your corporate view, at what stage in a documents life cycle should metadata be attached?

  If respondent feels metadata should be attached: In your corporate view what metadata is required to ensure that the information content of an electronic record you are preserving is authentic?

- What metadata format(s) does your organisation use?

  If respondent feels **one metadata format** should be used: Is there a particular metadata format that your organisation has adopted for the preservation of electronic records/data sets, and if so, why has this format been adopted?

  If respondent feels **several formats** of metadata should be used: What is/are the reason(s) for adopting more than one format?

  If an **in-house scheme** was developed: What were the main reasons behind adopting an in-house scheme?

- In your corporate view, do you see any areas where improvements could be made to this scheme?

- Does the type and amount of metadata vary, dependent on the purpose of the electronic records/data sets? (For example; those that are deemed to be of administrative, legal or permanent importance?)

- In your corporate view, does this metadata cater adequately for all the electronic records/data sets preserved by the institution/company?

  If respondent feels that only some electronic records/data sets are catered for: Which electronic records/data sets are not adequately catered for and why?

- In your corporate view, is it important to have interoperability between metadata formats, and if so, why?

Integrity and access

- In your corporate view, could you trust an entry for an electronic record on a world preservation database as proof that a work has been preserved, and discard your organisation's copy(s) on this basis?

  If respondent answers **no**: Why would this situation not be acceptable?

- If a record were to reach the end of its useful life (ceased to be of legal, scholarly or administrative value to your organisation), in your corporate view, would your organisation allow another institution or body to preserve the record?

  If respondent answers **no**: Why would this situation not be acceptable?

Balance between creators and users

- Who are the main users of the electronic records/data sets which your organisation preserves?

- Who are the main originators of the electronic records/data sets preserved by your organisation?

- Is there a policy to protect your organisation's rights, as a creator of electronic records/data sets, from external users of the data, and if so, what form does this take?

- Does your organisation balance the needs of creators of electronic records/data sets with those of users, and if so, how?

  If respondent answers **no**: Why is this not felt necessary?

Future Developments

- Are there any plans to extend the service and if so in what way will this occur?

Appendix C

Costing model for the super-metadata framework

C.1 Explanation of assumptions and variables

As a basis for the costs a number of assumptions were made and variables identified, as shown in Table C.1. The costs are based on a set of standard assumptions, a few of which originate from the Yale Project Open Book model (Task Force on Archiving of Digital Information, 1996) applicable to the digital archive as indicated in Table C.1. These costs can only be a guide to the true costs and their sensitivity to the standard costs has been tested by varying these assumptions within the range indicated in the table. It must be emphasised that the values aim to be realistic in order to identify the significant costs.

The cost model assumes the acquisition of 200,000 DDOs per year, each year over a ten year span. The assumed general rate of inflation over this period is taken to be three percent, with a corresponding borrowing rate of seven percent. The inflation rate applies to items such as salary and raw material costs. A ten year span is considered adequate for analysis since the initial outlay incurred by a repository is found to change very little when spread over a greater period of time (see Figure C.1[1]). The amortisation period is set at five years in line with the Yale model assumptions, similarly the period over which the hardware/software costs halve due to advances in technology is assumed to be five years. The implied hardware/software inflation rate is a fall of 13% per year, as expressed in the following formula.

$$-0.13 = (1/2)^{1/5} - 1$$

[1] The costs presented in this and the following bar charts are given in Figures C.1 to C.9 and Tables C.10 to C.15 at the end of this appendix.

Table C.1 Assumptions and variables used in the cost modelling of the super-metadata framework

Variable	Standard assumption	Range considered
General inflation rate	3%	-
Interest rate	7% (Y)	-
Hard/software cost halves in x years	5 years (Y)	-
Amortisation period	5 years (Y)	-
Implied hardware and software cost inflation (%)	-12.94 (Y)	-
Labour rate	£21.60 per hour	£14.40 - £28.80 per hour
Fraction of DDOs accessed	20% (Y)	10-20%
Percentage of DDOs selected automatically	75%	50-90%
Percentage of DDOs selected manually	25%	10-50%
Percentage of DDOs from offline sources	10%	5-10%
Percentage of living DDOs	2%	-
Percentage of DDOs from the Internet	30%	10-60%
No. of DDOs acquired annually	200,000 (Y)	-
Time taken to manually select	30 min	15-60 min
No. of unknown metadata formats	Halving each year, initially 5%	-
Cost of accessing DDO	£2.36 (Y)	-
Digital Public Lending Right (PLR)	£2.00	-
Percentage deposited with insufficient bibliographic details	5%	-
Access restrictions = ten percent access cost	£0.24	-
Storage cost per annum	£1.08 (Y)	-

Legend: Y = From Yale Project Open Book Model

The standard assumption sets the labour rate at £21.60 per hour or an annual salary of £30,000 for management and systems engineering staff, factored by 40% to include overheads. Like the Yale model, the standard assumption for the fraction of DDOs accessed per annum is set at 20%, that is 40,000 DDOs are assumed to be accessed every year. The figure is derived from a 5% increase on the circulation rate of 15% of volumes, achieved by the Yale depository library, assumed to be typical for a traditional research library. The five-percent increase is based on the assumption that easier access in the digital library will lead to an increase in rate of use. The cost of achieving access to individual DDOs is based on the Yale figure for the end of year four, converted from dollars to sterling giving a price of £2.36. This price excludes the cost of printing and delivery which is considered as a separate user incurred cost in the model. This is one of the major ongoing cost factors likely to be incurred and is primarily due to the cost of owning and maintaining computer servers. The cost of maintaining any access restrictions required such as those which may pertain to particular groups of end-users also needs to be considered. This is closely related to access costs and for the purposes of this research a figure of 10% of the access cost is assumed.

Although the Yale model recognises acquisitions, cataloguing and reference services it does not make an attempt to cost them. As the cost model developed by the researcher explores the costs incurred by selection, a number of additional variables to the Yale model needed to be considered. The process of selection proposed for this framework includes a substantial amount of automation, achieved by the use of intelligent software agents. Assumptions have to be made about what proportion of DDOs can be selected automatically and how many require manual intervention. The standard assumption for this cost study is that 75% percent of the DDOs will be selected automatically.

The acquired DDOs that are not online need conversion to online. For the purpose of this research it is assumed that only a small percentage of DDOs are received from offline sources and a figure of 10% is used for this study.

The percentage of living DDOs acquired by the repository is likely to be very small since they only represent a small proportion of available DDOs, a figure of two percent is used in this research. Even if this figure doubles to four percent the effect on costs is likely to be very small.

The percentage of DDOs collected from the Internet is another important parameter, since the selection procedures are different in this case. Unlike DDOs that have originated through a publisher and may have already been peer-reviewed, DDOs from the Internet have to undergo a thorough selection and review process. As the importance of global networks grows the percentage of DDOs retrieved from these sources are likely to increase. The standard assumption is set at 30% of DDO from global network sources.

For the DDOs not selected automatically, (which is assumed in the cost model to apply to 25% of DDOs) a manual process is required. The time estimated for this process, taking into account variations in the difficulty to select individual objects, has been taken to be 30 minutes, which has been chosen to keep manual selection costs in proportion to other costs.

An important aspect of super-metadata is the ability to absorb DDOs with any metadata format. This requires setting up a metadata registry. It is assumed that the percentage of DDOs entering the repository with a previously unknown metadata format is initially 5%, but this is assumed to halve each year. The associated set up costs fall each year accordingly.

Built into the costing model is a means of compensating owners for potential loss of revenue that may result from depositing material with a repository (see Section C.3.4). The model assumes that the percentage of DDOs from 'published' sources to be 65% of selected DDOs. Only 20% of these DDOs (access rate) incur this cost.

Another set up cost is the process of completing the bibliographic details, where necessary. It is assumed that the majority of DDOs will have sufficient bibliographic details attached and a figure of five percent is used for those with insufficient detail. Consultations with owners to acquire further details could represent costs for creators, repositories, ISPs and publishers.

The cost of storage per DDO per annum is set at £1.08 derived from the price stated in dollars in the Yale Model. Included in this cost is data migration, which Yale assumes to be double the storage cost. This covers the purchase of a new disk and the operations charge for migration of data.

The cost model attempts to cater for a greater diversity of stakeholders undertaking long-term preservation of DDOs and for a higher diversity of types of DDOs than in the Yale model. To achieve this a number of assumptions supplement those taken from the Yale model, and this section gives an explanation of each. Unless otherwise stated the costs are per DDO and incurred by the repository.

C.2 Component costs

Table C.2, shown below, defines individual items of cost. These items are coded to indicate whether the costs are linked to general inflation (rpi) or hardware and software cost inflation (hi). Additionally the code shows if the cost relates to; set up (t); selection (l); access (a); or storage (o) processes. The unit costs for items are listed and factored where applicable, by the quantity per DDO and in some cases by additional factors which can vary according to the stakeholder involved, since not all stakeholders are assumed to incur the same costs. Each stakeholder is identified under 'Factor' in table C.2, using the following codes; c, creator, p, publisher, s, Internet service provider, r, repository and u, end-user.

Where items are not specified below, a broad estimate of the unit cost and/or quantities have been given. Total costs for each item is the unit cost multiplied by the quantity and the factor appropriate for the given stakeholder. In many cases, the time quantities stated appear quite low since they are expressing an average value.

Table C.2 Cost breakdown for the super-metadata framework

Item No. (applicable nos. only)	Codes	Description of item	Unit cost £	Quantity per DDO	Stakeholder factor (see text)				
					c	p	s	r	u
1	rpi t	Create and attach metadata	s[1]	15 min	1				
2	rpi t	Application for unique identifier	s	5 min	1	1		1	
3	rpi l	Peer review	s	40 min		1			
4	rpi l	Reviewed by Internet Service Provider (ISP)	s	5 min			1		
5	rpi t	Legal deposit (where applicable). Cost of submission to depositor	10	1 unit		1	1		
6		Selection							
(a)	hi l	Intelligent software agent to retrieve DDOs	2.83	1 unit				1	
(b)	hi l	Intelligent software agent that selects DDOs deemed to be suitable for long-term preservation	2.83	1 unit				1	
(c)	rpi l	Manual selection when result of item (b) is marginal	s	30 min				1	
(d)	rpi l	Periodic refinement of intelligent software agent selection criteria	s	0.1 min				1	
(e)	rpi l	Periodic refinement of manual selection	s	0.2 min				1	

[1] s = labour rate in £/min

7	rpi a	Compensation to owner for DDO for lost commercial revenue	Set price 2.07	0.65 x n^2				1	
8	rpi l	Determination of 'snapshot' frequency for living DDOs only	s	5 min x 0.02				1	
9		Rights management							
(a)	rpi t	Negotiate access terms with rights holders	s	5 min	1	1	1	1	
(b)	hi a	Apply access restrictions	0.24	n				1	
10		Digital time-stamping (optional)							
(a)	hi t	Generation of a hash.	0.20	1 unit				1	
(b)	rpi t	Certificate of authenticity (post/electronic mail)	0.50	1 unit				1	

[2] n = rate of access of DDOs

11		Computer aided check of metadata							
(a)	hi t	Check metadata format against metadata registry	0.20	1 unit				1	
(b)	rpi t	Consultation with creator(s)/owner(s) where metadata is not recognised as a known type	s	15 min	1			1	
(c)	rpi t	Addition of previously unknown metadata formats to the registry.	s	15 min				1	
(d)	hi o	Maintenance of metadata registry (off site back up)	0.02	1 unit				1	
12	hi t	Attach unique identifier to form a link with the super – metadata	0.01	1 unit				1	
13		Life span							
(a)	rpi l	Determine life span – in perpetuity or specify review period	s	5 min				1	
(b)	hi l	Automatic counter to record amount of use. Used for life span review	0.01	1 unit				1	
14		Super-metadata formulation							
(a)	hi t	Automatic generation	0.05					1	
(b)	rpi t	Manual intervention when information for certain elements cannot be obtained from DDO or metadata	s	1 min	1			1	

15 (a) (b)	rpi a rpi l	Maintenance of a web site with criteria for deposit and subject coverage Proactive research by repository staff to determine what information is being produced (e.g. attendance at conferences)	s s	0.001 0.025				1 1	
16	rpi t	Recording disposal decisions	s	2 min				1	
17	hi a	Access cost (applies to r or u)	2.36	1 unit				1	1
18	hi o	Storage cost	1.08	1 unit				1	
19	rpi t	Where data not online convert to online	s	0.5 min x 0.1				1	
20	rpi t	Consult owner & complete bibliographic details	s	10 min x 0.05	1	1	1	1	
21	hi t	Check to see if DDO already exists in the repository. Assume same access costs as YOBP factored by 0.1[3]	0.24	1 unit				1	
22	hi a + rpi a	Printing cost (applies to r or u)	0.70	assuming only 10% of end-users require hard copy				1	1
23	rpi a	Repository staff aiding users.	s	0.1 min				1	

[3] Since this is a fully computerised process with exact bibliographic details to search, assume costs to be only 10% of normal user access costs. This cost does not increase since the bibliographic details are stored in alphabetic/numeric order, hence providing rapid access to the super-metadata. The bibliographic details are part of the super-metadata. The super-metadata/metadata database would have two versions, a static and living version. The living database could be continuously updated as new DDOs are entered into the repository. The static version of the database would be used by end-users to search for DDOs and could be updated on a daily basis.

Item 2 The application for a unique identifier is assumed to involve five minutes of labour time. For any given DDO this cost will be incurred by either a creator, the publisher or if neither of these have applied for a unique identifier the repository will have to incur this cost.

Item 5 Legal deposit is assumed to only apply to DDOs that have been published through recognised channels, although this requirement does not currently exist and a figure of £10 has been chosen. In reality, this could vary from a few pounds to several thousand pounds. The costs incurred by depositing DDOs with a repository is incurred by whoever publishes it, that is either an Internet Service Provider or traditional publisher.

Item 6a The cost of retrieving DDOs from networked sources uses the Yale access cost and factors this by 1.2 to take into consideration the extra searching needed for resource discovery.

Item 6b As shown in the framework, a second intelligent software agent selects the DDOs irrespective of their origin, therefore the selection costs for all material selected automatically is anticipated to be the same. The cost again is based on the Yale access figure and factored by 1.2 to allow for the increased work an intelligent software agent would undertake to select objects.

Item 6d Costs associated with the periodic refinement of the intelligent software agent and the manual selection and resource discovery criteria are distributed over the total number of DDOs selected per annum.

Item 7 The compensation payment to the owner of a DDO for lost commercial revenue has been set at £2.07 as discussed in Table C.2. This is then factored by 0.65 (the percentage of DDOs assumed to be received from traditional publishers). It is also factored by the rate of access to DDOs (specified as 20% as standard assumption in Table C.1). This cost is applied to the repository who administer the payments.

Item 8 Determining the snapshot frequency for living DDOs assumes five minutes of labour time, factored by 0.02, the assumed proportion of living DDOs.

Item 9a Negotiate access terms with rights holders. This assumes five minutes labour costs by the repository and the owner of the DDO. Thus, negotiation could involve creators, repositories, Internet service providers and publishers.

Item 9b	The application of access restrictions is a set price factored by the access rate and is assumed to be a cost incurred by a repository as part of its set up costs.
Item 15a	Maintaining a repository web site to communicate the existence of the holdings and requirements to potential depositors and end-users involves staff time, but the cost is small when factored by the number of DDOs acquired annually.
Item 15b	Repository staff time spent proactively seeking DDOs that are in process of being created is again a small cost per DDO as again this is factored by the number of DDOs acquired annually.
Item 17	The access cost as stated in Table C.1 is assumed to be a set price, which may be incurred by the repository, the end-user or the costs may be shared.
Item 22	The cost of acquiring a hard copy has been set at £0.70 per DDO, based on the Yale assumption that only 10% of the use of the repository results in end-users wishing to make a hard copy. The figure is therefore 10% of the actual cost, includes a charge for copyright clearance, and assumes this to remain stable in the future. The cost of printing, as with access, may be incurred by the repository, the end-user, or shared by both stakeholders depending on funding policies.

C.3 Results of the cost study

These are discussed in Section 8.2.12, in addition to the following sections which show the effects of varying the costing assumptions. The reader can explore the effects of changing these assumptions by using the interactive software version of the cost model enclosed in the pocket attached to the inside of the back cover of this volume (Appendix D[2]).

[2] This contains one diskette holding an interactive spreadsheet (file name: smcosts.xls), written in Excel 97, together with user instructions (file name: readme.txt).

C.3.1 The effects of intelligent software agent efficiency

The costing model standard assumption is that 75% of DDOs are selected automatically using an intelligent software agent, whilst the remaining 25% are selected manually. If the efficiency of the intelligent software agent is increased to 90%, the cost of preserving DDOs reduces from £1.92 for 75% efficiency, to £1.72 for 90% efficiency (see Figures C.1 and C.2). If the efficiency of the intelligent software agent is lowered to 50%, then the total cost at year ten increases to £2.26. Determining the optimum efficiency of an intelligent software agent performing retrieval or selection may involve cost benefit analysis if the costs of improving the agents are very high. However, since these costs can be spread over many thousands of selected DDOs, improving the intelligent software agent is highly likely to have a beneficial effect. Such refinement may involve co-operating with other repositories on intelligent software agent development. This could lead to the possibility of dividing tasks between agents based on their skills and pooling the results, concepts which are said to be in a development phase (Zarnekow, 1998).

C.3.2 Ten year projection based on the percentage of DDOs selected from public global networks

The standard assumption is set at 30% of DDOs selected from a public global network, such as the Internet, if this is doubled to 60% or halved to 30% there is little appreciable difference between the total costs at the end of year ten. (see Figure C.3). However, if the composition of costs for year one, with a 60% selection rate is examined selection costs are higher, £8.42, as opposed to £7.57, due to additional costs incurred from selecting material from the Internet as little peer-review of materials may have already been undertaken by stakeholders that a repository regards as being of 'recognised' standing.

C.3.3 Effect of labour rate

The standard assumption for labour rate is £21.60 per hour, or an annual salary of £30,000. This annual salary includes 40% to cover repository overheads. The projected labour costs have been factored by the assumed general inflation rate of three percent. Labour costs account for the largest proportion of the selection and set up costs (see Table C.3).

If the labour rate is increased by 33% to £28.80 per hour (see Figure C.4), this increases the overall costs by 16.1% at the end of year ten. Conversely, when the labour rate is reduced by 33% to £14.40 per hour the overall costs reduce by 16.1% at the end of year ten (see Table C.4). This shows that the costing model is sensitive to rises in labour rates. It is critical to offset potential increases by minimising time taken to perform manual processes and increasing the efficiency and effectiveness of computerised components of the framework.

Table C.3 Itemised costs per DDO

Codes & assumptions see below

Item no.	Description of item	Codes			Cost/ min	No. mins	Total cost
1*	Create and attach metadata	c	rpi	t	0.36	15	5.40
2*	Application for unique ID (c=10%, r=10%,p=80%)	cpr	rpi	t	0.36	5	1.80
3*	Peer review	p	rpi	l	0.36	40	14.40
4*	Reviewed (ISP)	s	rpi	l	0.36	5	1.80
5*	Legal deposit	ps	rpi	t	10.00		10.00
6a*	Intelligent agent (retrieval)	r	hi	l	2.83		2.83
6b*	Intelligent agent (selection)	r	hi	l	2.83		2.83
6c*	Manual selection	r	rpi	l	0.36	30	10.80
6d*	Periodic refine intelligent agent	r	rpi	l	0.36	0.10	0.04
6e*	Periodic refine manual	r	rpi	l	0.36	0.20	0.07
7	Digital PLR	r	rpi	a	2.07		2.07
8*	Snapshot frequency	r	rpi	l	0.36	5	1.80
9a*	Negotiate access (100% cost to each)	crsp	rpi	t	0.36	5	1.80
9b	Apply access restrictions	r	hi	a	0.24		0.24
10a*	Digital time-stamping hash	r	hi	t	0.20		0.20
10b*	Certificate of authenticity (post/electronic mail)	r	rpi	t	0.50		0.50
11a*	Check against metadata registry	r	hi	t	0.20		0.20
11b*	Consult creator unkown metadata (100% each)	cr	rpi	t	0.36	15	5.40
11c*	Addition of unknown metadata	r	rpi	t	0.36	15	5.40
11d	Maintenance of metadata registry	r	hi	o	0.02		0.02
12*	Attach unique ID to link with super-metadata	r	hi	t	0.01		0.01
13a*	Set life span	r	rpi	l	0.36	5	1.80
13b	Auto. count. to record use rate for life span	r	hi	l	0.01		0.01
14a*	Automatic generation of super-metadata	r	hi	t	0.05		0.05
14b*	Manual add. of info. to super-metadata c + r	cr	rpi	t	0.36	1	0.36
15a	Maintenance of web site	r	rpi	a	0.36	0.001	0.00
15b	Pro-active research for potential DDOs	r	rpi	l	0.36	0.025	0.01
16*	Recording disposal decisions	r	rpi	t	0.36	2	0.72
17	Access costs (assume 100% r)	ru	hi	a	2.36		2.36
18	Storage costs (YOBP)	r	hi	o	1.08		1.08
19*	Where data not online convert to online	r	rpi	t	0.36	0.5	0.18
20*	Consult owner & complete bib. details c + r,s or p	crsp	rpi	t	0.36	10	3.60
21*	Check if DDO already exists in repository	r	hi	t	0.24		0.24

Labour rate (£/min)		0.36
Cost of accessing DDO (YOBP)	Item 17	2.36
No. of DDOs aquired p.a. (YOBP)	Items 6d,e, 15a,b,	200,000
Digital PLR	Item 7	2.07
0.1 x cost of accessing DDO	Item 21	0.24
Access restrict. = ten percent access cost	Item 9b	0.24

Legend

* = one off cost (unstarred items annual costs)

c = creator; p = publisher; s = ISP; r = repository; u = end-user

rpi = general inflation; hi = computer inflation

One-off costs: l = selection; t = set-up

Annual costs: o = storage; a = access

Table C.3 continued
Repository costs for year 1 linked to computer inflation

Item no.	Description of item	Codes			Cost	Factor	Total cost
1*	Attach metadata	c	rpi	t	5.40		0.00
2*	Application for unique ID (c=10%, r=10%,p=80%)	cpr	rpi	t	1.80		0.00
3*	Peer review	p	rpi	l	14.40		0.00
4*	Reviewed (ISP)	s	rpi	l	1.80		0.00
5*	Legal deposit	ps	rpi	t	10.00		0.00
6a*	Intelligent agent (retrieval)	r	hi	l	2.83	0.3	0.85
6b*	Intelligent agent (selection)	r	hi	l	2.83	0.75	2.12
6c*	Manual selection	r	rpi	l	10.80		0.00
6d*	Periodic refine intelligent agent	r	rpi	l	0.04		0.00
6e*	Periodic refine manual	r	rpi	l	0.07		0.00
7	Digital PLR	r	rpi	a	2.07		0.00
8*	Snapshot frequency	r	rpi	l	1.80		0.00
9a*	Negotiate access (100% cost to each)	crsp	rpi	t	1.80		0.00
9b	Apply access restrictions	r	hi	a	0.24	0.2	0.05
10a*	Digital time-stamping hash	r	hi	t	0.20	1	0.20
10b*	Certificate of authenticity (post/electronic mail)	r	rpi	t	0.50		0.00
11a*	Check against metadata registry	r	hi	t	0.20	1	0.20
11b*	Consult creator unkown metadata (100% each)	cr	rpi	t	5.40		0.00
11c*	Addition of unknown metadata	r	rpi	t	5.40		0.00
11d	Maintenance of metadata registry	r	hi	o	0.02	1	0.02
12*	Attach unique ID to link with super-metadata	r	hi	t	0.01	1	0.01
13a*	Set life span	r	rpi	l	1.80		0.00
13b	Auto. count. to record use rate for life span	r	hi	l	0.01	1	0.01
14a*	Automatic generation of super-metadata	r	hi	t	0.05	1	0.05
14b*	Manual add. of info. to super-metadata c + r	cr	rpi	t	0.36		0.00
15a	Maintenance of web site	r	rpi	a	0.00		0.00
15b	Pro-active research for potential DDOs	r	rpi	l	0.01		0.00
16*	Recording disposal decisions	r	rpi	t	0.72		0.00
17	Access costs (assume 100% r)	ru	hi	a	2.36	0.2	0.47
18	Storage costs (YOBP)	r	hi	o	1.08	1	1.08
19*	Where data not online convert to online	r	rpi	t	0.18		0.00
20*	Consult owner & complete bib. details c + r,s or p	crsp	rpi	t	3.60		0.00
21*	Check if DDO already exists in repository	r	hi	t	0.24	1	0.24
					Annual + one off total cost =		5.30
					Annual cost total =		1.63
	Total at end of year 1 general + computer costs =						13.77
	Note: for PLR read digital compensation scheme						
	Note: for DOI read unique identifier						
	Assume period over which hardware & software halves in cost (yrs)						5
	Implied hardware & software cost inflation (%)						-12.94

Table C.3 continued Repository costs for year 1 linked to general inflation

Item no.	Description of item	Codes			Cost	Factor	Total cost
1*	Attach metadata	c	rpi	t	5.40		0.00
2*	Application for unique ID (c=10%, r=10%,p=80%)	cpr	rpi	t	1.80	0.1	0.18
3*	Peer review	p	rpi	l	14.40		0.00
4*	Reviewed (ISP)	s	rpi	l	1.80		0.00
5*	Legal deposit	ps	rpi	t	10.00		0.00
6a*	Intelligent agent (retrieval)	r	hi	l	2.83		0.00
6b*	Intelligent agent (selection)	r	hi	l	2.83		0.00
6c*	Manual selection	r	rpi	l	10.80	0.25	2.70
6d*	Periodic refine intelligent agent	r	rpi	l	0.04	0.75	0.03
6e*	Periodic refine manual	r	rpi	l	0.07	0.25	0.02
7	Digital PLR (assume applies to 65% of DDOs)	r	rpi	a	2.07	0.13	0.27
8*	Snapshot frequency	r	rpi	l	1.80	0.02	0.04
9a*	Negotiate access (100% cost to each)	crsp	rpi	t	1.80	1	1.80
9b	Apply access restrictions	r	hi	a	0.24		0.00
10a*	Digital time-stamping hash	r	hi	t	0.20		0.00
10b*	Certificate of authenticity (post/electronic mail)	r	rpi	t	0.50	1	0.50
11a*	Check against metadata registry	r	hi	t	0.20		0.00
11b*	Consult creator unkown metadata (100% each)	cr	rpi	t	5.40	0.05	0.27
11c*	Addition of unknown metadata	r	rpi	t	5.40	0.05	0.27
11d	Maintenance of metadata registry	r	hi	o	0.02		0.00
12*	Attach unique ID to link with super-metadata	r	hi	t	0.01		0.00
13a*	Set life span	r	rpi	l	1.80	1	1.80
13b	Auto. count. to record use rate for life span	r	hi	l	0.01		0.00
14a*	Automatic generation of super-metadata	r	hi	t	0.05		0.00
14b*	Manual add. of info. to super-metadata c + r	cr	rpi	t	0.36	1	0.36
15a	Maintenance of web site	r	rpi	a	0.00	1	0.00
15b	Pro-active research for potential DDOs	r	rpi	l	0.01	1	0.01
16*	Recording disposal decisions	r	rpi	t	0.72	0.05	0.04
17	Access costs (assume 100% r)	ru	hi	a	2.36		0.00
18	Storage costs (YOBP)	r	hi	o	1.08		0.00
19*	Where data not online convert to online	r	rpi	t	0.18	0.1	0.02
20*	Consult owner & complete bib. details c + r,s or p	crsp	rpi	t	3.60	0.05	0.18
21*	Check if DDO already exists in repository	r	hi	t	0.24		0.00
					Annual + one off total cost =		8.47

Percentage of DDOs disposed of per year	Item 16	5.00
X = frac. DDOs accessed (YOBP)	Items 7, 9b	0.2
Percent of DDOs selected automatically	Items 6b, d	75
Percent of DDOs selected manually	Items 6c, e	25
Percentage of material from off-line source	Item 19	10
Percent deposit with insuffic. bib. details	Item 20	5
Percent living DDOs	Item 8	2
Percent unknown metadata	Items 11b, c (halves each year)	5
Assumed general inflation rate (%)		3
Assume period over which hardware & software halves in cost (yrs)		5
Total cost less 11b, c		7.93
Items 11b + 11c		0.54

Table C.4 Sensitivity of the total cost per DDO after ten years to each of the identified cost variables

Cost variable	Cost variable value	Cost per DDO in £ after ten years	Percentage difference from standard assumptions
Intelligent software agent efficiency	50%	2.26	17.7
	75%	1.92	0
	90%	1.72	-10.4
Labour rate £ per hour	14.40	1.61	-16.1
	21.60	1.92	0
	28.80	2.23	16.1
Fraction of DDOs accessed	10%	1.67	-13.0
	20%	1.92	0
Percentage of DDOs from the Internet	10%	1.91	-0.5
	30%	1.92	0
	60%	1.95	1.6
Manual selection time	doubled	2.52	31.2*

*If the number of DDOs selected decreases sufficiently, the reduced storage cost will offset the increased selection costs.

C.3.4 Effect of access rate

The costing model assumes an access rate of 20% that concurs with the estimates of digital archive access costs quoted in the Yale Project Open Book. By decreasing the access rate to 10%, overall costs reduce for each year of the ten year projection (see Figure C.5). As the size of a repository collection grows over the years, it is possible that the access rate may reduce due to the growing age of the DDOs. Hence, it was considered reasonable to consider the effect of halving the access rate. This shows that access costs increase in direct proportion to the access rate.

If the breakdown of access costs is analysed at the end of year one (see Figure C.6), the largest percentage of costs (60%), are document server and software costs. This cost is expected to fall by the standard software/hardware inflation rate per annum.

The second largest percentage is the compensation payment made to owners of DDOs (34%). For the purposes of this model, the compensation rate is set at an average of £2.00 per annum per DDO. This figure originates from the Public Lending Right (PLR) figure of two pence, per loan of a paper-based book. It is assumed that each accessed DDO is looked at 100 times within the course of a year[3]. This amount is then assumed to apply to 65% of deposited DDOs (the percentage assumed in this model to originate from traditional publishers) factored by the 20% rate of access to DDOs[4].

C.3.5 Effect of varying the labour time devoted to selection

As manual selection accounts for 61% of the selection costs, which is split into time spent on determining life span and evaluating the information content of DDOs, compared with 39% for automated selection (see Figure C.7), it was felt worthwhile to explore the effect of altering the manual selection rate. This consisted of first doubling then halving the time allocated to items 6c,d, e, 8 and 13a in Table C.3. If the manual selection rate is doubled, then the total cost for long-term preservation at the end of year ten increases from £1.92 to £2.52 (see Figure C.9), a percentage increase of 31.2%. Conversely, if the time spent on manual selection is halved, the overall cost reduces by 31.2%.

C.3.6 Breakdown of set up costs incurred at year one

Figure C.8 shows that under the super-metadata framework set up costs can be sub-divided into eleven significant parts. It should be noted that costs of less than one percent are not considered great enough to be graphed successfully, but nevertheless are taken into consideration in the following analysis. This relates to the attachment of a unique identifier, automatic generation of super-metadata, conversion of 10% of DDOs from offline to an online source. It can be seen that where standard

---

[3] URL: http://www.earl.org.uk/partners/plr/payment.html

[4] A re-evaluation of compensation rates for owners of deposited objects is advocated as the anticipated inclusion of commercial databases could place a new perspective on the potential loss of revenue.

assumptions are applied, costs that involve computerised processes constitute a small proportion of the total set up costs (32%). The largest proportion of set up costs, are attributed to human interaction with other stakeholders to obtain details and access permissions for DDOs (52%). As with all the costs given in the cost model they can only be used as approximate indicators. Implementations of the proposed super-metadata framework may show different patterns resulting from repositories giving greater emphasis to certain components, or optional components such as time-stamping being substituted by other authentication systems, such as digital signatures, or omitted altogether. As the largest proportion of set up costs are linked to general inflation and are consequently expected to rise as time increases, there is a need for careful management of human resources. The only set up cost linked to general inflation that is not included in Figure C.8 is the cost of converting DDOs to online format, but since this relates to 10% (a percentage that is expected to decrease over time) this has not been included in the pie chart. Similarly, the small costs incurred per DDO for automated processes to attach a unique identifier (where one does not already exist) and automatically generate the super-metadata are not graphed, but do add small amounts to the total set up costs. Repository time spent on developing the efficiency of computerised processes would similarly aid the reduction in manual costs.

C.3.7 Itemised breakdown of costs for creators, ISPs, publishers, repositories and end-users

The possible range of costs incurred by creators

Table C.5 shows that creator could incur a cost of around £18 for each preserved DDO, assuming that the creator contributes to all the stated items. If the DDO requires minimal metadata and this is attached with little effort at creation the costs involved in consulting with a repository or a publisher at a later stage would be minimised. Similarly, if the creator does not wish to recoup revenue from a DDO (does not apply for a unique identifier to establish ownership) and has also declared

the access restrictions that apply to the DDO before it goes into a repository, this can reduce a creators costs to possibly only £5 per DDO.

Table C.5 Likely costs incurred by creators

Item no.	Description of item	Codes	Cost/min	No. mins	Total cost
1*	Create and attach metadata	c rpi t	0.36	15	5.40
(2*)	Application for a unique identifier	cpr rpi t	0.36	5	1.80
(9a*)	Negotiate access	crsp rpi t	0.36	5	1.80
(11b*)	Consult creator unknown metadata	cr rpi t	0.36	15	5.40
(14b*)	Manual add. of info. to super-metadata c + r	cr rpi t	0.36	1	0.36
(20*)	Consult owner & complete bib. details c + r,s or p	crsp rpi t	0.36 Total cost (max.) =	10	3.60 £18.36
			Total cost (min.) =		£5.40

See key to Tables C.5 to C.9 below.

Key to Tables C.5 to C.9

Codes and assumptions	
Labour rate (£/min)	0.36
Cost of accessing DDO (Y) (£/DDO) (Item 17)	2.36
No. of DDOs acquired p.a. (Y) (Items 6d & e, 15 a & b)	200,000
Digital PLR (£/DDO) (Item 7)	2.07
0.1 x cost of accessing DDO (£/DDO) (Item 21)	0.24
Access restrict. = ten percent access cost (Item 9b)	0.24
* = one off cost (unstarred items annual costs)	
c = creator; p = publisher; s =ISP; r = repository; u = end-user	
rpi = general inflation; hi = computer inflation	
One-off costs: l = selection; t = set up	
Annual costs: o = storage; a = access Brackets = costs that do not always apply and which are excluded from the minimised costs.	

The possible range of costs incurred by publishers

Table C.6 shows that if a publisher decides to deposit DDOs with a repository, under a negotiated agreement then the maximum cost could be in the order of £28.00 per DDO. A best case scenario could reduce costs to £10.00 per DDO, or even less if the cost of legal deposit can be reduced. In cases where a third party is managing a publisher's DDOs, a unique identifier has been attached by a creator and negotiation costs not incurred because of the existence of a clearly stated access policy for each class of DDOs. Similarly, peer review and access negotiations may be avoided in routine cases. If publishers decides to develop a digital archive of their own publications the costs would be similar to those incurred by a repository (see Table C.8).

Table C.6 Likely costs incurred by publishers

Item no.	Description of item	Codes	Cost/min	No. mins	Total cost
(2*)	Application for a unique identifier	cpr rpi t	0.36	5	1.80
(3*)	Peer review	p rpi l	0.36	30	10.80
5*	Legal deposit	ps rpi t	10.00		10.00
(9a*)	Negotiate access (100% cost to each)	crsp rpi t	0.36	5	1.80
(20*)	Consult owner & complete bib. details c + r,s or p	crsp rpi t	0.36	10 Total cost (max.) =	3.60 £28.00
				Total cost (min.) =	£10.00

For codes and assumptions see below Table C.5

The possible range of costs incurred by ISPs

The costs incurred by ISPs range from around £7 to £2. The best case scenario is dependent on creators providing sufficient metadata and access rights information, to ensure negotiations are reduced to a minimum. If the review procedures of ISPs become more stringent then the costs incurred will increase to approach similar figure to those incurred by publishers (see Table C.6).

Table C.7 Likely costs incurred by ISPs

Item no.	Description of item	Codes	Cost/min	No. mins	Total cost
4*	Reviewed (ISP)	s rpi l	0.36	5	1.80
(9a*)	Negotiate access	crsp rpi t	0.36	5	1.80
(20*)	Consult owner & complete bib. details c + r, s or p	crsp hi t	0.36	10	3.60
				Total cost (max.) =	£7.20
				Total cost (min.) =	£1.80

For codes and assumptions see below Table C.5.

Range of costs likely to be incurred by repositories

The total costs a repository is likely to incur is detailed in the itemised listing shown in Table C.8 with a total cost of about £14 per DDO for the end of year one. If the italicised items are not incurred by the repository the cost to the repository is reduced by approximately £1 per DDO, only a small fraction of the total. Unlike other stakeholders, costs incurred by the repository can be spread over the life span of the DDO. When one-off costs such as set-up and selection are reduced by spreading them over the length of time a DDO is kept (the model recommends that a DDO be retained for a minimum of ten years for it to be cost effective) then the total cost per DDO reduces to approximately £2 at the end of the tenth year in the repository. It should be noted that £2 is an average cost for long-term preservation and may vary for individual DDOs depending on their source of origin and the nature of the DDO itself. For example, the source of origin may be the Internet, a publisher, or direct from the creator and the nature of the DDO may be static, dynamic or living.

Table C.8 Likely costs incurred by repositories per DDO

Item no.	Description of item	Codes	Cost	No. mins	Factor	Total cost
2*	Application for a unique identifier (c=10%, r=10%,p=80%)	cpr	0.36	5	0.1	0.18
6a*	Intelligent software agent (retrieval)	r	2.83		0.3	0.85
6b*	Intelligent software agent (selection)	r	2.83		0.75	2.12
6c*	Manual selection	r	0.36	30	0.25	2.70
6d*	Periodic refinement of intelligent software agent	r	0.36	0.10	0.75	0.03
6e*	Periodic refinement of manual selection process	r	0.36	0.20	0.25	0.02
7	Digital Public Lending Right (rate per access)	r	2.10		0.13	0.27
8*	Snapshot frequency	r	0.36	5	0.02	0.04
9a*	Negotiate access (100% cost to each stakeholder involved)	crsp	0.36	5	1	1.80
9b	Apply access restrictions	r	0.24		0.2	0.05
10a*	Digital time-stamping hash	r	0.20		1	0.20
10b*	Certificate of authenticity (post/electronic mail)	r	0.50		1	0.50
11a*	Check against metadata registry	r	0.20		1	0.20
11b*	Consult creator unknown metadata (100% each stakeholder involved)	cr	0.36	15	0.05	0.27
11c*	Addition of unknown metadata	rpi	0.36	15	0.05	0.27
11d	Maintenance of metadata registry	r	0.02		1	0.02
12*	Attach unique ID to link with super-metadata	r	0.01		1	0.01
13a*	Set life span	r	0.36	5	1	1.80
13b	Auto. count. to record use rate for life span	r	0.01		1	0.01
14a*	Automatic generation of super-metadata	r	0.05		1	0.05
14b*	Manual add. of info. to super-metadata c + r	cr	0.36	1	1	0.36
15a	Maintenance of Web site	r	0.36	0.001	1	0.00
15b	Proactive research for potential DDOs	r	0.36	0.025	1	0.01
16*	Recording disposal decisions	r	0.36	2	0.05	0.04
17	*Access costs (best case scenario*	*ru*	*2.40*		*0.20*	*0.48*

	assumes 100% incurred by user)					
18	Storage costs	r	1.10		1	1.10
19*	Where data not online convert to online	r	0.36	0.5	0.1	0.02
*20**	*Consult owner & complete bib. details c + r, s or p*	*crsp*	*0.36*	*10*	*0.05*	*0.18*
21*	Check if DDO already exists in repository	r	0.24		1	0.24
*22**	*Printing cost (best case scenario assumes 100% incurred by user)*	*ru*	*0.70*		*0.20*	*0.14*
				Total cost	(max.)	£13.96
				Total cost	(min.)	£13.16

For codes and assumptions see below Table C.5

The possible range of costs incurred by end-users

Table C.9 Likely costs incurred by end-users

Item no.	Description of item	Codes	Cost/min	No. mins	Total cost
17	Access costs	ru hi a	2.36		2.36
22*	Printing cost	ru hi a	0.70		0.70
					Total cost (max.) = £3.06 Total cost (min.) = £0.00

For codes and assumptions see below Table C.5

Table C.9 shows that the costs incurred by an end-user vary from no charge being incurred, to one of approximately £3 per DDO if point of need charges were made for both the printing and access.

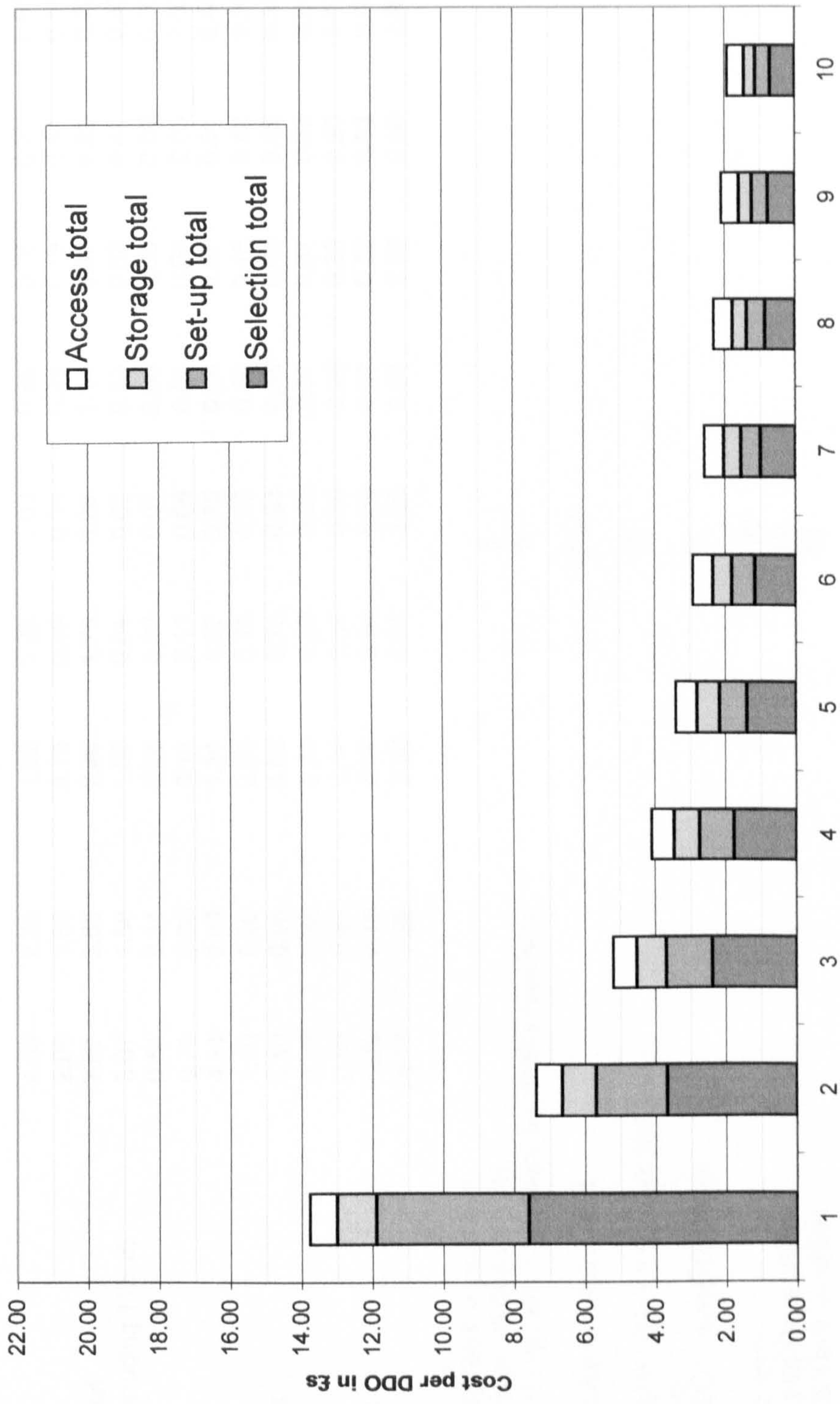


Figure C.1 Ten year projection based on standard assumptions

Table C.10 Costs in Figure C.1

End of year	1	2	3	4	5	6	7	8	9	10
Selection rpi	4.59	2.36	1.62	1.25	1.03	0.89	0.78	0.71	0.65	0.60
Selection hi	2.98	1.30	0.75	0.49	0.34	0.25	0.19	0.14	0.11	0.09
Selection total	**7.57**	**3.66**	**2.38**	**1.75**	**1.38**	**1.14**	**0.97**	**0.85**	**0.76**	**0.68**
Set-up rpi (excluding 11b, c)	3.07	1.58	1.09	0.84	0.69	0.59	0.52	0.47	0.43	0.40
11b, c	0.54	0.14	0.05	0.02	0.01	0.00	0.00	0.00	0.00	0.00
Set-up hi	0.70	0.30	0.18	0.11	0.08	0.06	0.04	0.03	0.03	0.02
Set-up total	**4.31**	**2.03**	**1.31**	**0.97**	**0.78**	**0.66**	**0.57**	**0.51**	**0.46**	**0.42**
Storage hi	1.10	0.96	0.83	0.73	0.63	0.55	0.48	0.42	0.36	0.32
Storage total	**1.10**	**0.96**	**0.83**	**0.73**	**0.63**	**0.55**	**0.48**	**0.42**	**0.36**	**0.32**
Access rpi	0.27	0.28	0.29	0.29	0.30	0.31	0.32	0.33	0.34	0.35
Access hi	0.52	0.45	0.39	0.34	0.30	0.26	0.23	0.20	0.17	0.15
Access total	**0.79**	**0.73**	**0.68**	**0.64**	**0.60**	**0.57**	**0.55**	**0.53**	**0.51**	**0.50**
Grand total	13.77	7.38	5.20	4.08	3.39	2.91	2.56	2.30	2.09	1.92

Standard labour time on selection	
Assumed general inflation rate (%)	3
Assume period over which hardware & software halves in cost (yrs)	5
Implied hardware & software cost inflation (%)	-12.94
Percent of DDOs selected by intelligent software agent	75
Percent of DDOs selected from the Internet	30
Labour rate (£/min)	0.36
X = fraction of DDOs accessed	0.2
Percentage DDOs from offline sources	10
Assumed interest rate (%)	7
Assumed amortization period (yrs)	5

Figure C.2 Ten year projection based on 90% intelligent software agent efficiency

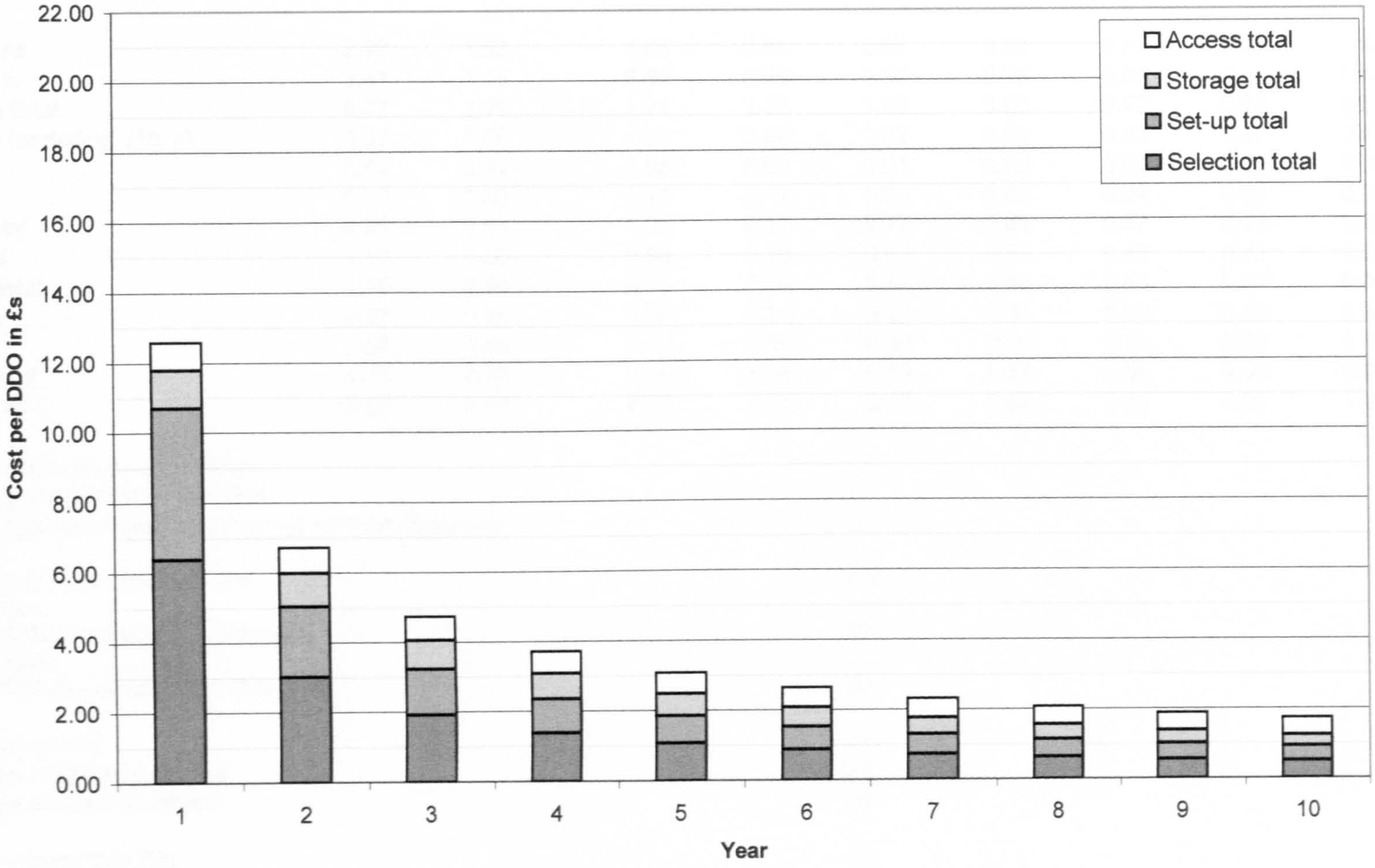


Table C.11 Costs in Figure C.2

End of year	1	2	3	4	5	6	7	8	9	10
Selection rpi	2.96	1.53	1.05	0.81	0.67	0.57	0.51	0.46	0.42	0.39
Selection hi	3.41	1.48	0.86	0.56	0.39	0.28	0.21	0.16	0.12	0.10
Selection total	**6.37**	**3.01**	**1.91**	**1.37**	**1.06**	**0.86**	**0.72**	**0.62**	**0.54**	**0.48**
Set-up rpi (excluding 11b, c)	3.07	1.58	1.09	0.84	0.69	0.59	0.52	0.47	0.43	0.40
11b, c	0.54	0.14	0.05	0.02	0.01	0.00	0.00	0.00	0.00	0.00
Set-up hi	0.70	0.30	0.18	0.11	0.08	0.06	0.04	0.03	0.03	0.02
Set-up total	**4.31**	**2.03**	**1.31**	**0.97**	**0.78**	**0.66**	**0.57**	**0.51**	**0.46**	**0.42**
Storage hi	1.10	0.96	0.83	0.73	0.63	0.55	0.48	0.42	0.36	0.32
Storage total	**1.10**	**0.96**	**0.83**	**0.73**	**0.63**	**0.55**	**0.48**	**0.42**	**0.36**	**0.32**
Access rpi	0.27	0.28	0.29	0.29	0.30	0.31	0.32	0.33	0.34	0.35
Access hi	0.52	0.45	0.39	0.34	0.30	0.26	0.23	0.20	0.17	0.15
Access total	**0.79**	**0.73**	**0.68**	**0.64**	**0.60**	**0.57**	**0.55**	**0.53**	**0.51**	**0.50**
Grand total	12.57	6.72	4.73	3.71	3.07	2.63	2.31	2.07	1.88	1.72

Standard labour time on selection	
Assumed general inflation rate (%)	3
Assume period over which hardware & software halves in cost (yrs)	5
Implied hardware & software cost inflation (%)	-12.94
Percent of DDOs selected by intelligent software agent	90
Percent of DDOs selected from the Internet	30
Labour rate (£/min)	0.36
X = fraction of DDOs accessed	0.2
Percentage DDOs from offline sources	10
Assumed interest rate (%)	7
Assumed amortization period (yrs)	5

Figure C.3 Ten year projection based on 60% of DDOs selected from the Internet

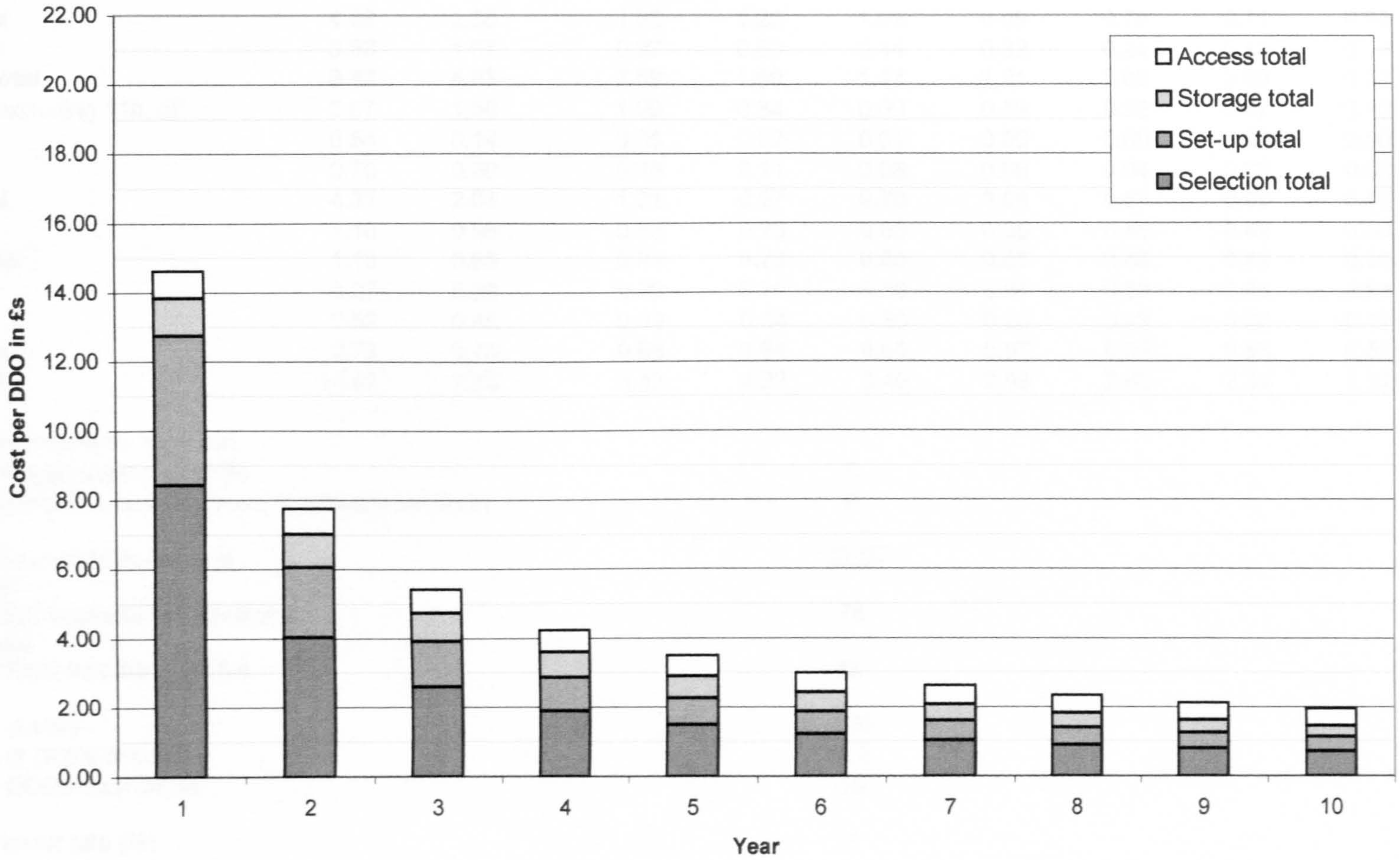


Table C.12 Costs in Figure C.3

End of year	1	2	3	4	5	6	7	8	9	10
Selection rpi	4.59	2.36	1.62	1.25	1.03	0.89	0.78	0.71	0.65	0.60
Selection hi	3.83	1.67	0.97	0.63	0.44	0.32	0.24	0.18	0.14	0.11
Selection total	**8.42**	**4.03**	**2.59**	**1.89**	**1.47**	**1.21**	**1.02**	**0.89**	**0.79**	**0.71**
Set-up rpi (excluding 11b, c)	3.07	1.58	1.09	0.84	0.69	0.59	0.52	0.47	0.43	0.40
11b, c	0.54	0.14	0.05	0.02	0.01	0.00	0.00	0.00	0.00	0.00
Set-up hi	0.70	0.30	0.18	0.11	0.08	0.06	0.04	0.03	0.03	0.02
Set-up total	**4.31**	**2.03**	**1.31**	**0.97**	**0.78**	**0.66**	**0.57**	**0.51**	**0.46**	**0.42**
Storage hi	1.10	0.96	0.83	0.73	0.63	0.55	0.48	0.42	0.36	0.32
Storage total	**1.10**	**0.96**	**0.83**	**0.73**	**0.63**	**0.55**	**0.48**	**0.42**	**0.36**	**0.32**
Access rpi	0.27	0.28	0.29	0.29	0.30	0.31	0.32	0.33	0.34	0.35
Access hi	0.52	0.45	0.39	0.34	0.30	0.26	0.23	0.20	0.17	0.15
Access total	**0.79**	**0.73**	**0.68**	**0.64**	**0.60**	**0.57**	**0.55**	**0.53**	**0.51**	**0.50**
Grand total	14.62	7.75	5.42	4.22	3.49	2.98	2.62	2.34	2.12	1.95

Standard labour time on selection	
Assumed general inflation rate (%)	3
Assume period over which hardware & software halves in cost (yrs)	5
Implied hardware & software cost inflation (%)	-12.94
Percent of DDOs selected by intelligent software agent	75
Percent of DDOs selected from the Internet	60
Labour rate (£/min)	0.36
X = fraction of DDOs accessed	0.2
Percentage DDOs from offline sources	10
Assumed interest rate (%)	7
Assumed amortization period (yrs)	5

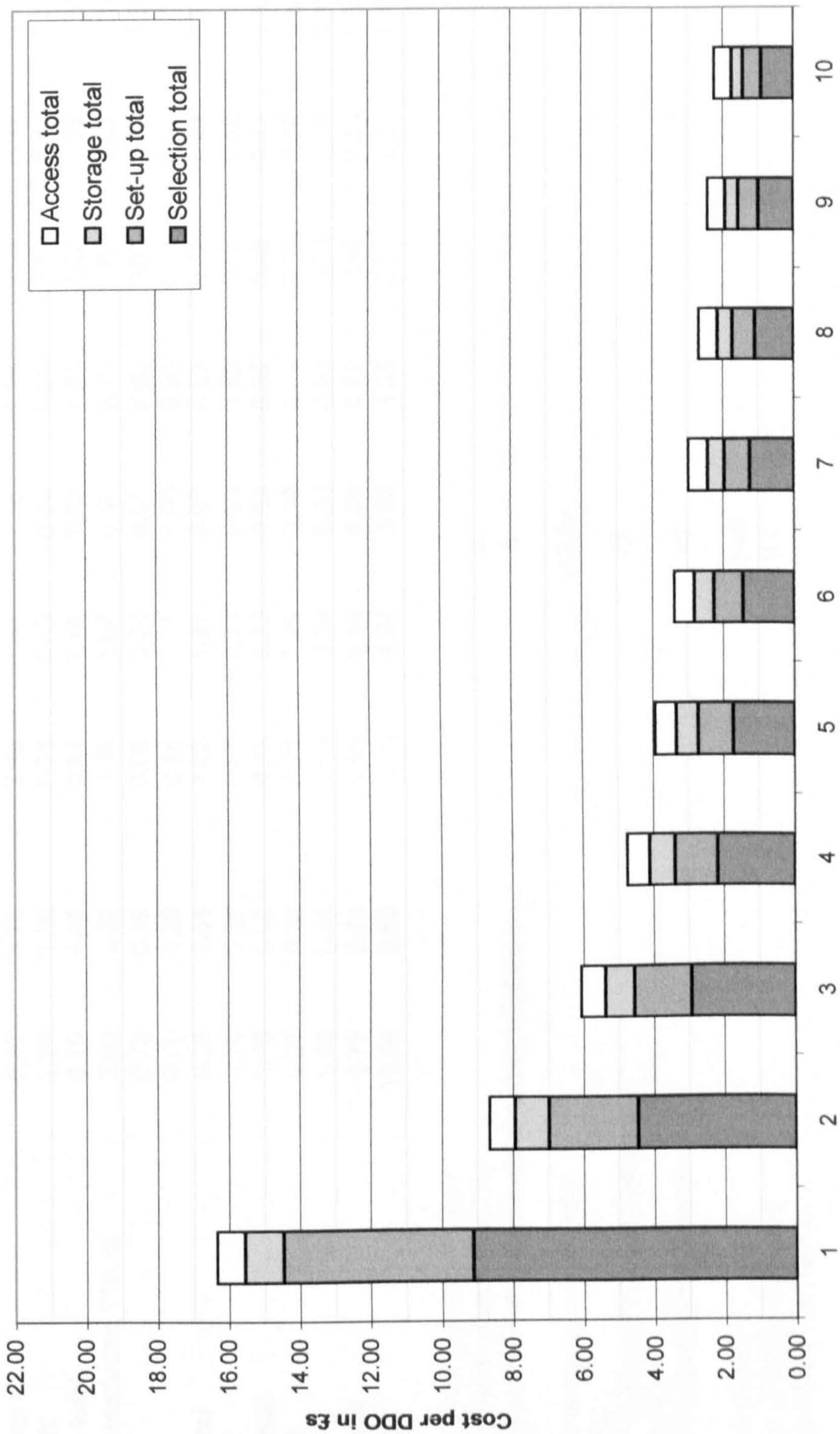


Figure C.4 Ten year projection based on labour rate of £28.80 per hour

Table C.13 Costs in Figure C.4

End of year	1	2	3	4	5	6	7	8	9	10
Selection rpi	6.12	3.15	2.16	1.67	1.38	1.18	1.04	0.94	0.86	0.80
Selection hi	2.98	1.30	0.75	0.49	0.34	0.25	0.19	0.14	0.11	0.09
Selection total	**9.10**	**4.45**	**2.92**	**2.16**	**1.72**	**1.43**	**1.23**	**1.08**	**0.97**	**0.88**
Set-up rpi (excluding 11b, c)	3.93	2.02	1.39	1.07	0.89	0.76	0.67	0.60	0.55	0.51
11b, c	0.72	0.19	0.06	0.02	0.01	0.00	0.00	0.00	0.00	0.00
Set-up hi	0.70	0.30	0.18	0.11	0.08	0.06	0.04	0.03	0.03	0.02
Set-up total	**5.35**	**2.51**	**1.63**	**1.21**	**0.98**	**0.82**	**0.72**	**0.64**	**0.58**	**0.53**
Storage hi	1.10	0.96	0.83	0.73	0.63	0.55	0.48	0.42	0.36	0.32
Storage total	**1.10**	**0.96**	**0.83**	**0.73**	**0.63**	**0.55**	**0.48**	**0.42**	**0.36**	**0.32**
Access rpi	0.27	0.28	0.29	0.29	0.30	0.31	0.32	0.33	0.34	0.35
Access hi	0.52	0.45	0.39	0.34	0.30	0.26	0.23	0.20	0.17	0.15
Access total	**0.79**	**0.73**	**0.68**	**0.64**	**0.60**	**0.57**	**0.55**	**0.53**	**0.51**	**0.50**
Grand total	16.34	8.65	6.06	4.74	3.93	3.38	2.97	2.67	2.43	2.23

Standard labour time on selection	
Assumed general inflation rate (%)	3
Assume period over which hardware & software halves in cost (yrs)	5
Implied hardware & software cost inflation (%)	-12.94
Percent of DDOs selected by intelligent software agent	75
Percent of DDOs selected from the Internet	30
Labour rate (£/min)	0.48
X = fraction of DDOs accessed	0.2
Percentage DDOs from offline sources	10
Assumed interest rate (%)	7
Assumed amortization period (yrs)	5

Figure C.5 Ten year projection based on access rate of 10%

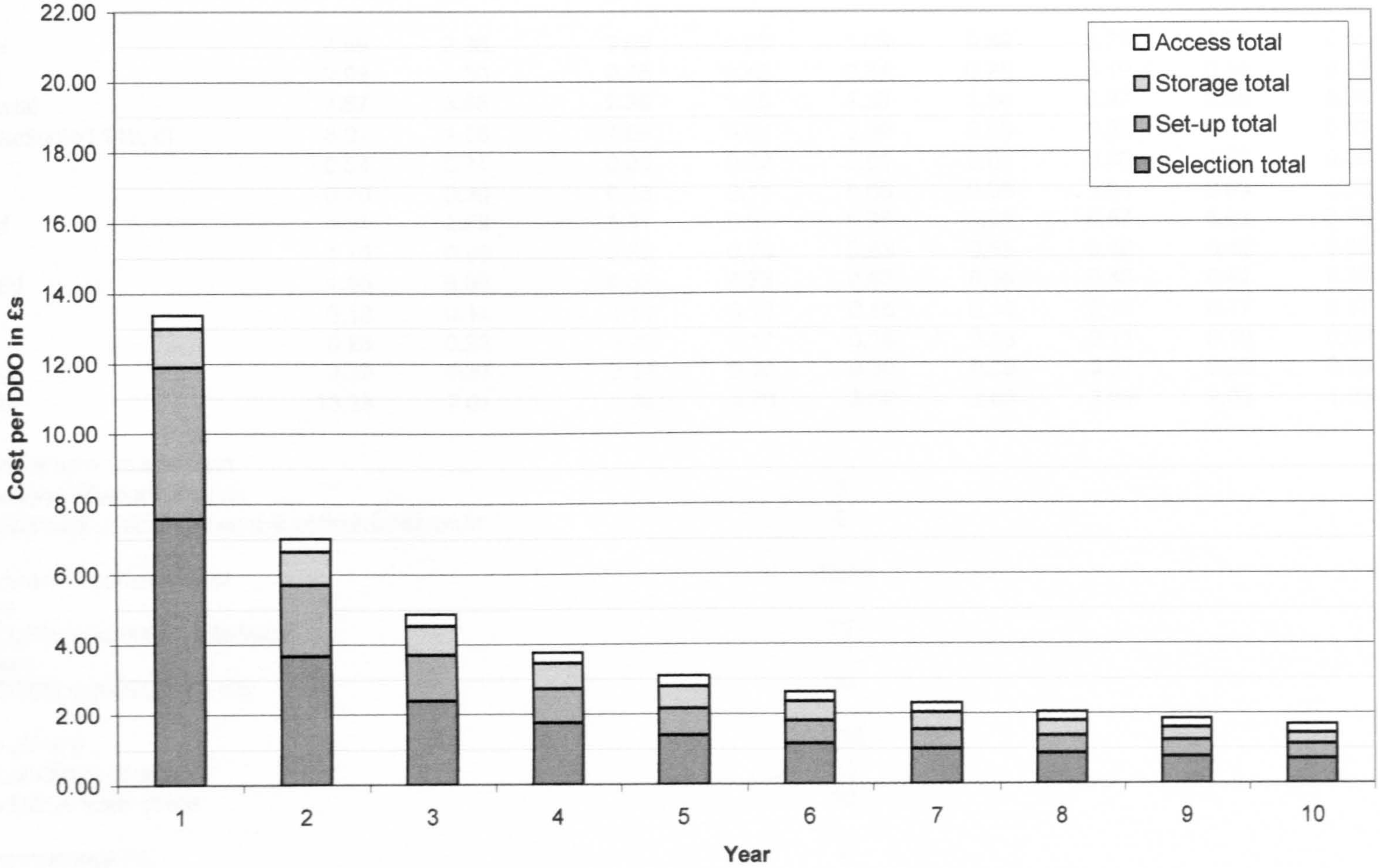


Table C.14 Costs in Figure C.5

End of year	1	2	3	4	5	6	7	8	9	10
Selection rpi	4.59	2.36	1.62	1.25	1.03	0.89	0.78	0.71	0.65	0.60
Selection hi	2.98	1.30	0.75	0.49	0.34	0.25	0.19	0.14	0.11	0.09
Selection total	**7.57**	**3.66**	**2.38**	**1.75**	**1.38**	**1.14**	**0.97**	**0.85**	**0.76**	**0.68**
Set-up rpi (excluding 11b, c)	3.07	1.58	1.09	0.84	0.69	0.59	0.52	0.47	0.43	0.40
11b, c	0.54	0.14	0.05	0.02	0.01	0.00	0.00	0.00	0.00	0.00
Set-up hi	0.70	0.30	0.18	0.11	0.08	0.06	0.04	0.03	0.03	0.02
Set-up total	**4.31**	**2.03**	**1.31**	**0.97**	**0.78**	**0.66**	**0.57**	**0.51**	**0.46**	**0.42**
Storage hi	1.10	0.96	0.83	0.73	0.63	0.55	0.48	0.42	0.36	0.32
Storage total	**1.10**	**0.96**	**0.83**	**0.73**	**0.63**	**0.55**	**0.48**	**0.42**	**0.36**	**0.32**
Access rpi	0.13	0.14	0.14	0.15	0.15	0.16	0.16	0.17	0.17	0.18
Access hi	0.26	0.23	0.20	0.17	0.15	0.13	0.11	0.10	0.09	0.07
Access total	**0.39**	**0.37**	**0.34**	**0.32**	**0.30**	**0.29**	**0.27**	**0.26**	**0.26**	**0.25**
Grand total	13.38	7.01	4.86	3.76	3.09	2.63	2.29	2.03	1.83	1.67

Standard labour time on selection	
Assumed general inflation rate (%)	3
Assume period over which hardware & software halves in cost (yrs)	5
Implied hardware & software cost inflation (%)	-12.94
Percent of DDOs selected by intelligent software agent	75
Percent of DDOs selected from the Internet	30
Labour rate (£/min)	0.36
X = fraction of DDOs accessed	0.1
Percentage DDOs from offline sources	10
Assumed interest rate (%)	7
Assumed amortization period (yrs)	5

Figure C.6 Breakdown of access costs at year one

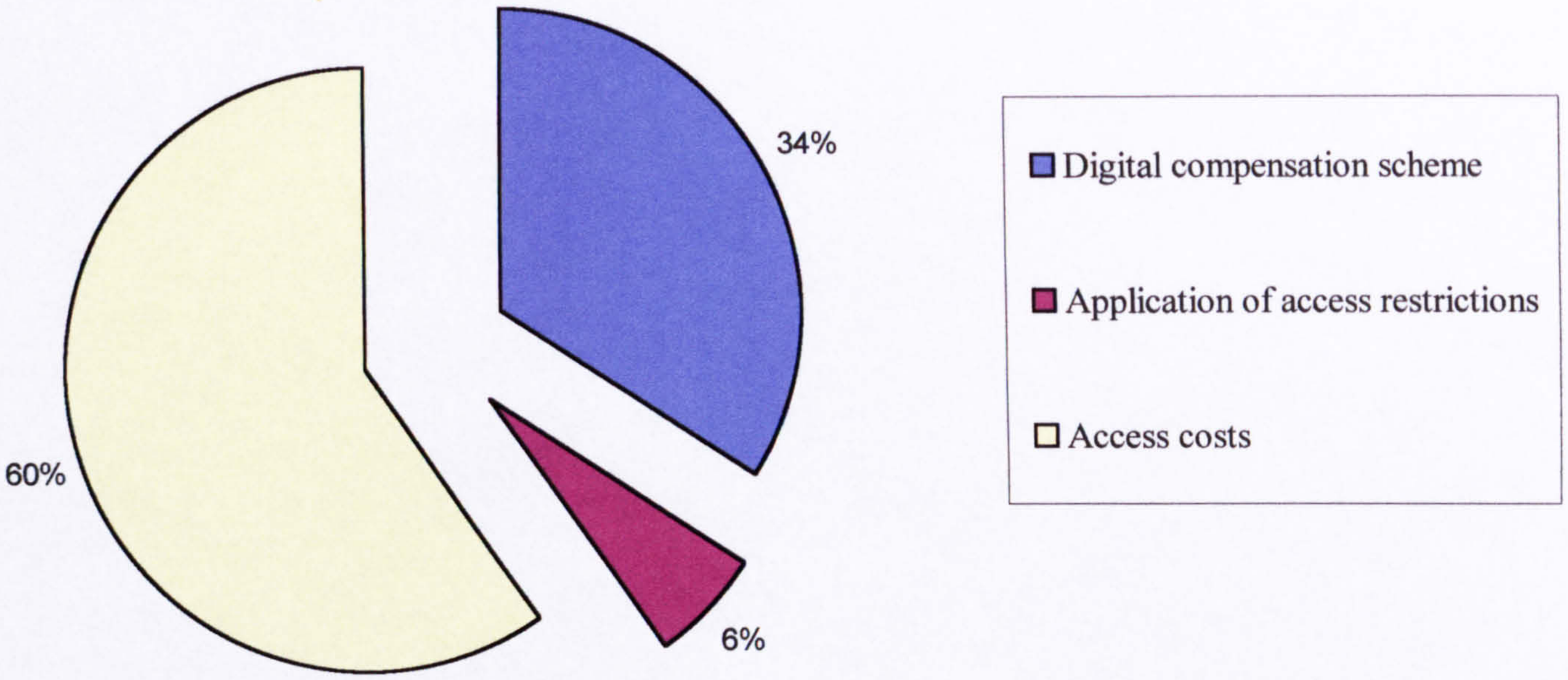


Figure C.7 Breakdown of selection costs at year one

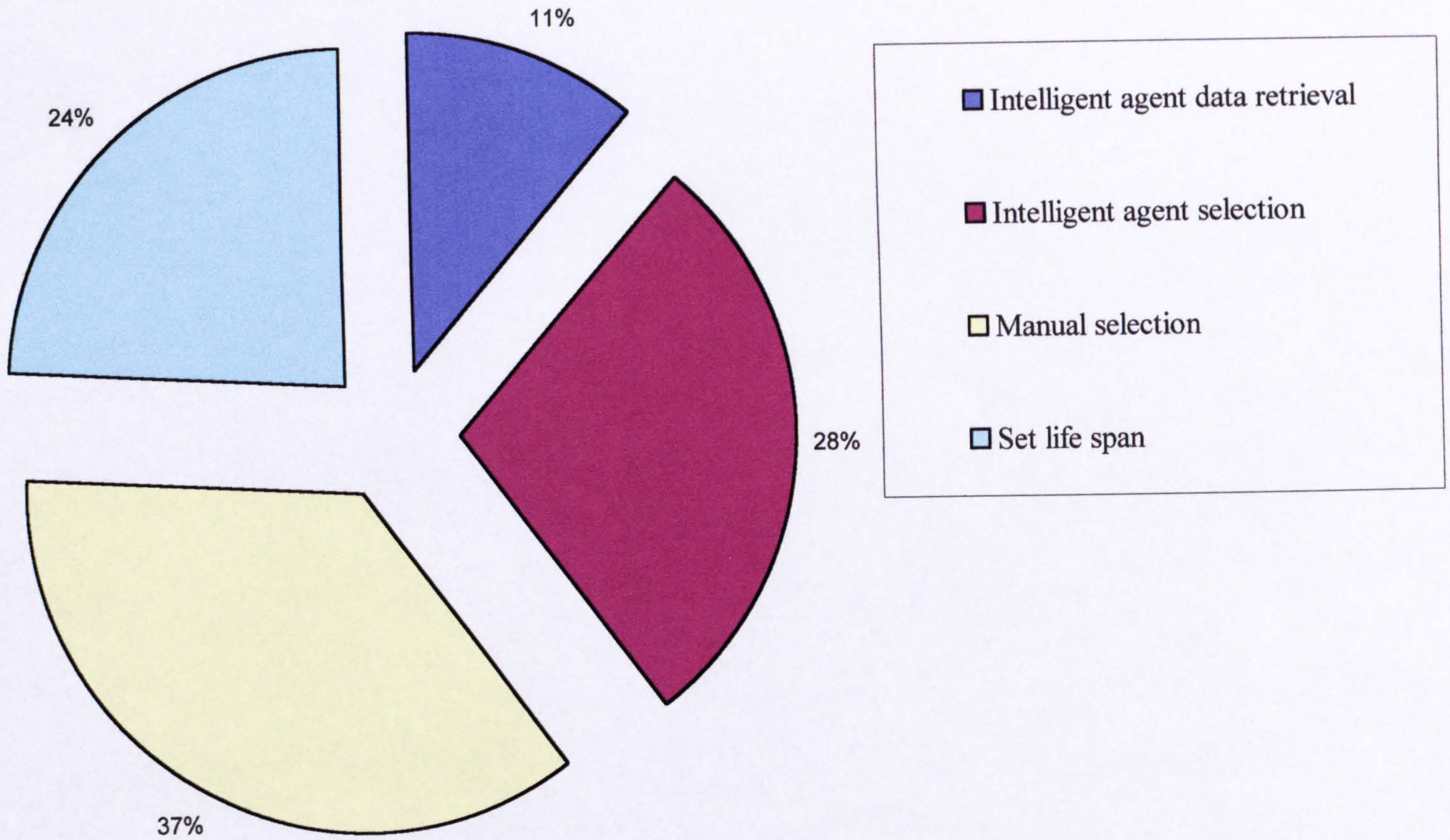
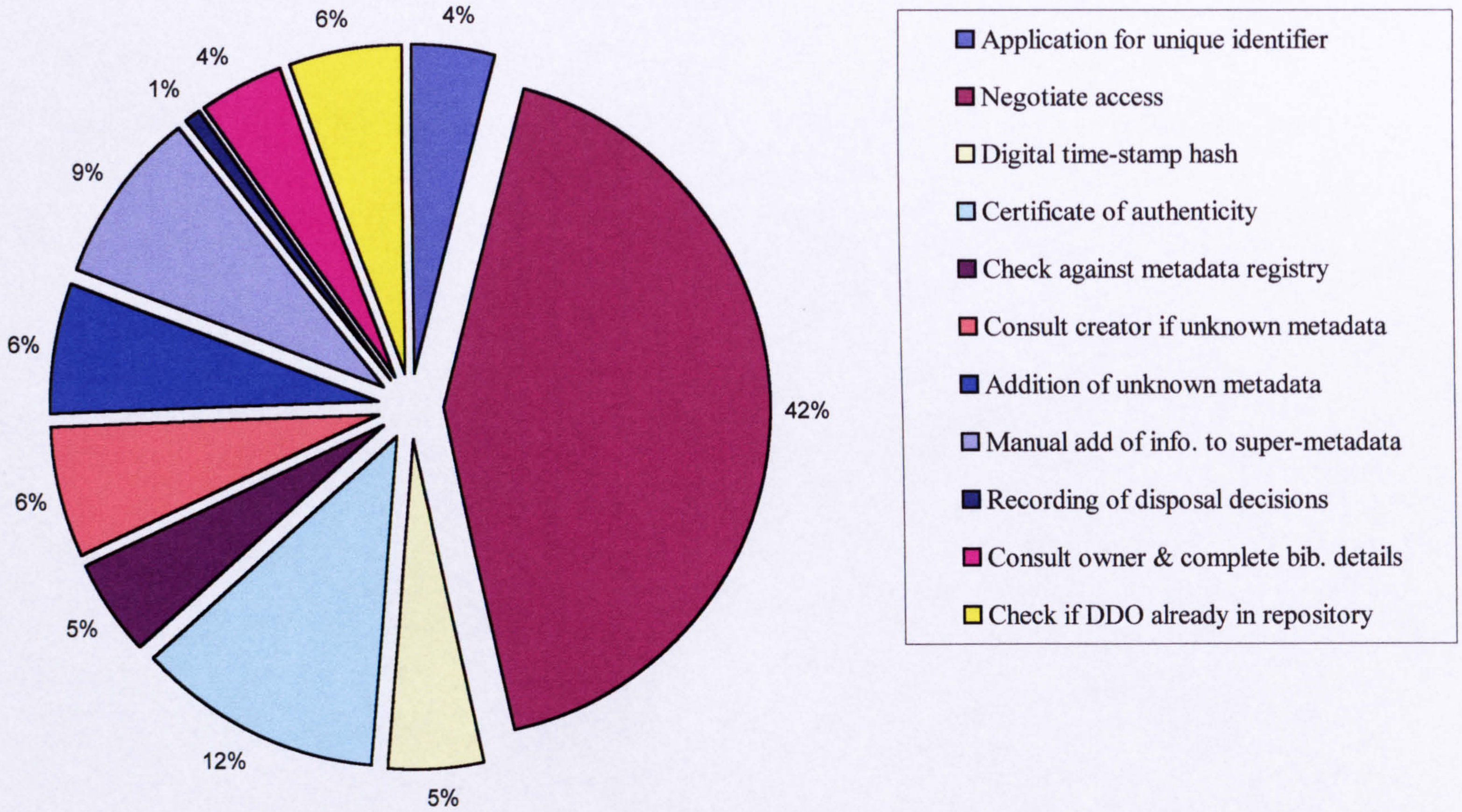


Figure C.8 Breakdown of set-up costs at year one



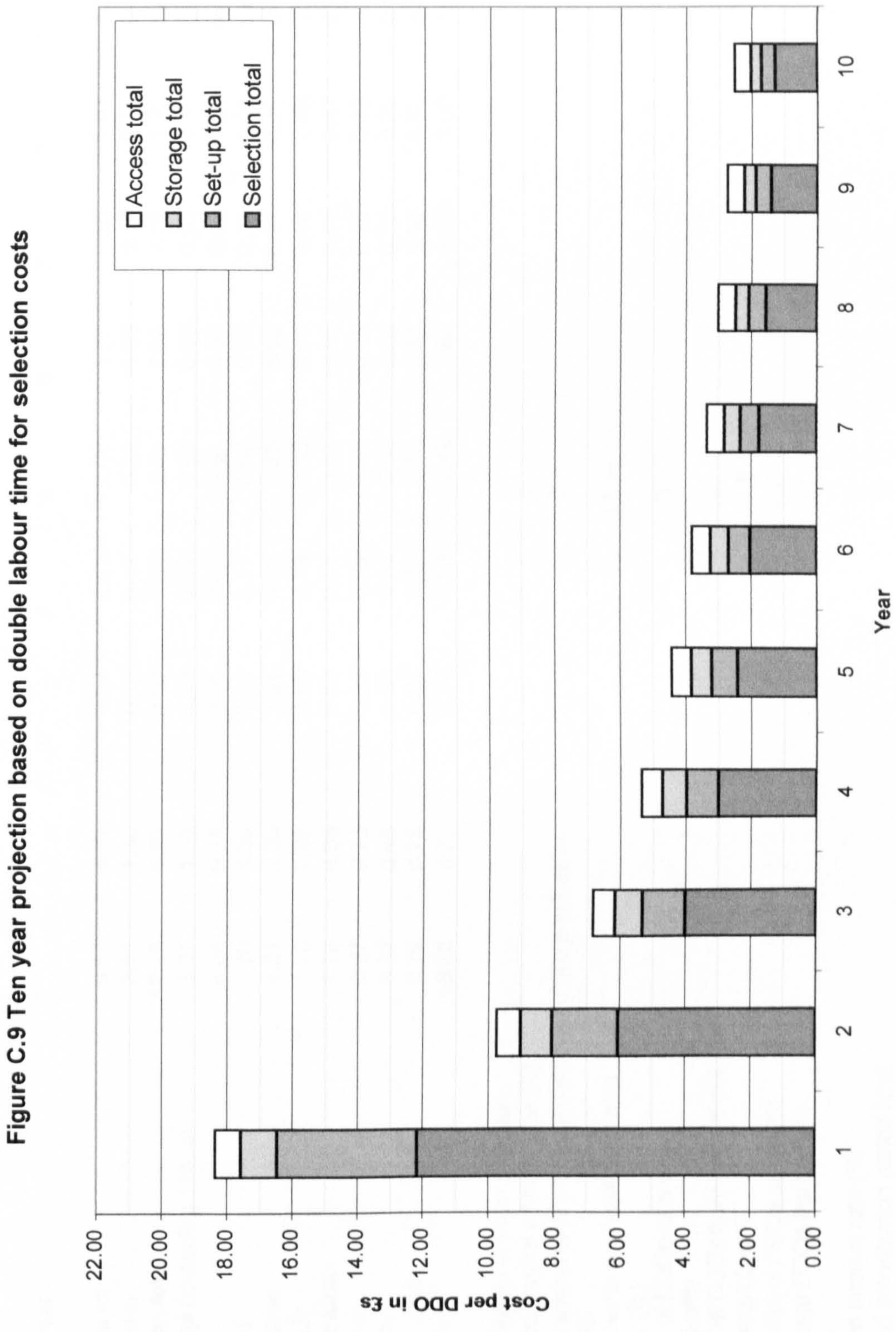


Figure C.9 Ten year projection based on double labour time for selection costs
Access total
Storage total
Set-up total
Selection total
Year
1
2
3
4
5
6
7
8
9
10
Cost per DDO in £s
0.00
2.00
4.00
6.00
8.00
10.00
12.00
14.00
16.00
18.00
20.00
22.00

Table C.15 Costs in Figure C.9

End of year	1	2	3	4	5	6	7	8	9	10
Selection rpi	9.17	4.72	3.24	2.51	2.06	1.77	1.56	1.41	1.29	1.20
Selection hi	2.98	1.30	0.75	0.49	0.34	0.25	0.19	0.14	0.11	0.09
Selection total	**12.15**	**6.02**	**4.00**	**3.00**	**2.41**	**2.02**	**1.75**	**1.55**	**1.40**	**1.28**
Set-up rpi (excluding 11b, c)	3.07	1.58	1.09	0.84	0.69	0.59	0.52	0.47	0.43	0.40
11b, c	0.54	0.14	0.05	0.02	0.01	0.00	0.00	0.00	0.00	0.00
Set-up hi	0.70	0.30	0.18	0.11	0.08	0.06	0.04	0.03	0.03	0.02
Set-up total	**4.31**	**2.03**	**1.31**	**0.97**	**0.78**	**0.66**	**0.57**	**0.51**	**0.46**	**0.42**
Storage hi	1.10	0.96	0.83	0.73	0.63	0.55	0.48	0.42	0.36	0.32
Storage total	**1.10**	**0.96**	**0.83**	**0.73**	**0.63**	**0.55**	**0.48**	**0.42**	**0.36**	**0.32**
Access rpi	0.27	0.28	0.29	0.29	0.30	0.31	0.32	0.33	0.34	0.35
Access hi	0.52	0.45	0.39	0.34	0.30	0.26	0.23	0.20	0.17	0.15
Access total	**0.79**	**0.73**	**0.68**	**0.64**	**0.60**	**0.57**	**0.55**	**0.53**	**0.51**	**0.50**
Grand total	18.35	9.73	6.82	5.33	4.42	3.80	3.35	3.00	2.73	2.52

Double labour time on selection	
Assumed general inflation rate (%)	3
Assume period over which hardware & software halves in cost (yrs)	5
Implied hardware & software cost inflation (%)	-12.94
Percent of DDOs selected automatically	75
Percent of DDOs from the Internet	30
Labour rate (£/min)	0.36
X = fraction of DDOs accessed	0.2
Percentage DDOs from offline sources	10
Assumed interest rate (%)	7
Assumed amortization period (yrs)	5

Appendix E

Standard letter

Dear

You may remember that I came to interview you in 1997/98 about the challenges facing organisations concerning the management of digital data. I have interviewed a range of stakeholder organisations (such as national archives, publishers, advisory bodies and science-based industry) and I am now proposing a framework for the long-term management of digital data. This framework, termed the **super-metadata framework**, aims to provide a systematic approach to managing a wide range of data in a variety of formats created/owned by a spectrum of information-dependent organisations.

As I am using Soft Systems Methodology which considers the involvement of people in the evolution of new systems, it is important that wherever possible I interview the same people, including yourself and ..., as I did in the initial stages of this research. It is anticipated that the interview would take between half an hour to an hour to complete. I appreciate that your time is valuable but I would be very grateful if you and ..... could participate in the final stage of this research.

This would involve looking at the proposed framework and the cost assumptions over the next week. Details of the proposed super-metadata framework and associated cost assumptions are given in the attached notes. The subroutines included in the framework, and a worked example of the costs can be provided at the interview if required. If you should have any queries about the information please do not hesitate to contact me.

This would be followed by an interview at a convenient time for you and .... within the next two to three weeks, to discuss your evaluation of the framework and the assumptions used to construct the cost model. This evaluation could, for example, consider the following questions.

- Would the framework meet your organisation's requirements?
- Are the itemised costs and standard assumptions realistic?
- Are there any other costs to be considered?

If you do not mind I will contact you by telephone some time next week to arrange a convenient date and time for me to come and conduct an interview.

Thanking you for your attention in this matter.

Yours sincerely

Section 1
A proposed new framework for the long-term management of Digital Data Objects

As a result of analysing the data from the in-depth interviews held with stakeholder organisations (which included your organisation) it has been possible to construct a framework, the **super-metadata framework.**

1.1 Recommendations

In order to develop a management system that supports the long-term access to Digital Data Objects (DDOs) originating from varied sources, this research makes the following recommendations:

- The creation of a conceptual framework (the super-metadata framework) for the long-term access to DDOs at a level higher than metadata which seeks to define stakeholder roles and suggests mechanisms to resolve issues raised by stakeholders. This framework is also needed to manage selection, rights management, review periods and physical preservation requirements.

- The process of selecting DDOs that meet the repository's criteria is viewed as an essential component of the super-metadata framework and aids end-user requirements.

- Both static (fixed data, fixed format) and dynamic (fixed data, variable format) DDOs could be selected for long-term preservation using similar criteria, but living DDOs (time varying data, variable format) require additional selection criteria to determine the frequency of 'snap-shots'.

- The implementation of the super-metadata framework could be aided by the recommendations listed below:

- Pilot projects to test the resilience of the framework to practical challenges.

- The development of standards through establishing the division of responsibility and degree of centralised control over the super-metadata framework.

- The development of a supportive legal structure that is conducive to the aim of providing long-term access to DDOs.

- Further research on emulation and migration strategies.

- Promotion of the spirit of co-operation between stakeholders and the co-ordination of revisions to the super-metadata framework.

- The development of staff training and re-training programmes.

1.2 The structure of the super-metadata framework

The super-metadata framework is designed to create a supportive structure to allow for past, present and future metadata developments within the information community. The new framework does not attempt to be prescriptive, but to provide a flexible structure that aims to cater for as wide a range of metadata formats as possible.

The entity-relationship diagram (Figure E.1) shows the structure of the super-metadata framework by following a DDO through the various processes necessary to ensure long-term access to the object. At the core of the framework is the concept of managing DDOs through the use of bi-level metadata, namely metadata and super-metadata (the latter containing the administrative data necessary to manage the DDO long-term). At the top of Figure 1 are a number of options open to creators wishing to make a DDO publicly available. As the interconnections between the stakeholders detailed in Figure 1 show, it is essential for the framework to function effectively that co-operation for mutual benefit is achieved. This research considers selection to be an important process in the long-term preservation of DDOs despite the fact that storage costs are low. It is seen as a benefit in saving end-user time searching for DDOs and in addition it provides a means of objectively evaluating these objects. Where possible computer-assisted mechanisms are employed to reduce costs. It should be noted that the remit of the collecting organisation (repository) and origin of DDOs can influence the degree to which stakeholders use computer-assisted mechanisms.

The entity-relationship diagram refers to a number of subroutines which are described below.

Subroutine 1
This is concerned with the processes of selection and rights management, all but the final stages of this subroutine form the generic structure on which manual and computer-assisted selection are based, to which subject specific selection criteria can be added. Within this subroutine are mechanisms to determine if the substantive content of a DDO with links to other DDOs is independent enough to be classed as a DDO in its own right if the links are severed. As ownership is seen to be a major influence on the nature of future access the subroutine is also concerned with rights management mechanisms to determine existing restrictions, negotiate future access, and determine the impact these have on future access.

Subroutine 2
This subroutine details the time-stamping method for authenticating DDOs. It should be noted that this is not the only method of authentication and it could be substituted by another method.

Subroutine 3
This subroutine performs bibliographic checks of the metadata and provides a detailed breakdown of the recommended components of the metadata.

Figure E.1 The proposed super-metadata framework

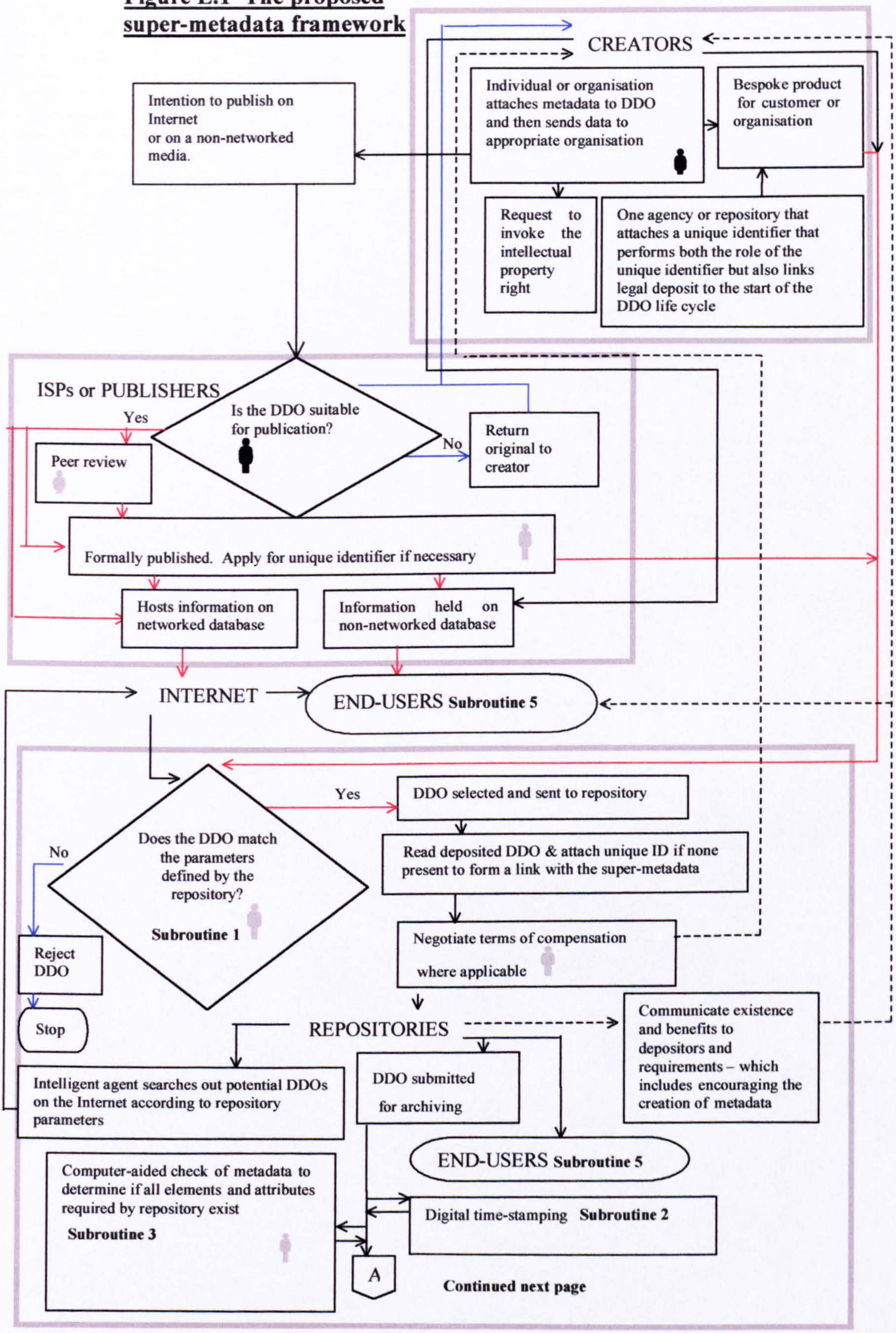
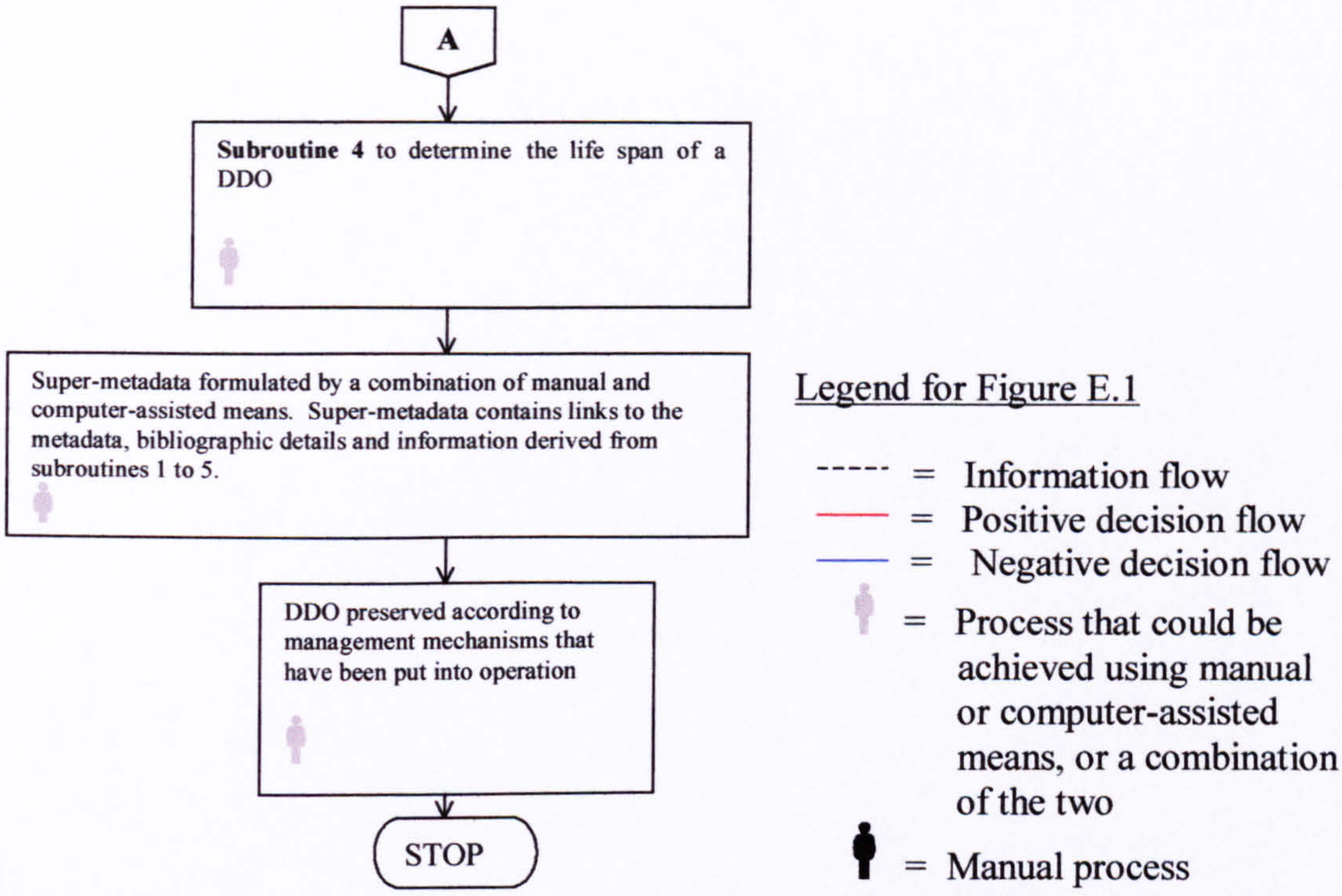


Figure E.1 Continued



Subroutine 4

Detailed in this subroutine are the mechanisms for establishing the time interval between reviews of the intellectual content (except in cases where periodic review is inappropriate) and for recording decisions at each review.

Subroutine 5

This details the processing of an end-user request to access a DDO. This considers rights of access, records any compensation payment due to the owner, and records the number of times the DDO is accessed.

Section 2

Costing model for the super-metadata framework

A costing model has been produced to estimate the cost of employing the super-metadata framework. The costs are based on 200,000 DDOs being acquired annually for long-term preservation, a figure derived from the Yale Project Open Book. This is

an electronic archive developed at Yale University[1]. A number of other assumptions have been taken from this project and the figures have been rounded up where applicable.

Selection

The following assumptions are made about the methods and sources of selection. The percentages are a percentage of the 200,000 DDOs assumed to be acquired annually by a repository.

- The percentage of DDOs selected by computer-assisted means for long-term preservation is assumed to be 75% per annum.

- The percentage of DDOs selected by manual means for long-term preservation is assumed to be 25% per annum.

- It is assumed that 10% of DDOs are received from offline sources.

- As the importance of global networks grow the percentage of DDOs retrieved from these sources are likely to increase. The standard assumption is set at 30% of DDOs from global network sources.

- For the purposes of this cost model the percentage of living DDOs is taken to be 2% of the acquired DDOs.

Set up

- An important aspect of super-metadata is the ability to absorb DDOs with any metadata format. This requires setting up a metadata registry. It is assumed that the percentage of DDOs entering the repository with a previously unknown metadata format is initially 5%, but this is assumed to halve each year. The associated set up costs fall each year accordingly.

- It is assumed that the majority of DDOs will have sufficient bibliographic details attached and a figure of 5% is used for those with insufficient detail. Consultations with owners to acquire further details could represent costs for creators, repositories, ISPs and publishers.

Access

- The cost of accessing a DDO is taken to be £2.40 based on Yale, to account for hardware costs and maintenance.

---

[1] Task Force on Archiving of Digital Information. 1996. *Preserving digital information. Report of the Task Force on Archiving of Digital Information commissioned by the Commission on Preservation and Access and the Research Libraries Group.* Washington, D.C.: Commission on Preservation and Access and the Research Libraries Group. (URL: http://www.rlg.org/ArchTF/).

- The application of access restrictions is assumed to add 10% to the access cost.

- It is assumed that 20% of DDOs are accessed per year for the first ten years of management by a repository. Access costs may be incurred by an end-user or a repository depending on the access policy.

Compensation payment

- Compensation to owners of DDOs for potential loss of revenue that may result from depositing material with a repository is set at £2.10 per DDO accessed.

Storage

- The cost of storage per DDO per annum is set at £1.10 derived from the Yale Model. Included in this cost is maintenance and supervision of hardware as well as data migration, which Yale assumes to be double the storage cost.

Labour Costs

- The standard assumption sets the labour rate at £22 per hour or an annual salary of £30,000 for management and systems engineering staff, increased by 40% to include overheads.

- The time estimated for manual selection of DDOs, taking into account variations in the difficulty to select individual objects, is assumed to be 30 minutes.

Hardware and Software Costs

- The period over which the hardware/software costs halve due to advances in technology is assumed to be five years in accordance with Yale.

The stakeholder organisations role determined which one of the following paragraphs was sent.

The estimated range of costs incurred by publishers

The costs given are not intended to be exact but to give an indication of the costs a publisher may incur per DDO published. The maximum estimated total cost that a publisher may incur by depositing a DDO with a repository under a legal deposit scheme would be, on average, approximately £30 per DDO. This maximum total cost is broken down into the following costs. For the purposes of this cost model it is assumed that a published DDO undergoes a process of peer review (£11), that a unique identifier has been applied for and attached (£2), an average loss of revenue per

DDO of £10 (which could be greater for some DDOs), the negotiation of access rights (£2) and the consultation with creators to complete bibliographic details (£4). A best case scenario could reduce costs to £10 per DDO or even less if the cost of legal deposit can be reduced. This could occur in cases where a third party is managing a publisher's DDOs, a unique identifier has been attached by a creator, and negotiation costs are not incurred because of the existence of a clearly stated access policy for each class of DDOs. Similarly, peer review and access rights negotiations may be avoided in routine cases. These costs are set up and selection costs and therefore only occur once in a DDO's life.

The estimated costs incurred by repositories

The costs for the repository are given per DDO managed, according to standard assumptions (see pages 379 to 381). The estimated cost is £14 after one year of being managed in the repository. When one-off costs such as set-up and selection are reduced by spreading them over the length of time a DDO is kept (the model recommends that a DDO be retained for a minimum of ten years for it to be cost effective) then the total cost per DDO reduces to approximately £2 at the end of the tenth year in the repository. It should be noted that £2 is an average cost for long-term preservation and may vary for individual DDOs depending on their source of origin and the nature of the DDO itself. For example, the source of origin may be the Internet, a publisher, or direct from the creator and the nature of the DDO may be static, dynamic or living.