Right to Preserve?

Copyright and licensing for digital preservation project

Final Report

Catherine Ayre and Adrienne Muir

March 2004

Ayre, Catherine & Adrienne Muir, 2004. Right to preserve? *The Copyright and Licensing for Digital Preservation project final report*. Department of Information Science, Loughborough University.

Project funded by the Arts and Humanities Research Board.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ADS	Archaeology Data Service
AHDS	Arts and Humanities Data Service
ALCS	Authors' Licensing and Collecting Society
ASCII	American Standard Code for Information Interchange
BLDSC	British Library Document Supply Centre
BNF	Bibliothèque Nationale de France
CAMiLEON	Creative Archiving at Michigan and Leeds: Emulating the Old on the New
CDPA	Copyright Designs and Patents Act 1988
Cedars	CURL Exemplars in Digital Archives
CHEST	Combined Higher Education Software Team
DCC	Digital Curation Centre
DCMS	Department of Culture, Media and Sport
DMCA	Digital Millennium Copyright Act
DOI	Digital Object Identifier
DPC	Digital Preservation Coalition
EPS Ltd	Electronic Publishing Services Ltd
IFLA	International Federation of Library Associations and Institutions
KB	Koninklijke Bibliotheek (Netherlands)
JISC	The Joint Information Systems Committee of the UK Higher and Further Education Funding Councils
LISU	Library and Information Statistics Unit
LOCKSS	Lots of Copies Keeps Stuff Safe
NEDLIB	Networked European Deposit Library
NESLi	National Electronic Site Licence Initiative
NIST	National Institute of Standards and Technology
OAI	Open Archives Initiative
OAIS	Open Archival Information System
OCLC	Online Computer Library Centre
ONIX	Online Information Exchange
RLG	Research Libraries Group
RoMEO	Rights Metadata for Open Archiving
SPSS	Statistical Package for Social Sciences
STM	Science, Technology and Medicine
UPF	Universal Preservation Format
UVC	Universal Virtual Computer

ACKNOWLEDGEMENTS

The authors would like to thank everyone who participated in this research and to the Arts and Humanities Research Board for funding it.

The members of the project Advisory Board provided us with very useful advice and support. We also have received a great deal of support from our colleagues in the Department of Information Science at Loughborough University, in particular from colleagues in LISU.

We are also grateful to Rachel Hardy, Michael Norris, Margaret-Mary O'Mahony, Rebecca Smith, Rachel Spacey, Gemma Towle and Charlotte Webb for their assistance in the research. Mark Bide did a great job in facilitating the project seminar.

Much of the content of the literature review in this report is adapted from work carried out by Adrienne Muir as part of her doctoral research.

EXECUTIVE SUMMARY

THE RESEARCH

The Copyright and Licensing for Digital Preservation Project ran from September 2002 to March 2004 and was funded by the Arts and Humanities Research Board. The aim of the research was to investigate whether and how copyright legislation and licensed access to digital content affect the ability of libraries to provide long-term access to that content, and to suggest ways in which any problems can be overcome.

The project included a review of the library, legal and related literatures. Questionnaire surveys were then used to explore the views of libraries, publishers and authors. These were supplemented by twenty in-depth, face-to-face interviews with librarians, publishers, legal experts, digital preservation experts and representatives of rights holder organisations. An invitation-only seminar was held, at which delegates discussed possible solutions to the issues identified.

DIGITAL PRESERVATION COPYING

Implementing digital preservation strategies will involve activities such as:

- repeated replication of material
- reformatting material
- saving software that allows information to be accessed and used
- developing new software to allow digital information to be accessed and used

It is unlikely that the copying actions required to preserve digital material are legal under UK copyright law. However, there is no case law, and further clarification is needed. Other countries' laws which were investigated also do not address the rights issues of digital preservation adequately, if at all. Digital preservation strategies may also raise moral rights issues if they result in changes to the material.

PRESERVATION NEEDS AND RESPONSIBILITY FOR DIGITAL PRESERVATION

A large majority of library respondents already have some digital materials in their collections. More than half of the respondents to the library questionnaire said that they are already taking responsibility for the preservation of their digital materials, or expect to do so in the future. However, few of the libraries with digital collections are currently undertaking preservation activities or have preservation policies. While some publishers are taking responsibility for preserving their own material, they may be taking a more short-term view of preservation than libraries. Awareness of digital preservation is low among both libraries and publishers, and needs to be raised.

The Legal Deposit Libraries Act 2003 will allow UK legal deposit libraries to copy the digital materials deposited with them for preservation. However, legal deposit will not cover all the material that will have to be preserved by UK libraries and the Act does not allow other libraries to undertake preservation copying.

PRESERVATION OF REMOTELY ACCESSED MATERIAL

There is some evidence that librarians are concerned about how they can ensure their users have access to remotely accessed material for as long as it is needed, since they may have to rely on publishers to preserve it. The publisher licences investigated vary considerably in whether they address long-term preservation adequately. There are existing model licences that do address the issues, but they are not always accepted by publishers. Even where publishers do address this issue in licences, it is not clear how they will honour them. None of the library respondents said that they have yet had to test such provisions.

ADDRESSING THE ISSUES

Decide on Responsibilities

There is an urgent need to make decisions on who takes responsibility for digital preservation, how they can obtain the necessary rights to preserve, and how access to preserved material can be controlled.

Changing Copyright Law Not Helpful

Project participants were not opposed to changing the law to allow preservation copying by all libraries. However, they felt that this is unlikely to happen in practice, and would happen too slowly to be useful.

Finding Models for Preservation

While legal deposit libraries have a role in carrying out preservation, arrangements between rights holders and other types of libraries also have to be made. This could involve the use of trusted repositories. There are some examples of potential models in this area, such as the JSTOR initiative and agreements between publishers and the Royal Library of the Netherlands. However, there are various difficult funding and management issues that would need to be addressed in order for this approach to be feasible. Models for the implementation and funding of digital preservation need to be investigated.

Licensing Preservation

Project participants felt that licensing solutions to the rights issues of digital preservation would be preferable to legal solutions. They felt that it would be too time-consuming for libraries to have to seek individual licenses from publishers to preserve digital materials, and a collective licensing approach was favoured. A

possible solution would be a hybrid licence with both fixed and optional clauses to cover the preservation and access issues. A UK scheme could potentially be managed by existing reproduction rights organisations.

Other possible solutions include rights metadata and digital rights management systems. Project participants agreed that these would be useful in principle, but felt that further development would be needed before these solutions would be widely accepted.

Cooperation Between Rights Holders and Preservation Bodies

It is not yet clear who will be taking responsibility for digital preservation in the UK, but it is clear that stakeholders need to work together to ensure that digital publications are preserved.

Re-evaluation of Copyright in the Digital Environment

A bigger question that emerges from this research and from other digital library activities is whether there is a need to reassess the concept of copyright in the digital environment.

1 INTRODUCTION

1.1 Background

Digital preservation is an increasingly important subject of research, development and discussion. There is a general perception that the preservation of digital information is more problematic than print or other formats. This is increasingly a cause for concern as the proportion of information being made available only in digital form increases, because there is no hardcopy version that can be preserved.

Many of the problems are technical in that digital media are likely to degrade relatively quickly compared to more traditional media, and the speed of change in electronic publishing means that digital information may become stranded in an obsolete technological environment. Other properties of digital information also pose barriers for its preservation including the complexity of digital information products, the ephemeral nature of some types of digital information and the prevalence of linkages between different bits of digital information. Information is being disseminated in new ways and business models are also changing. Increasingly, information is not purchased but "rented" through licensed access rather than physical ownership of an information artefact. Libraries have traditionally had a role in preserving publications and if they are to continue to do this in the digital environment, there are a number of problems they will have to solve.

To preserve digital information, it is not just a case of solving the technical problems. There are also other issues, managerial, economic and legal, that have to be dealt with. A major legal issue is that of intellectual property rights, especially copyright. Digital preservation strategies will involve copying digital information many times, and migration to new technological environments over time. Preservation strategies may also involve some loss of the intellectual content, functionality and look and feel of digital information products and services. Current UK legislation allows limited copying for the creation of preservation copies. However, this privilege is restricted to so-called prescribed libraries (in practice public sector and learned society) and before the research was carried out, it was unclear whether any digital preservation copying would be allowed. Even if the law gives libraries the rights they need to preserve

digital information, increasingly, libraries do not physically own the digital information they provide access to. The model for access to digital information in libraries is now licensed access. The information itself does not necessarily reside in the library and libraries may not have any access to older information. In fact, they may not have access to any information at all if they cease to subscribe or if an information provider goes out of business. This is in contrast with print-based subscriptions where the libraries control how long information survives and who retains access to material they have already purchased even when they cease subscribing.

Legal deposit legislation in the UK has recently been amended to provide for extending its scope to non-print publications, including digital publications. Legal deposit is a statutory obligation on publishers to deposit a copy of all material made publicly available in the UK with a number of designated libraries. It is not clear whether and how it will be possible or even acceptable to publishers of publications such as frequently updated and/or large databases or on demand publications, to physically deposit them. In any case, the preservation of the UK digital output in deposit libraries would not mean that all the digital information library users require long-term access to will be preserved. In addition, the level of access to digital legal deposit material is, in the short-term at least, likely to be just as restricted as print material. Legal deposit material is currently only available within the reading rooms of the UK deposit libraries.

UK legal deposit libraries are likely to benefit from explicit exceptions to copyright law in order to acquire and preserve digital information, but this does not help other libraries who want to preserve digital information. If other libraries do not have physical ownership of their digital collections, the onus for the provision of long-term access is with information providers. There is some movement towards including long-term access provision in licensing arrangements. However, it is not clear how information providers will implement this in practice. A major problem is the complexity of rights ownership in digital information. The organisation providing the information product or service may not own all the rights in that product or service. In which case, they cannot easily grant permission to libraries to preserve or guarantee to preserve access themselves.

Copyright and licensing for digital preservation has been identified as a problem by research and development projects. Various issues required clarification, including the copyright implications of digital preservation strategies, what is allowed under copyright law and what preservation provisions are included in licence agreements for digital publications. There was also a need to investigate how important the issues are, for example will libraries other than legal deposit libraries want to preserve digital content they purchase, rent or create, and are publishers giving and able to implement guarantees on longer-term access to digital content to libraries. The research therefore also looked at wider issues of responsibility and cooperation in digital preservation management.

1.2 Aim and objectives

The aim of this research was to investigate whether copyright legislation and licensed access to digital content threaten the ability of libraries to provide long-term access to that content and to suggest ways in which the problems can be overcome.

Specifically, the objectives were to:

- Assess whether the provisions of present and proposed UK copyright legislation meet the digital preservation needs of UK national, academic, public and special libraries
- Investigate to what extent licensed access to digital material in libraries takes account of preservation needs and identify examples of best practice in the UK and from the literature
- Investigate how publishers and information providers are planning to achieve the provision of perpetual access to digital material for libraries, including how access to third party information and software is being dealt with
- Identify if and how copyright legislation in other countries addresses this issue
- Make recommendations for amendments to UK legislation if appropriate
- Develop model licences for long-term access

- Make recommendations on how legislators, information providers and libraries can work together to ensure long-term access to digital information

It was clear from an initial review of digital preservation literature that there was a need for clarification of the rights and licensing issues involved in preserving digital material. There was also a need to identify possible ways of dealing with these issues.

The Arts and Humanities Research Board provided a grant to investigate these issues. The Copyright and Licensing for Digital Preservation Project started in September 2002 and concluded in March 2004. The aim and objectives were addressed by carrying out desk research, questionnaire surveys and semi-structured face-to-face interviews. A seminar was held towards the end of the project both to disseminate the findings of the research and to elicit the opinions of stakeholders on the findings and ways forward.

1.3 Limitations of the research

The main limitation of the research is the response rates to the questionnaire surveys, particularly by publishers and authors. The author surveys were made available electronically only. This was because we could not find an easy way to sample authors and were not able to use organisations representing authors to distribute questionnaires. There was a risk of bias in any responses we did receive because it was likely that only authors with experience and an interest in electronic communication would respond. The findings of the author survey are included in this report for completeness, but we cannot claim that the views expressed are in any way representative or even indicative of the views of authors. Unfortunately, we were only able to supplement these findings to a limited extent with the views of organisations representing author interests because of lack of interest. While the publisher and library response rates were also low, we were able to supplement this with findings from the literature, interviews and participation of publisher representative organisations in the seminar. We were not unable to include the views of intermediaries, such as content aggregators, on issues such as whether they might have a role in preservation in the research. However, research carried out by Muir (2004) suggested that aggregators see their role as secondary publishers that add value

to pre-existing content and therefore not involved in carrying out preservation or rights management for preservation.

Surveys were the most obvious way to gather the baseline data on publishing activities, collections and preservation activities in libraries and publishing organisations. However, the questionnaires were much more difficult to design than originally envisaged because the issues involved were so complex. With hindsight some questions that should have been asked were not, and some questions could have been phrased differently. Some of the responses received were difficult to interpret because despite explanations and definitions being provided, it was not always clear what the respondents' understanding was. Very few of the respondents agreed to follow-up interviews, so clarification of responses was limited.

The complexity of this issue caused difficulties in other areas of the research. The various strands of the issue, that is physical ownership and rights ownership, responsibility and licensing, are so closely intertwined that it is almost impossible to consider each in isolation. This made it difficult to frame the seminar questions and to structure accounts of our findings. However, the research has value in that it brings together in one place all the relevant work in this area and can be used to raise awareness of the issues, problems and potential solutions. Therefore the findings of this research will be widely disseminated in appropriate forms to the relevant stakeholder groups.

1.4 Structure of the report

The methods used to achieve the aims and objectives of the research are set out in chapter 2. This is followed by the review of the literature. The findings of the surveys, interviews and the seminar are set out in four thematic chapters dealing with digital publishing, digital preservation activities and policies, rights issues and policies and views on possible solutions to the problems identified. This structure was selected to provide a more integrated and coherent account of the project findings, rather than ordering material by research instrument, stakeholder or chronological order of the research. The final chapters draw together the findings of the desk and primary research in a set of conclusions and recommendations.

The target audiences for this report are legislators and policy makers, managers of national, academic, public and special libraries and providers and publishers of information.

2 METHODS

2.1 Introduction

The aim of this research was to investigate whether and how copyright legislation and licensed access to digital content affect the ability of libraries to provide long-term access to that content and to suggest ways in which any problems can be overcome.

Specifically, the objectives were to:

1. Assess whether the provisions of present and proposed UK copyright legislation meet the digital preservation needs of UK national, academic, public and special libraries
2. Investigate to what extent licensed access to digital material in libraries takes account of preservation needs and identify examples of best practice in the UK and from the literature
3. Investigate how publishers and information providers are planning to achieve the provision of perpetual access to digital material for libraries, including how access to third party information and software is being dealt with
4. Identify if and how copyright legislation in other countries addresses this issue
5. Make recommendations for amendments to UK legislation if appropriate
6. Develop model licences for long-term access if appropriate
7. Make recommendations on how legislators, information providers and libraries can work together to ensure long-term access to digital information

These objectives were addressed using a several methods. These were a comprehensive literature review, questionnaire surveys, in-depth interviews and a seminar.

2.2 The literature review

A review of the library, law and related literatures was carried out using a variety of bibliographic tools. The aim of the literature review was to investigate the following:

- Different approaches to digital preservation and their copying implications
- Rights issues in complex digital content
- Preservation provisions of copyright legislation in the UK and overseas
- Other legal provisions relating to the archiving and preservation of digital information
- Issues relating to licensed access to digital information
- Non-legal approaches to dealing with copyright and licensing issues
- Publisher and blanket licences and any relevant research in this area

The sources searched included library catalogues, bibliographies, abstracting and indexing services, Internet search engines, paper and online directories and discussion lists. These included general sources such as *Walford's guide to reference material*, *OCLC First Search* databases and *Google*; library and information sources such as *Library and Information Science Abstracts* and legal databases. Search terms used included *digital preservation*, *digital archiving*, *copyright* and *intellectual property*.

The literature search revealed much material on digital preservation and digital copyright and licensing. However, there seems to be a paucity of literature on the copyright and licensing issues in digital preservation.

Additional surveys were carried out to investigate two specific aspects of the issues being explored. The copyright laws of twenty-five countries were studied to see whether they contain clauses that allow the copying of copyright material to aid its preservation. The texts of these laws were taken from the World Intellectual Property Organisation Website (http://www.wto.org). Other clauses which may hinder or help preservation were also identified. These were principally clauses which concern the adaptation of computer programs and the bypassing of technological protection measures.

A second study investigated archival and preservation clauses in publishers' licences. Publishers' Websites were identified using publisher directories; 1070 Websites were identified, of which only 411 were found to be actual publisher Websites. These were then searched for information about the publishers' archiving and preservation

policies. In some cases, the text of the publishers' licence agreements was found; in others, relevant information was taken from the 'terms and conditions' section or elsewhere on the Website. Relevant information was only found on twenty-two publishers Websites, and two of the licences found were duplicates as they were from different branches of a major publisher.

2.3 Questionnaire surveys

Questionnaire surveys were carried out in order to obtain an overview of stakeholder activities, plans and perceptions in connection with digital preservation. Specifically, the aim of the questionnaires was to gather data on:

- Current and planned digital publishing and preservation activity
- Awareness and experience of digital preservation issues, including copyright, and issues related to preservation of digital materials
- Experience of problems with long-term access to digital material
- Opinions on possible solutions to the problems posed by copyright and licensing restrictions

The content of the questionnaires was informed by the findings of the literature review. Separate questionnaires, with some parallel content, were developed for libraries and publishers and distributed by post. A total of 1,600 paper questionnaires were distributed: 1,000 to libraries and 600 to publishers. A questionnaire for authors was also designed and mounted on the project Web site.

2.3.1 Issues covered by the questionnaires

The majority of the questions were closed single or multiple choice. There were also opportunities for respondents to make further comments. The library questionnaire was divided into the following four sections:

- Extent of digital collections
- Digital preservation practices

- Rights and licensing issues for digital preservation
- Solutions to copyright and licensing issues in digital preservation

The publisher questionnaire was divided into the following four sections:

- Publishing activities
- Digital preservation practices
- Rights and licensing issues for digital preservation
- Solutions to copyright and licensing issues in digital preservation

Libraries and publishers were asked different questions in each of these areas appropriate to their roles in digital publishing and preservation. Both questionnaires had identical questions on responsibility for archiving and preservation and potential solutions to issues to allow comparison of attitudes in these areas. See Appendix 1 for the library questionnaire and Appendix 2 for the publisher questionnaire.

The author questionnaire was divided into the following five sections:

- Types of materials published
- Digital publishing practices
- Digital preservation
- Rights and licensing issues for digital preservation
- Solutions to copyright and licensing issues in digital preservation.

Some of the questions were the same as, or similar to, questions contained in the other questionnaires, to allow for comparisons. However, the author questionnaire covered some different areas which were only appropriate to authors; for example, it asked how they would prefer to submit their publications to a library or other agency for preservation. The author questionnaire can be found at Appendix 3.

2.3.2 Sampling

The library sample was drawn from:

- Academic libraries (both higher and further education)
- Public libraries
- Special libraries
- National libraries

The Library and Information Statistics Unit (LISU) mailing lists were used as the basis of the sampling frame. Entries were updated where possible. On the advice of the LISU statistician, a sample of 1,000 libraries was drawn. This included 50% of all UK academic libraries, and 50% of all public libraries. This accounted for 419 libraries with the remainder of the sample being made up of special and national libraries. Questionnaires were sent to all the UK national libraries, since they are all likely to be involved in digital preservation. The samples were selected alphabetically (i.e. one library starting with the letter A, then one B and so on until the end of the list, this was then repeated). The sample breaks down as follows:

Library Type	**Sample size**
University and other higher education	88
Further education	206
Public	125
National/special	581

Table 2.1 Library questionnaire sample

The sample for the publisher survey was taken from various sources, as shown below. This was necessary to comply with UK database right restrictions. The Publishers Association Website carries a copyright disclaimer allowing the use of all the contents of the Website including their online directory of members. Permission was granted to take data from the Directory Publishers Association Web site.

Source	No. sampled	% of source
Whitakers Directory of Publishers 2003	315	10
Publishers Association Directory 2002	63	10
Publishers Association Web site	162	Permission given
Directory Publishers Association Web site	60	Permission given

Table 2.2 Publisher sample sources

Sampling of the printed sources was alphabetical until the sample reached 10% of each source. The first source used was the Whitaker directory. In the second and subsequent sources publishers already selected from Whitakers were ignored.

2.3.3 Piloting and distribution of questionnaires

Limited piloting of the drafts of the questionnaires was carried out at the end of January 2003. The pilots involved one academic author, one publisher, two librarians and one legal expert. Four of the individuals piloting the questionnaires were members of the project Advisory Board.

Paper questionnaires were circulated to libraries and publishers at the end of February 2003. The final version of the author questionnaire was uploaded to the Web on the 28th February 2003. Each paper questionnaire was accompanied by a covering letter outlining the issues being addressed by the project; the letters sent to libraries and publishers can be found be found at Appendices 4 and 5. The libraries and publishers that had not responded by the closing date of the 31st March were contacted by telephone at the beginning of April. Of the 1,000 questionnaires sent to libraries 168 (16.8%) were returned. Eighty-two out of 600 publishers returned questionnaires giving a response rate of 13.7%. Twenty-five library and 18 publisher questionnaires were returned to sender, no action was taken with these.

Initial responses:	138
Follow up responses:	30
Total responses	**168**

Table 2.3 Library questionnaire response rate

Initial responses:	74
Follow up responses:	8
Total responses	**82**

Table 2.4 Publisher questionnaire response rate

The online author questionnaire was publicised on ALPSP-alert and PALS@JISCMAIL.AC.UK. Several author's agents were also contacted, but none responded.

2.3.4 Response Rates

Responses were received from a range of library sectors, but by far the best response came from academic and research libraries. This perhaps reflects the awareness of and interest in digital preservation and rights issues in these sectors.

Library sectors	**No. of Responses**	**% Return Rate**	**% of all Responses**
Academic	68	77.3	40.5
Public	36	28.8	21.4
Special	46	12.3[1]	27.4
Other	13		7.7
National	5	66.7[2]	3.0
Total	**168**		**100**

Table 2.5 Responses from libraries

Eighty-one of the eighty-two publisher respondents indicated their area of business. Some of the respondents were publishing in more than one of the categories included in the questionnaire.

[1] Includes libraries of research councils, government libraries, non-departmental public bodies, professional bodies and commercial organisations and counted as "special" for sampling purposes.
[2] Questionnaires were sent to the three UK national libraries, the British Library, the National Library of Scotland and the National Library of Wales. Responses were received from two of these libraries, the other responses were from libraries based in national institutions or with a national remit of some kind.

Publisher	No. of responses	% of Responses	% of Respondents
STM	6	5.4	7.4
Professional	20	18.0	24.7
Trade	20	18.0	24.7
Education	43	38.7	53.1
Other	22	19.8	27.2
Totals	**111**	**100**	**137**

Table 2.6 Responses from publishers

The education sector was most heavily represented, but there was a spread of responses from different publishing sectors. The twenty two respondents who selected 'other' published material including: academic material; market reports; biographies/autobiographies; children; humour; fiction; consumer publishing; reference; health; humanities; social science; English language teaching; Islamic manuscripts; legal and regulatory; local history and guides; magazines and books for primary teachers; media and news; music; politics; philosophy; current affairs; public policy; sport; and directories for local telephone dialling code areas.

The response rate for the author questionnaire was very low. Initially, only nine authors responded. A delegate at the project seminar, who is involved with the Authors' Licensing and Collecting Society (ALCS), offered to help publicise the questionnaire within ALCS, the Writers' Guild and the Society of Authors. This only brought an additional four responses, making a total of 13 responses.

Authors were asked to state which subject area(s) or publications genre(s) they write in. The most frequently mentioned subject was archaeology: four respondents gave their subjects as: 'archaeology', 'archaeology and history', 'archaeology, local history, genealogy', and 'archaeology / IT'. Since three of these responses were received consecutively, it is likely that these respondents all know each other. This would explain why such a large proportion of respondents (30.8%) were from the same subject area. A further three respondents were from the Library and Information Science area. The high participation from this subject are may be explained by the fact that this is the same subject area in which the research is being carried out. The

other subject areas mentioned were theoretical linguistics, technology, philosophy / ethics and 'the study of the Irish Diaspora'. One respondent gave no response to this question. Three respondents gave some indication of the genre of their writing. One is a technical writer, one produces school textbooks, and the third said:

> *My work appears regularly in academic journals and in other journals. I have some major book contracts in place. I am also active in drama, for radio and theatre. And I write song lyrics and short stories*

There is therefore a heavy bias towards arts and humanities subjects among the respondents, with almost no representatives of the Science, Technology and Medicine (STM) subject areas. This is interesting, since it might have been expected that there would be more interest in digital publishing, particularly self-publishing, from the STM areas. The limited range of subject areas means that it was not possible to draw any conclusions about how different subject areas are using digital publishing.

Authors were also asked about the types of publication in which their works have been published. Most of the respondents publish in 'academic / scholarly publications' and / or 'non-academic / scholarly' publications. Only one respondent published in 'official / government publications', and this author, an archaeologist, in fact publishes in all three categories. Almost half of the respondents (46.2%) publish in two of more of these categories.

Publication type	**No. of responses**	**% responses**	**% respondents**
Academic / scholarly	10	50	76.9
Non-academic / scholarly	9	45	69.2
Official / government	1	5	7.7
Total	**20**	**100**	**153.8**

Table 2.1 Type of publication in which respondents' work is published

2.4 Interviews

Semi-structured face-to-face interviews were carried out to explore the issues covered by the questionnaires in more depth. They gave respondents the opportunity to

explain the reasons for their views, to describe any relevant experiences, and to raise any other issues they considered to be relevant. Interviewees were told that their comments would remain anonymous. However, three interviewees specifically asked to have their comments ascribed. This has been taken into account in the results chapters.

2.4.1 Issues covered by the interviews

Separate interview schedules were drawn up for the different groups involved. These were based on generic schedules. The schedules were modified for each interview, based on questionnaire responses, where available, and information gained from company or organisational Websites. For example, questions about e-prints were added for two interviewees since they were known to be working on a project about the preservation of these.

The interview schedules for librarians were divided into the following sections:

- Extent of digital resources
- Digital preservation activities
- Licensing of digital resources
- Views on legal and preservation issues

The interview schedules for publishers were divided into the following sections:

- Publishing activities
- Licensing of digital resources
- Digital preservation activities
- Views on legal and preservation issues

A further schedule was developed for use with the digital preservation experts, legal experts and representatives of reproduction rights organisations interviewed. This needed to be adapted to suit each interviewee's particular areas of expertise. This was divided into the following sections:

- Technical preservation strategies
- Copyright issues
- General and metadata issues

As with the questionnaires, the different groups were asked different questions which reflected their roles and expertise in the different areas being investigated. However, some issues were raised with all groups: these included views on responsibility for digital preservation and on the implications of copyright law for digital preservation.

The interview schedules are included in Appendices 6-8.

2.4.2 Sampling

Potential interviewees were suggested by project team members and members of the Advisory Board. These individuals were selected on the basis of their interest in or knowledge of the issues, or their work in areas related to the project. Some of the interviewees were selected because they had given interesting responses and had identified themselves on their returned questionnaires. In all, 20 one-hour interviews were carried out, involving 24 people. The full list of interviewees can be found at Appendix 9; all interviewees were asked and agreed for their names to be mentioned in this way.

The stakeholder group which was most willing to be interviewed was publishers. In all, 11 publishers were interviewed; these represented nine publishing companies. This means that a disproportionately large percentage of the interviewees (45.8%) were publishers. However, it meant that the opinions of different types of publishers could be explored. The publishers interviewed represented a broad range of types of publishing, including digital-only (3), open access (1) and database publishing (1). They also represented educational (3), academic (2), STM (2) and trade and professional publishing (1). A further interviewee works for Electronic Publishing Services Ltd, which carries out research into electronic publishing issues and advises the publishing industry.

Four librarians were interviewed, two representing legal deposit libraries; a further two from academic libraries (2). One of the librarians has responsibility for digital preservation, one is involved in licensing, and one manages a digital library. Unfortunately, no public or special librarians were willing to be interviewed.

Following the poor response to the authors' questionnaire, no authors were interviewed, although some interviewees were themselves authors and made some comments from an author perspective.

Other types of stakeholders were also interviewed. These included representatives from reproduction rights organisations for publishers and visual creators; again, no author representatives agreed to be interviewed. Three legal experts and three digital preservation experts were also interviewed.

2.5 Data Analysis

2.5.1 Questionnaire data analysis

The postal questionnaire responses were initially entered into Excel spreadsheets. Data input was then checked against the completed questionnaires and any errors were corrected.

It was necessary to 'clean' some of the data. Of the 1,000 questionnaires sent out, three were sent to national libraries, namely the National Library of Scotland, the National Library of Wales and the British Library. However, six of the respondents designated themselves as national libraries. Since two of the three National Libraries responded so it must be assumed that some respondents interpreted the term "national" as the library of a national body or a library with a national remit.

Some library respondents gave multiple responses when asked to state type of library. These responses were treated as follows:

No. of respondents	Response	Action
2	Academic and Public	Treated as Academic
1	Public and special	Treated as special
1	Academic, Public, Special	Treated as special
1	All options	Treated as Other

Table 2.7 Cleaning library response data

Basic counts for all responses to all questions were carried out in Excel. This was done using Excel's 'sort' function. The data for both questionnaires was then exported from Excel to SPSS (Statistical Package for Social Science) software and frequency counts and cross tabulations were carried out.

2.5.2 Interview data analysis

The interviews were all recorded and transcribed. Interviewees had been asked if they wanted to see a transcript of their interviewee, and transcripts were sent for checking to the interviewees who had requested this.

The interview data was coded and analysed using the Atlas/ti software package. Initially, general codes were used. However, the codes were modified as more transcripts were coded to ensure that the code accurately described the attached quotations. Lists of quotations attached to each code were checked periodically to ensure that the new segments being assigned were relevant. Codes which had particularly large numbers of quotations were subdivided, while codes with very few quotations were sometimes merged.

2.6 Project seminar

An invitation-only seminar was held at the Policy Studies Institute in London on 19th November 2003. Mark Bide of Rightscom Ltd was invited to chair the seminar. A report of the day and the recommendations made can be found at Appendix 11.

Everyone who was interviewed for the project was invited to the seminar, as well as the project Advisory Board members. About 20 others were also invited. These were people from the different stakeholder groups who were known to have some knowledge of or interest in the issues being discussed.

A total of 23 delegates were present at the seminar. As well as the project team, these included six librarians and six publishers, three delegates with legal knowledge, three who are involved in digital preservation and two representatives of reproduction rights organisations.

All delegates were sent a briefing document and questions for discussion before the seminar. These can be found at Appendix 10. It proved difficult to develop suitable questions for discussion because, as was stated in the Introduction, the different issues relating to rights issues for digital preservation are so closely linked. The initial plan was to give different questions to each of the three groups, so that each group could discuss a few issues in depth. However, it was impossible to separate the different issues into three distinct groups, of a similar size. Following discussion with the Seminar chair, a decision was therefore taken to give the same, broad list of questions to all three groups.

3 REVIEW OF THE LITERATURE

This research investigated two separate but related issues: how libraries can preserve digital content that is only available remotely; and whether digital preservation is possible under the current UK copyright regime where libraries hold a physical copy of digital material. This chapter provides background information on digital preservation, including problems identified and potential strategies to address these problems. The likely copying implications of the different strategies are summarised. The provisions of relevant rights law in the UK is discussed and there is a preliminary assessment of whether current provisions are likely to meet the needs of libraries wishing to preserve digital information. Since commentators assume that this is not the case, the issues arising from the need to seek permission to preserve digital information are explored. This is followed by a discussion of provisions for archiving and preservation of licensed digital material. The last section of this chapter focuses on possible approaches to address the issues and problems identified, including changes to copyright law, model licence clauses and mechanisms for their implementation, and the use of rights metadata.

3.1 Summary of preservation problems

3.1.1 Media instability

Estimates of the likely life expectancy of various storage media vary from around 1 to 100 years. Rothenberg gave some low estimates, including as little as two years for magnetic tape in some circumstances (Rothenberg 1999a, p. 3). Unfortunately he gave no explanation for the low estimate and did not source his figures. The US National Media Laboratory contested Rothenberg's estimates; it cites a 5-30 life expectancy for magnetic tape and between 5 and 100 years for optical discs. Even so, these projections do not compare well with established archival media such as permanent paper or preservation microfilm. For these carriers, life expectancy is hundreds of years with optimal conditions.

As well as having inherent instabilities, the physical carriers used for digital information also react to environmental factors. These factors include both extremes of and fluctuations in temperature and relative humidity. Physical media also suffer from wear and tear and incorrect handling (Kranch 1998, p. 138, Feeney 1999b). Van Bogart produced a report on the storage and handling of magnetic tape, which is widely quoted in the literature (Van Bogart 1995). The National Institute of Standards and Technology (NIST) in the US are also carrying out some longevity testing and investigation of the deterioration process in optical discs (Lee et al 2002, pp.101-102). Byers (2003) of NIST summarised current knowledge on the care and handling of optical media. Other relevant studies at NIST include testing of interchangeability and interoperability of optical discs for use in high density storage systems and the suitability of different types of high capacity systems for different applications, including preservation. The Turbo coding system could be used to facilitate the recovery of information for failed discs.

3.1.2 Technological obsolescence

While the relatively short life expectancy of digital media is acknowledged, developing archival quality media will not ensure the preservation of digital information. In the print environment, preserving the medium or the artefact effectively also preserved the information scratched, written or printed on it. Media instability is not the main problem as far as the preservation of digital information is concerned. The main problem is that viewing and using digital information requires the aid of equipment. The biggest threat to long-term survival is that of technological obsolescence of the hardware and software used to create, view and use digital information.

Recognition of technological obsolescence as the main threat to the long-term survival of digital information becomes prominent in the library and information science literature from the mid-1990s. Lehman (1996) sets out some of the aspects of technological change. These include changes in coding and formats, software, operating systems and hardware. These changes can render digital material unreadable. Granger (2002) has commented on the actions of technology vendors in creating hardware and software obsolescence requiring customers to constantly

upgrade technology and perhaps locking them into proprietary technology. While customers may resist this if taken too far by vendors, Granger points out that preservation problems will be created before that happens.

3.1.3 Complexity of digital information

Digital information can be complex in various ways. Text, sound, still and moving images can be created in, or converted to, digital form and combined. However, these different types of digital information are encoded in different formats. For example, text formats include ASCII, RTF, various word processor formats, and markup languages. Visual material can be encoded in different formats such as jpeg, tiff, and bitmaps. There are many different standards, which may not be implemented uniformly or may change rapidly. Specialised symbols such as mathematical symbols, chemical formulae or musical notations do not have standard digital representations.

In addition to the plethora of standards, complex digital material may also be dependent on software for search and retrieval and other functionality.

3.1.4 Security and copyright protection

Digital information may be surrounded by technology designed to protect it from unauthorised copying and redistribution. This may inhibit or prevent preservation actions. As will be discussed later (Section 3.4.6) the 2003 UK copyright regulations give legal protection to technological protection measures, and also to digital rights management information (Great Britain 2003a).

3.2 Summary of preservation approaches

There is confusion in the terminology used for the preservation of digital information. Digital archiving and digital preservation are used interchangeably and mean different things to different communities. From the library perspective, digital archiving is

more likely to refer to the capture of material, whereas digital preservation refers to the actions taken to make sure this information remains accessible. Another example - and a regular source of confusion - is that there is a difference between digital preservation and preservation digitisation (Russell 1999). Digital preservation is "the storage, maintenance, and accessibility of a digital object over time". Preservation digitisation involves digitising a fragile object to preserve its intellectual content. Preservation digitisation produces a surrogate for the original object. It should be noted that creation of digital surrogates of non-digital material is not currently recognised in the UK as an acceptable preservation strategy because of the uncertainty over how to ensure the survival of the resulting digital objects over time. However, many libraries are digitising their collections to improve access for users. Given the level of investment this requires, libraries are likely to want to ensure the resulting digital collections remain accessible over time.

There are a number of potential preservation strategies that address different preservation requirements and timeframes. A key question in deciding what strategy to use is what is to be preserved. In the rapidly changing digital environment, preserving the physical carrier of information does not necessarily mean that the information itself is also preserved because of technological obsolescence. Wheatley (2001) suggests that a number of questions need to be considered when deciding how and what to preserve. These are,

- Who are we preserving for?
- Why are we preserving?
- What are the relevant significant properties of the object, given the answers to the first two questions?
- Which method of preservation will most accurately preserve these significant properties within time and cost constraints?

The Cedars project (CURL Exemplars in Digital Archives) has investigated the idea of Significant Properties in digital preservation decision making. Significant properties refer to the level of content and functionality to be retained (Cedars 2001, p. 14). When identified, the Significant Properties “determine the underlying

technical components that need to be documented and supported to ensure preservation of those Significant Properties". The Cedars project calls this the Underlying Abstract Form (Cedars 2001, p. 14). It may be difficult to specify exactly what significant properties are and what to save in order to preserve them. This is particularly the case for complex digital information such as multimedia or highly interactive information. Text, sound and pictures may be integrated; the software associated with the information may allow interaction between the user and the information. The retention of *all* aspects of look and feel and functionality may only be appropriate for items that are considered to be core to a particular collection (Cedars 2001, p. 14).

The main possible preservation strategies at the time of writing are technology preservation, migration and emulation. The focus of technology preservation and emulation are on preserving the technological environment, while the focus of migration is on changing the information so it can be rendered in a new technological environment. Technology preservation involves no change to the digital information other than refreshing it. While technology preservation is largely discounted as nothing more than a stop-gap measure, there has been ongoing work on migration and emulation over the last few years, most notably by the Cedars (http://www.leeds.ac.uk/cedars/) and CAMiLEON (http://www.si.umich.edu/CAMILEON/) projects. The Koninklijke Bibliotheek (KB) in the Netherlands has been working with IBM Netherlands on the Universal Virtual Computer approach, which looks like a way of implementing migration and emulation strategies.

3.2.1 Technology preservation

This involves preserving the information in its original form and also the original software and hardware used to create and access the information. The strategy is likely to also involve media refreshment, especially for information stored on media with very short lifetimes (Hendley 1998, p. 17). This can only be a short-term solution for a variety of reasons, including space, maintenance and costs. Hardware can only be maintained in working order for a finite period. Documentation and expertise in old systems disappears over time (Lee et al 2002, p. 95). Storage media

become obsolete and unavailable over time, so the information cannot be refreshed and still be readable in the original disc drives. Access is dependent on proximity to the hardware (Cedars 2001), so the digital material is not portable.

3.2.2 Migration

The Task Force on Archiving of Digital Information was set up in 1994 by the Commission on Preservation and Access and the Research Libraries Group (RLG) , to investigate and make recommendations about how to ensure "continued access indefinitely into the future of records stored in digital electronic form" (Research Libraries Group 2002). The Task Force favoured the migration approach to digital preservation. The Task Force report defines migration as "the periodic transfer of digital material from one hardware/software configuration to another, or from one generation of computer technology to a subsequent generation." (Waters & Garrett 1996, p. 5)**.**

Feeney (1999b) provided an overview of different migration strategies. A very basic migration strategy is "refreshing" information by copying it from one type of physical carrier to a new one. Another simple strategy is changing media. This may involve moving the information from a less stable medium, such as magnetic tape, to a more stable medium, such as optical disk. A rather ironic version of this particular strategy is printing out digital information onto archival quality paper. This may be useful for documents that have similar attributes to paper documents, such as letters and text-based reports. However, this strategy is not suitable for more complex digital information, as much of the functionality of the publication will be lost. However, conversion to non-digital formats is being used to facilitate a digital migration strategy. According to Thibodeau (2002, p. 24-25), Rajasekar is developing the Rosetta Stones Translation approach. There are no other references in the literature to this approach, but Thibodeau states that it involves assembling a representative sample of objects of a particular type. The sample would include all the significant features of a particular format. Each of the items in the sample is duplicated in a format that is easily read by humans, such as paper or microfilm. This second set of documents is the reference set. Alongside the target format type, these two sets of documents can be used to define the rules for translating from the original to the

target format. Rather than migrating already migrated material, the original digital documents and the hard copy reference set can be compared with new target formats as they are introduced and new migration rules developed.

Another migration strategy relies on the backward compatibility of application software. Information created in an older version of the software is loaded into the new version and saved in this format (Feeney 1999b, p. 44). However, this would require regular migration of material because backwards compatibility may only last for a few generations of software. A similar alternative relies on interoperability between competing applications. Information is saved in one application in a common interchange format and is then imported into a rival application.

A final migration strategy involves conversion to standard formats (Feeney 1999b, p. 45). The aim of this strategy is to tackle the problem of the large numbers of formats used for creating digital publications. Collection managers specify a manageable number of preferred formats and publications in other formats are converted. XML is widely viewed as a good standard to use, although concern is being expressed that the multitude of XML specifications being implemented may hinder the exchange of data and lead to wasted money and effort (Donoghue 2003). Work is also underway by AIIM International (2003) to develop a PDF format (PDF-A) specifically for use in long-term preservation. According to Thibodeau (2002, p. 24), in principle, the standard format should be a "superclass of the original data types – one that embodies all essential attributes and methods of the original formats. Alternatively, collection managers only accept publications in one of the preferred formats. While this may simplify the migration task, migration is still likely to be required because even standard formats change over time and in any case, they may not be implemented uniformly. Thibodeau (2002, p. 24) refers to Typed Object Model (TOM) conversion. Digital objects can be grouped into types according to values such as "attributes, methods, or semantics" and the essential properties of a type of object can define "respectful conversions". Respectful conversions mean that although the format is changes, the essential properties stay the same.

Rothenberg equates migration to translation "not only does each translation lose information, but translation makes it impossible to *determine* whether information has

been lost, because the original is discarded." (Rothenberg 1999a, p. 11). More recent approaches to migration address this issue as is discussed below. According to Rothenberg, migration cannot deal with paradigm shifts in technology, such as the moves from hierarchical to relational to object-oriented model databases. The reason for this is that "paradigm shifts do not necessarily provide upward compatibility" (Rothenberg 1999a, p. 13). Rothenberg made similar comments in 2000 (Rothenberg 2000, p. 1).

Wheatley (2001) considers the migration strategies described by Feeney (1999a, 1999b) as “traditional”. In 2001, he published a paper with the aim of clarifying the meaning of migration and set out different migration practices and the issues arising from these. He mentions the categories of migration included in the Open Archival Information System (OAIS) international draft international standard (Consultative Committee for Space Data Systems, 2002). These are refreshment, replication, repackaging and transformation. Wheatley considers the first three to be processes related to the management of a digital archive, used to preserve bit streams and maintain a “reliable package” in the OAIS archive. Wheatley thinks that only transformation equates to migration in that the bit streams are actually transformed. The OAIS model also specifies that migration can be reversible or non-reversible. From a preservation point of view, reversible migration would be preferable, because the success of a migration action could be tested through reversing the action and assessing whether any information has been lost. The reversibility of preservation actions is also part of the traditional ethical code of conservation. However, there is a problem with this in that it may not be possible to reverse a migration that was carried out in the past because the original migration tool and platform may not be available (Cedars 2001, p. 64)

Wheatley (2001) breaks down migration into specific cases. He calls the basic activity of preserving the bit streams minimum preservation. His migration categories are: minimum migration; preservation migration; recreation; human conversion migration and automatic conversion migration. Minimum migration involves a minimal amount of technical work to make material viewable. This could be a version of the conversion to standard formats mentioned by Feeney above. It looks as if this would be appropriate when only the raw content is required because it involves

stripping away features such as formatting. Preservation migration combined with the minimum preservation option allows some aspects of the look and feel to be preserved, but in a non-technical way. This could include recording screen shots or the key processes involved in using the material. These actions could be accompanied by textual descriptions of the look and feel. Wheatley sees recreation as a separate, but related, process. Recreation creates a new digital object, which represents the significant properties of the original, but does not incorporate any elements of the original digital object. This is a costly process. Human conversion migration recreates the software elements of the digital object but reuses as much of the data of the original object as possible, creating a more accurate reproduction of the digital object than recreation. Automatic conversion migration involves software tools to convert the digital object to a current environment. Wheatley considers this approach to be an example of the traditional view of migration described above.

The Cedars migration on demand approach has been further developed using ideas developed through the CAMiLEON project emulation work, described below. In theory, the CAMiLEON migration on request tool gets over some of the disadvantages of the strategy identified by Rothenberg and others (Mellor, Wheatley & Sergeant 2002). First of all, all migrations are performed on preserved identical copies of the original bit streams; digital objects are maintained in their original formats. The bit streams are converted from their original formats to new formats as needed in the future using input modules that can run on current platforms. Each class of original format requires one input module, these are developed as new formats appear. The output modules are amended to run on new platforms as technology moves on. So, the problem of introducing errors that are then compounded through successive migrations is removed. Migrations are only carried out as needed and the coding of the input module never has to be rewritten, saving resources. The approach has been tested using Windows Meta Files, Draw and Scalable Vector Graphics files and the conclusion was that the migration tool developed successfully imports, converts and exports a number of vector graphic formats. An apparently successful reverse migration was also carried out. The tests did raise problem areas to be explored. Although this looks like it addresses many of the disadvantages of periodic migration, over the long-term the migration tool could become very large as more and more input modules are added.

3.2.3 Emulation

The aim of emulation is to allow long-term preservation of digital material by the use of software to allow new technology platforms to mimic the behaviour of older platforms. The aim is to retain the functionality and look and feel of the material. Granger has briefly described different approaches to emulation (2001).

Jeff Rothenberg (1995) has been a major proponent of emulation as a preservation strategy for digital information since the mid-1990s. Emulation can be carried out at different levels, including application software used to create and view material, operating system or hardware. According to Rothenberg's initial thinking, the behaviour of the software or hardware can be described and the description saved so that its behaviour can be re-created in the future. This would need the development of emulators, or software programmes to mimic this behaviour based on the descriptions or emulator specifications (Rothenberg 1999a, p. 15). For Rothenberg, hardware emulation is potentially a simpler proposition than software emulation. The reasons he gives for this is that there are fewer hardware platforms than operating systems and application software, so fewer emulators would have to be specified. Secondly, writing specifications for hardware is a better-developed practice than for software, so it would be easier to do (Rothenberg 1999b, p. 22). The requirements for this approach would include saving the digital documents, the programmes that were used to create the documents and all software required to run the documents. Rothenberg set out the requirements for implementing emulation of hardware (1999b). These include:

- techniques for specifying emulators
- techniques for saving the necessary metadata (for finding, accessing and recreating documents) in human-readable form
- techniques for encapsulating documents, attendant metadata, software, and emulator specifications in a coherent and incorruptible way

Rothenberg identifies what he calls an ancillary issue that would arise from the use of emulation as a preservation technique. This ancillary issue is intellectual property rights, but Rothenberg does not elaborate.

Rothenberg's original emulation approach had its critics. One of these is David Bearman who set out his objections to Rothenberg's ideas in an opinion piece (Bearman 1999). Bearman accuses Rothenberg of misunderstanding the preservation needs of electronic records. He says "Rothenberg is fundamentally trying to preserve the wrong thing by preserving information systems functionality rather than records. As a consequence, the emulation solution would not preserve electronic records as evidence … and is serious overkill for most electronic documents where preserving evidence is not a requirement". In making the comments about records and evidence, Bearman himself has misunderstood Rothenberg's purpose. His misunderstanding may have arisen because of the confusion about the meaning of "archives" and "archiving" in the digital environment. Rothenberg does not talk about electronic records; he talks about the type of digital documents held in libraries. So the preservation of functionality could be a legitimate aim of preservation strategies. Bearman may well be right in his claim that emulation may be overkill in certain situations. However, digital publications are increasingly rich in functionality, so the accusation of overkill for most electronic documents in the library environment is probably not accurate in the library context.

Gilheany (1998) has described his version of emulation. This will involve saving the original item and creating an emulator that will run on any standard model computer instruction set. The emulator is "the binary embodiment of a specific configuration of hardware in binary form" along with the original item. According to Gilheany's approach there would only ever be two layers of emulation: "the original emulator that runs on a Turning [sic] Machine … and "a Turing Machine Emulator" that runs on some future computer (Gilheany 1998, p. 4). A Turing Machine Emulator is only created when access is needed from some future platform.

Rothenberg further developed his approach through an experiment in emulation with the Koninklijke Bibliotheek in the Netherlands (KB). The aim was that the experiment would go through several iterations throughout the period 1999 to 2001.

The plan was that the first stage would result in a design for the whole experiment, a plan for testing and comparing the results of the emulations with the original works and a framework of preservation criteria and authenticity characteristics. The second stage involved modelling the emulation process and identifying metadata and functionality requirements. At the time the NEDLIB (Networked European Deposit Library) project, led by the KB, was building a demonstrator digital deposit system. Publishers participating in the project were providing material to feed into the system to test the handling functions. The last stage of the emulation experiment was to be the implementation and evaluation of the emulation process in the testbed.

The first stage of the emulation experiment was completed in 1999. In the resulting report, Rothenberg set out his current thinking of how emulation for the preservation of digital documents would work. Like Gilheany's approach layers of emulation would be required (Rothenberg 2000, pp. 7-13). The so-called "emulation virtual machines" that would run on a future platform would have to be emulated in turn when the these platforms themselves become obsolete. The reason for taking this approach is apparently to minimise the amount of work needed for emulation since nothing would have to be rewritten. All that would be needed is a new emulator specification each time a platform becomes obsolete to allow the previous emulator to work on the next new platform.

Rothenberg concluded that: "The results of this study suggest that using software emulation to reproduce the behavior of obsolete computing platforms on newer platforms offers a way of running a digital document's original software in the far future, thereby recreating the content, behavior, and 'look-and-feel' of the original document" (Rothenberg 2000, p. 83). This claim seems somewhat inflated since the actual experiment actually involved running Windows 95 publications on an Apple Mac using Connectix VirtualPC software as the emulator. The most Rothenberg can claim is that this particular software does what it says it does. The NEDLIB project has now ended and there is no sign of any reports on the other stages of this experiment, so it is likely that it was never completed. In fact, the KB is now working with IBM on long-term preservation and Rothenberg has some involvement in this.

Rothenberg is not the only researcher working on emulation. Holdsworth and Wheatley (2000) of the CAMiLEON project are nervous about Rothenberg's reliance on the production of specifications for emulators that can be built at some time in the future. No matter how carefully the specification is written, there is no guarantee that it can actually be used to build a future emulator and no way of testing that the emulation is successful, because it will not be possible to compare it with the obsolete original technological environment. Holdsworth and Wheatley suggest an alternative approach involving developing an emulator at the point of obsolescence and designing the emulator in such a way that it should only require minimal amendment to run on the next generation of computers. They have also considered the level at which emulation should be carried out. Whereas Rothenberg advocates emulation at the hardware level, Holdsworth and Wheatley recommend at the Application Programming Interface level - the interface between the application software and the operating system. They identify a number of factors that should influence the choice of emulation interface. These are lack of complexity, availability of documentation, mapping of peripherals to easily specified abstractions and retention of the significant properties of the digital object. When discussing levels of emulation, Thibodeau (2002, p. 19) considers that emulation of application software would be more difficult than operating systems, basically because there are more of them. Whether the CAMiLEON approach is a simpler and more efficient approach than the one advocated by Rothenberg remains to be seen. Neither approach has been tested by practical experience over the long-term. However, the CAMiLEON team have at least successfully emulated an obsolete system, unlike Rothenberg who only demonstrated emulation on contemporary but non-compatible platforms.

It is worth noting that Thibodeau argues that the founding principle of emulation is that "all computers are Turing machines and that any command that can run on one Turing machine can run on any other Turing machine" and that this principle breaks down at an empirical level (Thibodeau 2002, p. 20). This would make emulation unreliable.

3.2.4 The Universal Virtual Computer approach

Lorie (2001) suggests that whilst emulation could mimic obsolete technologies to allow display of the original, it will not allow future users to manipulate the information. The aim of the Universal Virtual Computer (UVC) idea is to use software engineering to develop "…virtual machines that can execute essential functions on a variety of platforms" (Thibodeau, 2002, p.22). The UVC will be a program comprising low-level instructions. The theory is that "The rules are written in machine language that is…so simple that it can be interpreted to run on any computer in the future. (Thibodeau 2002, p.22)"

Lorie distinguishes between data archiving and program preservation.

> *For data archiving, we propose to save a program P that can extract the data from the bit stream and return it to the caller in an understandable way, so that it may be exported to a new system. The program P is written for a Universal Virtual Computer (UVC). All that is needed in the future for executing P is an interpreter of the UVC instructions. The execution of P in the future will return the data with additional information, according to the metadata (which is also archived)* (Lorie 2001).

> *In the future a restore application program reads the bit stream and passes it to a UVC interpreter, which executes the UVC program. During that execution the data is decoded and returned to the client according to a logical view (or schema). The schema itself must also be archived and easily readable so that a future client may know what a schema is. In our approach, the logical schema is an XML-like structure. When read, all data items are returned to the user, tagged with a semantic label. The structure of the returned data and the tags are defined in the metadata.* (Lorie 2001).

Data archiving sounds like a form of migration, in fact it also sounds similar to the Cedars / CAMiLEON approach in the sense that it involves a tool to extract

information from its native format using a logical representation of the information. It looks like the difference is that the migration on demand tool then converts data to the target format, whereas data archiving presents the data to the user, then they can decide whether to import it. The other difference is that migration on demand would not require an intermediary UVC, as the migration tools would be maintained over time. In the UVC approach the interpreter is crucial. Lorie (2001) argues that the interpreter should be separate from the preserved information and it must also be universally accessible. The UVC program will remain the same, but the interpreter could be changed to accommodate new technologies. IBM has carried out an apparently successful proof of concept experiment on PDF files (Lorie 2002)

According to Lorie, under the UVC approach, emulation will still be necessary to preserve "programs", or the functionality of complex digital information. Given that Lorie is actually working on this approach it must be assumed that Thibodeau misunderstands the UVC approach because he says that "the virtual machine approach avoids the need for an emulator by providing required functionality in a virtual machine that, in principle, can be implemented on a great variety of computing platforms indefinitely into the future." (Thibodeau 2002, p. 22).

3.2.5 Other suggestions to aid digital preservation

In addition to the more mainstream preservation strategies, there are some other suggestions for facilitating the preservation of digital information. Thibodeau discusses the configurable chip idea where an old computer is recreated on a chip. When the chip deteriorates or becomes obsolescent, its contents are just copied onto a new chip (Thibodeau 2002, p. 20). It is not clear what the practicalities of implementing this strategy would be.

Robertson's (1996) Digital Rosetta Stone model works with encapsulation techniques. It involves storing the information needed to interpret digital objects separate from encapsulated objects. Because many objects will be created by or run on popular application or operating software, the Rosetta Stone repository will avoid duplication of effort and promote efficient use of storage space. The Universal Preservation Format (UPF) was being developed upon the theory of encapsulation (UPF Home

2000). The UPF idea involves a wrapper to hold the essence (the digital object) and the metadata together. The wrapper is a "file format that has a framework structure" (Shepard & MacCarn 1998). The UPF could be a "self describing" format through use of identifiers, metadata and wrappers. The UPF would use a "digital Rosetta stone" as a key, for defining data types and encapsulating algorithms, and perhaps also as a local registry for unique identifiers. In order to be truly self-describing there should be some non-digital element. This could perhaps involve print, but Norsam Technologies have developed a nickel-based disc that has a long life expectancy and also stores information created digitally in analogue form. All that would be required to read the contents on the discs would be magnification.

Lynch (1999) discusses the role of canonical formats and canonicalisation in digital preservation. This would facilitate preservation strategies such as migration and help ensure that Significant Properties are retained. The approach assumes that a canonical form that captures the "essential characteristics" for a class of digital objects can be defined. There could be a hierarchy of canonical forms with some forms providing "more detail or richer semantics" than others. It should be possible to translate all data formats used to encode a given type of object to the canonical form. The canonical form will not include irrelevant data included in some data formats, so a reverse translation may not result in a identical object at the bit level. If reversibility of migration actions is important, then there is the question of what would be an acceptable definition of reversible. Lynch suggests that canonical formats may be useful for authenticity purposes. Rather than digitally signing the original digital object, a depositor would sign the canonicalised object. Hashes computed over this object would apparently remain valid. The results of a migration can be verified not by reversing the migration, but by checking that the hashes for the original and new object match. Lynch thinks that there would be little difficulty in developing canonical formats for image data, but research would be required to do this for other types of information. There is already work being carried out on XML structured objects.

Various groups have recognised the importance of retrieving and preserving information about file formats; these are needed to develop digital preservation tools and strategies. The National Archives in the UK has produced the PRONOM

database of file formats (Darlington 2003), and the Representation and Rendering Project (2003) has assessed different sources of file format information. Both have found that file format information is difficult to acquire, particularly for formats which have already been discontinued, or if a vendor has been taken over. Documentation which is available may be inaccurate or incomplete.

An initiative that has the potential to assist publishers and libraries implement archival clauses in licence agreements is LOCKSS (Lots of Copies Keeps Stuff Safe http://lockss.stanford.edu/). LOCKSS is a distributed digital archiving system, involving (apparently) low cost persistent digital caches of electronic journal content maintained in institutions subscribing to the journals. Libraries take custody of material in all formats delivered via HTTP, rather than have publishers deliver the material on an offline medium. Material is collected as it is published. Permission to do this is acquired at the point of subscription, with publisher indicating whether their material is "LOCKSS compliant", so libraries do not need to negotiate permissions on an individual basis. LOCKSS caches cooperate to detect and repair preservation failures. The technology will be supported by the LOCKSS Alliance which will provide services to libraries and publishers wishing to use the LOCKSS system. LOCKSS has the potential to address the problem of continuing access to material, according to licences agreements, when it is not available from the publisher or a subscription ends. What is not clear is how this initiative will be sustained in the long-term when different libraries are likely to have different versions of material if they carry out preservation actions.

A further relevant initiative is JSTOR, the Scholarly Journal Archive (http://www.jstor.org/). Digitising old editions of journals is advantageous for libraries, since it saves shelf space and improves access, but it is too expensive for each individual library to do this. JSTOR provides a central service which digitises journals from many libraries and makes them available to all, so that the costs and benefits can be shared. JSTOR is aware of the need for digital preservation and has a dedicated 'Electronic-Archiving Initiative'. JSTOR offers a possible model for the preservation of born digital materials which libraries access but do not own.

3.3 Copying implications of preserving digital information

Each of the preservation strategies discussed above involves some kind of copying activity. These range from the straight replication of digital information to changing the formatting in a way that may change the look and feel. Digital preservation strategies may also involve developing new software to emulate original technological environments or even writing new software to recreate original software environments. The digital preservation actions that will have copyright implications are summarised in the table below.

STRATEGY	COPYING ACTIONS REQUIRED	REQUIREMENTS AND OUTCOMES OF PRESERVATION ACTIONS WITH POSSIBLE RIGHTS IMPLICATIONS
Media refreshment Media change	1. Copying from an old medium to a new one of the same type 2. Copying from a digital medium to paper or microform 3. Copying from an one type of digital medium to another	
Migration Universal Virtual Computer	1. Media refreshment and media change 2. Conversion of content formats • Conversion to standard/interim/target formats • Repeated migration • Migration on demand 3. Recreation of interfaces; reverse engineering of software 4. Recording of "look and feel" 5. Recreation of content	Depending on success of strategy, loss of look and feel and functionality Acquisition and preservation of copyrighted format specifications, software and software specifications
Emulation Universal Virtual Computer	1. Media refreshment and media change 2. Encapsulation of content, application software, operating software, hardware specifications, software specifications 3. Reverse engineering of software to create emulators 4. Developing software to allowing older software to run in a new technological environment	

Table 3.1 Preservation strategies and copyright implications

3.4 Copyright Law in the UK

This section concentrates on the relevant UK legislation and its implications for digital preservation. In the UK, the relevant law is the *Copyright Designs and Patents Act* (Great Britain, 1988). A new statutory instrument, the Copyright and Related Rights Regulations 2003 (Great Britain 2003a), was published on 31st October 2003 and came into effect at the end of 2003. This implements the European Union Copyright Directive (European Parliament and Council of the European Communities 2001).

3.4.1 The Copyright Designs and Patents Act (1988)

Copyright comes into being at the moment of creation of a work, and no formal procedure to register a copyright is required, or available, in the UK. The copyright owner has the exclusive right to reproduce a work, issue copies to the public, rent or lend, perform, show or play the work in public, broadcast the work or include it in a cable programme service, or make an adaptation of the work. The copyright owner also has the right to prevent third parties from carrying out these "restricted" acts without prior permission. To carry out a restricted act on all or a substantial part of a copyright work without the permission of the owner, or authorise someone else to do so, is infringing the copyright in that work. Infringement of copyright can carry both civil and criminal penalties, depending on the nature of the infringement (Great Britain 1998, s.16)

There is copyright protection for specific classes of works. The different classes of work that are likely to be found in digital library collections include:

- Literary works, including novels, poetry and non-fiction and other written works that are original. Their literary merit is unimportant. Computer programs and code are also protected as literary works. Letters, memoranda, e-mail messages and Web pages are protected.
- Dramatic works must include some spoken words or actions to perform to distinguish them from literary works

- Artistic works include graphic works, photographs, sculptures, collages, maps, charts and plans, regardless of artistic merit
- Musical works and sound recordings recorded on any medium and musical scores including any annotations and directions. Lyrics are protected as literary works
- Films, including any medium from which a moving image may be reproduced
- Broadcasts, including any transmission by wireless telegraphy that is capable of lawfully being received by members of the public. This includes satellite transmissions

Published editions are also protected; there is copyright in the typography and layout of a literary, dramatic or musical work (Great Britain 1988, s.3 - s.8).

3.4.2 Copyright terms

Copyright exists for a limited period only. *The Copyright Designs and Patents Act 1988* (CDPA) was amended in line with a European Directive (European Parliament and Commission of the European Communities 1993) that harmonised the basic term of copyright in the EU at 70 years from the end of the year the author died (Great Britain 1988, s.12-15, Great Britain 1995). However, there are still some differences in lengths or terms of copyright protection between the different classes of work.

- Literary, dramatic and musical works are protected for the duration of the author's life until 70 years after his/her death.
- Works of joint authorship are protected for 70 years from the death of the last author
- Artistic works are protected for the duration of the author's life plus 70 years after his/her death.
- Published anonymous works are protected for 70 years from first publication
- Copyright in unpublished literary, dramatic and musical works which have been created by an known author known who died before 1 January 1969 expires on 31 December 2039. Copyright in works created by a known author

who died on or after January 1969 expires 70 years after the death of the creator.

- Copyright in unpublished anonymous or pseudonymous works created before 1969 expires on 31 December 2039. Copyright in anonymous or pseudonymous works created on or after 1969 expires 70 years after the creation date.

Films are protected for 70 years from the death of the last to survive of the principal director, the author of the screenplay, the author of the dialogue and the composer of the music specially created for the film.

While software is protected under UK copyright legislation as a literary work, there is an allowance for one backup copy for software to be made .

Copyright protection for some other works is set at 50 years.

- Sound recordings are protected for 50 years from first publication, but 50 years from fixation, if unpublished during that time
- Broadcasts and cable programme services are protected for 50 years from when broadcast first made or programme included in a cable service
- Computer generated works are protected for 50 years from first creation

Published editions are protected for 25 years from first publication of that edition. Publication or communication to the public of a previously unpublished literary, dramatic or musical or artistic work or film in which copyright has expired will result in 25 years of protection from first publication.

3.4.3 Exceptions to copyright

While copyright law provides very clear rights to copyright owners, it also provides limited exceptions to those who might wish to use the copyrighted work for legitimate purposes that do not damage the copyright owners' legitimate commercial interests. These are provided in Chapter III of the CDPA 1988 : "Acts Permitted in relation to

Copyright Works". Two key exceptions provided in the CDPA 1988 are "fair dealing" and "library privilege". Sections 37-42 of the Act deal specifically with library privilege. The library privilege of interest to this project is one which permits copying for purposes of preservation or replacement (Great Britain 1988, s.42). This permits a librarian or archivist of a *prescribed library* (including school, university, further education, public and government libraries) or *archive* to make a copy from any item in the permanent collection in order to preserve or replace that item, providing that the prescribed conditions are complied with. It also allows for the copying in order to replace an item in the permanent collection of another prescribed library or archive. Copying is only permitted where it is not reasonably practicable to purchase a copy of the item. Materials can only be in the permanent collection in the libraries of both the donor and the receiver and must be for reference use only. Materials on temporary loan such as interlibrary loans are not eligible and so the status of digital material that is subscribed to but remotely accessed is not clear. This exception only applies to literary, dramatic and musical works, not artistic works.

3.4.4 Databases

Much digital information is made available via one form of database or another. A European Directive on databases was issued in 1996 (European Parliament and Council of the European Communities 1996) and a Statutory Instrument was subsequently passed in the UK to implement the Directive and protect databases (Great Britain 1997). Databases, under some circumstances, can enjoy double protection: the database or *sui generis* right, and copyright. The database right applies where there has been a substantial investment in obtaining, verifying or presenting the contents of the database. The term of protection in this case is only 15 years, but may be renewed if there is a substantial change to the database. The database right prevents the unauthorised extraction and re-utilisation of material from a database, whether it enjoys copyright or not. As a result, both the copyright residing in the structure of the database and the database right restrict the transference of databases to another medium. However, it is not an infringement if a person, who has a right (by licence or otherwise) to use the database (databases as literary works), exercises their right to access the database and to use its contents by whatever means necessary. Any

term in the contract or licence that prohibits this is irrelevant (Great Britain 1988, s.50 (d1-2)).

3.4.5 Moral Rights

While this research is focused on copyright and licensing issues, it is worth noting that creators of material have moral rights under UK law that are distinct and separate from property rights. These include:

- The right of an individual author of a work to be acknowledged as the author or creator
- The right not to have his or her work subjected to "derogatory" treatment
- The right of an individual to refuse to be associated with something he or she did not create.

Moral rights cannot be transferred, but can be waived. Moral rights do not apply to creators of :

- computer programs
- the design of a typeface
- any computer-generated work
- the creation of any work reporting current events
- works that have appeared in newspapers, magazines or learned journals[3]
- other collective works
- most employee-created materials (Great Britain 1988, s.79)

3.4.6 EU Copyright Directive and digital preservation

The European Union Copyright Directive (European Parliament and Council of the European Communities 2001) was designed to harmonise various aspects of copyright law amongst the Member States. It was implemented in the UK on 31st October 2003

[3] It is worth noting that some people believe that this does not apply to scholarly journal articles.

as *The Copyright and Related Rights Regulations 2003* (Great Britain 2003a). The Directive provides extra protection for works communicated over networks and also provides extra protection for digital rights management systems. The Directive allows Member States some discretion with regard to what libraries within their jurisdiction are permitted to do in order to preserve copyright items. The directive recommends that exceptions and limitations should be defined more harmoniously. However, it does not make it *compulsory*. The existing exception for preservation copying (Section 42) has not been altered by the new regulations.

Recital 44 and Article 5.5 of the Directive confirmed that all exceptions to copyright are subject to the Berne "three step" test. This is mentioned in the explanatory note that follows the UK Regulations. This test is used as a standard in framing exceptions to rights and ensures that the exceptions are not in conflict with the normal exploitation of the work and do not unreasonably prejudice the legitimate interests of the right holder.

Section 24 of the new UK Regulations (Great Britain 2003a, s.24) provides legal protection against the circumvention of technological measures which provide legal protection for copyrighted works. It also provides legal protection against the manufacture of products which could enable such circumvention. The expression "technological measures" is defined in section 296ZF as:

> *any technology, device or component which is designed, in the normal course of its operation, to protect a copyright work other than a computer program* (Great Britain 2003a, s.296ZF)

Effective technological measures include the application of an access control or protection process, such as encryption, scrambling or other transformation of the work or other subject-matter or a copy control mechanism.

The Directive also states that Member States must ensure that people can make copies of works protected by effective technological measures, if there is an exception that permits them to do so. Section 296ZE of the UK Regulations states that a user may appeal to the Secretary of State if effective technological measures prevent them from

carrying out a 'permitted act'. The regulations specifically state that preservation copying by libraries and archives is included in this. It is illegal to produce or distribute tools which enable the circumvention of such measures (296ZB). This could hinder legitimate copying for preservation, since it would be difficult to prove that any such tools were only intended for use in creating copies under exemptions, and not infringing copies.

Section 25 of the Regulations deals with obligations concerning rights management information. This provides legal protection against the removal or alteration of any electronic rights management information. It also instructs against the distribution, importation for distribution, broadcasting, communication or making available to the public of works or other subject matter from which electronic rights management information has been removed or altered without authority. Rights management information is defined in Section 25 as:

> *… any information provided by the copyright owner or the holder of any right under copyright which identifies the work, the author, the copyright owner or the holder of any intellectual property rights, or information about the terms and conditions of use of the work, and any numbers or codes that represent such information.*

3.5 Access rather than ownership issues

For traditional media, libraries acquire and physically own a discrete physical information object. This is not the case for much digital information where the model is paying for access to information held remotely. Access may be directly from the publisher or through some sort of intermediary, such as an aggregator. Libraries may or may not have access to material that is not current. With print material, cancelling subscriptions would only result in loss of access to future material; previously purchased material would still be available on library shelves. This may not be the case with digital material. The question of responsibility for the preservation of information arises here. If libraries do not physically own digital material, they cannot preserve it. Publishers may or may not have a commitment to preserving their own

information; this is likely to depend to an extent on the type of publisher and its mission.

The ephemeral nature of much online information has implications for digital preservation. According to Brewster Kahle, founder of The Internet Archive, the average life of a Web page is 100 days (quoted in Mayfield 2001). One of the great characteristics of the World Wide Web is the facility to create links between distributed pieces of content. Online material can be added to or amended quickly and easily. Web pages disappear every day as their authors revise them or servers are taken out of service; but users become aware of this only when they click on a hyperlink or a URL. The extent of a hyperlinked digital publication is also difficult to identify when deciding what to preserve. While some types of organisation may take responsibility for preserving their material, others may not.

Other new ways of disseminating digital information, for example Open Archives, may also present preservation problems. The Open Archives Initiative (OAI) is developing and promoting interoperability standards to enhance access to e-print archives (Open Archives Initiative [n.d.]). While some see preservation as a distraction from the OAI's primary purpose of improving the availability of scholarly communication, the preservation of e-prints is now being addressed (James et al, 2003). Pinfield and James (2003) argue that e-prints should be preserved as well as published journal articles; this will ensure that access and references will be maintained for the long-term, and is particularly important where e-prints differ in some way from the published article.

3.6 Analysis of the extent to which copyright law and licensed access to digital information meets digital preservation copying needs

The only type of copying that seems to be clearly legal under the preservation exception in UK law is the first act of media refreshment or migration. If digital information is copied to paper or microform, then users would be able to access this version of the information because use requires no further copying. It is not clear whether the law would allow the following:

- The creation of multiple copies for the purposes of redundancy. (Only one copy of software can be made under the law)
- Periodic refreshment and media migration

It could be argued that these actions fall within at least the spirit of the law although they require multiple acts of copying of the same material.

While researchers are now looking at how to minimise preservation interventions and avoid repeated conversions or emulations of emulations, it really is not clear whether these technical strategies and the gathering and preservation of documentation that would be needed to implement them would be allowed under current provisions. This also seems to be the conclusion of the final report of the Cedars project, although, strangely, library and archive preservation copying is not mentioned in its consideration of copyright issues. The report provides guidance on how preserving institutions should go about seeking permission to copy for preservation purposes, but does not actually give reasons why existing preservation and archiving provisions are not considered sufficient (Cedars Project 2001, p. 51).

Charlesworth's assessment of the copyright issues arising from the emulation of the BBC Domesday Project as part of the CAMiLEON project concluded that the copying and emulation envisaged for the project might be defended under Section 42 of the CDPA 1988 and was unlikely to be seen by rights holders as detrimental to their commercial interests. However

> … *further use of those works during their term of copyright, including their use in published materials or in public displays, without further permissions from rightholders being obtained and/or royalties being paid, clearly may attract complaints, and even legal action, from some rightsholders.* (Charlesworth 2002, p. 12)

The whole point of preserving information is to make it accessible in the future. Charlesworth does not make entirely clear the rights implications of making an

emulated work such as the Domesday Project available in a library. However, it looks as if developing an emulation may be allowed under Section 42 of the CDPA 1988, but making emulated material accessible will infringe copyright (Charlesworth 2002, p. 7). The UK National Archives has been granted permission by BBC Worldwide to make its re-engineered version of the Domesday Project available in its reading rooms (Darlington, Finney & Pearce, 2003). However, the CAMiLEON Project was refused permission to allow access to its emulated version of the Domesday Project. It is not clear why this is the case.

In an interview with Abbott (2003), Wheatley argues that:

> *in some cases emulation may involve a simpler copyright question than migration, simply because less is actually changing when preservation action is taken.*

Emulation involves making the original software and data work on a modern computer, so that the original data can be accessed. However, some other preservation methods actually change the original data. This could be viewed as creating a completely new version rather than just a copy, and Wheatley suggests that this could cause greater problems with rights holders. However, there is no legal precedent for this, and it is difficult to see how precedents from the print environment could be applied here.

In addition to the copying activities required to implement technical strategies, Mauritzen & Solbakk (2000, p. 12) point out that there may be moral rights issues arising from migration activities if they result in changes to the migrated material. Another possible issue that is not mentioned in the literature is whether changes to look and feel might also infringe the rights publishers have in published editions.

There is also the question of bypassing technical protection measures to carry out these actions. The new UK copyright regulations (Great Britain 2003a) should allow libraries to circumnavigate copyright protection to exercise their privileges, since proposed amendments to exceptions, as well as other unamended exceptions to copyright in the UK legislation, are considered to comply with the three-step test.

Libraries must first apply to the publisher for permission to bypass copyright protection measures to make a copy allowed under an exception. If permission is refused, they can appeal to the Secretary of State. Since the preservation exception does not allow much copying for preservation purposes, this may not be particularly useful anyway.

Copying of material that libraries can access, but do not own, does not seem to be allowed under library privileges. However, the meaning of "permanent collections" in the digital environment is not clear. There is also the possibility of copying publicly accessible material in order to archive and preserve it. Harvesting is used for Internet information; software is used to identify and pull in information from sites. According to Charlesworth (2003, p. 8)

> *… with regard to digital archiving, especially Web archiving, the legislation as currently worded is not terribly helpful. If rights owners control access and use through licence agreements and access is remote, libraries are dependent on them to continue to provide access to the material.*

If current law does not allow copying for digital preservation, the most obvious solution is to change the law. If libraries want to preserve information, they need to be able to carry out the required activities on material in their physical possession. However, the interests of rights holders need to also be considered. While they may agree with the principle of preservation of digital material, they may have concerns about the implications and safeguards on the use of preserved material would need to be in place. Legal deposit libraries have a role to play in the long-term preservation of digital information. If copyright law does not allow preservation copying, an alternative is to ask for permission to copy for preservation purposes. For material that is accessed remotely on a licensed basis, then licence agreements could provide provisions for archiving and preservation. These approaches are discussed below.

3.7 Rights for digital preservation

3.7.1 Legal Deposit

The Legal Deposit Libraries Act 2003 (Great Britain 2003b, s.6) requires UK publishers to deposit print materials with six legal deposit libraries. This Act also enables the Secretary of State to make Regulations extending legal deposit to non-print materials. IFLA (the International Federation of Library Associations and Institutions) has issued guidelines on legal deposit legislation (Larivière 2000), including dealing with digital information. Legal deposit is a statutory or other type of legal requirement that everything "published" in a particular country is deposited with one or more designated depositories who take on long-term preservation responsibility for the information. Various countries have, or are planning, to extend their legal deposit legislation to include digital publications. Extending legal deposit to cover digital publications may get over the problem of access to rather than ownership of digital information. However, this will only address the issue to a limited extent. Legal deposit only covers the national digital output and libraries may wish to preserve material that originates elsewhere.

At present, and until the Regulations for the legal deposit of non-print material in the UK are drafted, the UK legal deposit libraries are operating a voluntary scheme for handheld digital material. The British Library has also been involved in limited experimentation of gathering Internet-based material. The UK Voluntary Deposit Scheme has been monitored by the Joint Committee on Voluntary Deposit, which included representatives of the legal deposit libraries and the publishing community in the UK. The expanded Joint Committee for Legal Deposit has replaced the JCVD. The new Committee's remit is to discuss the establishment of the future statutory Advisory Panel for legal deposit of digital materials (AOP and British Library Discuss Archiving the Web 2003) and to develop a policy for access to materials deposited voluntarily (British Library 2003). The statutory Advisory Panel will be involved in developing and monitoring the Regulations for legal deposit law (The United Kingdom Parliament 2003, column 712).

Other countries have been increasingly extending their legal deposit legislation, initially often to cover only digital publications on physical carriers or online information that does not change (for example Denmark 1997). Legislation in some countries, for example Norway (1989) or South Africa (1997), theoretically extends to all digital publications.

Publishers can provide the information on a physical medium. They can arrange to transfer, or "push", information to depositories via networks. Alternatively, libraries can "pull" information from publishers' servers, with their permission. Harvesting can be carried out selectively or comprehensively. For example Scandinavian national libraries are harvesting their "national" portion of the Internet and the Internet Archive (http://www.archive.org) is trying to gather the whole Internet. Other libraries such as the British Library and the National Library of Australia are initially taking a selective approach to gathering this type of digital information. The selective approach involves making arrangements with publishers for the deposit of material. This is either because of the volume involved or the inability of harvesting software to gather from the "deep Web". The deep Web is material that cannot be found or accessed by automated harvesting tools for various reasons, for example a user ID and password is required for access.

Comprehensive harvesting of Internet material is likely to infringe copyright legislation. The Norwegian National Library's Paradigma project will, amongst other things, explore the legal issues of collecting online material and making it available to users (Nasjonalbiblioteket, [n.d.]). The Joint Information Systems Committee of the UK Higher and Further Education Funding Councils (JISC) and the Wellcome Trust commissioned Charlesworth to study the legal issues relating to Web archiving, including copyright (Charlesworth 2003). The primary target audience for the report is archivists working in UK research institutions. However, Charlesworth acknowledges that the issues have a much broader scope. He concludes that as things stand, the only way to archive Web pages and conform with copyright law would be to obtain permission from the relevant rights holders. This would become more and more difficult the more comprehensively the deposit library wanted to collect material because of the difficulty of tracking and recording the rights holders (Charlesworth 2003, p. 8). Charlesworth assesses two approaches to overcome this problem. These

involve the rights holders either providing legal metadata, including ownership and level of permission to copy, and Web archives providing an opportunity to opt out of having their material included in the Web archive. The latter can be accomplished by rights holders indicating through codings that they do not want their material to be harvested or giving rights holders the opportunity to ask for material that has been collected without their permission to be removed from the archive (Charlesworth 2003, p. 9). Charlesworth points out that rights holders are under no obligation to do this under copyright law and copying material in this way without prior permission may result in legal action.

An option not considered by Charlesworth is amending copyright legislation to allow Web harvesting. Some countries are issuing or amending laws to allow deposit institutions to gather Internet material. The Swedish government issued a decree relating to the Kulturarw³ Heritage Project at the Royal Library. The Decree now authorises the Royal Library to not only collect Swedish Websites on the Internet but also to allow the public access to it within the library premises (New decree for Kulturarw3 2002). The French government has adopted a law requiring every French Web site to be archived (République française 2001). While many organisations are expected to do this, the Bibliothèque Nationale de France (BNF) and Ina, the national audiovisual depository, harvest material at regular intervals. The Legal Deposit Libraries Act 2003 (Great Britain 2003b, s.8) includes provision for amending rights laws to allow legal deposit to harvest, preserve and give access to Websites.

However, this leaves the question of how other types of libraries can legally preserve digital material.

3.7.2 Copyright law and preservation in other countries

A survey was carried out to investigate whether other countries' copyright laws contain clauses relevant to preservation copying by libraries. Many countries' laws do contain relevant clauses; several countries surveyed, mainly in Europe, do not make any such provision. In some countries, for example Hong Kong (Hong Kong 1997), these clauses are modelled on UK law. A few of the laws surveyed contain

clauses which may be more beneficial to libraries wishing to carry out preservation copying than the UK law.

In the United States, the Digital Millennium Copyright Act (DMCA) expressly allows authorised institutions to make up to three digital preservation copies of an eligible copyrighted work. It allows the institution to loan those copies to other institutions and permits preservation, including by digital means, when the existing format in which the work has been stored becomes obsolete (United States 1998, Title IV). This law seems useful, but having to wait until works have become obsolete to take action to preserve them is very unsatisfactory. Hirtle (2003) argues that preservation by individuals might be allowed under the Fair Use provision in US copyright law; however, there is no case law on this issue.

The Canadian Copyright Act allows electronic publications to be converted to a contemporary format if necessary for preservation purposes. The work must be in the library's permanent collection and the copy must be made in order to maintain the collection (Institute for Information Law 1998).

> *30.1 (1) It is not an infringement of copyright for a library, archive or museum or a person acting under the authority of a library, archive or museum to make, for the maintenance or management of its permanent collection or the permanent collection of another library, archive or museum, a copy of a work or other subject-matter, whether published or unpublished, in its permanent collection.*
>
> *30.1 (1) (c) in an alternative format if the original is currently in an obsolete format or the technology required to use the original is unavailable" (Canada 1997, Ch. 24).*

The latter clause implies that action can only be taken when the work is already obsolete. Again, this this is not ideal, since preservation experts agree that waiting until a work is obsolete to preserve it is too late. It is also not clear from this how 'obsolete' is to be defined.

Australia's copyright law (Australia 1968) allows libraries to make and communicate copies of copyright works that have been damaged or have deteriorated. No mention is made of the format of either the original or the copy. However, digital copies made under this exception may be '[made] available online to be accessed through the use of a computer terminal installed within the premises of the library or archives'. The conditions attached to the use of preservation copies of 'original artistic works' are more stringent than those for other types of work; it is not clear why this is the case. The relevant clause states that the computer terminal in question may not be 'able to be used to make or distribute copies of the original'. This may only be done if:

> a) *the work has been lost, or has deteriorated, since the preservation reproduction of the work was made; or*
> b) *the work has become so unstable that it cannot be displayed without risk of significant deterioration.*

Depending on interpretation, this law may only refer to the digitisation of material for preservation purposes.

In New Zealand, a recent Cabinet Paper (Tizard 2003) outlines proposed changes to the Copyright Act (Consolidation) 1994 (New Zealand 1994). This is intended to clarify confusion about how the existing, 'technology-neutral' archiving exception applies in the digital environment. It recommends that the relevant clause, Section 55, be clarified:

> *to enable preservation by digital means, including format shifting (for example from print to digital).*

Unfortunately, the document does not specifically mention copying of born digital materials to new digital formats. Hopefully, any eventual legislation will make this clearer. However, the context suggests that such copying would be allowed.

3.8 Licences and perpetual access

Information is preserved for a reason. The reason is usually to allow access to material beyond the short-term. Making preserved digital material available also involves copying. Access is a separate issue that also has to be dealt with.

If libraries do not have the legal right to copy for preservation purposes, then one approach they could take is to ask rights holders for permission. This was a recommendation from the Cedars project (2002, p.16). The Cedars Project also recommended that if the preserving institution seeks a licence it should ensure that the licence allows it to take all necessary steps to preserve the intellectual content of the digital object. The rights in the different elements of complex digital material, the content and any software may belong to a number of different individuals or organisations, rather than the publisher. If the institution physically owns the digital object, there should be a clause included in the licence agreement to allow making preservation copies. If the digital object is accessed remotely, there should be a clause that allows the institution to access the content for preservation purposes. Alternately, there should be an agreement clearly stating who has archival responsibility.

Some licences for digital material may be negotiated on the basis of perpetual access. In the case of electronic journals, this would provide for continuing access to back issues even when a subscription has lapsed. However, provision for perpetual access is not the same as a licence for preservation purposes. It relies on other players, such as information providers, publishers, hosts or other intermediaries taking responsibility for and ensuring preservation. Even if guarantees are given, organisations go out of business or just may not have the resources to do the job.

Clearing rights for the preservation of digital materials is likely to be resource intensive and difficult. The complexity of rights holders and regimes associated with many complex digital resources can make clearing rights problematic. The publisher may not hold the rights to all the content or any related software, but may have licences with third parties. Preservation institutions may have to spend a lot of time and resources locating and identifying rights owners and asking permission. The

institution would then have to keep records of all the various agreements reached with the rights holders, which would also be a drain on resources.

If a publisher is willing to negotiate a licence for preservation copying, it might be possible for the publisher to indemnify the library against infringement of third party rights.

3.8.1 Publishers' licences

A survey was carried out of archival and preservation clauses in publisher licences. 1070 publisher Websites were identified using publisher directories. Of these, only 411 were found to be actual publishers' Websites. These Websites were then searched to find the publishers' licence agreements. Relevant information was found on 21 Websites. This was either the actual licence, or was taken from a 'terms and conditions' or other relevant section.

Of greatest interest were the specific archiving and preservation clauses that these contained. Five did not contain specific archiving and preservation clauses, although some of these did contain other clauses of interest.

One point of particular interest was whether publishers allow their users to access their publications once they have terminated a subscription. Seven licences said that this is the case, with two stating that continued access is only available on request, and another two stating that continued access is only available for some publications. Three said that continuing access was not available; one explained that this is because the publication in question is a current awareness service, and only one year's content is ever available. Seven licences did not state whether continuing access is available. Two licences specifically mention that this access will be free. Others state that there may be costs, particularly if content is accessed via an intermediary.

Ten licences explained exactly what users would have access to once they have cancelled their subscription. In all cases, they will only have access to the materials that were published during their subscription. They will not, therefore, continue to have access to any backsets to which they had access during their subscription period.

Ten licences explained how continued access would be provided:

- Four publishers will either continue to provide access from their servers or will give the files to the user. They may give the files in different ways, either on a physical storage medium such as a CD-ROM or by electronic transfer (e.g. FTP).
- Two publishers will continue to provide access from their servers.
- One will initially provide access from its server, but may move to a third party server later.
- One says that access will be provided via an intermediary or third party.
- One says that users may retain electronic copies of its publications.
- One publisher says that it reserves the right to change the way in which it provides such access over time.

There is currently no standardisation amongst these publishers of the way in which continuing access will be provided. A majority of the licences state more than one way in which access could be provided. The publishers who may give the files to the user do not state definitively how this would be done. One publisher states that it is the publisher who will decide how this is done, but in other cases, it is not clear whether it would be the publisher or the subscriber who would decide.

Several of the licences found indicated how long publishers are intending to keep their materials available for. Some are guaranteeing to keep their material accessible in perpetuity, while others only guarantee they will endeavour to do so. Some licences state that the publisher is only guaranteeing to keep materials available for a fixed number of years after the end of a subscription. The lengths mentioned were 4, 5 and 10 years. Others are not making any guarantees about the long-term availability of their materials, because of the technological difficulties and costs of doing this.

Some publishers indicated which methods they are using to ensure that their information remains accessible. These include:

- Using '*rigorous industry standards*' for archiving and storage
- Archiving their materials in multiple locations
- Participation in schemes such as JSTOR and LOCKSS
- Giving responsibility for preservation to a third party (former subscribers may have to pay for this)
- Setting up an escrow fund to pay for converting materials to new formats

Another potential issue is whether libraries have access to back files during their subscription period. Only five of the licences examined mentioned that access to back files is included in the subscription. One of these states that the subscription includes four years of back files, which may well be a 'rolling wall' type of arrangement.

The licences examined contained several other clauses that may affect the ability of libraries to preserve the electronic materials that they subscribe to. Some clauses that relate to publishers' responsibilities have some relevance to digital preservation. For example, some licences state that the publisher may change features such as the delivery format, access method or display if necessary, since technological formats may change over time. Many licences contain clauses about continuous access to online materials. Usually, the publisher only accepts liability for restoring access as quickly as possible, although publishers who do not host their own content are unable to guarantee this. Some clauses which outline users' responsibilities are also relevant. Some publishers allow users to make back-up copies as necessary, while others will provide physical copies of material for backup purposes, perhaps on payment of a fee. However, licences usually restrict the making and storing of copies of multiple extracts or entire journal issues. Customers usually also need permission to modify licensed materials in any way, or to make derivative works from them. Such clauses may well mean that making preservation copies would not be allowed.

3.8.2 Existing model licences

The JISC / National Electronic Site Licence Initiative (NESLI) Model Licence for Journals was approved in September 2002 (NESLI2 2002). Clause 2.2.2 states that:

> *After termination of the Licence, the Publisher will provide the Licensee and its Authorised and Walk-In users with access to the full text of the Licensed material which was published and paid for within the subscription period, either by continuing online access to the same material on the Publisher's server or by supplying an archival copy in an electronic medium mutually agreed between the parties which will be delivered to the Licensee, or to a central archiving facility operated on behalf of the UK HE community without charge.*

Other model licences also address the preservation of remotely accessed materials. Clause XIII of the Liblicense Standard License Agreement (Liblicense 2001) states that:

> *Except for termination for cause, Licensor hereby grants to Licensee a nonexclusive, royalty-free, perpetual license to use any Licensed Materials that were accessible during the term of this Agreement. Such use shall be in accordance with the provisions of this Agreement, which provisions shall survive any termination of this Agreement. The means by which Licensee shall have access to such Licensed Materials shall be in a manner and form substantially equivalent to the means by which access is provided under this Agreement.*

Similarly, the John Cox Associates' licence for single academic institutions (Licensingmodels 2000) contains an optional clause that states that:

> *The Publisher undertakes to [use reasonable endeavours to] provide or to make arrangements for a third party to provide an archive of the Licensed Materials for the purposes of long term preservation of the Licensed Materials, and to permit Authorised Users to access such archive after termination of this License.*

Jones (2003) has studied issues surrounding the archiving of electronic journals in the UK academic sector. This study was commissioned by JISC, and views of the archiving clauses of the JISC model licence for journals were part of the research. While the scope of this study was much narrower than the research reported here, the results are of interest.

Jones concluded that although there is no imminent danger of loss of content to licensed e-journals, there is an urgent need to provide a co-ordinating archiving service for the UK that can develop in stages. She found that concerns about continued archival access and VAT were the two most cited barriers to moving to e-only access with any degree of confidence. Of the three options for continued access referred to above in the Model Licence, only two - continued access from the publishers server or an archival copy delivered to the Licensee - are currently available. The publishers are most offering a CD-ROM to libraries, but this is the least favoured option for the libraries. The third option - archiving by a central facility operated on behalf of UK HE - does not at this stage exist. The consultancy suggested that it is premature to build a monolithic facility as other developments are not yet clear and the costs may not be justified for the UK market alone. Contract law governing these licences also means that individual institutions are currently responsible for enforcing them. This is particularly problematic in cases where renewal contracts override the access obligations in previous contracts. Developing the capacity to handle complex rights was seen as an essential development by both libraries and publishers at a JISC Workshop held during the consultancy.

3.8.3 Central rights clearance for preservation

An alternative to individual libraries seeking permissions is some form of central rights clearance operation. There are a number of licensing bodies in the UK, for example the Copyright Licensing Agency (CLA), the Newspaper Licensing Agency (NLA) and the Design and Artists Copyright Society (DACS). HERON, a division of Ingenta Plc, is operating a transactional copyright clearance service for digitisation for Higher Education Institutions. This is done under the terms of an Agency agreement with the CLA, and applies to the UK publications for which the CLA is mandated to

do this. It is not clear from the literature whether any of the existing licensing bodies are involved in licensing for preservation copying.

3.9 Metadata

Metadata seems to be the key to the preservation of digital information and there has been much work going on in this area. Yet another possible alternative to clearing rights could be including rights information in metadata associated with digital material. While it is clear that the library community has incorporated rights into its preservation metadata schemes, this area is not well-developed. Activity in the libraries field came together in the OCLC/RLG Working Group on Preservation Metadata (2002) framework. The framework is based around the OAIS information model, uses elements from various existing schemes and also adds new elements. The OCLC/RLG framework is a high-level scheme, so further work is needed to implement it. Further work is being carried out here by OCLC and RLG. The framework includes a rights-related element. The aim of this element is to set out what a preserving institution can do to preserve and disseminate content.

While some of the metadata relates to internal processing and preservation activities and would have to be populated by the libraries, cooperation with publishers would be needed to acquire much of the metadata. Publishers have been involved in Online Information exchange (ONIX) (Martin 2001). This is an emerging standard for the exchange of product information between different players in the information industry. ONIX has some rights-related elements, but these are designed for commercial purposes and may not be helpful for preserving institutions. The status of material with regard to rights may change over time. The temporal aspects of rights management have not been dealt with at all. The use of persistent identifiers such at Digital Object Identifiers that could point to related metadata both on the rights holder and library side may be useful here (The Digital Object Identifier System 2004).

The RoMEO (Rights Metadata for Open Archiving) Project (Project RoMEO 2002) investigated rights issues relevant to self-archived research. Although the project was only dealing with Open Archiving and is not specifically looking at preservation, the

project's conclusions about rights metadata may be helpful to the digital preservation community. The RoMEO Project (Gadd, Oppenheim and Probets 2003) concluded that the Creative Commons scheme (Creative Commons, [n.d.])is the most appropriate of the existing digital rights expression languages, for this purpose, and recommended some amendments to its set of 'licences' to make it more suitable.

4 DIGITAL PUBLISHING AND DIGITAL COLLECTIONS

4.1 Publishers, authors and digital publishing

4.1.1 Digital publishing by publishers and authors

The questionnaires aimed to identify what problems libraries are likely to encounter in preserving digital publications. So publishers and authors were asked about the publications they are creating and preserving. Librarians were asked about the nature of their collections and preservation.

Eighty-one of the 82 publisher respondents indicated whether they publish in digital formats. Thirty-four of them (42%) currently do. A further eight had firm plans to publish digitally in the next 12 months. It should be noted that with a response rate of 13.7% from the 600 publishers surveyed, the questionnaire results only give a broad indication of publisher views and activities.

Authors were also asked whether any of their work is published digitally. Only one respondent has not yet published anything digitally, and this author indicated that they may do so in future. However, it is probably to be expected that it is predominantly authors who are already involved in digital publishing who answered the questionnaire. Authors who dislike or do not use technology would have been very unlikely to respond, since the questionnaire was only available online. A significant majority of respondents publishing digitally published their own work. Unfortunately, those respondents who have not yet done this were not asked whether they plan to do so in future.

4.1.2 Types of material published digitally

The types of material that publishers produce in digital form vary, reflecting the varied nature of the publishers surveyed.

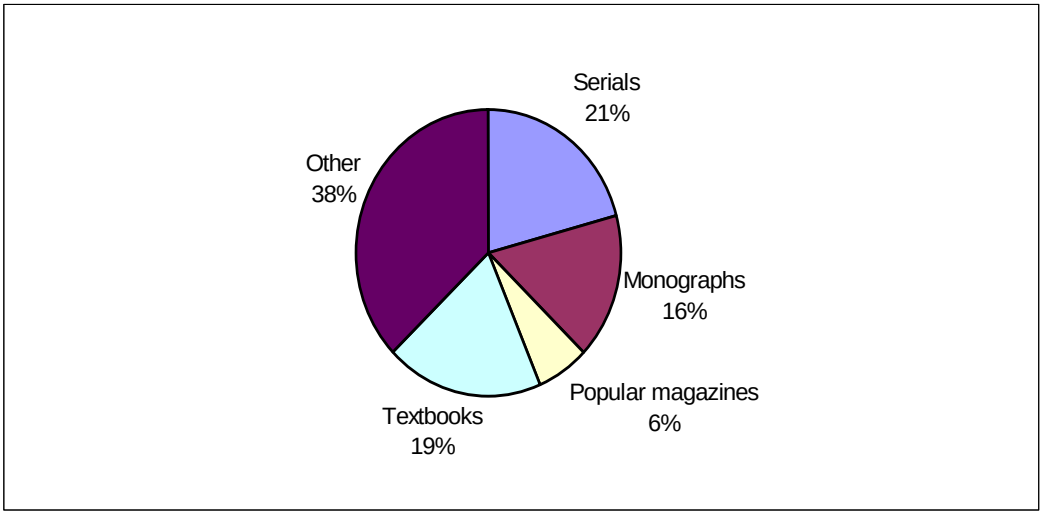

Figure 4.1 Types of digital publication

Thirty-eight per cent of the respondents to this question indicated that they publish material that did not fit into any of the suggested categories. The "other" types of publication specified by respondents were:

- Additional material to accompany text books.
- Audio documentaries, news bulletins (text)
- Course listings, institution profile advertising.
- Databases (3 respondents)
- Directories
- Editions of Buddhist texts in Pali
- Handbook data
- Law reports, practitioner texts
- Literary texts
- Manuscripts and rare printed sources; archival materials
- Market reports
- Reference works
- Sheet music
- Software
- Teacher / classroom resources (4 respondents)
- Trade books

One of the publishers interviewed produces an online magazine with an integrated discussion forum, while another has an online news service. Other types of digital publications mentioned include special issues of journals made available as e-books, and 'Web-focuses' that stem from journal articles.

Authors were also asked about the types of digital publication in which they have published their material. The second most frequent option was e-journals, which was selected by six respondents. This is to be expected, given that ten respondents publish in academic or scholarly publications.

Publication type	No. of respondents	% responses	% respondents
Personal Websites	10	40	83.3
E-journals	6	24	50
Online databases	4	16	33.3
Open Archives	3	12	25
Other	2	8	16.6
E-zines	0	0	0
Weblogs	0	0	0
E-books	0	0	0
Total	**25**	**100**	**208.2**

Table 4.1 Types of digital publication in which respondents publish

The respondents who selected 'other' explained that they have written CD-ROMs to accompany books, and an online course for the Open University. No respondents selected e-zines, Weblogs or e-books. This is not surprising, since these are not yet as widespread as the other types of publication mentioned. Three-quarters of the respondents selected more than one response. One author selected four options (e-journals, online databases, personal Websites and CD-ROMs accompanying books).

The author questionnaire also asked whether the materials that respondents publish digitally are just textual, or include images or other types of information. The most frequent response was 'text and images' (six respondents). The one author who selected 'other' explained that his or her work includes 'some simple diagrams'. Four respondents chose more than one option.

Format	No. of respondents	% responses	% respondents
Text and images	6	40	50
Text only	4	26.6	33.33
Multimedia	4	26.6	33.33
Other	1	6.7	8.33
Total	**15**	**100**	**125**

Table 4.2 Content of digital publications

Four respondents selected 'multimedia'. This suggests that authors are taking advantage of the opportunities offered by digital publishing.

4.1.3 How long publishers and authors have been publishing digitally

The publisher and author questionnaires asked respondents how long they have been involved in digital publishing, in order to get an idea of how likely material is to affected by technological obsolescence.

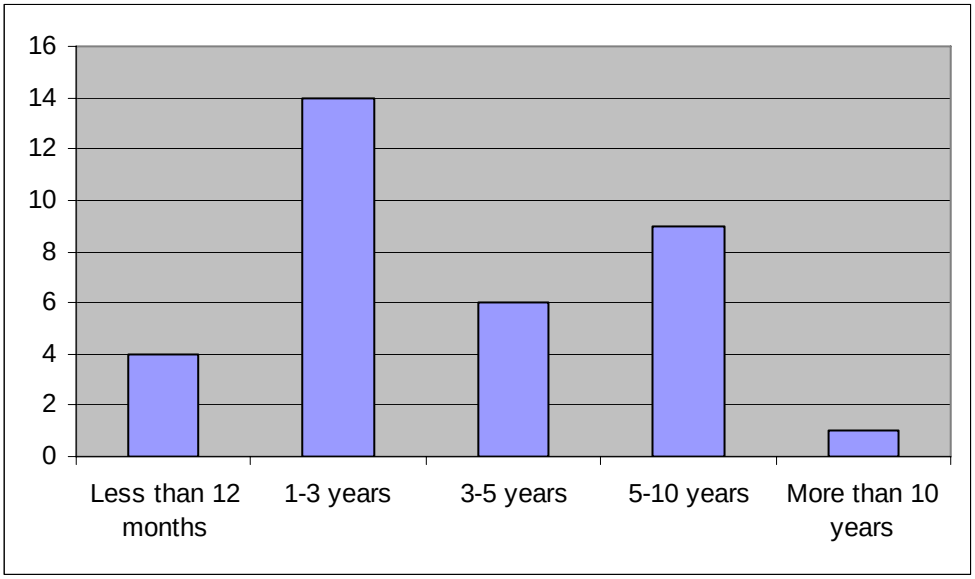

Figure 4.2 How long publishers have been publishing in digital formats

While more than half of the respondents had been involved in digital publishing for three years or less, some had been publishing for some time. Virtually all author respondents have been publishing digitally for between one and six years, with five selecting '1 – 3 years' and six selecting '4 - 6 years'.

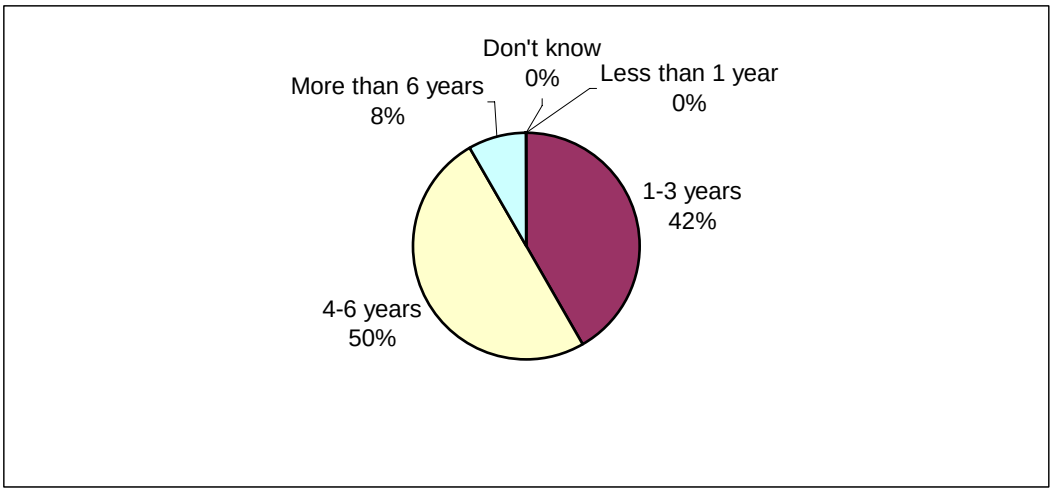

Figure 4.3 Number of years respondents have been publishing digitally

There is no real correlation between the number of years an author has been publishing digitally and his or her subject area. Perhaps surprisingly, the author who has been publishing digitally for the longest gave his or her subject area as 'archaeology', while the one respondent who writes in a technical subject (technology and writing) has only been publishing digitally for 1-3 years.

4.1.4 Benefits of digital publishing

The publishers interviewed gave various reasons why they publish digitally. Several of them explained that digital publications offer much better functionality than print ones: they are easy to search, can be interactive, and can contain links to related materials. Another publisher said that publishing digitally cuts costs. Not all publishers interviewed were completely in favour of digital publishing. One expressed concern that the ease of publishing digitally would compromise the quality of publications. This is particularly true of Web publishing, for which there is not necessarily any quality control.

4.1.5 Distribution formats used

The publishers surveyed distribute their digital material in a variety of formats. More than half the respondents use both offline and online distribution formats, and even the least frequent response, 'online (intermediary/aggregator)' was selected by 17 of the 41 publishers who answered this question (41.5%).

Distribution format	No. responses	% responses	% respondents
Offline (e.g. CD-ROM)	27	37%	65.9%
Online (direct)	27	37%	65.9%
Online (intermediary / aggregator)	17	23.3%	41.5%
Other	2	2.7%	4.9%
Total	**73**	**100**	**178.2**

Table 4.3 Distribution format of digital publications

Some of the publishers interviewed do not host their online materials themselves, but give their content to others, usually aggregators, to host. Users therefore access the publisher's content via third parties. This may affect the preservation of these materials, as the subscriber may not have a direct relationship with the original publisher and/or rights holders. One publisher commented that some large research libraries in the US choose to host publishers' content themselves, so that they can use their own, single interface for all their content. This might make it possible for the library to preserve these materials itself, although it is not clear whether this is in fact the case. The publishers who use intermediaries explained their reasons for doing so. One publisher began doing this when 'technology was more in its infancy and still developing', so that the company could focus on their content, and not have to train and pay IT staff. This publisher did, however, feel that they, rather than the aggregator, had had to take the lead with this. They also suggested that they may not continue to work in this way in the future. Another publisher felt that giving content to others to host might be useful as a backup:

> *we're very aware that if our servers get blown up, we have a problem. And it's kind of reassuring that some of our content is hosted by*

> *OCLC, some by Ebsco, some by Ovid. And maybe we could go round and collect it up again.*

Another publisher suggested that aggregators might generate additional income. The use of intermediaries may also benefit libraries, since it may be advantageous to have a single point of contact for all digital resources. Intermediaries may also provide a standard interface for all the materials they host, although one publisher believed that this would probably have less functionality than the publisher's own interface.

The interviewees largely agreed that while CD-ROMs used to be important, they are now 'dropping out of circulation' in many circles, in favour of online publishing. A publisher and an academic librarian explained that libraries do not like having to network CD-ROMs themselves, and therefore prefer online, remote access. This reflects the results of Maggie Jones's consultancy into the archiving of electronic journals (Jones 2003). Standalone CD-ROMs that accompany books may also present libraries with preservation difficulties, since books and CD-ROMs need to be stored and preserved in different ways. One interviewee said that another publisher was intending to stop producing its materials as CD-ROMs as well as online, because it was too expensive to continue to publish its material in two formats.

Several interviewees also commented on the future of print as a publication medium. Print-only journals seem to be declining: only one publisher interviewed is still publishing its journals in print form only, and even this publisher is already producing other materials digitally. Two academic publishers interviewed commented that the demand for print has not diminished as much as they had expected: when asked whether they expected print publication to cease, one answered:

> *Not in the foreseeable future … If you'd asked us that question five years ago, we'd probably have said yes.*

The publishers interviewed felt that there is still a strong demand for print. This is not the case with libraries: libraries were said to favour digital delivery of journals, and one publisher suggested that print would become increasingly irrelevant as today's students are likely to use electronic rather than print journals. It is not clear that this is

in fact the case. However, interviewees felt that most people will probably continue to print articles from electronic journals rather than reading them online. They also thought that personal subscribers to journals will still want to be able to browse print copies, that and authors and editors will still want to see their work in print. The publishers interviewed were therefore not planning to stop print altogether in the near future, although there is an increasing trend towards producing digital-only publications.

Lastly, more than one publisher uses the World Wide Web to update the materials it has already published in other formats. In one case, these are printed materials, and in the other, CD-ROMs.

4.1.6 Digital-only and parallel publications

Publishers were also asked what proportion of their digital material is available in parallel versions.

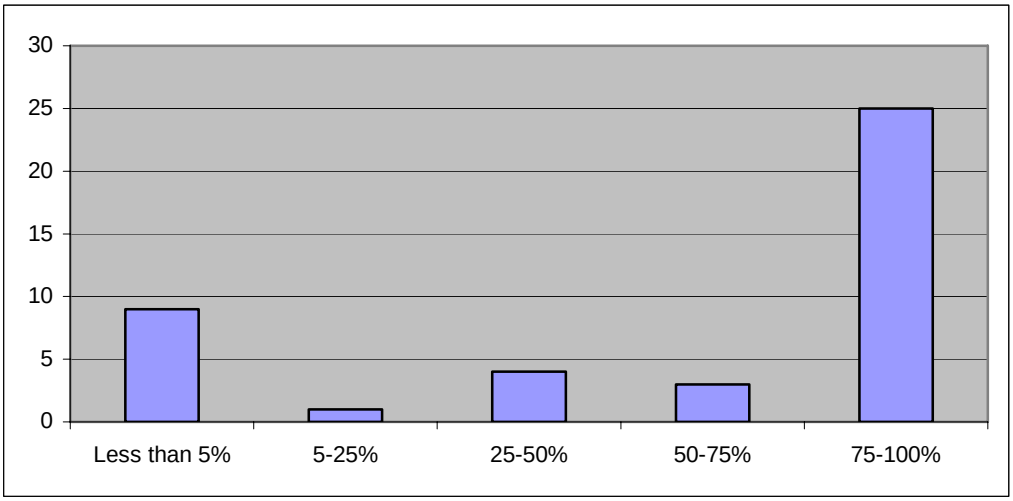

Figure 4.4 Digital publishing with print equivalents

Twenty-five of the 42 publishers who answered this question (59.5%) stated that at least 75% of the digital materials that they publish are also available in print form. However, nine of them (21.4%) stated that less than 5% of their digital material also exists in print format. The number of publishers who responded to the questionnaire

is too low to be able to draw any conclusions about the publishing industry as a whole.

The publishers interviewed described a wide range of relationships between the different publication formats they produce. In some cases, print and digital versions of publications are substantially the same except for the increased functionality of the digital version. However, online products sometimes contain extra information. One professional publisher interviewed makes part of its magazine contents available freely online, probably to attract new subscribers. Some publishers only produce digital materials to support their print products; for example, one publisher supports its textbooks with online materials both for students and for lecturers. Conversely, a few publishers produce only a small number of print publications, and these only to support their digital output; for example, an Open Access journal publisher makes all its content available freely online, but makes a charge for any paper copies it supplies.

The author questionnaire asked respondents whether they publish their work in print form as well as digitally. Four respondents publish their works in digital format only, and six publish parallel print and digital versions. They were also asked to state what percentage of their publications is published in each format. Of the two respondents who answered this question, one currently publishes over half of his or her works as simultaneous print and digital publications, and one currently publishes over half his or her works in digital format only. This respondent indicated that he or she publishes 76 – 100% of his or her work in this format.

Authors were also asked whether the print and digital versions of their parallel publications differ in any way, and how different the two versions are. Five respondents said that there is no difference, one said that the versions differed a little, and one said that they differed considerably. Several respondents commented on this question. One said that the print and digital versions are identical because the online publications are PDF versions of the printed academic journals. A final comment gave one reason why print and digital versions of publications may differ:

> *Reading from paper is quite different from reading from the screen and one has to write differently for the medium.*

4.1.7 Effect of digital publishing on the publishing industry

The growth of digital publishing has led to the emergence of new types of publications and new publication models. Custom publishing is now being introduced, because it is simple and cheap to carry out short-run digital printing from print copies of materials. One example of this is that a publisher interviewed allows lecturers to choose relevant chapters from different publications and have a new book published from these to support a particular course. These publications are assigned ISBNs, but are not currently being deposited with legal deposit libraries. Whether they should be deposited would depend on whether they were considered to be "published" or not.

A particularly significant new publishing model is Open Access publishing. Interviewees had differing views as to whether Open Access publishing will succeed. Its supporters believe that they are offering a much-needed alternative to traditional publishing models; its opponents doubt that its business model is viable. There is currently tension between those who support and oppose Open Access publishing. In general, Open Access publishing is gaining support. However, one mainstream publisher interviewed suggested that it will be 'ten years' before it becomes significant and an Open Access publisher interviewed thought that while it will become the 'prevailing model for primary research information', it is unlikely to supplant traditional journal publishing completely.

Interviewees said that traditional publishing models are gradually disappearing. For example, one publisher felt that the traditional journal would gradually be replaced:

> *Why would the print publishing model for journals continue in an electronic environment? Why would you have a release once a month of a new issue? Why, when that article is published, is it then set in stone forever more? You can update, you can publish continuously, you can link in ways that are impossible in a print environment.*

This last point indicates a potential problem with digital publishing from a preservation perspective: it may be difficult to define the boundaries of publications

when these contain links and therefore difficult to decide what is the "publication" to be preserved.

Another consequence of the changes to the publishing industry is that new subscription models are emerging. For example, one publisher interviewed explained that some of its large databases can be purchased outright as well as subscribed to. Some large libraries prefer to purchase materials outright, even if they do not have the storage capacity to host them themselves, while other libraries cannot afford this and prefer to pay annual subscription charges. This may affect who is able or willing to preserve this material. Another model mentioned was transactional charging, whereby publishers may sell access to individual articles rather than to complete journals or databases.

4.2 Digital collections in libraries

4.2.1 Current and future digital collections in libraries

The library questionnaires asked about the extent of digital holdings in libraries. Most of the libraries responding to the questionnaire said that they have some sort of digital content in their collections.

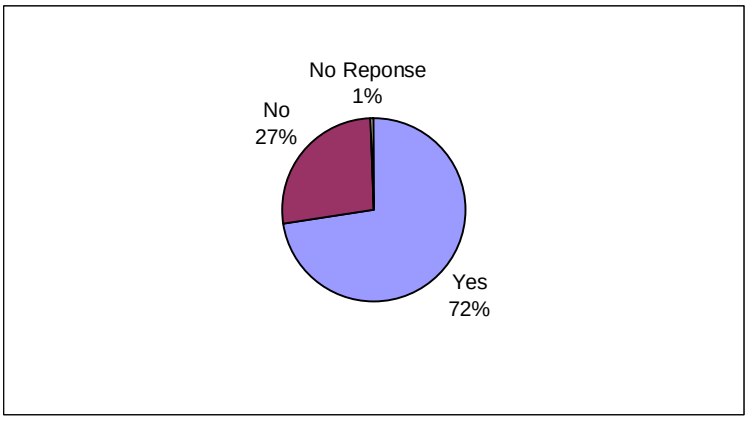

Figure 4.5 Libraries with digital content

The proportion of libraries with digital content was fairly high for all sectors. However, it is surprising that only 71% of academic libraries claimed to have digital

collections, since other surveys have shown that all further and higher education libraries in the UK have at least some digital content. It is possible that this survey may have produced such anomalous results because the person who answered it did not have the requisite knowledge. Respondents could also have misunderstood the question, perhaps thinking that the survey only wanted to know about online materials. Furthermore, with a response rate of 16.8% from the 1,000 libraries surveyed, the library questionnaire results only give a general indication of the digital collections in libraries. Digitisation initiatives in the public library sector are likely to explain why a large proportion of public library respondents also have digital content (75%). In fact, 15 out of 27 public libraries with digital content had created at least some of that content in house (see Table 4.3).

Sixty-one libraries responded to a question asking whether they are expecting to acquire digital material within the next five years. Eighteen of the libraries that already have digital material said they were planning to acquire it in the near future, despite being directed to ignore this question. Of the remaining 43 libraries that answered this question, 20 (46.5%) are expecting to acquire digital material in the next five years, with a further 17 (39.5%) unsure. Only six library respondents (14%) neither have digital content nor are expecting to acquire any in the next five years.

A result that is perhaps surprising is that academic and special library respondents without any digital content were less likely to be expecting to acquire it than were public libraries. No public libraries said that they were not expecting to acquire any digital content, with the vast majority (86.7%) saying that they were expecting to. By contrast, half of the academic and special libraries that do not currently have any digital content either said that they were not expecting to acquire any, or were unsure whether they would do so.

Library Sector	Yes	No	Don't know	Total
Academic	12	3	9	24
Public	13	0	2	15
Special	8	2	6	16
National	2	0	0	2
Other	2	1	1	4
Total	37	6	18	61

Table 4.4 Libraries planning to acquire digital material in the next five years by sector

The library questionnaire results therefore suggest that many libraries already have digital content, and that the number of libraries with digital content is likely to increase in the next five years. Unfortunately, the questionnaire did not ask libraries what percentage of their collections currently consists of digital material. However, the librarians interviewed said that the size of their digital collections is likely to increase.

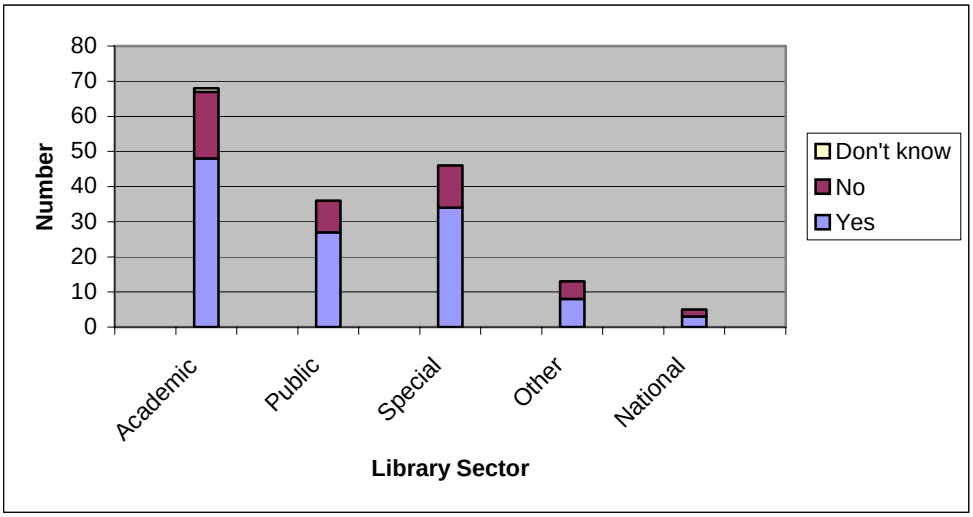

Figure 4.6 Libraries with digital content by sector

4.2.2 How libraries acquire their digital content

The library questionnaire asked how libraries acquire their digital material. One hundred and twenty-three respondents answered this question, although only 122 had earlier indicated they had digital material. Individual respondents acquire their digital

content in a variety of ways. While licensing content from a vendor was the most frequent response, a significant proportion seems to have physical ownership of content. Libraries physically acquire material by purchasing it outright, creating digital material themselves and through deposit or donation.

Method of acquisition	No. of Responses	% of Responses	% of Respondents
Licensed from vendor	81	29.1	65.9
Purchased outright	68	24.5	55.3
Created in-house	55	19.8	44.7
Harvested from Web	29	10.4	23.6
Donation	23	8.3	18.7
Voluntary deposit	14	5.0	11.4
Other	8	2.9	6.5
Total	**278**	**100**	**226**

Table 4.5 Methods of acquisition of digital material

Twenty-nine respondents stated that they harvest content from the Web. This is perhaps surprising, since there are copyright issues involved and the technologies to do this are still in development. It is not clear whether these respondents are really downloading material from the Web, what the nature of the material they are downloading is, or whether they are really just linking to Web material through their library catalogues or Websites. The "other" responses included materials obtained as part of a project to digitise existing microfilms, and another response really related to voluntary deposit.

The most common method of acquisition for academic, special and other libraries with digital materials was 'licensed from vendor'. This option was selected by 72.9% of academic libraries, 67.6% of the special libraries and 62.5% of the 'other' libraries that have digital content. The most frequent option for public and national libraries with digital content was 'created in-house', which was selected by 59.3% of the public libraries and all five of the national libraries with digital content. 'Outright purchase' is also a common option, which was selected by 55.3% of respondents. The librarians interviewed explained that this generally applies for digital publications on physical storage media such as CD-ROMs. The British Library also purchases a

lot of material, because the voluntary deposit scheme only covers materials published in the UK, and material published abroad is essential for its research collection.

Method of acquisition	Academic	Public	Special	National	Other	Total
Licensed from vendor	35	14	23	4	5	81
Purchased outright	25	14	22	4	3	68
Created in-house	16	16	14	5	4	55
Harvested from Web	9	2	13	3	2	29
Donation	6	5	9	3	0	23
Voluntary deposit	4	2	4	3	1	14
Other	0	4	2	1	1	8
Total	**48**	**27**	**34**	**5**	**8**	

Table 4.6 Methods of acquisition by library sector

The least common options were 'donation' and 'voluntary deposit'. An interviewee at the British Library explained that much of the donated digital material the library receives is from manuscript collections. Again, since it is known that only two of the national libraries responded to the questionnaire, it is surprising that 14 libraries said that they receive digital materials by voluntary deposit. It would be interesting to know what sort of materials are being deposited and under what circumstances, or whether voluntary deposit has just been confused with donation.

Library respondents were asked to estimate what proportion of their digital collections had been acquired by each method of acquisition. Only 106 of the 122 libraries with digital material answered this question (86.9%). Those that did gave the following responses:

Method of acquisition	<5%	6-25%	26-50%	51-75%	>75%	Total no. of respondents
Licensed from vendor	5	16	12	6	33	72
Purchased outright	3	23	12	5	17	60
Created in-house	5	21	8	3	10	47
Harvested from Web	1	12	9	0	2	24
Donation	2	16	3	0	1	22
Voluntary deposit	3	6	2	0	1	12

Table 4.7 Percentage of digital collection acquired by different methods

Of the 72 respondents that gave an estimate of the proportion of their collections licensed from vendors, 39 (54.2%) said that this material made up more than half of their digital collections. If these libraries are accessing their licensed material remotely, they are potentially totally dependent on the vendor for the preservation of this material. Thirty-two of the 72 libraries with licensed material (44.4%) were academic libraries. Of these, 25 (78%) said that over half of their digital collections were acquired in this way. Licensed material made up a much lower proportion of the digital collections of other types of library, with the exception of "other" libraries. Two of these libraries had licensed material; in one library licensed material made up more than 75% of the collection and in the other it made up 25% or less of the collection.

4.2.3 Types of digital material held by libraries

The libraries represented by interviewees hold a wide variety of digital materials. Some of these materials are digital versions of the types of publications traditionally held by libraries, for example, electronic journals, abstracts and indexes, reference works, and to a lesser extent e-books. All the libraries interviewed also have some digitised materials, which may include digitised audio files and images. One academic library has a dedicated electronic library for materials for teaching and research that contains digitised book and journal extracts, student projects, images, and audio and video materials. In addition to these, one university library is being asked for advice about preserving electronic datasets held within the university. It is also involved in developing an e-print server and has an e-theses project. The

libraries interviewed treat these different types of digital material as a single group when considering digital preservation. The non-traditional materials may be considered to be a higher priority for preservation than the traditional ones, particularly if they are unique, as is the case with unpublished datasets and manuscript materials.

4.2.4 Reasons for acquiring digital collections

Out of interest, the library questionnaire asked respondents about their reasons for acquiring digital content. Many of the respondents to this question gave multiple responses.

Reasons	No. of Respondents	% of Responses	% of Respondents
Access	85	35.9	60.3
Functionality	34	14.3	24.1
Demand	31	13.1	22
Space	26	11.0	18.4
Not available in hardcopy	23	9.7	16.3
Preservation	22	9.3	15.6
Cost effective	5	2.1	3.5
Necessary to maintain collection regardless of format	4	1.7	2.8
Legal deposit obligations	3	1.3	2.1
Ease of reproduction and dissemination	2	0.8	1.4
Donations	1	0.4	0.7
Results of digitisation projects	1	0.4	0.7
Total	**237**	**100**	**168.1**

Table 4.8 Reasons for acquiring digital material

Unsurprisingly, the most common reasons for acquiring digital material were improved access and functionality, with space savings another frequent response. Twenty-three respondents indicated that they were acquiring material in digital form because it is not available in any other format; the responses to specific questions on this issue are given below (Tables 4.11 and 4.12). Interestingly, 22 respondents said

they acquired digital material for preservation purposes. Most of these are from public libraries. Only one respondent specifically mentioned digitisation, but answers to later questions suggest that more libraries than this are involved in digitisation, so it is likely that such material is covered by other answers here. The table below shows the responses broken down by sector.

Reason for acquiring	Academic	Public	Special	National	Other	Total
Access	34	20	24	3	4	85
Functionality	13	10	9	1	1	34
Demand	18	4	6	1	2	31
Space	9	7	8	1	1	26
Not available in hardcopy	6	5	5	1	6	23
Preservation	1	14	3	2	2	22
Cost effective	4	0	1	0	0	5
Necessary to maintain collection regardless of format	2	1	1	0	0	4
Legal deposit obligations	0	0	1	2	0	3
Ease of reproduction and dissemination	0	1	1	0	0	2
Donations	0	0	1	0	0	1
Results of digitisation projects	0	1	0	0		1

Table 4.9 Reason for acquiring digital materials by sector

Three libraries mentioned legal deposit obligations here. This is surprising, since it is known that only two legal deposit libraries responded to the questionnaire. The third library that selected this option is probably the Advocates Library in Edinburgh, which receives some legal deposit materials.

Some of the libraries that do not yet have digital content gave reasons for this.

Reasons	No. of Respondents	% of Responses	% of Respondents
Cost	12	50.0	66.7
No demand	4	16.7	22.2
Not relevant to needs at present	3	12.5	16.7
Storage and preservation concerns	2	8.3	11.1
Incompatible with current systems	1	4.2	5.6
Copyright issues	1	4.2	5.6
No staff support	1	4.2	5.6
Total	**24**	**100**	**133.3**

Table 4.10 Reasons for not acquiring digital information

Concerns over copyright and preservation issues were not prominent reasons for not acquiring digital information. In fact, the most common reason was cost, followed by a lack of demand.

4.2.5 Digital-only publications in libraries

The issue of preservation is most pressing for digital material that does not have a print equivalent. The library questionnaire therefore asked respondents what proportion of their digital material exists exclusively in digital form.

Library Sector	No. of Respondents	% of respondents in sector with digital-only content
Academic	24	50
Public	11	40.7
Special	22	64.7
National	2	40
Other	5	62.5
Total	**64**	**52.5%**

Table 4.11 Respondents with digital material that does not have a print equivalent

Sixty-four of the 122 library respondents who currently have some digital material said that some of this material does not have a print equivalent. Those who did not respond to this question may well not have been able to provide this information without some research. The most that can be said is that of all the respondents with

digital material, at least 52.5% have material that only exists in digital form. Perhaps surprisingly, the sector with the fewest libraries with digital-only material is national libraries (40%). Public libraries were also relatively unlikely to have digital-only material (40.7%), with 'other' (62.5%) and special libraries (64.7%) the most likely.

While about half of respondents with digital collections do have digital-only material, for the majority of them (75%) this type of material makes up less than 50% of their digital collection.

Percentages of digital-only material	No. of Respondents	% of Respondents
Less than 5%	11	17.2
5-25%	24	37.5
25-50%	13	20.3
50-75%	4	6.2
75-100%	12	18.8
Total	**64**	**100**

Table 4.12 Percentage of digital material that does not have a print equivalent

Only twelve libraries (18.8%) stated that more than three-quarters of their digital material has no digital equivalent. It would be interesting to know if any of these only hold digital-only material, and what exactly this material is. The results indicate that 'other' and academic libraries are most likely to have a high percentage of digital-only material. Forty percent and 37.5% respectively stated that at least 50% of their digital material has no print equivalent. Public libraries are least likely to have digital-only material, with 81.8% of them stating that less than 25% of their digital material is of this type.

Library Type	Less than 5%	5-25%	25-50%	50-75%	75-100%	Total
Academic	3	8	4	3	6	24
Public	4	5	1	0	1	11
Special	4	8	6	1	3	22
National	0	1	1	0	0	2
Other	0	2	1	0	2	5

Table 4.13 Percentage of digital material without print equivalent by sector

One library interviewed has increased its journal holdings substantially by acquiring online-only titles, and has stopped subscribing to print abstracting and indexing services. This is illustrative of the way in which academic library collections are developing. The digital materials being deposited under the voluntary deposit scheme currently in place at the British Library generally do not have print equivalents, since for parallel publications where the two versions are substantially the same, the publisher only has to deposit the print version.

5 PRESERVATION ACTIVITIES AND POLICIES

5.1 Awareness of digital preservation

The preservation experts interviewed confirmed that awareness of digital preservation is generally low. They reported encountering a lack of awareness in a wide variety of sectors, including among academic authors, university administrators, local government and charity bodies. One legal expert suggested that awareness also needs to be raised among Members of Parliament if changes to the law are to be made to facilitate preservation.

Some interviewees seemed completely unaware that digital publications are under threat. For example, one publisher commented, 'I don't see what the issue is', while a representative of a rights holders' organisation suggested that:

> *in reality it's not just going to fall apart, because … technology will evolve to find ways of rescuing it and reconfiguring it in some other form..*

Other interviewees said they were aware that they know little about the issues. One publisher said: 'I don't think as a whole we know enough on these topics'.

Several interviewees gave their views on the level of awareness among authors. Their comments suggest that the lack of author interest in the project could well be because of a lack of awareness among this group. As already stated, only 13 authors responded to the online questionnaire, and none were available for interview. While some interviewees thought that there would be a range of views among authors, others believed that:

> *a lot of authors, a lot of academics have just not thought about digital preservation, and often assume if something's available electronically that's fine, it's no problem to preserve it.*

One publisher suggested that:

I think they, by and large, would assume that the publisher or a library was somehow taking care of it.

Other publishers interviewed thought that authors are generally not checking to see whether this is in fact the case. Visual creators are also evidently unaware of the issues surrounding digital preservation. A representative of a rights holders' organisation said that some creators perceive digital publications as having a longer lifespan than print ones.

Those interviewees who did show some awareness of digital preservation often had misconceptions about the extent of the problem, or about how best to solve it. A common problem is the failure to understand the difference between 'preservation and backup'. As one digital preservation expert said:

The main thing I'd be concerned about is the extent to which people still think that all you have to do is to take the stuff and put it on a couple of discs.

Others seemed only partially aware of the problems. For example, one interviewee suggested that different versions of Web browsers and plug-ins might lead to problems accessing digital audio and video in the future, yet seemed unaware that this might also be a problem with ordinary Websites and online materials.

Some interviewees commented on the possible implications of this lack of awareness. For example, one of them expressed the fear that:

we're in danger of creating a situation where we can't access material in ten, twenty years' time

Several interviewees mentioned instances where digital information has already been lost. For example, many early television broadcasts have been lost, and research data is being lost in academic disciplines where there is nobody to take responsibility for it. Some publishers interviewed were confident that they could still access all their

digital publications. However, others were unsure whether they had kept these at all, or had transferred them from obsolete storage media in time.

Several interviewees commented on the need to raise awareness of digital preservation issues. One publisher suggested how this might happen:

> *that is probably a line that libraries should develop a bit more. Because on the various library lists, you don't really see much about that, about long-term preservation.*

He evidently thinks that the responsibility for raising awareness should come from within the library community. Seminar delegates thought that the Digital Preservation Coalition (DPC) should be involved in raising awareness, which is in fact one of the DPC's objectives (see Section 5.7).

5.2 Preservation plans

5.2.1 Rights holders views on what should be preserved, and for how long

Many publishers are reluctant to preserve their publications for the long term because they do not see a need to do this; this is an area of tension between librarians and publishers. One publisher interviewed said 'we get far too hung-up on preserving things in perpetuity'. He feels that:

> *we need to get away from the view that the printed word per se is in some form inherently worth preserving.*

More than one publisher felt that trying to preserve everything that is published would be completely impractical:

> *I'm not sure that seeking to preserve every single last jot and tittle is necessarily viable, apart from anything. … Yes, it's a pity we've lost books and manuscripts over the years, and built and knocked old*

buildings down and built new ones, so it's sort of part of life, isn't it, really. I can see … a situation arising where there's no room for anybody because it's full of libraries!

This is particularly true because the 'information explosion' has made it 'so easy to put out literature now'. In fact, librarians do not try to preserve absolutely everything: even legal deposit libraries do not take all publications, and they realise that they will need to be even more selective with digital publications.

The publishers interviewed agreed that some types of publications are worth preserving, while others are not. They felt that libraries should make a case for preserving particular items, rather than automatically preserving everything. This would depend on a number of factors; two important criteria mentioned were the quality and value of a publication. They felt that particular care is needed in deciding whether or not to preserve materials published on the World Wide Web, since this is a particularly impermanent, ephemeral medium. The interviewees' main concern about preserving Websites was that they are often of poor quality:

80% of Websites people create are so short-lived and the content is unverifiable, or of low quality. They have no standing, citations or anything to really give it credibility. It is just noise.

An author respondent said that: 'preservation of the digital form is not significant for personal Web-site material', and one interviewee thought that there is no point in preserving old Websites at all. Interviewees therefore thought that anyone seeking to preserve Websites must realise that they are not all 'the same' and that they are not all worth preserving.

The publisher questionnaire asked respondents how long they are planning to preserve their digital publications for. More than half of the 29 publishers who replied indicated that they would try to preserve material for as long as possible. Depending on how long "as long as possible" is, this could be encouraging. However, almost 45% said they would preserve material only as long as it is commercially valuable.

	No. of Responses	% of Respondents
As long as commercially valuable	13	44.8
As long as possible	15	51.7
Other	4	13.8
Total	**32**	**110.3**

Table 5.1 How long material will be preserved by publishers

Among the publishers who selected "other", one said that digital material should be preserved for as long as it is useful. The question that arises here is whether "useful" means for the publisher or for users. This suggests that while publishers have some involvement in preservation activities, which contradicts the common view, they may not be taking a long-term view.

The publishers interviewed also commented on how long they think their publications should be preserved for. Some of them thought that their publications are worth preserving for the long term. One said: 'we assume our content has some sort of value over an extended period of time'. This is particularly true for 'refereed research papers' and other academic literature including journals. Humanities publications were said to have longer-term value than those from some other disciplines. The publisher of a digital-only magazine explained that:

> *we believe that we have very high quality debates, on issues of ongoing importance. A key to knowledge is the ability to learn, especially from past events and previous lessons. Our accumulation of content, of important debates and ideas, is of value.*

This type of publication clearly has value beyond the immediate present. However, in some cases publishers are only intending their publications to have short-term value. For example, one publisher explained that the tables of contents that they mount on their Website 'aren't being archived anywhere, they're seen as temporary'. Certain types of publication become obsolete over time because 'knowledge moves on'. A publisher of professional journals believed that 'nobody who's actually in the field' would 'ever' want the information they published five years ago. This publisher also produces a database of suppliers' details that, again, becomes obsolete quickly:

> *I can't conceive a reason why you'd want an old database. If somebody changes their address, there's no commercial need to have a record of old addresses … Maybe lawyers or somebody would need them, but that's not the audience we're talking about.*

This publisher is therefore aware that his material could be of interest beyond its target audience. Schools' publishing also has a short shelf life since both the curriculum and styles of teaching change regularly. Children's fiction also loses popularity over time. An educational publisher interviewed explained that it does not need to preserve the online materials that it produces to support its textbooks for very long:

> *Our books are revised on average about once every three years. And what we will do is usually update or do a new site for the new edition. But we will keep the old site available for up to 12 months for those students who might still be using the old edition. And then, after that period, we would take it down.*

Again, however, this type of material may have some value to historians of education or social historians. Some publishers interviewed conceded that it might be worthwhile for libraries to continue to preserve their materials. However, those who held this view clearly felt that if libraries wanted such items to be preserved, they had to take responsibility for doing so themselves.

The author questionnaire also asked respondents how long they want their digital work to remain accessible. Six of the 13 respondents said that they want their work to be available for as long as possible. The respondent who selected 'other' agreed, stating: 'perpetuity: what's the point of scholarship that is not accessible in the long term?'

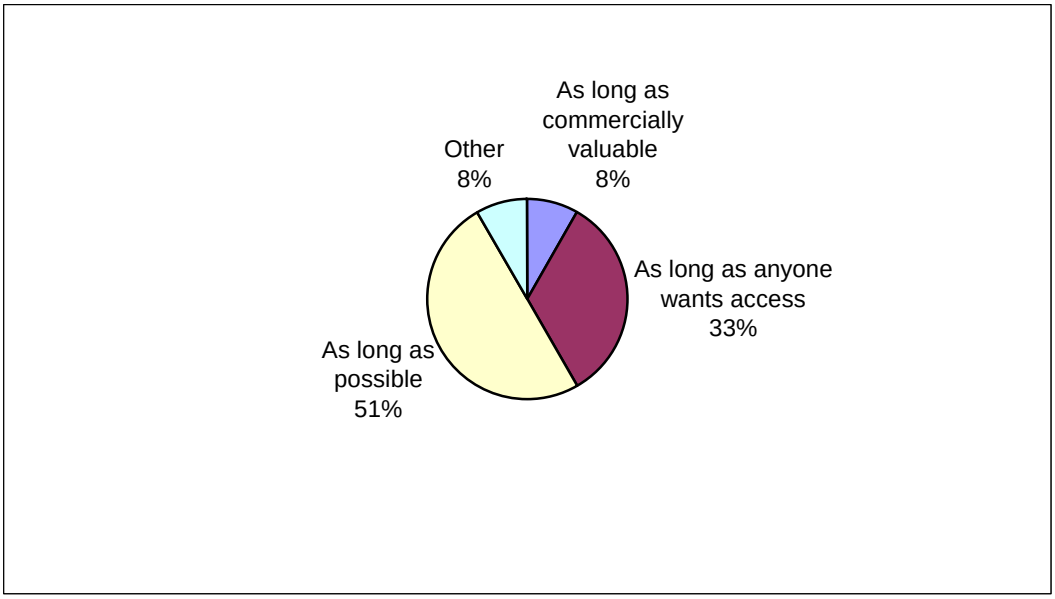

Figure 5.1 How long authors want their work to remain available in digital format

Four respondents said that they want their work to be available 'for as long as anyone wants access'. One of these is writing in philosophy / ethics. It is likely that people will continue to want to access this author's material for some time, since information in these subject areas does not become outdated as quickly as, for example, scientific information. However, one of these respondents writes school textbooks that are likely to go into new editions every few years, and the remaining two gave their subject areas as 'information science' and 'technical writing', which may go out of date quite quickly. Only one respondent said that they only wanted to keep work available for as long as it is commercially valuable. This respondent gave their subject area is 'technology and writing', which of all the subject areas represented is the one likely to lost its commercial value the most quickly. These results contrast with comments made about authors by interviewees. One publisher suggested that academic authors are primarily interested in:

> *being published within a quality, peer-review journal, and get[ting] as much quality dissemination of their work as possible. And that's it.*

This publisher feels that their emphasis is therefore on the short to medium term. Another publisher suggested that authors do not always want their work to be preserved for posterity:

in some cases, they wouldn't want it to be held at all. It's like, "oh no, not that book"

This suggests that the type of publication they are producing determines the author's views on this issue.

The publisher and author questionnaires also asked respondents whether their digital materials are still accessible. Only 29 publishers responded to a question asking whether they can still access their earliest material. Of these, 22 (75.9%) can still access all their earliest material; this only represents 27% of all publisher respondents. One further respondent (3.4%) said that they were able to access some of their material, and six could not do this. The author questionnaire asked respondents what percentage of their digital materials is still accessible. Over half of the 13 respondents stated that at least three-quarters of their digital work is still accessible; it would be interesting to know whether all of their work is in fact still accessible. Only one respondent believed that less than a quarter of their work was still accessible.

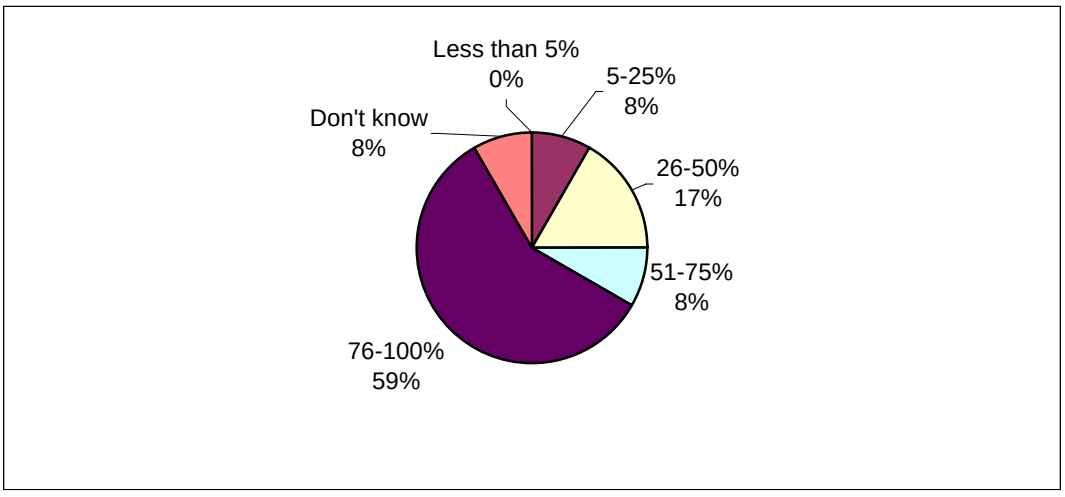

Figure 5.2 Proportion of authors' digital work that is still accessible

Interestingly, the one respondent who has been publishing digitally for more than six years stated that 76 – 100% of this material is still accessible. Of the six respondents who have been publishing digitally for 4 – 6 years, only one can now access less than a quarter of his or her material, while three can still access at least three-quarters of

their material. This suggests that the length of time that an author has been publishing digitally is not a significant factor in determining how much of that work is still accessible. It is perhaps more likely that an author's ability to access his or her digital materials is influenced by the type of hardware and software he or she is using.

5.2.2 Preservation plans of libraries, publishers and authors

The library questionnaire asked respondents whether or not they have a digital preservation policy. Of the 69 libraries that responded to this question, only four currently have a policy. This may indicate that only four of the 122 libraries with digital collections have a preservation policy. Even more worryingly, only four of the 51 libraries taking responsibility for the preservation of their digital resources have a policy to help them do this.

Do you have a digital preservation policy?	**Academic**	**Public**	**Special**	**National**	**Other**	**Total**
Yes	1	1	1	1	0	4
No	20	18	21	1	2	62
Don't know	1	0	2	0	0	3
Total	**22**	**19**	**24**	**2**	**2**	**69**

Table 5.2 Libraries with a digital preservation policy

The questions also asked about respondents' intentions to develop a preservation policy in the next twelve months. Of the 85 respondents to this question, only 27.1% are planning to develop a policy with the rest uncertain or with no plans.

Interviewees were also asked about their preservation plans. One librarian who said that his library sees itself as having responsibility for preservation explained that:

> *saying that we have a responsibility doesn't necessarily mean we have a systematic policy.*

Another librarian interviewed does not have a formal policy for preservation, but has a policy for backing up digital material. He views this as contributing to preservation:

That's partly preservation, because a lot of the content will be static, only really needs to be backed up once.

This will only help until the format in which the content is stored becomes obsolete. The publisher questionnaire asked whether publishers have formal long-term digital preservation policies. Almost 70% of those responding to this question do not have a formal strategy at the moment, but a larger proportion of publisher respondents than library respondents do have a formal policy. Thirteen publisher respondents already have policies, and a further six respondents have plans to develop a policy in the next 12 months.

The publishers interviewed were also asked about their preservation plans. One was unsure whether they had a plan and one had 'no specific plans'. One was not doing anything 'systematically' and another thought that preservation should be addressed on an 'as needs' basis. One publisher admitted that his company is 'not very good at even archiving our print material', and said that they were 'looking to start something' like this. Another felt that 'this is an issue we do need to just make sure we've thought through'. Two publishers stated that they see it as their customers' responsibility to keep the digital materials they have purchased accessible. One of these said that backing up was done for the company's benefit since users 'have got the product already, so they don't need anything else'. Despite this, most interviewees said that their intention was not to lose any of their digital materials. One publisher stated that 'our policy is that we try not to lose anything. That's about it, really', while another said: 'we have absolutely no intention of getting rid of anything that we've published'.

The author questionnaire asked whether authors take publishers' preservation policies into account when deciding where to publish their work. Of the ten respondents who replied to this question, equal numbers replied 'yes' and 'no'. One explained: 'I am first concerned with dissemination'. Another agreed that other factors are more important:

> *Most work in my area is published in journals. These vary considerably in specialisation and in status and breadth of dissemination. These have to be the deciding factors, given career needs.*

A third respondent, who had answered 'yes' to this question added: 'BUT I have never actually seen a publisher publish a policy on preservation'.

The one organisation interviewed that does have a formal preservation policy is the British Library. This is to be expected, since the British Library is closely involved with digital preservation and is likely to play an even greater role in it once legal deposit has been fully extended to digital publications. Deborah Woodyard, the British Library's Digital Preservation Coordinator, explained that:

> *I'm trying to make sure that the whole of the library has a consistent approach to the preservation of digital materials, and also that the preservation of the digital material that we're collecting is taken into account in areas where it may not have been considered.*

In keeping with its role as an archive of the nation's output, the British Library is planning to preserve its digital holdings, 'for the long term / indefinitely'.

The British Library's preservation policy, which has just been published, is very general:

> *It basically contains very broad, high-level statements, say, 'we intend to preserve the digital materials in the collection'. And a couple of basic principles, like 'we will always keep an original copy' … however it came in, although to preserve it we may need to migrate it to another format'.*

The British Library is taking a 'whole life-cycle approach' to digital preservation, so it views preservation as including collecting material, producing metadata and making storage decisions. It is aware of the range of preservation strategies currently

available, and is planning to use different strategies as appropriate, rather than using the same strategy for everything. Its policy therefore:

> *lists a whole range of different preservation strategies that are possible, and says that we will continue to examine them and use whichever one is appropriate as we see fit.*

However, the policy does state that there are some strategies that will not be used:

> *We do say that as a strategy, 'do nothing' is not acceptable. And technology preservation is not suitable, either. So trying to keep all the different machines is not really an option.*

This plan therefore reflects the fact that digital preservation strategies are still being developed.

5.3 Preservation experts and digital preservation

5.3.1 Technical preservation methods used by preservation experts

The digital preservation experts interviewed confirmed the findings of the literature review (Section 3.1) about the reasons why digital preservation will be necessary. One of them added that an important difference between preserving print and digital materials is that digital materials have to be managed actively if they are to survive:

> *Anything digital requires some degree of active ownership constantly. If you don't have active ownership of it, then Computer Services or someone are going to reclaim that service space or decommission the server or so on, then it's gone. It doesn't survive … neglect very well.*

This is another reason why the time frames for the preservation of digital materials are much shorter than those for print.

The digital preservation experts interviewed were asked about the technical methods that they are developing to preserve digital materials. Again, many of their comments confirmed the findings of the literature review (section 3.2), but they also highlighted particular problems associated with the different methods.

One of the digital preservation experts interviewed explained that migration may be difficult to do because:

> *you need to know the design of these different file formats to be able to write a tool that will migrate between them.*

As will be seen (Section 6.3.2), documentation describing file formats is not always available. Once the design of a file format is known, preservation experts need to create a software tool that can extract information from it. These tools need to be written in such a way that they will 'last as long as possible'. One interviewee explained that migration to standard formats hardly differs from 'conventional' migration, since preservation experts still have to overcome this first step of extracting data from a proprietary format. However, once the data has been migrated to 'something sensible', subsequent migrations should be simpler.

A digital preservation expert interviewed explained how emulation works. With this method:

> *you keep the data as it was, and when you can no longer find a program to run it, you … go back to the documentation for how that program was written, and create a new one which will work on new technology*

However, he confirmed that there are difficulties with emulation. It is not enough just to emulate an application. Hence, to emulate Microsoft Word, a preservation agency would also have to emulate the Microsoft Windows Operating System to run it on. This is difficult to do, since it is 'a bit unrealistic to assume that you can just emulate the entire original environment'. Interviewees also explained that emulators are

complex to write, and will become increasingly complex as computers become more advanced. Emulators will themselves have to be preserved over time.

The interviewees explained that the Universal Virtual Computer approach is considered to be less well developed than the aforementioned approaches. They explained that the UVC is broadly similar to emulation, but incorporates an 'extra level':

> *The UVC is a sort of a raft on which you produce different emulators … You can migrate the raft over time, drive that into the future, and that's the only thing you need to preserve over time. … And as soon as you migrate your raft to a new computer, whatever replaces PCs, for example, then all your emulators will run on it.*

The UVC approach was said to be an extension of the principle of using standard formats for publications:

> *We say … use XML, because … then we'll all understand it and it'll survive, and even if we don't have the software, it'll be quite easy to write something which can understand that. UVC is saying, do the same thing for the code.*

The UVC approach was also compared to the programming language Java, since a program written in Java can be run on any platform for which a Java engine has been written.

The digital preservation experts interviewed confirmed that technical preservation is not a suitable method of digital preservation. As one explained, 'trying to keep all the different machines is not really an option'. One interviewee suggested that this method would only be used to rescue important materials:

> *You wouldn't continue to preserve it that way but you'd use that as the means of getting access to it and then taking it off to something else.*

Another method of digital preservation that was mentioned is to 'keep the object in its original form, and maintain a way of just displaying that'. Viewers for different file formats are being developed; these generally only permit items to be viewed and perhaps printed, but not manipulated in any way.

One interviewee explained that there are '101' other methods of digital preservation being developed. In general, these fall into the range:

> *between taking the data and changing it, … taking the format, changing it so it works in newer technology, or taking the program code and changing it so it works.*

They are therefore similar to some of the methods already outlined.

The digital preservation experts interviewed confirmed that it is not yet clear what the best method or methods for digital preservation may be. They commented that 'no-one quite knows how to do it yet' and that it is 'really still emerging as to what is the best practice'. One reason for this is that:

> *certain types of material are going to be suited to migration, certain types are going to be suited to emulation, or a combination of those things.*

An interviewee explained that in some cases, a preservation agency might begin to preserve an item using one method, then realise that another method would be better. However, once a method has been tried and has worked for one item, it can be used for other items of the same type:

> *if we're going to keep … one digital item in a particular format, it's going to be reasonably straightforward to apply the same strategy for the preservation of that to … hundreds or thousands of the same kind of thing*

There should therefore be economies of scale.

The digital preservation experts interviewed also described some of the individual processes that they carry out as part of the work of preservation. One interviewee explained that it is necessary to undertake some work upon receipt of digital materials, 'to make sure that they're suitable for preservation long-term'. As well as keeping the original, this preservation agency creates two new versions of each digital object it receives. One of these will be in a format that is suitable for migrating to other formats in the future, while the other is in a format that is 'convenient' for users to access. As this interviewee explained:

> *if we get an Access 95 database in, we'll put that onto our system with some metadata about it. We'll then export the data, … move it to text file with some sort of … description of the structure of the database, which tells you what you need to recreate that … And then we'd also probably create, say, an Access 2000 or XP version, because that's likely to be the software people will have use of.*

This preservation agency makes it clear to creators that 'we're basically going to extract data from whatever system they have with it' rather than keeping it exactly in its original form.

The preservation experts interviewed agreed that it is essential to keep the original publication received, for number of reasons:

> *It's good in the short term, because if you make mistakes in some other version, you can go back to what you were given. It's good in the short term, because you can give a depositor back exactly what they gave you, which, if they lose it six months later, that's what they'll want.*

The original can also be used to check whether the new versions made are exact copies. The preservation experts interviewed also agreed that it is good practice to keep the 'raw data' from the original bit streams, rather than keeping the actual storage media. This is done because 'none of the media particularly are long lasting'. Once they have been extracted, bit streams can be copied to disc or put into a

repository, backed up and refreshed, and stored in multiple locations. One interviewee said that this is one of the few aspects of digital preservation where he is able to say, 'yes, that's the way to do it'. Interviewees from outside the specialist preservation community also agree that this is good practice. One publisher interviewed recommends that schools that purchase its CD-ROMs with network licences 'copy the whole thing to their server, and just put the CD-ROM in a cupboard'. However, extracting data from physical storage media is not always practical. Deborah Woodyard said that the British Library is not currently doing this with CD-ROMs attached to books, because it receives too many of them.

The preservation experts interviewed also agreed that it is important to act as early as possible in the life of a digital object (see also Section 3.2.3). They believe that emulators need to be tested straight away, so that any errors can be corrected while the original software is still available:

> *That tends to be a lot easier if you're doing it at the time the software's still around and in use, so you can still have a go and compare it. Now obviously it would have been a lot easier to unravel the hieroglyphics 2000 years ago when you could have tried it out with an Egyptian.*

One preservation expert interviewed stated that if an agency is managing its digital materials actively, it should always have a version that it can access:

> *You can't just let it sit for ten years and then think about it. You've got to be aware of the state it's in.*

5.3.2 Difficulties with technical preservation activities

As has been seen, digital preservation experts need documentation describing file formats so they can produce tools to preserve their formats. However, documentation is not always available for proprietary formats. Ideally, such information would be available immediately, but one interviewee felt that:

> *it's much better to have something ten years old than to have no documentation at all and just not be … able to do it.*

One digital preservation expert interviewed hopes that software companies may be persuaded to release documentation about their products after a period of time had elapsed. If one or more new versions had been released, this should not damage the company's commercial interests. Another problem preservation experts have encountered is that even when documentation is available, it is not always accurate. For example, an attempt to emulate a 1970s computer initially failed because the documentation failed to describe a bug in it. The software written for this system had taken account for the bug, so the bug had to be included in the emulator before the software would run on it.

Digital preservation experts interviewed added that problems may also arise when authors and creators do not give adequate information about their digital materials. This type of information is needed to enable decisions to be made about the best way to preserve an item. The preservation agency needs to know information such as:

> *What it is, where it's from, how you created it, why you created it, what decisions you made which might [enable] someone to use or interpret it*

This is more necessary with materials such as datasets than with published research output, since this type of information needs some explanation.

In the light of all these practical difficulties, it is perhaps not surprising that one of the digital preservation experts interviewed said that 'we don't guarantee to maintain access to everything we're given'. He commented:

> *I'd be willing to go about as far as saying, 'if you give us a plain text file in ASCII, 7-bit ASCII, then I could present that'. But even there I can't guarantee that the font will be the same*

This is partly because further work is needed to develop preservation methods, but also because of a general lack of faith in IT. As one interviewee said, 'my computer crashes twice a day, so what are my chances of getting it to run in fifty years?'. Agreements between preservation agencies and depositors generally reflect this, and do not guarantee retention periods. These agreements are generally licences, and, as a digital preservation expert explained, they need to be worded carefully. They need to be precise enough that depositors can understand what may be done, but not so precise that any future technological developments are excluded. The digital preservation experts interviewed hoped that better preservation methods would be developed in the future so that they would be able to make guarantees about what they can do.

Another issue facing preservation agencies is how they should decide what to preserve. They will need to be selective because 'we don't have resources in terms of people or costs to preserve everything'. As one librarian explained, this is not a new problem for libraries:

> *Librarians have traditionally not preserved everything, … the fact we have collection acquisition policies shows that we're selecting certain stuff. And then even within our collections we may prioritise certain things for preservation in the print environment.*

However, while legal deposit libraries have traditionally preserved almost everything published in print, Deborah Woodyard explained that the British Library is likely to have to be more selective about what it preserves in the digital environment. Interviewees felt that the need to be selective is greater now than ever before, since it is so easy to publish digitally and this often has an adverse affect on the quality of publications. One publisher felt that:

> *one should be making a positive case to preserve things rather than the default assumption that things should be preserved unless they're clearly worthless.*

Seminar delegates thought that preserving libraries, in line with their collection policies, should ultimately make these decisions. However, there should be some input from other stakeholders.

The preservation experts interviewed mentioned several criteria for selection. They felt that the output of research funded by funding councils would be worth preserving, since it will already have undergone 'quality review'. Items that are unique or only available digitally were also said to be priorities for preservation, as were items that are in great demand from users. The views of creators and rights holders should also be taken into account, although, as has already been seen, publishers do not always see the value of preserving their publications for the long-term. Repositories may need to make two decisions about an item's value, firstly about 'acquisition and access' and then about preservation. This may not be easy: a preservation expert suggested that it may be dangerous to make such decisions, since you may find out too late that you have preserved the wrong items.

5.4 Preservation activities of libraries, publishers and authors

5.4.1 Stakeholders' current involvement in preservation activities

The questionnaire results indicate that some libraries are already taking responsibility for the preservation of their digital materials. However, while 73% of library respondents have digital collections, less than half of these (41.8%) said that they are currently taking responsibility for the preservation of this material. A further 24.5% of libraries with digital material are planning to take responsibility for its preservation in the future. The library questionnaire also asked respondents whether they are currently taking any preservation action. Eighty-two libraries responded to this question. Interestingly, 26 (55.3%) of the 47 respondents who are currently taking no action had indicated that they take responsibility for the preservation of digital materials in their collections.

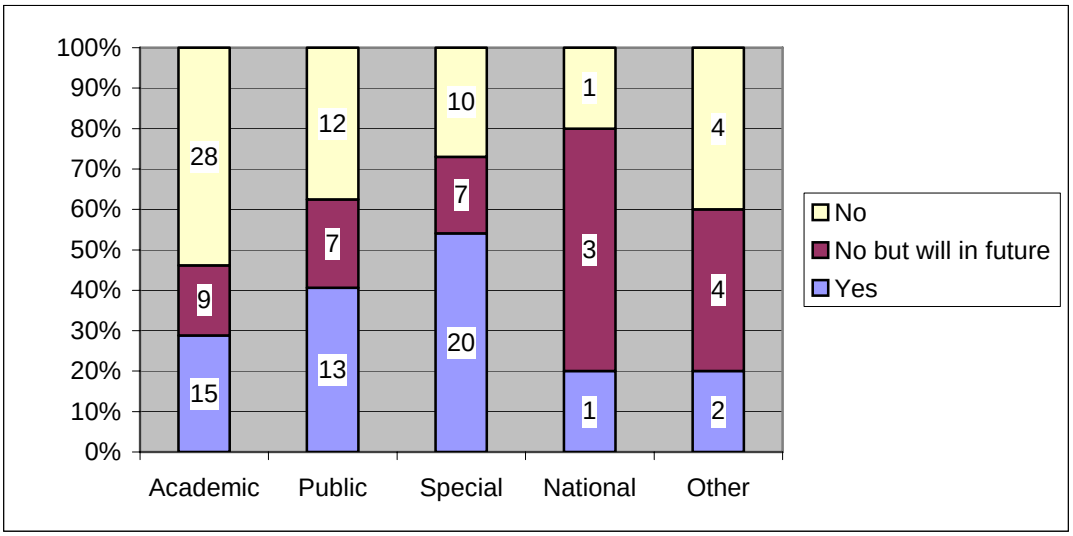

Figure 5.3 Libraries Taking Responsibility for Preservation by Sector

When the responses are broken down by sector, it is clear that a greater proportion of special libraries are taking responsibility for preserving digital material (54%) than libraries in any other sector. When future plans are taken into account, 80% of national libraries and 62.5% of public libraries will take responsibility, while only 46% of academic libraries will take on this role. The reasons for these responses are not clear, but there may be several explanations. For example, public libraries may be creating their own digital content through digitisation and therefore wish to protect their investment. While the response from special libraries might be unexpected, it may depend on the types of special library that responded. For example society and government libraries are more likely to be interested in preservation than work place libraries. Some respondents did give some reasons for their answers. Reasons given for not taking responsibility included:

- No need to preserve (have print copies, material publicly available, alternative department's responsibility, issue has not yet arisen).
- Cannot preserve (licensed/remote access only)
- Do not preserve (the nature of the library; focus on current material, do not preserve anything)
- Too expensive, time consuming, difficult
- Have not yet addressed the issue

Of the library respondents who do take responsibility for preservation, most do so for material they physically own, either through purchase, donation or deposit. However, fourteen (17.3%) claimed to take responsibility for the preservation of licensed material. These respondents did not give details of what this actually means.

Method of acquisition	No. of responses	% of responses
Purchased outright	23	13.7
Other	22	13.1
Licensed	14	8.3
Donated	13	7.7
Deposited by publishers	9	5.4

Table 5.3 Responsibility for preservation by method of acquisition

Of the 22 respondents who selected "other", 20 take responsibility for material created in-house, one takes responsibility for student work and one takes responsibility for material of unknown provenance. What is meant by "taking responsibility" needs to be explored further, but if these responses are representative of the library sector and taking responsibility means copying material for preservation purposes, then it will be important to sort out the legal issues.

Publishers were asked whether they are undertaking short- or long-term preservation. More than half of the 42 publishers who responded to this question indicated that they are undertaking long-term preservation, with some respondents giving more than one response.

	No. of Responses	% of Responses	% of Respondents
Short Term	15	30.6	35.7
Long Term	23	46.9	54.8
Neither	8	16.3	19.0
Don't know	3	6.1	7.1
Total	**49**	**100**	**116.7**

Table 5.4 Publisher preservation activities

Nearly 15% of all publisher respondents are taking some sort of action every 12 months or less, with a further 8.5% taking action every one to five years and one respondent taking action every five years or more. However, almost a quarter of respondents did not answer this question.

Authors were also asked who currently backs up or preserves their digital work. The most frequent option was 'self', which was selected by nine respondents. It would be useful to know what action these authors are taking, since this would show how aware they are of the issues surrounding digital preservation, and whether the action they are taking is sufficient to prevent their work becoming obsolete. Unfortunately, the questionnaire did not ask this question. Publishers are also widely relied on for preservation and backing up; seven respondents selected this option. A few respondents stated that their employer currently does this for them; this may refer to employers' back-up procedures. Only two respondents are currently submitting their works to e-print archives. This reflects the fact that such archives are still being developed.

	No. of respondents	**% responses**	**% respondents**
Self	9	39.1	69.2
Publisher	7	30.4	53.8
Employer	3	13.1	23.1
E-print archive	2	8.7	15.4
Other	2	8.7	15.4
Library	0	0	0
Legal deposit library	0	0	0
Other voluntary deposit scheme	0	0	0
Total	**23**	**100**	**176.9**

Table 5.5 Who backs up or preserves respondents' digital work

No authors stated that their works are currently being preserved or backed up by libraries. This is perhaps surprising, given that seven authors thought that libraries should be responsible for long-term preservation, and five thought legal deposit libraries should do this.

5.4.2 Preservation methods used

The library and publisher questionnaires asked respondents about the preservation methods they are using. No library respondents are undertaking emulation, which is not surprising given that this strategy is still the subject of research and development rather than a working tool. Some of the libraries are relying on preserving the original technology, but some are carrying out data refreshing and migration, both of which require copying.

Preservation strategy	No. of responses	% of responses	% of respondents
None	47	48.5	57.3
Refreshing	17	17.5	20.7
Migration	15	15.5	18.3
Don't know	7	7.2	8.5
Other	6	6.2	7.3
Technology preservation	5	5.2	6.1
Emulation	0	0	0
Total	**97**	**100**	**118.3**

Table 5.6 Use of preservation strategies by libraries

Of the library respondents who selected 'other', one prints out internally produced digital documents and databases and stores them in hardcopy form. Another said that their priority was to preserve original materials and conservation microfilm, not digital versions. One respondent explained why few libraries are using technological preservation methods:

> *Digital preservation (especially using techniques such as emulation) is at present far too specialised for most libraries. At present it would be best handled by the British Library (or other copyright libraries, or other specialist libraries).*

Twenty-six publishers answered a question on the preservation strategies they are using. It is very difficult to draw any conclusions on publisher preservation activities because of the low number of responses to the questions. More than half are

refreshing material. Three said they are using emulation. It would be interesting to know what they meant by this and why they are using emulation.

Preservation strategy	No. of responses	% of responses	% of respondents
Refreshing	14	43.8	53.8
Technical preservation	6	18.8	23.1
Other	5	15.6	19.2
Migration	4	12.5	15.4
Emulation	3	9.4	11.5
Total	**32**	**100**	**123.1**

Table 5.7 Use of preservation strategies by publishers

Again, few respondents are using technical preservation strategies such as migration and emulation. One publisher interviewed explained why this may be: he felt that this is 'the language of a completely different community'. It is possible that although definitions of preservation and preservation strategies were given on the questionnaire, the respondents did not know the answers to the questions either because of lack of awareness or activity in the area.

Several publishers interviewed are relying on keeping multiple copies of their publications, either within their own organisation or elsewhere. These publishers either keep extra copies of physical objects such as CD-ROMs, or hold the data in storage systems. Interestingly, more than one publisher is relying on keeping print copies rather than digital ones. As one said:

> *We do see the continuation of publishing print as a simple solution to making sure everything is archived.*

Another publisher stores its materials on microform for preservation purposes, because 'microform does have a preservation standard in a way that digital clearly doesn't'.

Several interviewees felt that storage is an important part of digital preservation; some suggested that good storage must be in place before digital preservation

methods can be used. The British Library is has been working towards implementing a Digital Object Management System for some time. Deborah Woodyard explained that:

> *we need to actually store the stuff first, which is actually such a large job, that's going to take quite some time before we can really consider the rest.*

Adequate storage space was viewed as extremely important; one interviewee commented that 'in a sense, preservation's got a lot to do with storage capacity, I suppose'. Storage space may be less of a problem with digital materials than with print, although one interviewee suggested that even the British Library would not have sufficient storage capacity for digital materials once legal deposit is extended to include them. However, one publisher did mention that storage for digital materials is becoming a lot cheaper. Interviewees also agreed that digital materials need to be stored in temperature and humidity controlled conditions, and be protected from fire and radiation to prevent disintegration. More than one interviewee is storing multiple copies of their digital material in different places as another safeguard.

However, contrary to what several interviewees implied, good storage alone is inadequate to ensure the long-term preservation of digital materials. As one interviewee commented:

> *There's no problem with storage capacity, storing bits is no problem. It's being able to associate those with metadata, and being able to associate those with policies to resurrect and migrate material which is the problem.*

Digital repositories are therefore using specialised storage systems to assist with this. One academic library has purchased such a system for its digital library: the system is secure enough to be used in the defence sector, and all its content is backed up on more than one server to prevent loss. A further security feature is that once files have been entered into this system, they cannot be changed or deleted. Preservation copies therefore need to be entered separately, but the system can then point users towards

the most up-to-date version of a document, while also providing links to earlier versions. Again, however, while these systems help with digital preservation, they cannot prevent digital obsolescence.

More than one interviewee, both publishers and librarians, mentioned the LOCKSS system (Lots of Copies Keep Stuff Safe, http://lockss.stanford.edu/). This initiative has the potential to assist publishers and libraries to implement archival clauses in licence agreements. LOCKSS is a distributed digital archiving system, involving (apparently) low cost persistent digital cashes of electronic journal content maintained in institutions subscribing to the journals. Libraries take custody of material in all formats delivered via HTTP, rather than have publishers deliver the material on an offline medium. Material is collected as it is published. Permission to do this is acquired at the point of subscription, with the publisher indicating whether their material is "LOCKSS compliant", so libraries do not need to negotiate permissions on an individual basis. LOCKSS caches cooperate to detect and repair preservation failures. The technology will be supported by the LOCKSS Alliance, which will provide services to libraries and publishers wishing to use the LOCKSS system. LOCKSS has the potential to address the problem of continuing access to material, according to licence agreements, when it is not available from the publisher or a subscription ends. What is not clear is how this initiative will be sustained in the long-term when different libraries are likely to have different versions of material if they carry out preservation actions. Deborah Woodyard explained that:

> *it manages to keep online journals alive for now, in the current format. But it cannot take them forward into the future, because it has no long-term strategy. There's no migration, there's no emulation, or similar, or alternatives to those. It's just to keep things alive now, and to make sure the copies that exist aren't corrupted. But it's not a long-term strategy.*

However, she thought that it might be useful as part of a 'life-cycle' approach to digital preservation.

One publisher interviewed explained that their participation in the LOCKSS scheme is part of their strategy of storing multiple copies of their content in different places:

> *we feel that that is the way of preserving, or making sure that at least it's somewhere.*

However, another publisher interviewed was reluctant to participate in the scheme because it does not support their policy on access to back sets after a subscription has been terminated:

> *The technology cannot distinguish between material subscribed to, and archive material. If a library cancelled their subscription, they would still have access to archive material to which they were no longer entitled.*

This publisher was also concerned about how secure the LOCKSS system is. They felt that the system would work well if the people involved were trustworthy; however, since materials are being shared, there is always a risk that untrustworthy people could acquire them.

Some publishers are interested in making arrangements whereby a specific library takes responsibility for the long-term preservation of the publisher's entire output; more than one publisher mentioned the agreements that have been made by Elsevier and Kluwer with the Koninklijke Bibliotheek (KB) in the Netherlands. The KB has made deals with the Dutch publishers, Elsevier Science and Kluwer, and the UK-based publisher BioMed Central to archive these publishers' entire outputs 'in perpetuity'. It will update the publications as technology advances, to ensure that they are always accessible. Elsevier's material will initially only be made available to those who visit the library, or are allowed access to its collections. However, should Elsevier cease to make its publications available commercially, the KB will be able to provide remote access to them to anyone (New e-preservation deals, 2002). Since BioMed Central is an Open Access publisher, its publications will be available via the KB Website from the start (BioMed archive deal, 2003). For all these publishers, the

KB's archive will be used to provide backup should the publisher's own server ever be out of action.

5.4.3 File formats and preservation

The file format in which a digital publication is created may affect its long-term survival. The publisher and author questionnaires therefore asked respondents about the formats they currently use, and in particular whether they use open or proprietary formats and why. Interviewees were also asked about this, and were asked whether they had encountered any difficulties in accessing and preserving the formats they use.

Forty-one publishers indicated what formatting and encoding standards they use. Since only 34 indicated that they are publishing digitally, some of these respondents were presumably referring to the internal version of material published in other formats. The majority of these respondents use a mixture of open source and proprietary standards and a further substantial proportion use only proprietary standards.

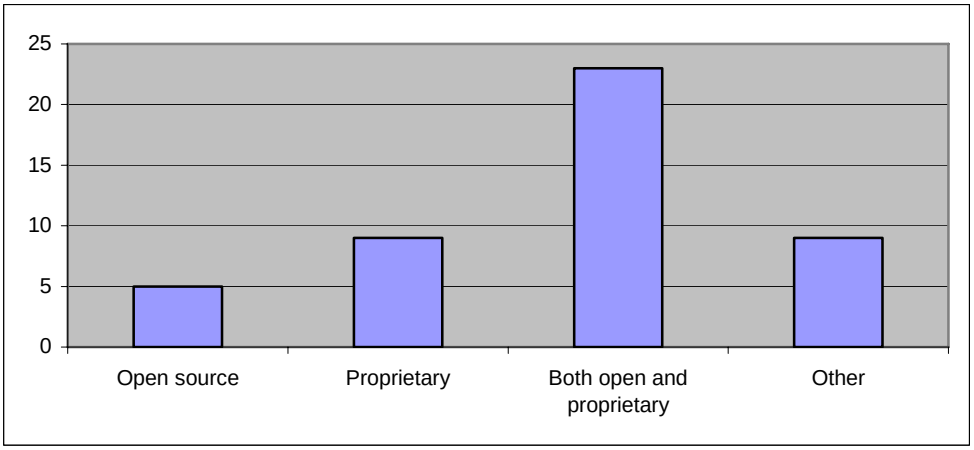

Figure 5.4 Formatting and encoding standards

Among the "other" formats or encoding standards selected by respondents were Adobe Distiller and PDF, which are both proprietary standards, and XML which is open source.

Authors were also asked which publication formats or encoding standards they use in their digital publications. Eight of the 12 authors who answered this question use open source standards such as the mark-up languages HTML language and XML. Five respondents use proprietary formats, with three using both types of format.

Format	No. of responses	% responses
Open Source only	5	41.6%
Proprietary only	2	16.7%
Open Source and proprietary	3	25%
Other	2	16.7%
Total	**12**	**100**

Table 5.8 Formats of digital publications

One of the respondents who selected 'other' explained that he or she uses postscript files. This is a programming language, used to describe graphics and text for printing purposes. The other explained:

> *The Websites that I most use nowadays make use of software called Cold Fusion, which is basically a database which can present material to the Web.*

Half the respondents selected more than one option for this question. It would therefore be useful to know exactly which file formats, and how many different formats, respondents are using.

Several publishers interviewed use proprietary formats such as Microsoft Office products, while other avoid proprietary formats altogether. More than one interviewee described situations where users had experienced difficulties accessing digital materials because of the differences between different versions of proprietary software. For example, one explained that a Website they had developed did not work with earlier versions of a proprietary browser. Such problems are serious for publishers, since they clearly want and need customers to be able to access their materials. An Open Access publisher interviewed therefore deliberately uses standard formats, to ensure that it can provide as wide access as possible.

Interviewees mentioned several other proprietary formats they use. Three publishers interviewed use the mark-up language SGML; one commented that this is a 'is a fairly easily transferable format from one platform to another'. HTML was also mentioned by three publishers; in one case, this was just used for delivery, not for storage. Other formats mentioned include Linux formats and RTF. SCORM (Sharable Content Object Reference Manager) may be used for creating e-learning materials, and TeX and LaTeX are typesetting systems that may be used for scientific articles including formulae. JPEG, GIF, TIFF and RasMol (a molecular graphics program) files may be used for images. One publisher writes its own software so as to be platform independent; this could cause difficulties for a library trying to preserve its publications, if the agency does not have access to documentation describing the software.

Proprietary formats may present particular problems for long-term preservation. One preservation expert described Microsoft Word as 'contentious':

> *Word [is] obviously proprietary, which means … not a good format for preservation, because you don't know how to decode it if Microsoft disappears. And the counter-argument is: 'yes, but it's used so widely that there's very, very little chance of ever getting caught in that situation as long as you're actively managing things... As long as you don't have a twenty year gap where you forget about it and then notice that things have moved on and all the migration tools were ten years ago'.*

As this interviewee indicates, a further problem with proprietary formats is that they are generally updated every two or three years. A digital preservation expert therefore argued that using standard formats should delay the need to take preservation action.

While PDF is a proprietary format, one interviewee described it as 'a kind of a de facto standard'. This is probably because it is so widely available; comments from interviewees confirmed that it is very widely used. Several publishers said that they use PDFs for their electronic journal articles, while one interviewee said that the

British Library uses a customised version of this format for its document delivery. A librarian interviewed explained that it uses PDFs for its digitised materials because PDF viewers are freely and widely available. PDFs are also liked by rights holders, since they are a 'a true representation of the original'.

PDF files were said to be ambiguous from a preservation point of view. At present, Adobe makes the format specifications available on its Website. However, as one preservation expert observed, they only make the specifications of the most recent version available, which could cause problems with older documents. Another potential danger is that:

> *Adobe …may change their licensing agreement, and it may be that you can't use PDF any more.*

Similar problems have already arisen: a digital preservation expert explained what happened when the company that designed GIF files patented the compression element in them and tried to demand payment for their use. In this case, the company was forced to back down, although work had also been done in the meantime to develop a similar format to replace it. One interviewee had heard about a 'rival' to PDF, but did not know how this would be different, or whether it would be better.

The interviewees mentioned a number of ways in which standard file formats may facilitate preservation. Standard formats are easier for preservation agencies to work with than proprietary formats, since the source codes needed to create tools for preservation are widely available. It is also less labour-intensive to preserve a small number of standard formats than to preserve a large number of proprietary ones. Despite these advantages, one digital preservation expert felt that standard formats would only have limited usefulness, since:

> *most of the standard formats that people suggest aren't actually going to be around for that much time anyway.*

They will therefore only delay obsolescence, rather than preventing it. However, the preservation experts interviewed generally felt that it is best to try and preserve materials in standard formats.

According to Nick Dempsey of EPS Ltd, XML is the 'only really … standard format'. The use of XML is becoming increasingly widespread in the publishing industry; Nick Dempsey said that in the last four years,

> *XML has gone from being on the wish list to being something that people feel they've got to grips with.*

Only one publisher interviewed is actually using XML at the moment, with another piloting it, and a third considering using it in the future. A current barrier to the acceptance of XML is that it is expensive to implement; it is therefore mainly large publishers that are using it at present. Interviewees mentioned several advantages of XML. It is a standard, neutral way of storing information that can be understood easily by people as well as being machine-readable. It uses a series of tags that define what is in the information, meaning that XML-tagged information can easily be re-purposed. This feature is particularly useful for publishers: one suggested that XML tags might be useful in producing translations of publications, and others explained that XML-tagged material could be exported easily to other types of files or re-rendered automatically.

Some interviewees believed that XML would ensure the longevity of digital publications:

> *there will always be a way of reading the codes, reading the actual letters and the actual structure codes. It's inconceivable that… I mean, yes, if all of a sudden we … go over to a completely different way of writing, a completely different alphabet and so on, then we might get into difficulties. But that's not really foreseen.*

Another interviewee described XML as 'eminently future-proof'. However, some publishers suggested that XML might not always be seen as the best format for

preservation. One said: 'we believe that this is really the best we can do at this stage of the technology' while another thought that:

> *if something better and longer-lasting than XML came on the market, and was represented as the new technology to use, then we would consider that, on the benefits of adopting that over any other system that might be available.*

A preservation expert agreed that relying on XML would be foolish, since: 'if you look back in history, you'll see that most standards don't actually last very long at all'. He gave the example of ASCII (American Standard Code for Information Interchange), which was seen as a fundamental standard, but is now being replaced by several different versions of Unicode. This may well become a problem since, as Nick Dempsey explained, XML is already being used in a variety of ways, with different industries developing their own 'flavours' of XML.

The benefits of standard formats mean that some preservation agencies try to persuade depositors to use standard formats. This is particularly true of agencies preserving datasets and e-prints. One preservation agency advises academics about file formats and metadata for the datasets they are creating. As one interviewee said:

> *there ought to be some kind of understanding that they've got to prepare, or produce, their digital artefacts in such a form that makes preservation easier. You've got to use standards, for example, where possible.*

However, this is not easy to do since:

> *Generally, users create documents in the formats that they want to, it's impossible to force them to do it the sensible way.*

An interviewee involved with e-print archives explained that repository managers are currently reluctant to insist that academics submit their e-prints in particular formats, as this may discourage them from submitting altogether. Preservation agencies may,

however, refuse to accept some formats, such as 'really proprietary' formats or formats that are already obsolete, unless the item is considered to be particularly valuable. Another solution mentioned is for e-print repositories to accept articles from authors as Word documents, but migrate them to PDF or another format on receipt. A third possibility is for e-print archives to accept proprietary formats for access purposes, but not guarantee to preserve them in the long-term. This reflects the fact that e-print archives are primarily about 'immediate access rather than about preservation'.

A final recommendation made by an interviewee involved in the Open Archive movement was that preservation agencies should keep a list of the file formats they hold, so that they can keep track of them. This is particularly important if the agency holds a large number of formats.

6 RIGHTS ISSUES OF PRESERVATION COPYING

6.1 Rights issues and digital preservation

6.1.1 UK copyright law and digital preservation

One of the objectives of the Copyright and Licensing for Digital Preservation Project was to find out whether existing UK copyright law allows libraries to undertake preservation copying of digital materials.

As a legal expert commented, Section 42 of the Copyright Designs and Patents Act 1988 reads as if it has been drafted with print materials only in mind:

> *It makes more sense as drafted, I think, if you're thinking of print books. But it doesn't actually limit it to the medium on which these particular copyright works are recorded.*

However, it does not specifically exclude digital materials, since it does not mention the formats of the materials that may be copied. This interviewee therefore thought that this might apply to digital materials, although the courts can only ultimately decide this.

However, two other parts of Section 42 were thought to be likely to affect libraries' ability to preserve digital materials. Firstly, libraries may only copy materials from their *permanent collection* for preservation. Interviewees were asked whether they thought this applies to materials that a library subscribes to and has access to, but does not physically own. Two of the librarians interviewed felt that this is ambiguous. As one said:

> *'hold' is itself an ambiguous term, isn't it, because many of them we will have access to, but won't hold them.*

The legal experts interviewed felt that this clause will apply for digital publications on physical storage media, e.g. CD-ROMs, since libraries own copies of these.

However, they did not think that this clause would apply to publications that libraries license but do not own. One legal expert acknowledged that this was a debatable point, for which there is no relevant case law. This interviewee said that the law had clearly been written with print publications only in mind. They concluded:

> *This works right if you're thinking of a book. Otherwise you have to be a little bit inventive in how you interpret it, but you can't obviously invent things that aren't correct. So you can't… twist the wording and make it match something that really isn't true. And … you don't actually have the database in your premises. I think that's a little odd to say it's in your collection.*

A seminar delegate thought that materials that libraries subscribe to but do not own cannot be seen as part of their permanent collections because the subscription could end at some point. It seems likely, therefore, that copyright law does not allow the preservation copying of licensed materials.

Section 42 also contains a condition that states that 'you can't do anything until … it's not reasonable to actually purchase another copy'. A legal expert interviewed interpreted this as meaning that with digital materials:

> *you certainly, therefore, are not going to be able to purchase a copy and immediately preserve it, because you could have purchased two copies.*

This shows a misunderstanding of the issues: buying an extra copy would not remove the need for preservation activities. However, it does appear that this clause in current UK copyright law will at least restrict, if not completely prohibit, digital preservation activities. Not being able to make preservation copies until an item is no longer available would be seen as unsatisfactory by preservation experts, since this could mean that the material was already obsolete, and that it was too late to take action to preserve it.

6.1.2 UK database law and digital preservation

Interviewees thought that database law might also affect the ability of libraries or others to preserve digital materials. One preservation expert felt that database law could inhibit preservation activities. This is because it is difficult to preserve the design of a database when copying it to a different format, but the law protects the design of databases. One legal expert suggested that some preservation copying might be allowed, since:

> *lawful users of a database are able to extract and reutilise insubstantial parts of the database for any purpose*

This would not, however, allow whole databases to be copied. UK database law does contain some exceptions, which differ from those in copyright law. One interviewee thought that copying a whole database for preservation might be considered to be fair dealing, but only 'if you got to the point where there's no alternative but to do something to preserve it'. She thought, however, that this would only apply once a database already was, or was about to become, obsolete, which might be too late.

As with copyright law, it is not clear exactly what effect this law may have on preservation copying. Again, this is partly due to a lack of case law. One legal expert interviewed was not aware that this issue had ever been discussed properly in the UK. The UK database law is based on the EU Database Directive, so the case law on database extraction from Germany and the Netherlands might be helpful. A legal expert interviewed advised that because the UK law is based on an EU directive, any changes needed to preserve electronic databases would have to be made at European level.

6.1.3 Rights implications of preservation activities

The digital preservation experts interviewed said that the rights implications of preservation activities are as yet badly understood, and may become an issue for digital preservation before technological issues. There is no relevant case law, and

preservation agencies do not want to become test cases. Interviewees therefore felt that these issues need to be resolved. As a publisher representative explained:

> *it seems to me that there are some major issues which probably haven't been addressed*

One such issue identified is that 'it is not always libraries who are preserving'. Even if the law allows libraries to carry out preservation copying, this may not help other organisations to preserve.

Preservation activities involve copying, and in some cases, it is not clear whether they may be carried out legally. Preservation agencies routinely copy bit streams from digital materials to facilitate long-term storage and preservation. As one interviewee said:

> *We don't necessarily have the right to copy it for preservation. Which, basically, is what you have to do for digital materials. You have to copy it at least onto another storage medium.*

Re-copying will also almost certainly need to take place over time, for example with repeated migrations. A further issue is that many preservation agencies want to keep multiple copies of digital materials for purposes of redundancy.

A digital preservation expert thought that there could be significant problems with copyright if a preservation agency were to write completely new software to 'mimic' another computer. Rights holders might see writing new software, perhaps to give a digital item a new interface, as 'republishing' their material. The original designers might complain if their interface were replaced with a new one and the preservation copy looked completely different from the original. The main issue here is 'whether a preserved version … constitutes a copy or a new version'. There could well be legal difficulties if it were viewed as a new version, but if it could be argued that it is just a copy, then 'maybe that's something that you're just taking action to preserve'. In the opinion of the interviewee who raised these matters, it would be 'difficult' to argue that a re-created version was just a 'copy'. Again, he suggested that keeping an

original copy might help with this. He therefore argued that emulation might therefore be the best preservation method to use from a copyright point of view, 'because you can argue that it is just a copy, it's not a new version'.

The digital preservation experts interviewed described some of the practical copyright difficulties that they had encountered in their work. One interviewee who had been involved in preserving the BBC Domesday discs described this as a test case for preserving copyright digital material:

> *With Domesday, we had to tread very carefully in that it was such a new area. We didn't want to come up against a brick wall and have one of the original copyright owners who contributed to Domesday saying, 'no, you can't do this', and then have Domesday disappear and be lost.*

This interviewee felt that it had been easier to gain permission to preserve the BBC Domesday project than might have been the case for other publications, because the BBC Domesday project is so well known. Many of those who had contributed were keen for it to be revived and wanted to see the preserved version. It is therefore possible that there would have been a public outcry against any company that refused permission for its material to be used in this way. Even so, permission was only obtained to preserve one of the two original discs, and permission was not granted to provide access to the preserved version.

Another example mentioned was the Internet Archive, which does not seek permission to archive every Website, but states clearly that it will remove anything if the rights owner is unhappy about its inclusion. This was seen as being a pragmatic approach; the Internet Archive has apparently been threatened with legal action, but has never actually been sued.

A further issue mentioned is that publishers may protest if the appearance of their material is changed because they have copyright in the typography and layout of their publications; as one publisher explained:

> *publishers cannot really claim … copyright on the data or the knowledge embedded in articles… So there's a bit of a move to claim copyright for the exact wording and the exact rendition.*

A librarian interviewed explained that publishers already insist that digitised copies must be an 'exact representation of the original'. He felt that they might also expect this of preservation copies.

6.1.4 Copyright and licensing issues with software

The preservation experts interviewed have also experienced practical difficulties in their work because of copyright and licensing issues with software. For example, UK copyright law allows anyone purchasing computer software to make one backup copy, but would not allow libraries to make and store multiple copies, as they would want to.

Interviewees were concerned that preservation copying could be hindered if the conditions of software use changed in the future. Software is also governed by licence agreements, and these, too, may affect libraries' ability to preserve digital publications. A digital preservation expert expressed concern that the conditions of use of software may change. This might make it difficult, impossible or illegal to use preserved digital materials in the future. One interviewee suggested that licences are already beginning to affect people's ability to preserve digital materials:

> *I always understood that if you had some computer software you can make a backup copy of it, that's fair, that's part of the [deal]. But the sort of things that are coming into licensing agreements are starting to impact on that.*

He believes that licences should allow users to preserve software that they have bought:

really, there should be a right to preserve something that you've bought. If you've bought a license to be able to use it, then you should be able to take action to ensure you can.

One preservation expert had heard rumours that some software companies may stop granting use in perpetuity as at present, and start issuing limited licences for their products. This could affect emulation, since if a computer is emulated at the hardware level, the preservation agency will need a valid licence to use the original operating system software. Changes to the conditions of use of software could also cause problems with peripheral software such as plug-ins or print drivers.

Interviewees also mentioned the possible effects of encryption and technological protection measures. These are a particular concern now they are protected under UK copyright law. Systems are being developed and implemented that, for example, prevent people playing music CDs on PCs, and prevent them making more than three digital copies of minidiscs. Preservation experts are therefore concerned that this type of measure may prevent them from preserving digital materials, although one of them concluded that:

There's nothing there that is really massively worrying at the moment, but it shows future trends.

Another result of software being subject to copyright is that software companies do not release documentation describing their file formats. This is another important concern for digital preservation experts, since:

you need to know the design of these different file formats to be able to write a tool that will migrate between them.

One of the preservation experts interviewed had carried out some research into file format information. This documentation is very hard to obtain for proprietary formats, since:

> *A lot of companies feel that they don't want to release documentation about the structure of their file formats, because… you'd lose the commercial advantage, and they want to tie users into it.*

This preservation expert is arguing that there is no commercial advantage in denying access to this documentation in perpetuity, since 'they won't be selling it twenty years later, there won't be a market for it'. However, it is not clear how likely it is that software companies will accept this argument.

6.1.5 Moral rights, publisher rights and digital preservation

A number of interviewees raised the issue of moral rights, particularly the moral right of integrity. The moral right of paternity – the right to be identified as the author of a work - was clearly less of a concern to rights holders. Only one interviewee mentioned this right, explaining that it may be a particular problem in the digital environment, since: 'digital's really quick, you often lose the attribution'.

The comments made by interviewees suggest that publishers and creators interpret the right of integrity to mean that their digital publications should not be changed in any way as a result of preservation activities. A legal expert suggested that such concerns were actually a 'red herring', since:

> *The integrity right is not a right to object to any change whatsoever, it's the right to object to derogatory treatment of the work. Which is … much narrower than being able to object to the fact the colour's changed slightly.*

In this interviewee's view, it seems unlikely that the minor changes that might result from preservation activities could be viewed as 'derogatory treatment', since:

> *if you're doing your best to make it the same, I think it's quite unlikely that you've ended up doing anything derogatory.*

There might be a few rare cases where a creator could claim derogatory treatment, for example:

> *if the whole essence of that artwork were, that shade of pink was that shade of pink, and … that was your signature.*

Otherwise, it is 'unlikely' that courts would agree that minor changes constituted 'derogatory treatment'. One of the legal experts interviewed also questioned how likely it was that anyone would actually complain about this type of change. However, one of the preservation experts interviewed thought that creators might well complain about changes to their works, even if the only alternative was for their work to become obsolete.

Even if rights holders would have no legal basis for complaint if their publications were changed as a result of preservation copying, the results of the author questionnaire and comments made by interviewees show that they are nonetheless concerned about the changes that may result from preservation copying. The author questionnaire asked respondents how important it is to them that preserved copies of their works remain identical to the original digital version. Almost all the respondents (11 out of 13) said that it was 'important' or 'very important', although significantly more respondents selected 'important' (7) than 'very important' (4). It would be interesting to know why the remaining two respondents felt that this is not important, and what types of changes they would be happy with.

The author questionnaire also sought to clarify whether there are particular types of changes that authors would object to more than others. The authors surveyed agreed that the content of their work must not be changed if their work is copied for preservation; all respondents gave 'content' a maximum rating of '5'. One author respondent commented on this:

> *It's the content that is important. Look and feel are less of an issue, but I would be scandalised if the content were to be altered.*

Another agreed, saying that he or she would allow content to be moved 'from word to PDF or whatever as long as the words appear in the same order'.

Interviewees also agreed that content is the most important aspect of a publication. One expressed concern that some of 'the naked content' might be lost during preservation copying, or that 'somewhere along the line, a vital word is going to change, and alter the meaning of someone's paper'. He described incidences in the past where governments with 'revisionist motives' had 'nullified' historical texts or 'subtly amended' them; he pointed out that such changes could potentially be made deliberately in the course of preservation activities, although this was unlikely to happen.

A legal expert felt sure that content would not change during preservation copying:

> *I don't see why any words should end up being changed, I really don't see why that should be a necessary or even likely consequence of an attempt to preserve something. You should have something that's trying to preserve all words, surely, otherwise what are you doing? You're not preserving it, are you? You're changing it.*

A digital preservation expert admitted that such changes might occur quite innocently, particularly if a large number of items were migrated simultaneously. Such changes could also happen if the person carrying out the preservation did not understand the item being preserved adequately.

Authors generally did not feel that the look and feel and functionality of their work were as important as the content. They gave functionality a mean rating of 3.9 (out of five) and look and feel a mean rating of 2.9. One respondent explained that functionality might only be important to those authors whose digital publications are text-based; one author commented:

> *were my illustrated print material to be digitised, the appearance would be of prime concern*

Perhaps surprisingly, a representative of a reproduction rights organisation suggested that artists might not expect copies of their works to be perfect, since they have long been aware that no print copy of their work is ever perfect. They would, however, still want copies to be 'as good as possible'. A few of the publishers interviewed also thought that the look and feel of their publications are not important. They believe that it is the content, not the 'format' or the 'way it's rendered' that preservation experts should be seeking to preserve:

> *What they should be doing is looking at the intellectual property in there … and preserving that in a format that may be relatively easily transferable from one retrieval engine to another.*

This is especially true for dynamic databases, where the content was said to be far more important than 'snapshots' of how that content was presented at one particular point in time to one particular user.

These concerns may cause problems for digital preservation experts since, as they explained, preservation activities always change digital objects in some way. As one interviewee described a Dutch project that investigated the changes that take place with migration:

> *Effectively, what you do is you dig up some old computers and some old software and you create something in it, and you just push it through all the versions up to the current day, and you discover that it's not quite what you started with.*

Similar problems may arise with emulation. Work has been carried out to emulate modern Windows programs on Macintosh computers, and here, too, 'you can see it's not perfect … even when you've got the other system sitting there running to check it'. Some types of digital publication were said to be particularly difficult to preserve. One example given was databases:

> *if you give us a database … any forms, front end, coding, anything like that … is not a significant property as far as we're concerned,*

> *because it's difficult to preserve and we can't guarantee to do it in a fashion which would replicate what we've been given. The data, the fields, the columns, would pretty much be preserved there … We certainly wouldn't go around conflating two fields into one … without good reason.*

CD-ROMs were also said to be difficult to preserve, since:

> *you could copy the files off the disc or the CD or whatever, but not get all the important information you might need*

Copying might not capture the structure of the files, any audio tracks that it contains, or features such as auto-start-up. Taking an image of the CD-ROM against which to compare any later versions can help this. Because of these difficulties, the preservation experts interviewed said that they do not currently make any guarantees about the digital objects they are preserving.

The preservation experts interviewed explained what they do to ensure that preservation copies are as similar to the original digital works as possible, and to avoid conflict with authors or creators. This should satisfy a publisher representative interviewed, who thought that rights holders would want to 'take an interest' in the preservation of their materials to ensure that they are being preserved in a 'specific way'. One agency represented makes an initial preservation copy on receiving a digital work, and shows this to the author to check whether it is acceptable. Even before taking action to preserve something they will try to determine:

> *the aspects of the digital resource that are important enough to need benchmarking, so that you can define the perceptible drift. You can say, 'it needs to be like this, and it can't change more than so much'.*

However, the preservation experts interviewed admitted that at the current stage of the development of preservation techniques, they couldn't always manage to preserve even these aspects of a publication. Difficulties may therefore still arise with authors and creators, particularly if they see the functionality and appearance of their work as

being an integral part of its intellectual content. As one interviewee explained, changes that seem minor to the preservation agency may be much more significant to the creator:

> *It's not just a technicality of delivery. Looks like [it] on the outside, but it's not. Just technicalities of, 'oh, we're going from one medium to the next'. It's, fundamental decisions are taken about what is actually worthy of being preserved. It's like taking a computer game and saying, 'oh, we're not going to keep the sound, we're just going to keep the pictures'.*

One interviewee suggested that some rights holders would prefer things not to be preserved at all if they are not preserved exactly. Asked what creators would think faced with a choice between an inferior copy and no working copy at all, a preservation expert commented that 'it won't stop people complaining'!

6.1.6 Authenticity and digital preservation

Both interviewees and seminar delegates expressed concern about the authenticity of digital objects that have undergone preservation copying. This may be an issue with migrated digital objects, as a digital preservation expert explained. A publication that has undergone multiple migrations, may look like the original, yet have an underlying bit stream that is completely different from the original:

> *Say it's a textual document, when you actually present it to the user, it might look similar. But in terms of the bit stream, the byte stream that represents it, it might be completely different because it's stored in a different format.*

This may make it difficult for preservation agencies to prove that their preserved copies are actually exact copies of the originals:

> *The actual data changes if you change the format. It's not the bits so there's no way of proving it's the object you started with, and you can't*

really say that someone didn't just quietly alter the vital [content] of something in the middle and move it along.

It is therefore is not clear what the legal status of migrated copies is:

does a migrated object, been migrated several times, is it really the original, therefore does it still count in law as the original?

This preservation expert felt that this is not at all clear, and is 'a long way away from being explored in a court of law'. This is one reason why preservation agencies will usually keep the original version of a digital object: so that they can prove that the migrated version is a copy of the original.

Even if a migrated object is legally counted as an original, rights holders may still have concerns about the authenticity of the migrated copies. Interviewees and seminar delegates therefore called for good version control so that everyone knows what the 'authentic' version of a publication is. This could be recorded in metadata.

There may also be authenticity issues with the preservation of e-prints. One interviewee argued that it is that e-prints are preserved, since the aim of e-print archives is to provide wide access to research output. He believes that this 'means not just immediate access, it means long-term access, open access'. Some in the e-print community apparently believe that only the original, published journal article needs to be preserved. However, one interviewee argued for preserving an e-print as well as the related published journal article. This would be sensible:

where the e-print is something other than the published paper. And that could be that it's got a particular dataset associated with it, which is not in the published paper, or it could be where the e-print itself goes into more detail or has more data.

This raises questions as to which is the 'authentic' version of the paper.

The author questionnaire asked respondents how they manage authenticity issues with works they self-publish. One question asked whether they preserve all versions of works that they self-publish then update. Of the ten respondents to whom this question applied, six said that they only keep the most recent version of their work, while only two said that they preserve all versions of their work. The remainder preserve 'certain' versions of their work. It would be interesting to know exactly what these authors meant by this, and the basis on which they decide which versions to keep.

Authors were also asked whether they notify their readers if they change URLs, internal or external links or the structure of their digital materials. This is significant, since it may affect how useful digital publications are to users. Only three of the ten respondents who answered this question currently do this. It would be interesting to know why they do or do not do this. Authors were then asked how often they tell users about such changes. Only two respondents answered this question. One of these answered 'every 12 months or less' and the other answered 'between 1 and 5 years'. It is not clear what types of changes are being notified to users in these cases. A further question asked how authors notify users of such changes. Two respondents indicated that they use email for this; one stated that such emails are sent to 'registered users'. Another respondent wrote: 'if they need me they will find me…' This either indicates that he is used to being contacted by users, or that he thinks that he does not have a responsibility to inform users of changes.

6.1.7 Access to preservation copies of materials

A major issue for many publishers is whether libraries should be allowed to give access to preservation copies of materials that they are preserving. One publisher representative felt that 'the issue of access is everything' and that this is a more important issue than copyright. He also suggested that preservation and access are separate issues as far as publishers are concerned, concluding that:

> *I would be an advocate of the most liberal and generous exception if I understood and was clear about, in the next fifty years, of who would get access and under what circumstances.*

However, it is actually impossible to separate copyright and access in the digital environment, since a copy of a publication has to be made to allow it to be viewed. Interviewees' comments about this issue support the view given by a legal expert that publishers would welcome 'dark archives'. However, not all the publishers interviewed would prohibit all access to their digital publications. Several interviewees suggested that access should be embargoed until 'something has ended its commercial shelf-life', the publisher goes out of business, or the material is no longer in copyright. Another possibility would be to restrict access by only allowing legal deposit libraries to give access via CD-ROMs on standalone machines, and not networking them at all. This would have the disadvantage for the publisher that they would have to deposit six copies, rather than just one. Seminar delegates thought that libraries might need to licence access to materials that they are preserving. One digital preservation expert suggested that:

> *maybe you could actually stick it all on the Web, and with an emulator, and say 'you can download this music, but only if you've got a copy of the original disc'.*

It seems unlikely that publishers would view this suggestion as either workable or secure. Lastly, some publishers, notably open access publishers, are happy for their material to be made widely available by libraries.

Not surprisingly, the views of librarians and digital preservation experts on this issue are very different from those of the publishers. They generally felt that digital publications should be held at a 'more public level', since, as one interviewee said, we could otherwise end up with:

> *too much data locked into proprietary formats for private institutions, private copies, that might start charging for it.*

Several interviewees questioned whether there was any point in preserving materials without providing access to them. One legal expert felt that it would be particularly difficult to acquire funding to preserve materials that would not be made available to

the public until they were out of copyright, because 'people providing funding for preservation would not understand this'. A digital preservation expert said that they would wonder how long they should preserve materials for, if 'you're never going to have access to it'.

Other interviewees, however, felt that libraries and digital preservation experts should preserve digital materials anyway. One preservation expert explained that this should be done so that the materials can be made available to the public once they are out of copyright. He explained that copyright law is a 'bargain': the law protects intellectual property for a certain amount of time, but copyright works have to become 'public knowledge' after this time. He argued that publishers who protest about their output being preserved should be reminded of their responsibility to ensure that this is freely available once it is out of copyright.

Another interviewee felt that it is a shame when access may not be given, since:

> *one of the great things, supposedly, about the digital material is that we can provide better access to it or more access to it.*

She hoped that legal deposit libraries would eventually be able to provide online access to the materials that they are preserving.

Some publishers are also concerned that there may be commercial implications if their publications are changed as a result of preservation copying. One publisher interviewed was concerned that those carrying out digital preservation might start 'moving something from one format to another and then … reselling the new format'. One publisher said that they would be particularly concerned if:

> *people were … doing stuff with our content that was somehow making it more valuable than what we had done.*

This may in fact happen, since preservation activities may actually add functionality to some digital materials, particularly those produced with formats that are already obsolete. For example:

> *on the modern computer, because the resolutions are so much better and you can put so much on the same screen, you can actually get both views on the same screen*

Similarly, a digital preservation expert suggested that since a lot of effort and 'creative activity' goes into preserving digital materials, those carrying out preservation could technically claim some rights over the new format they produce. However, they do not actually do this. These issues may be a cause for concern among publishers, although their main concerns about this are commercial, and it seems unlikely that preservation agencies would try to profit from their preservation activities.

A publisher representative concluded that it is vital that the commercial environment is taken into account when preservation is carried out, since:

> *otherwise, you're creating an area of tension which damages both the preservation function, because people with valuable stuff will try to evade it, or the commercial market.*

Digital preservation experts will need to take rights holders' concerns into account when preserving, as well as the law.

6.1.8 Rights holders' views of the rights implications of digital preservation activities

The interviews revealed a variety of publisher attitudes towards digital preservation. Some of them seemed unconcerned by it:

> *my guess is that … moving something from one format to another, for the sake of preservation, would be not something we would have a problem with.*

This publisher felt that the library community is right to be raising this issue with publishers. Another publisher thought that publishers would view this type of copying as 'fair dealing'. He explained why:

> *if somebody's print copy disintegrates, we don't expect them to therefore have to buy it again from us. … We do think that they've got that content and they would have every right … to copy it for their own use.*

Other interviewees were concerned about preservation copying. Such concerns generally stem from their concern to 'protect their own and their authors' intellectual property'. An author who responded to the questionnaire expressed concerns about the use that may be made of the preservation copies of his or her work:

> *I think every writer is happy to be an entry on the British Library catalogue ... and this would be simply a further development of this. The library's right does not affect the writer's rights. But you'd have to think through the consequences of a lending right, or a duplication right.*

Rights holders seemed particularly concerned about changes to their publications that might result from preservation copying. This will be discussed in Section 7.3.

6.2 Licensing issues and digital preservation

6.2.1 Licence agreements between libraries and publishers

6.2.1.1 Types of licence used

The questionnaire results showed that licensed material forms a significant part of the digital collections of a large number of libraries, and that a significant proportion of publishers licence digital materials to libraries. The terms of the licence agreements for digital publications may have a significant impact on their long-term preservation.

The questionnaires therefore asked librarians and publishers about the types of licences they have.

For librarians, individual licences with suppliers were the most frequent option; 87.7% of the 81 libraries with licences had these. Agreements based on model licences were also frequent, especially with academic libraries. Forty-six per cent of academic libraries responding to this question have agreements based on model licences, compared with 11.4% of public libraries, 27.3% of special libraries, 20% of national libraries and 18% of libraries in the "other" category.

	No. of Responses	**% of Responses**	**% of Respondents**
Individual licence	71	36.0	45.5
N/A	59	29.9	37.8
Model-based licence	47	23.9	30.1
Shrink wrap	16	8.1	10.3
Other	4	2.0	2.6
Total	**197**	**100**	**126.3**

Table 6.1 Licensing options in libraries

Of the four respondents who selected 'other', one uses Copyright Licensing Agency agreements, and another uses HERON. Presumably these will be for digitisation purposes or for access to digitised material. Two respondents said they use Combined Higher Education Software Team (CHEST) / Joint Information Systems Committee licences or model licences. Some respondents selected more than one option.

When asked whether they used model licences as a basis for their agreements with customers, 20 publishers responded. Seventeen indicated that they do. Three respondents said they use a JISC model licence and one uses the Liblicense model. Other respondents indicated awareness of these models, or had developed their own agreements after looking at these models. One respondent said they licence their material via Netlibrary. Another said that while they are aware of model licences and support them in principle, they prefer to negotiate their own terms rather than use a model licence.

	No. of Responses	% of Responses	% of Respondents
Other	13	48.1	50
Don't know	10	37.0	38.5
JISC	3	11.1	11.5
Liblicense	1	3.7	3.8
Total	**27**	**100**	**103.8**

Table 6.2 Use of model licences by publishers

Thirteen respondents selected "other" when asked what type of model licence they use. These included HERON and a license for journals that is derived from the John Cox, liblicence and to some extent from PA/JISC models. Five respondents use their own licence agreement.

An academic librarian interviewed explained that some licensing in the UK is negotiated on a national basis. This is partly since, unlike in the USA, UK academic libraries do not employ specialist staff to negotiate licences for individual libraries or consortia. The British Library does have staff dedicated to the licensing of digital content. It has its own licence that it offers to publishers for electronic journals, and has a licence for other materials, although it normally uses the publisher's licence for these. JISC plays a significant role in negotiating national licences, since it coordinates digital information provision in the Higher Education sector in the UK. JISC has been involved in developing the National Electronic Site Licence initiative (NESLi) model licence for journals.

However, while one librarian explained that his library tries to subscribe using the national deal, this is not always possible. He explained that some publishers do not accept the national deal, which leaves libraries no option but to accept what the publisher offers. This probably explains why the most frequent response from libraries was 'individual licences'. As a result, libraries have to deal with a wide variety of licences. John Sweeney, who is involved with licensing digital resources at the British Library, commented that this creates difficulties for libraries:

> *It doesn't really help … that so many licences are so different. … I think one would want to see some sort of cooperation to get that*

standardised, have model licences and so on. … There will, of course, always be publishers that want to do their own thing, but the more that could be doing something reasonably similar, the better

However, this would be difficult to achieve, partly because the law of the country affects licence terms or state in which the publisher is based. One interviewee commented that a lot of time is wasted negotiating with publishers whose licences omit important clauses.

Shrink-wrapped licences were said by one interviewee to be a particular problem. Publishers state that opening the wrapping implies acceptance of the licence, but the library cannot always tell whether a licence is suitable until they have opened the wrapping, by which time it is too late to reject the terms of the licence. It is not clear whether it is in fact a legal practice.

6.2.1.2 Preservation clauses in existing licences

An important issue for preservation is whether licence agreements permit libraries to copy their digitised and born digital works for preservation. One publisher interviewed thought that preservation copying would almost certainly contravene libraries' existing licence agreements. This view was confirmed by a librarian, who said that:

Many licences probably wouldn't allow you to make copies, I would doubt very much that it allowed you to do that. So you would have to have special provision in most licences to preserve stuff. And … I haven't checked many very recently, but I don't think any licences mention specifically about preservation.

Where works are digitised using schemes such as HERON, the rights libraries are given are determined by the terms and agreements that HERON has made with publishers. These may stem from either agreements such as the Higher Education Digitisation Agreement or from arrangements made between individual publishers and Higher Education Institutions and brokered by HERON. For example, libraries

are only granted permission to keep the digitised copy for one year at a time, meaning that they cannot take any action to preserve these for the long term. This need not be a problem as long as the print original is being preserved somewhere. One interviewee mentioned that HERON allows digitised materials to be backed up onto floppy disc, but not to CD-ROM, because the latter medium is seen as more permanent. This suggests that libraries are not expected to want to preserve digitised materials. An academic librarian explained that it is possible for libraries to preserve digitised copies of manuscript or special collection materials, since these are generally out of copyright. In such cases, the digitisation is planned so as to make future preservation easier, for example by using standard file formats.

Only 5.3% of 152 library respondents to a question on the inclusion of provision for preservation copying in digitisation agreements said that they included such provision, while 13.2% did not.

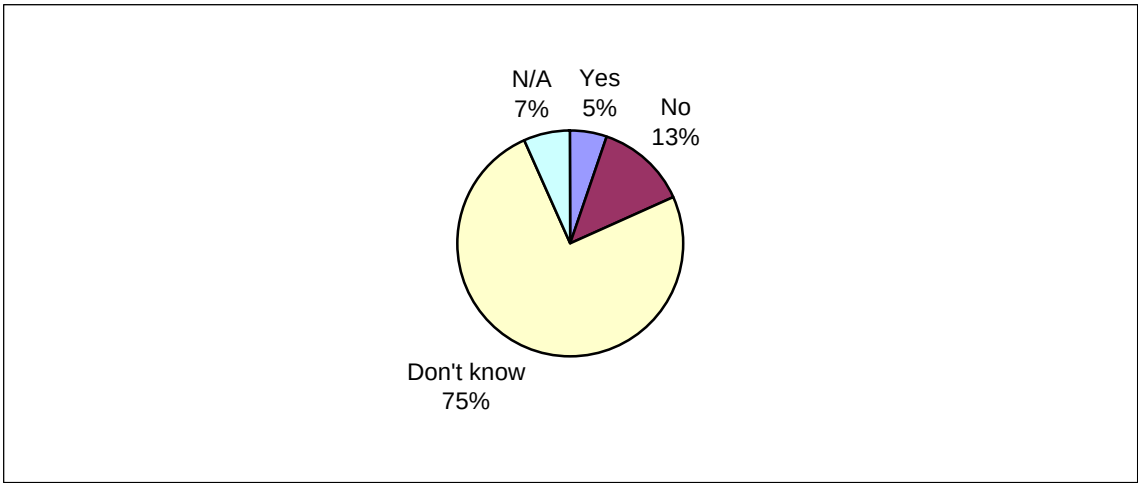

Figure 6.1 Copying for preservation included in digitisation agreements

Ten per cent of respondents said that licences for included this provision for born digital material.

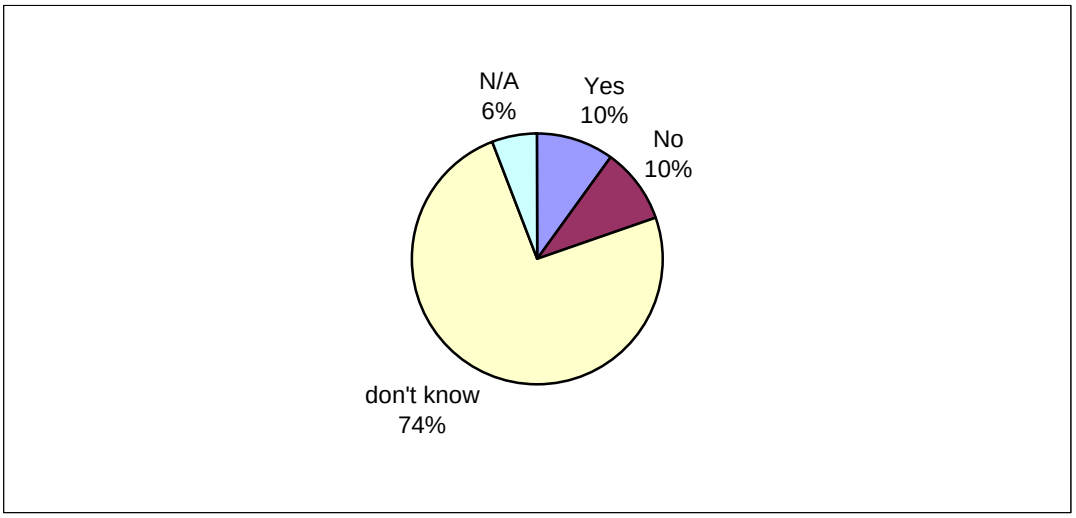

Figure 6.2 Copying for preservation included in agreements for preservation of born digital material

Unfortunately, three quarters of the 152 library respondents to a question on whether they include permission to copy for preservation purposes in digitisation agreements did not know the answer. This was the case for a similar proportion of respondents when asked if they sought permission to copy born digital material for preservation purposes.

6.2.1.3 Libraries' experience of clearing copyright

If libraries' licences do not allow them to make preservation copies of their digital materials, they may need to approach publishers directly to seek permission for this. The library questionnaire asked respondents time they currently spend each year clearing rights for digitisation or preservation, to get some idea of their current administrative load. Only 41 libraries responded to this question, but the response indicated they do not spend much time on this at the moment.

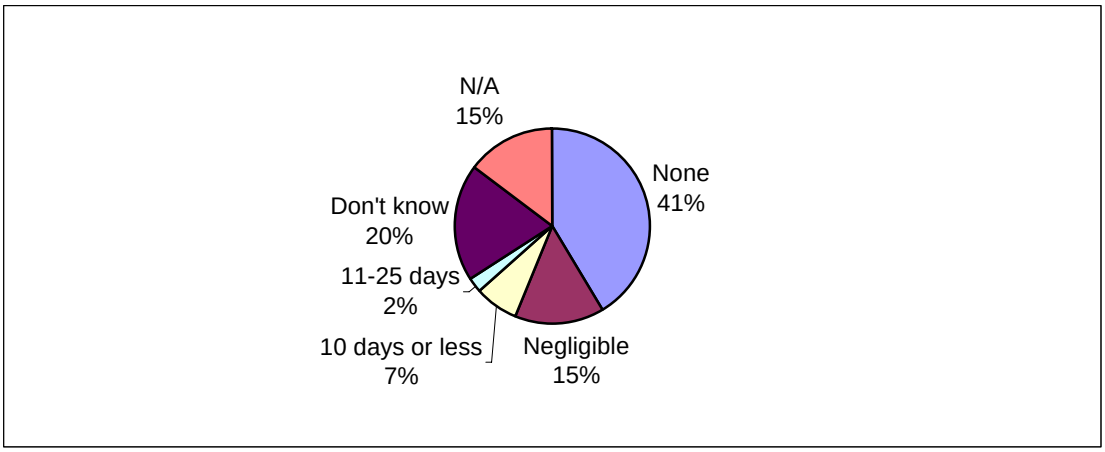

Figure 6.3 Time spent clearing copyright per annum

Deborah Woodyard of the British Library explained that clearing copyright to preserve each individual digital item would be extremely time-consuming for legal deposit libraries, because of the size of their digital collections. She explained that:

> *we don't have the luxury of doing that, we have so much stuff to deal with that it would just be too resource intensive.*

Some rights holders and rights holder representatives interviewed also spoke of their experience of granting licences for the use of their works. One of the publishers interviewed said that they generally look at such requests 'on a case-by-case basis'; this publisher might therefore not want to be involved in a blanket licence.

Interviewees gave various explanations as to why publishers are often slow to give copyright clearance. Two interviewees with an overview of the publishing industry explained that such requests would not be 'very high up the publishers' list of priorities' since they would not bring the publisher much money and were not central to their business. A representative from a reproduction rights organisation also suggested that rights holders may seem to be slow in granting permission because users only approach them at the last minute. This organisation is trying to encourage people to approach them during the planning stages of a publication, as it is difficult for them to come up with a suitable solution quickly, when reacting to a near-finished proposal:

> *there's nothing worse than a publisher coming along, or another consumer, any kind, a university, saying, 'OK, we're ready to go with this, this is what we want'. And we have to say, 'we can't say yes to that'.*

The rights holders interviewed also explained that they are still trying to decide how to deal with licensing in the digital environment. One problem is that factors traditionally taken into account in granting licences no longer apply in the same way as for print. For example, licences were traditionally issued for particular geographical territories, but it is now so easy to email materials around the world that these territories have become impossible to enforce. This may also mean that it is slow and difficult to develop a suitable licence for digital preservation.

6.2.1.4 Access or ownership of digital publications

A significant issue with the preservation of digital materials is whether and how libraries can copy and preserve material that they do not actually own. As a publisher who was interviewed explained:

> *in most cases … we are not delivering anything physically to the library, so there's a huge question about how you would intend to capture all of that anyway.*

If libraries are unable to preserve licensed materials themselves, they may be dependent on publishers undertaking to preserve their own materials. The findings reported in Chapter 5 suggest that by no means all publishers are doing this, or doing it effectively, which may mean that the long-term preservation of licensed materials is in doubt. A librarian who responded to the survey commented on this issue:

> *In all cases long-term access cannot be guaranteed despite assurances from the publisher.*

This suggests that libraries do not trust publishers to preserve their materials adequately. Nick Dempsey of EPS Ltd (one of our respondents who wanted to be identified) added that publishers would be unlikely to invest in preserving materials that no longer have any commercial value. Conversely, one of the publishers interviewed said that they see themselves as having 'an obligation' to preserve their materials so that libraries can access them in perpetuity.

6.2.1.5 Access to back files

Another important issue with licensed access to materials is whether libraries have access to back files – material that was made available before the start of the subscription – during their subscription period, and whether there is an additional charge for this. The library questionnaire asked respondents' experience of this.

Ninety-three libraries responded to this question. Almost 70% of the respondents have agreements that allow access to all material during the subscription period. Thirty-two per cent of respondents said they have deals allowing access to all material published during the subscription period as long as the subscription lasts. Presumably this means that for at least some of the respondents there are different levels of access depending on the deal. This was reflected in the "other" responses. Respondents who selected "other" said that the provision for access to back files in licence agreements depends on the type of material or publication and varies enormously between publishers and from licence to licence.

	No. responses	**% responses**	**% respondents**
Yes - during subscription period, to all material	65	36.9	69.9
Yes - during subscription period, to material published during the subscription period	30	17	32.3
Paid access only once subscription ends	10	5.7	10.8
No access once subscription ends	41	23.3	44.1
No access to back files during subscription period	21	11.9	22.6
Other	9	5.1	9.7
Total	**176**	**100**	**189.2**

Table 6.3 Provision in library licence agreements for access to back files

The respondents were not given the option of "free access to material available during subscription period once subscription ends" and no respondents gave this response under "other". Unfortunately, the questionnaire did not ask libraries to specify the proportion of agreements falling into each category. This is an issue that needs to be explored further.

Twenty-seven publishers answered a question on their provision of access to back files. The most frequent response was access to all material during the subscription period. Eight of the respondents provide no access at all to back files and five only provide access to these during the subscription period. No publisher not currently providing access had definite plans to do so in the next twelve months, four did not know and sixteen were definitely not going to provide access.

	No. of responses	**% responses**	**% respondents**
Yes during subscription period to all material	14	41.2	51.9
Yes during subscription period to material published during subscription period	2	5.9	7.4
No access once subscription ends	5	14.7	18.5
No access to back files	8	23.5	29.6
Other	5	14.7	18.5
Total	**34**	**100**	**125.9**

Table 6.4 Provision of access to back files by publishers

Among the "other" responses, one publisher said that access depends on whether the material is primary or secondary, but did not elaborate.

Comments made by interviewees explained why publishers might be reluctant to make their back files available to libraries as part of a subscription. Digitising old material and making it available is expensive, and publishers want to be sure that their investment will bring a satisfactory return. A publisher that includes full-text articles from 1994 in its subscriptions is not intending to digitise any earlier material since there is currently little demand for earlier articles to justify this expenditure.

Conversely, a small, digital-only publisher is intending to start charging for access to back copies of its online magazine, as a means of generating income over the long term, while keeping the most recent content free. This publisher will not have the costs of digitisation, but it will be interesting to see whether this unusual business model is successful. A further difficulty identified is that some publishers want to charge an annual fee for access to their back files, so that they can continue to gain income from them, while libraries believe that they should only have to make a one-off payment for them.

The library questionnaire asked respondents whether they are satisfied with the provision for access to back files that is made in their licences. Only 62 respondents commented on whether these guarantees are implemented satisfactorily, but the majority (83%) agreed that this was the case. It is not clear from the responses whether these libraries have had to test the publisher guarantees yet.

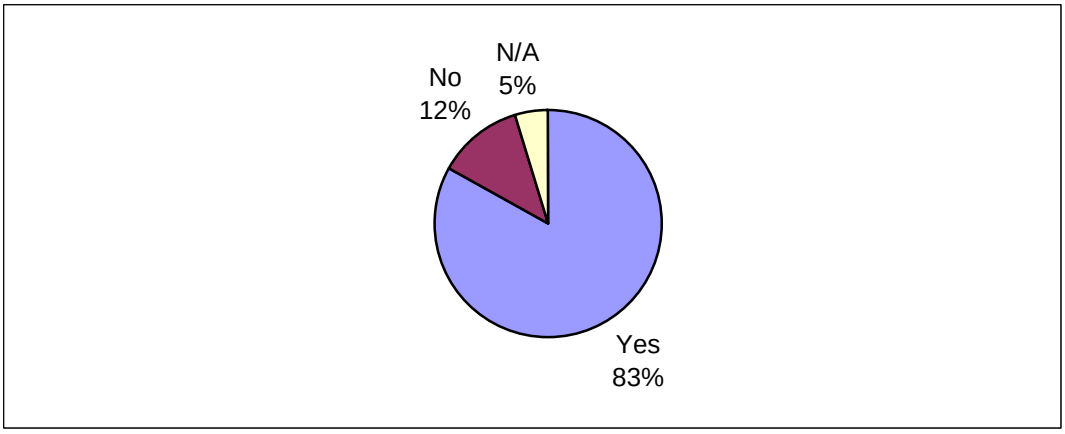

Figure 6.4 Library satisfaction with publishers' provision of access to back files

Table 6.5 shows the responses broken down by type of provision. The two categories with more dissatisfaction than satisfaction are, not surprisingly, "no access to back files" and "paid access only once subscription ends". On the other hand more than twice as many respondents were satisfied with the "no access once subscription ends" option than were dissatisfied.

	Satisfied	Not satisfied	Total
Yes - during subscription period, to all material	46	6	52
Yes - during subscription period, to material published during the subscription period	16	6	22
Paid access only once subscription ends	2	3	5
No access once subscription ends	18	7	25
No access to back files	2	6	8
Other	1	3	4
Total	**54**	**8**	**62**

Table 6.5 Provision of access to back files and satisfaction with that provision

The six libraries that said they were dissatisfied made further comments. Most of these were not directly related to the clauses about access; for example, more than one complained about problems accessing material even during a subscription. Two libraries commented that they had 'not had occasion to test provision yet'.

One librarian interviewed complained about the 'rolling wall' approach used by some publishers. With this model, subscriptions include access to a fixed number of years' worth of back files, so that after a certain amount of time, even material originally subscribed to is no longer available. This librarian complained that this model is 'alienating the library community':

> *I think once you've paid for it, you've paid for it, and you should gain access to it. … And there's no way that your subscription should drop off the bottom because we've moved on a year.*

6.2.1.6 Access in perpetuity

A further issue with licensed digital materials is whether libraries will have any access to materials if they cancel their subscriptions. If such access is allowed, libraries then want to know how long this access is guaranteed for, and how it will be provided. Interviewees were asked whether their licences contain clauses about access in perpetuity. The British Library's own electronic journals licence does include such a clause, although staff there still feel that they will need to discuss this issue with

publishers more thoroughly in the near future. Three of the publishers interviewed have access in perpetuity clauses in their licences, at least for some of their products. One of them said that this was deliberately intended to 'mirror … what happens in a print environment'. Some publishers who did not include such clauses in their licences explained that this was because they are secondary publishers, and their licences from the original rights holders would not permit them to license in perpetuity.

When asked how publishers ensure access to back files, 81 libraries responded.

	No. of Responses	% of Responses	% of Respondents
Publisher undertakes to provide remote access	52	49.5	64.2
Don't know	23	21.9	28.4
Publisher provides copies of material to the library	15	14.3	18.5
Publisher relies on a third party to provide remote access	14	13.3	17.3
Other	1	1	1.2
Total	**105**	**100**	**129.6**

Table 6.6 Libraries' experience of how publishers ensure access to back files

The most frequent response was that the publisher undertakes to provide remote access. Nearly 19% of respondents said that the publisher would provide copies to the library for access purposes. No respondents commented on whether they were then granted permission to preserve this material.

Twenty-two publishers answered a question on responsibility for provision for access to back files. Of these, more than half said that they would take responsibility themselves, and 13.6% said they would use a third party. The same percentage said they would provide physical copies of materials to libraries.

	No. of responses	% responses	% respondents
Will take responsibility for preservation	12	48	54.5
Don't know	7	28	31.8
Will use a third party for preservation	3	12	13.6
Will provide physical "copy" of material	3	12	13.6
Total	**25**	**100**	**113.6**

Table 6.7 Responsibility for the provision of back files

Interviewees described a number of ways in which publishers may provide libraries with access to materials to which they no longer subscribe. Some publishers give the files to libraries to host themselves. This may be achieved by giving the library the files on a physical medium such as a tape, a CD-ROM or a disc, or by sending them by FTP. Another method mentioned was for libraries to continue to access the materials from the publisher's Website, as they did during their subscription. One publisher stated that 'we would consider giving them a backlog of print'.

6.2.1.7 Access to materials if a publisher is taken over or ceases to publish

A final issue with licensed digital materials is whether publishers guarantee to maintain access to them if their company merges with another or ceases operations entirely. Only 19 publishers answered a question on this.

	No. of responses	% of responses
Don't know, haven't thought about it	6	31.6
Give over or sell to 3rd party	4	21.1
No access will be provided	3	15.8
Will still continue to provide access but don't state how	3	15.8
Depends on other company or circumstances	2	10.5
Not applicable	1	5.3
Total	**19**	**100.0**

Table 6.8 How publishers would guarantee access to back files if they merge or cease operations

The responses to this question were not encouraging from the library perspective. Most of the publishers who responded either had not thought about it, could not guarantee access, or said it was dependent on circumstances. It is not clear from the

responses whether any guarantees made would be honoured if material was sold or given over to another organisation. Only three respondents stated that access would continue, but they did not say how this would be achieved.

6.2.2 Licence agreements between rights holders

Both publisher and author questionnaire respondents have also had to seek permission from other rights holders to include copyright material in their publications. The agreements that they make with these rights holders may affect their ability to grant libraries permission to copy their works for preservation.

Some publishers, particularly secondary publishers, include in their materials text or images that have already been published by other publishers. They will need to licence this material from the original publisher, which may affect their ability to grant libraries permission to preserve their materials. Thirty-three publishers answered a question on the licensing of intellectual property from third parties. Of these, around half license text and images and just over 21% license software.

	No. of Responses	**% of Responses**	**% of Respondents**
Text	17	30.9	51.5
Images	16	29.1	48.5
Software	7	12.7	21.2
Multimedia	5	9.1	15.2
Don't know	5	9.1	15.2
Other	5	9.1	15.2
Total	**55**	**100**	**166.7**

Table 6.9 Licensing of third party material by publishers

As more than one publisher interviewee explained, if a publisher has to obtain a licence to use particular materials itself, this will be for a specific purpose only. If anyone else wishes to re-use or copy that material, or use it in a different way, they must seek permission from the original publisher. This is likely to make copyright clearance for preservation slow, as several rights holders may need to be traced and contacted.

One publisher interviewed thought that similar problems could arise where publications include photographs of clearly identifiable trademarked materials such as toys or sweets. This publisher has to obtain permission to use such photographs, and there are often strict conditions placed on their doing so, since 'a lot of companies are very, very touchy about the use of their trademark's name'. This interviewee said that they would not be able to give libraries permission to make copies of publications containing these, and felt that the companies in question would be unlikely to grant permission for preservation copying.

The author questionnaire investigated the extent to which authors seek permission to include materials owned by other rights holders in works that they produce. A significant majority of respondents (9) have had to do this. Authors were also asked what they usually use others' copyright material for. Six respondents normally request permission to include third party copyright material in one particular edition of a publication. No respondents normally ask permission to use material in more than one edition of a publication. This may be because publishers might be unwilling to grant permission for possible future uses of their materials. The author who selected 'other' explained that he or she asks for 'permission to include links to other on-line sites'.

Permission sought	**No. of responses**	**% responses**	**% respondents**
Permission to reproduce material for inclusion in one particular edition of a publication	6	54.5	66.7
Permission to reproduce material for inclusion in several editions of a publication	0	0	0
Permission to reproduce material in perpetuity	2	18.2	22.2
Permission to reproduce material in the context of the preservation of overall publication	2	18.2	22.2
Other	1	9.1	11.1
Total	**11**	**100**	**122.2**

Table 6.10 Types of permission usually requested

Two respondents indicated that they seek permission to reproduce material in the context of the preservation of an overall publication. It would be interesting to investigate this further, to find out exactly what they are doing with these publications. One respondent made a further comment:

> *So far, if the material is for a journal or a book, I have left the contract details to the publisher. But I am going to have to be clearer about that.*

This suggests that this author has in the past left negotiations about copyright to his or her publisher. It would be interesting to know why he or she believes that he needs to be 'clearer' about contract details in the future.

Publishers' agreements with authors and creators whose work they publish may also affect their ability to grant permission to preserve their publications. Forty-two publishers gave details of their rights agreements with content creators. Of these, 51% said authors assign all rights, with only three respondents dealing specifically with preservation rights. Presumably, these publishers would be in a position to grant libraries permission to copy for preservation if they wished. Comments from the six respondents who selected 'other' include that authors do not have rights in databases; that the authors are company staff, so presumably the organisation is the rights holder; and that there are different policies for different titles because the publisher also publishes on behalf of learned societies.

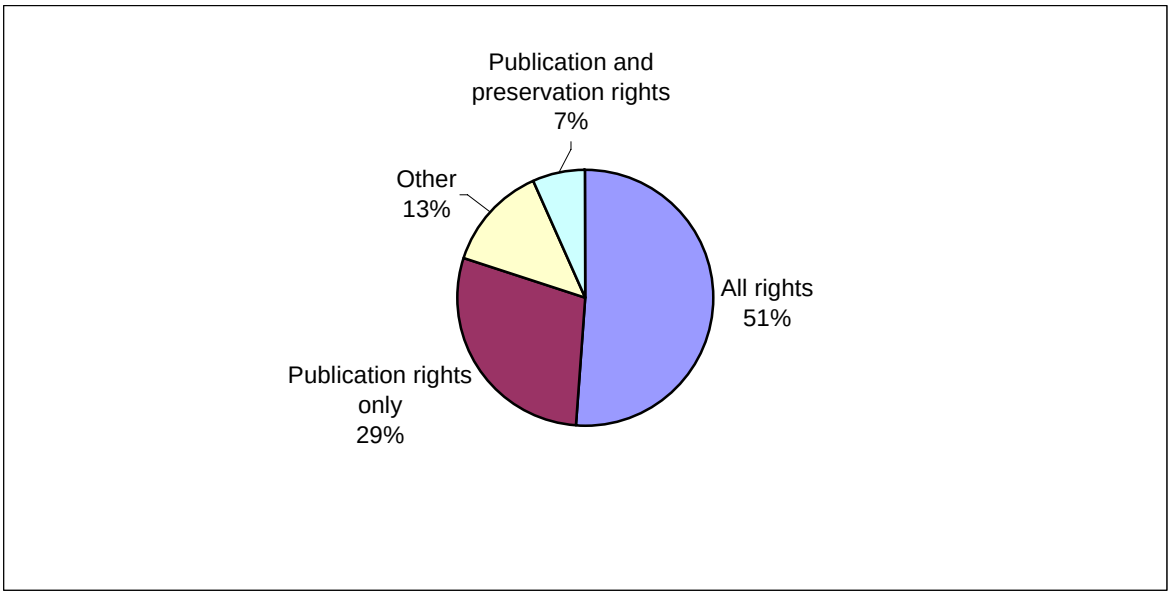

Figure 6.5 Assignment and licensing of rights by authors / creators

Authors were also asked about their rights agreements with publishers. The most frequent option selected was 'don't know' (five respondents). Two authors were not sure whether or not they offer preservation rights to their publishers; one of them said:

> *I am not sure of the policies of all the relevant journals. In some cases I have not been able to understand the wording of the copyright agreement with respect to digital archiving.*

These comments suggest that the authors surveyed are unaware of the issues, although it may be that the publishers in question are vague about their undertakings with respect to preservation. Slightly more authors have licences that include a clause that allows copying for preservation than do not (four respondents as compared to three).

Format	No. of responses	% responses	% respondents
Assign all rights	0	0	0
Assign rights for copying for preservation only	0	0	0
Assign some rights not including right to copy for preservation	0	0	0
Exclusive licence including a clause that allows copying for preservation	1	6.7	8.3
Exclusive licence without a clause that allows copying for preservation	0	0	0
Non-exclusive licence including a clause that allows copying for preservation	3	20	25
Non-exclusive licence without a clause that allows copying for preservation	3	20	25
Don't assign or license any rights at all	2	13.3	16.7
Don't know	5	33.3	41.7
Other	1	6.7	8.3
Total	**15**	**100**	**125**

Table 6.11 Preservation rights assigned or licensed by respondents when they license rights to their own work

None of the authors interviewed assign rights to publishers. This is interesting given that 59% of the publishers surveyed have rights assigned to them by authors. Six respondents have non-exclusive licences and only one has exclusive licences. Two respondents do not assign or licence any rights at all; it would be interesting to know why this is the case. Commenting on this question, one author admitted that he or she 'never think[s] to offer preservation rights'. The questionnaire then asked those respondents who have not yet licensed or assigned the right to copy for preservation purposes whether they would be willing to do so in the future. All the authors who answered this question said that they would be willing to do this. This is encouraging, but it is not clear whether they are intending to ask specifically about preservation rights, or whether they would just sign a licence that mentioned this if the publisher offered it.

One publisher interviewed insisted that it can only make full use of articles if it owns the copyright for them. By contrast, open access publishers argue that full

exploitation is best achieved by making copyright 'as irrelevant as possible' so that articles can be placed practically 'in the public domain'. Two authors expressed their views about this. One of them wrote at length about how copyright 'is becoming a major problem for all writers'. This author is not convinced by publishers' arguments that they can only 'safeguard further uses or displays of the material' if they have had copyright assigned to them. He or she concluded:

> *EVERYONE's [sic] lawyers is giving them the same advice - grab copyright and secure an interest in further development. But the writer can only develop that further development if he has held on to copyright...*

This is one reason why Open Access publishing may be advantageous in terms of long-term preservation. Author agreements do not present any problems, since authors publishing in Open Access journals license publishers and users to make any legitimate use they like of the material. Since the Open Access model means that articles are effectively 'commercially worthless' as soon as they have been published, neither publishers nor authors will object to additional copies being made for preservation, since this will not result in financial loss. Nor will they object to libraries giving access to preserved copies, since Open Access publishing aims to give as wide access as possible.

Interviewees also mentioned situations where poorly written licence agreements with authors have created difficulties for publishers. For example, problems have arisen where publishers have wanted to re-publish print works in digital form, but their licences with authors have not provided for this. In such cases, the publisher would have to seek permission from the author, which would be time-consuming. As one publisher explained:

> *very often you can't contact them. … academic authors tend to move, corporate authors probably even more so move. You know, you can't catch up with them.*

This suggests that where publishers have author agreements that mention specific publication formats, it could be difficult for them to give libraries permission to make copies for preservation. Interviewees suggested two possible solutions to this: either to ask for assignment of all rights, or to draft licences so that they will not exclude any future forms of technology or any changes that might need to be made to publication formats.

7 SOLUTIONS TO THE RIGHTS ISSUES OF DIGITAL PRESERVATION

7.1 Responsibility for digital preservation

7.1.1 Stakeholder views on who should take responsibility for digital preservation

The questionnaires asked respondents who they thought should take responsibility for digital preservation and why. Opinions varied widely. One hundred and sixty libraries responded to this question.

	No. of Responses	**% of Responses**	**% of Respondents**
Legal Deposit Libraries	86	36.0	53.8
Publishers	50	20.9	31.3
Libraries	35	14.6	21.9
Don't know	29	12.1	18.1
e-print archives	25	10.5	15.6
Other	12	5.0	7.5
Authors	2	0.8	1.3
Total	**239**	**100**	**149.4**

Table 7.1 Library views on responsibility for preserving digital publications

Legal deposit libraries was the most frequent option, selected by over 53.8% of respondents. The second most frequent option was publishers, which was selected by 31.3% of respondents. However, there was a spread of opinion and 38% of the respondents selected more than one option. The majority of these selected two or three options. The combination of publishers and legal deposit libraries was the most frequent (18 responses), then legal deposit libraries and e-print archives (6), then legal deposit libraries, libraries and e-print archives. Perhaps surprisingly, 28 libraries did not know who should be responsible for preservation.

Legal deposit libraries were also the top choice for publishers when asked about responsibility for long-term preservation. Of the 80 publishers responding to this question, 42.5% selected this option, compared with 53.8% of library respondents. Proportionally more publisher than library respondents thought publishers should

have a role with 38.8% of publishers compared to 31.3% of libraries. Only 10% of publishers thought that other libraries should have a role, compared with nearly 22% of library respondents. E-print archives were also a more frequent option with libraries than publishers, with 15.6% of libraries selecting this option compared to 6.3% of publishers. Sixty-six or nearly 81% of publishers selected a single option. Legal deposit libraries were marginally more frequent (24 respondents) than publishers (23 respondents). Of those that selected more than one option, the most frequent combination was legal deposit libraries and publishers, but this only accounted for four respondents.

	No. of Responses	**% of Responses**	**% of Respondents**
Legal Deposit	34	35.1	42.5
Publishers	31	32	38.8
Don't know	11	11.3	13.8
Libraries	8	8.2	10
E-print archives	5	5.2	6.3
Other	4	4.1	5
Authors	4	4.1	5
Total	**97**	**100**	**121.3**

Table 7.2 Publisher views on responsibility for long-term preservation

It is not really surprising that publishers are slightly less favourable towards libraries and e-print archives taking responsibility for preservation, because giving over such control may be perceived as risky.

Authors were also asked who they thought should be responsible for preserving digital materials. The most frequent responses were 'libraries (general)' and 'publishers', which were both selected by seven respondents. A similar number of respondents (6) thought that authors themselves should be responsible. This contrasts with the results of the library and publisher questionnaires. Interestingly, more respondents thought that libraries in general should be responsible for this (7) than legal deposit libraries in particular (5).

	No. of responses	% responses	% respondents
Libraries (general)	7	25.95	53.8
Publishers	7	25.95	53.8
Authors	6	22.2	46.2
Legal deposit libraries only	5	18.5	38.5
Don't know	1	3.7	7.7
Other	1	3.7	7.7
Total	**27**	**100**	**207.7**

Table 7.3 Author views on responsibility for preservation

One respondent selected authors, publishers and legal deposit libraries, and explained that 'at various points in the chain it is sensible to be future-proof'. Other respondents clearly agreed with this. Only five respondents selected just one option, and one of these did not know who should be responsible. Of the others, two selected 'legal deposit libraries only', and one selected 'libraries (general)'. The fifth thought that the Archaeology Data Service (ADS) should be responsible for preserving his or her works, since they 'are the only people with a track record in this for Archaeology'. The ADS will be discussed in more detail in Section 7.1.7. It is positive that this established digital preservation agency is so well regarded by archaeologists. Four respondents selected three responses, and one selected four (authors, publishers, legal deposit libraries and libraries in general).

Authors were also asked how they would prefer to submit self-published works for others to preserve them. Respondents expressed a preference for methods that do not involve them having to transfer their work to a physical storage medium and send this to a preservation agency; only one author selected this option.

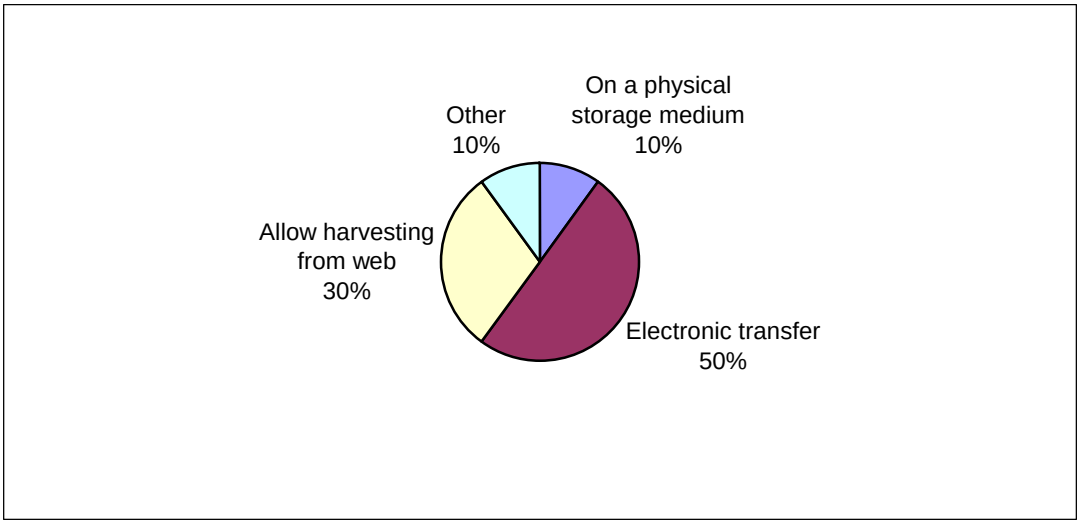

Figure 7.1 Preferred method of submitting self-published digital works for preservation

Only three respondents expressed a preference for their works to be harvested from the Web. This suggests that the majority of authors want to have control over submitting their works for preservation. The respondent who selected 'other' commented: 'I use both Web publication and electronic transfer'. It is not clear how Web publication could be a method of transferring materials for preservation.

7.1.2 Factors that will determine how preservation is organised

Interviewees suggested several ways in which preservation could be organised, and mentioned a number of factors that they felt would determine who eventually takes responsibility for digital preservation. As one of them said:

> *I … think it's a question which doesn't actually depend much on the issues of digital preservation, it depends on organisational matters, legal matters, funding arrangements.*

Some interviewees said that nobody currently seems to want to take responsibility for preservation. They suggested that responsibility might therefore fall to those who are seen as having the greatest responsibility for preservation, or those who have had the

greatest responsibility in the past. More than one interviewee felt that legislation would need to be passed to set out who should take responsibility.

A very important issue, mentioned by a number of interviewees, is funding. Seminar delegates suggested that this issue might ultimately determine who takes responsibility for digital preservation. Funding is essential if adequate digital preservation is to be undertaken. As one interviewee said:

> *Ultimately … if you have enough money, then clearly you can preserve anything, because you can simply continue to produce the hardware and software that read it. Ultimately, yes, there's no problem, but it's a resource issue.*

One librarian interviewed commented that some people apparently believe that digital preservation 'is a no-cost option, that it doesn't cost anything to hold things'. Awareness of the need for digital preservation may therefore need to be raised before funding can be obtained. Another problem is that cost models have still to be developed. Until these have been developed, it will be difficult for those intending to carry out preservation activities to bid for the money they need. Interviewees also felt that funding models will be difficult to develop: they will be complex because of the different aspects of preservation that will require funding:

> *we need to be aware of exactly what's involved in getting these things, preserving them, making them available to the public and so on.*

Interviewees felt that it is by no means certain that the necessary funding will be available. More than one interviewee expressed doubts about this; one said: 'I think that there's a funding hole here'. Interviewees generally felt that nobody wants to pay for digital preservation. As one of them said:

> *it's a resource issue … I would guess that's why publishers might tend to say libraries should do it, libraries might, I suspect, tend to say publishers should do it.*

The interviewees generally assumed in their comments about funding that it would be libraries who would carry out digital preservation. However, they were unsure how they would be funded to do this:

> *Whether that's a commercial arrangement, or whether it's a nationally funded, publicly funded arrangement, I don't know at this stage.*

Other comments made suggested that neither of these groups might actually be willing to pay for preservation. Nick Dempsey of EPS Ltd thought it was doubtful that public money would be made available for preservation:

> *I think it's going to be very hard to persuade governments of the merits of spending a lot of money on this.*

This is probably true, despite the fact that legislation has been passed allowing preservation in the context of legal deposit.

The publishers interviewed were clear that they do not want to have to pay for preservation. They are reluctant to do this since it will be expensive, and they do not expect that preserved materials will bring them much income. As one interviewee said:

> *I don't think there's much money to be had in this. There's money to be spent, but there's not much money to be made.*

However, seminar delegates thought that publishers should be required to contribute to the cost of preserving their publications, perhaps by paying a lump sum or a fixed percentage of the cost, or by charging lower subscription fees to preserving libraries. Nick Dempsey explained that publishers are particularly reluctant to pay for preserving their older materials:

> *There's a difference between material which has ended its economic shelf life and material which hasn't ended its economic shelf life. The material that really has ended its economic shelf life, I don't think*

publishers see their role in any way as archivists of that, simply because they are commercial entities, they're there to make money, they are not investing in something which is dead commercially.

Publishers will, however, 'invest in retaining something that we're going to get the return on'. One publisher explained why:

The company has to preserve its electronic content, because that's where much of the value of the business lies. And that electronic content has long-term value. So it's important to the whole future of the business that we preserve it well, because the use you make of it today may not be the use you want to make of it tomorrow

Publishers who see their material as having potential commercial value in the future may be reluctant to let anyone else take responsibility for preserving it.

Another suggestion made by seminar delegates is that funding for preservation could be raised by charging library users on a pay-per-view basis for access to preserved materials. This could be integrated into license agreements. The charges paid could be divided between the library (to pay for preservation activities), the publisher (as an incentive to deposit its materials for preservation) and the author. It would need to be clear who has to pay, since there could be legal difficulties if libraries charge those who are allowed to view a publication under an exception to copyright law. Good administrative procedures would be needed to enable libraries to operate this system successfully. If a pay-per-view system is adopted for access to preserved publications, some initial investment will still be needed to set up preservation activities.

A further potential issue is what happens to published materials if a publisher ceases to publish them, or a library ceases to subscribe to them. Funding models will need to be developed so that the preservation of these materials will be secure even if this happens.

A lack of funding could lead to the failure of any attempt by libraries to undertake digital preservation, which was a matter for concern with one of the publishers interviewed:

> *I am sceptical that the British Library is currently adequately funded if … the extension to legal deposit legislation goes through. Whether it's adequately funded to preserve everything that it would then find being delivered to it. I think the library might say it is, DCMS [the Department of Culture, Media and Sport] might say it is. I am sceptical, because I think the cost would be staggering.*

There may be a contrast here between what these organisations would say in public, and how much funding they actually have. This again points to the problems that may arise if the costing models are not clarified. One issue raised with respect to both public and private funding is whether libraries have the power to 'persuade' others to give them money, or to 'squeeze' money from them; this is another reason why awareness of digital preservation issues needs to be raised.

7.1.3 Centralisation of preservation

Several interviewees felt that preservation should be carried out centrally, rather than being undertaken by each library and publisher. Their reasons for this are explored here; suggestions about individual organisations that could do this will be discussed in Section 7.1.7.

One of the publishers interviewed suggested that it would be 'nonsense' or 'insane' for everyone to attempt to carry out digital preservation. It would be 'more efficient' if fewer organisations were involved as there would be 'economies of scale'. This could also lead to cost savings. Seminar delegates therefore suggested that consortia arrangements might be beneficial. A further barrier to everyone carrying out preservation is that few libraries would have sufficient storage capacity to preserve significant amounts of digital materials; Deborah Woodyard of the British Library explained that:

> *I think that's why a lot of institutions are now starting to work together to come up with these kinds of solutions, because it does take a very large place and a very large system to accommodate a lot of this material. Which … one institution can't really manage on their own, easily, unless they're enormous.*

Another significant reason for centralising preservation is that it is felt to be too complicated for everyone to undertake:

> *the level of complexity involved and the level of expertise required means that very often individual institutions can't cope with large-scale digital preservation and it should be carried out by supra-institutional agencies*

The technical expertise needed to carry out digital preservation was also said to be in short supply:

> *it's very scarce skills. We don't have the skills available yet to share them out for every institution to have their digital preservation technician, digital preservation officer. Yes, so my view is that it would be at this stage unrealistic to expect institutions to be doing all this kind of thing. So it's got to be done by other agencies on behalf of the institution.*

One interviewee therefore suggested that it would be best to concentrate the skills in a few institutions, 'rather than all institutions trying to compete, or trying to train these people up'. A suggestion made was that preservation could be undertaken by a different institution from the one providing access to the digital materials, or by a specialist company. Some e-print repositories were said to be considering this model. An alternative suggestion made was that 'tools and techniques' for preservation could be developed centrally and made available for individual institutions to use.

Several interviewees mentioned that each individual digital publication only needs to be preserved by one institution, as long as other institutions can access it. As one librarian said:

> *a lot of the material, published material at least, is not unique. If somebody preserves Web of Science or Science Direct or whatever, that's fine and that's the case for everybody.*

We therefore do not need to have copies being preserved 'all over the place' as long as 'it is happening somewhere in the chain'. This was not possible in the print environment, but is possible with digital publications:

> *Print material, you did need more people to have it up front, because only one person can ever look at the book at the same time, and this is true forever. … But as soon as you've got something that is non-print … in 200 years it can be networked so everyone can see it … From one copy originally, and very easily.*

However, it would not be wise to rely on preserving one copy only; a certain amount of redundancy is still needed to ensure that nothing is lost. This may become an issue as libraries increasingly move from a 'just in case' to a 'just in time' set up.

Finally, seminar delegates suggested that some international cooperation is needed for digital preservation. This is because libraries will continue to want access to materials that were published in other countries, and which they therefore have no right to preserve. One suggestion is that libraries access these materials from other national libraries under licence. Another is that there could be a designated centre for the preservation of international materials. However, they thought that progress needs to be made on a local and national level first, since developing international solutions may be very slow.

7.1.4 Legal deposit libraries and responsibility for digital preservation

As has been shown, many of the respondents to the publisher questionnaire thought legal deposit libraries should have a role in long-term preservation of digital material. The questionnaire asked publishers whether they currently participate in the voluntary deposit scheme currently in operation in the UK. Only 14 out of 76 publishers who responded to this question (18.4%) said that they were participating in the voluntary scheme. Of these, ten are providing material on physical storage media. Nearly 70% of the respondents to this question are not participating in the scheme. Some publishers who are not currently participating gave reasons for this.

Reasons	No. of Responses
Do not publish in digital format	4
Didn't know it existed	2
Deposit of CD titles, not online	1
Our publications all go to the British Library as hard copy	1
Support the guidelines but do not have specific digital materials that fall within the scope of digital deposit.	1
Only just started publishing digitally	1
Not participating for the time being	1
Will participate. Current projects source material from the British Library (and other libraries). As per contract, the British Library will get free access.	1
Total	**12**

Table 7.4 Reasons for not participating in UK voluntary deposit scheme

Two respondents said that they were not aware of the scheme, and one said they did not have any material that came within the remit of the scheme. The voluntary scheme will continue for some time until the implementation of the new Act is negotiated.

Several interviewees agreed with questionnaire respondents that legal deposit libraries are the best place for digital preservation to take place. This reflects the feeling that it would be sensible to centralise preservation (Section 5.4). The seminar delegates also agreed with this, although they felt it was unrealistic to expect legal deposit libraries to undertake preservation alone. As John Sweeney of the British Library said, the British Library has national and international standing and contacts:

> *They can act nationally … or even internationally, collaboration between major national libraries.*

Deborah Woodyard added that:

> *they have the responsibility for the print material, they have a lot of experience, and I believe they're trustworthy.*

Interviewees also thought that legal deposit libraries 'have the will to do it'.

It seems likely that legal deposit libraries will play a large part in the preservation of digital publications in the future, because of the Legal Deposit Libraries Act 2003, which will be discussed later (Section 7.3.2). This Act was passing through parliament at the time of the interviews, and interviewees were asked their views about the extension of legal deposit to non-print materials.

The publishers interviewed had mixed views about legal deposit. Nick Dempsey said that:

> *They are fairly apathetic about this, because they're publishers, and legal deposit is just a kind of chore. What they're excited about is selling things to people, they don't get excited about things like legal deposit.*

A digital-only publisher interviewed knew almost nothing about legal deposit; this is perhaps not surprising from a publisher that does not produce print publications, and therefore currently has no legal deposit obligation. In general, the publishers interviewed were in favour of legal deposit, since 'it ensures that there's some copy of something somewhere, that hopefully will always be available'. Several interviewees were aware of, or had been involved in the voluntary deposit scheme already in place; one has been involved in negotiations about this.

Several interviewees felt that legal deposit of non-print materials would be very difficult to achieve in practice. From the legal deposit libraries' point of view, they felt that libraries would have to decide 'what actually is worth archiving' and which formats should be deposited, as well as how 'British' publications should be defined. Interviewees also questioned whether the British Library would have adequate storage space.

Publishers interviewed mentioned several concerns they have about the feasibility of depositing digital publications. A database publisher questioned whether they would legally be allowed to deposit their publications, given that these contain other publishers' copyright material:

> *We have a licence from the copyright owner to deliver their content, usually in some sort of aggregated form with other content, to libraries. We don't necessarily have a right with that content to deposit it.*

It is not clear that they would actually have an obligation to deposit their publications. This publisher also questioned whether there would be problems with depositing third party software used in digital publications. Section 9 of the Legal Deposit Libraries Act 2003 (Great Britain 2003b) appears to say that the secondary publisher depositing its materials would not infringe contracts between primary and secondary publications. Several interviewees were concerned about the practicalities of depositing materials. For example, one asked how the British Library would deal with material to which access is controlled by password. Of particular concern was the question of how dynamic publications can be archived and preserved. Some databases change daily or monthly, and publishers questioned how anyone would decide when to archive a particular publication. They also questioned whether there is actually any value in preserving historical versions of these, since most publishers only see the current version of their dynamic publications as relevant, and do not preserve every version. While two publishers thought that some people might be interested in comparing different versions in the future, one of them felt that:

libraries would be massively overextending themselves if they attempted to … preserve every iteration of one of these databases.

One publisher added that it would also be a huge burden for publishers to have to deposit multiple versions of their publications. Another thought that there is a flaw in the whole concept of legal deposit for non-print materials:

You're trying to apply a set of guidelines on preservation that were established for a printed book, and very much a fixed object, and one that wasn't very frequently updated, to … a dynamic database that changes frequently.

This reveals a lack of understanding of the law; the law does not yet apply to this type of publication, and may never do so.

Similar issues were raised with respect to Web harvesting. Like databases, many Websites are updated frequently, and interviewees questioned how often snapshots should be taken, and whether each version needs to be preserved. One interviewee felt that to do this would be:

a horrible, horrible nightmare, which no-one will ever properly solve, and will cost a lot of money

A further problem with preserving Websites is that links would not remain active over the long term. As one publisher said:

Who's responsible for maintaining those links? How important is it to maintain those links? Is it integral to the article? I don't know… The resources involved would be vast.

Interviewees were asked whether they though that the British Library should be able to harvest material from the World Wide Web. Some publishers felt that it would be better for them to send their Websites to the British Library, rather than allowing them to harvest it directly. One of the publishers interviewed had been involved in

discussions with the British Library about the voluntary deposit of online journals. This publisher commented that these discussions had enabled each party to understand the other's concerns much better. Had the British Library merely harvested the journals from the publisher, this would not have happened. This has also resulted in a closer relationship than would have existed with the deposit of print materials.

Many publishers are also concerned about the commercial implications of depositing their digital publications. The cost of actually depositing them was generally felt to be insignificant: to send a file by email costs nothing, and the costs of producing and sending one extra CD-ROM are small. Nick Dempsey said that a survey carried out by EPS Ltd had concluded that this was not a significant concern for publishers, except perhaps for some very small publishers with small profit margins. However, one publisher felt that the costs involved in depositing databases would be substantial because of the software applications they include and, again, because they change so frequently. A greater concern for many publishers is the potential loss of sales if they have to deposit their publications. While some publishers see this merely as 'a mild irritant, because you'd much rather sell it', this is again a significant issue for database publishers. Since they may only sell a few copies of their large databases, including some to the existing legal deposit libraries, the loss of even six sales could have a significant impact on their profit margins. Conversely, one publisher felt that legal deposit could add value, since it might reduce pressure on publishers' servers, and provide a backup service if they went down. Another publisher thought that legal deposit could help to increase use of its free material.

One of the main concerns of the publishers interviewed is how much access legal deposit libraries will give to their publications. They are also concerned about whether libraries would give access to preserved copies of their publications as well as or instead of originals. They do not appear to have investigated this. Publishers interviewed feared that if libraries allow users to access their materials freely, this might prejudice sales. Some therefore wanted to control this access by licence. One publisher, whose commercial model depends on selling access to its archive, felt that allowing libraries to give access to this material 'would undermine the revenue model for online publishing completely'. Another felt that there would be no point in the

company digitising and licensing access to older material if legal deposit libraries were also providing access to it. Some publishers were also concerned that people who accessed their materials in legal deposit libraries could make and distribute infringing copies of them. Two interviewees with an overview of the publishing industry did not think that this would be a major concern. However, one publisher felt that there would be a difference between print and digital publications:

> *people can go to the stacks and pull out a deposit copy and have a look at it if they really want, so there's a sort of parallel with looking at a digital copy. But the physical copy then goes back on the shelf, I'm always just a bit concerned about what happens to the digital copy.*

This is not strictly true, since access to legal deposit libraries is very limited, and materials generally have to be fetched from closed stacks. One interviewee reported hearing the view that while legal deposit libraries are only supposed to provide access as a last resort, the British Library's central London location makes it too 'convenient' for people to access its stock.

Publishers interviewed were generally unhappy about libraries being able to provide access to their materials since they fear that libraries may try to 'rival' their services or 'compete' with them. These fears stem in part from existing concerns about the British Library's Document Supply service; publishers feel that libraries are using their copyright exceptions to 'get around budgetary constraints', and anecdotal evidence suggests that libraries are cancelling journal subscriptions since they can obtain the articles they need more cheaply from the British Library. Publishers see these 'services' as 'effectively sub-publishing' or 'republishing' their materials, and they do not want their digital publications to be used in this way. They feel that the British Library's role has become confused and they want to know 'what is the purpose of the preservation that the library is doing'; they seemed unwilling to believe that the British Library does not use legal deposit material for the purposes of Document Delivery. As one publisher explained:

> *if [legal deposit] were about preservation, and very limited access during the period of copyright, commercial publishers would have far fewer concerns.*

While they seem to accept that legal deposit is intended to preserve the national published output, publishers clearly fear that libraries will also use their materials for other purposes. Publishers want access levels and permitted uses for legal deposit items to be defined clearly, and interviewees felt that these should be defined either by legislation or by a code of conduct for legal deposit. Interestingly, while publishers were concerned about the British Library's activities, they did not mention the commercial services that are currently undercutting the British Library's services.

7.1.5 Other types of libraries and responsibility for digital preservation

Interviewees agreed with questionnaire respondents that libraries should be involved in digital preservation. They gave a number of reasons for this, many of which are similar to the reasons given for legal deposit libraries. One library questionnaire respondent thought that libraries might need to take responsibility for 'local history / data'. Seminar delegates thought that libraries should be involved in preserving any rare or unique materials or special collections that they hold Librarians are considered to be much more excited about preservation than publishers, and 'in a better place' to do it. As Nick Dempsey said:

> *It's the thing they're most evangelical about. You see a strange light in a librarian's eyes when they talk about the good they're doing for the country in keeping things.*

This contrasts strongly with publishers, of whom he commented that 'it's not what they do'.

Both librarians and publishers commented that libraries have traditionally been much more involved in preservation than publishers:

The issue of preservation has traditionally been farmed out to the libraries in many ways, [by] traditional publishers. And it was always considered a fine thing for them to be doing.

There are a number of reasons why this is the case. Several interviewees commented that libraries do not trust publishers to preserve their own publications. For example, a librarian related that when Elsevier promised to guarantee long-term access to its publications, the library community said that they did not trust them to do this, and requested that they deposit their material with third parties instead. As one interviewee said: 'publishers have not been very good at preserving their own material in the print environment'; one of the publishers interviewed admitted that his company has not got an archive even for its print material. Given that this has been the case for print, interviewees felt that they are even less likely to undergo the additional work and expense of preserving digital publications. A further concern of librarians is that publishers may go out of business or be taken over. This could put long-term access to their publications at risk. As one interviewee explained:

If one company says 'we will guarantee to preserve this stuff', but then they're taken over by another company that doesn't have the same policies, what happens then?

This will be discussed further in Section 9.8. Interviewees therefore expected that libraries would end up preserving digital publications, as they have preserved print publications.

7.1.6 Publishers and responsibility for digital preservation

The results of the publisher questionnaire indicate that some publishers are undertaking some short- or long-term preservation activities, which suggests that they see themselves has having some responsibility for preservation (see section 6.3.3). Some of the publishers interviewed agreed that preservation should be their own responsibility. This was mainly the case with new, digital-only publishers, who may have little to do with libraries. As one such publisher commented:

my beliefs are very firmly that, particularly as we're working with new technology in a new domain, that individual organisations are responsible for their own archiving.

He believes that publishers undertake their own preservation because they 'understand what's to be done'. This publisher also feels that it would be foolish not to carry out preservation activities:

We've invested a considerable amount of money in the products, and so to have a technology come along that renders them unusable is … silly.

Interestingly, this publisher is not currently relying on good storage and multiple backups to preserve its publications.

However, most of the publishers interviewed did not see preservation as being their own responsibility, and Nick Dempsey confirmed that 'the will isn't there' with publishers. This reflects the views of librarians reported in the previous section (5.5). Interviewees made several suggestions as to what would be needed to motivate publishers to preserve their own materials. One librarian suggested that publishers would be motivated to do this if their users demanded it, while Nick Dempsey felt that:

They probably won't do it off their own bat, they'll probably have to be kicked hard by someone to do it.

Other interviewees felt that legislation would be needed to motivate publishers to take responsibility for preservation. More than one also thought that publishers would follow what the industry leaders decided to do in this area; as one of them commented:

I think to some extent we would go with the flow there, we would think that bigger publishers than us would dictate what becomes the norm.

This view reflects publishers' concern not to damage their business in any way. Larger publishers were thought to be likely to lead the way because:

> *many publishers apart from those in the STM area are still struggling with the issues of how to publish in that area, let alone preserve.*

Smaller publishers therefore tend to be:

> *… focusing forwards rather than worrying about what happens looking back, preserving things.*

They are unlikely to have given much thought to digital preservation.

7.1.7 Other organisations that may be involved in digital preservation

Project participants mentioned a number of bodies and organisations that they felt could or should play a role in digital preservation. In keeping with the views already expressed about how it would be sensible for preservation to be centralised, these are all regional, national or international groups.

Two of the publishers interviewed were interested in making this type of arrangement with the British Library. However, one complained that this was not currently possible:

> *Actually we've been a bit disappointed with the British Library. Because we would have liked to deposit stuff much earlier, but they seem to be less interested in it. Or maybe that's the wrong word, but they seem to make less progress with it than, for instance, the Dutch National Library. The Dutch Royal Library jumps on it whenever given the chance. And the British Library we've offered it to, but they've never come back to us.*

It is not clear why this is the case; unfortunately, this question was not put to the British Library interviewees.

The Digital Preservation Coalition was felt by seminar delegates to have an important role to play in coordinating preservation in the UK. The Digital Preservation Coalition (2002) was founded in 2001 to promote joint action towards preserving digital resources. It is a consortium of UK organisations including legal deposit libraries, publisher and library organisations, the Arts and Humanities Data Service (AHDS), JISC, OCLC and Resource. It intends to work collaboratively with industry and research institutions and other bodies, both nationally and internationally. The DPC aims to raise awareness of the issues of digital preservation, including the need for funding, to provide 'a common forum for the development and co-ordination of digital preservation strategies', and to disseminate information about these.

Two interviewees mentioned the forthcoming Digital Curation Centre (DCC), which will be funded by JISC. One of them was aware that the DCC is intended to take the lead in digital preservation in the UK; the other was unaware of its actual role, thinking that it would be a museum of old computers to read obsolete formats. In fact, the DCC's stated aims include the following:

- *Establish a vibrant research programme into the wider issues of data curation*
- *Become an international centre for developing tools and techniques for long term, secure data curation*
- *Develop a reliable, sustained repository of generic tools, software, and documentation, to support curation, preservation and use of digital resources*
- *Pilot development of services for recording and monitoring file formats and preservation planning tools utilising these services*
- *Provide advisory services on curation 'best practice' and to be pro-active in raising awareness of curation issues.* (JISC 2003)

The Digital Curation Centre will not itself be a repository for digital materials, and is intended to work together with other organisations.

Three interviewees suggested that JISC itself should be involved in digital preservation. One of these thought that it should work with other groups such as the Higher Education Computing Group. However, another commented that to take complete responsibility for digital preservation, and that it would be unfair for it to be expected to do so. It seems unlikely that JISC could take complete responsibility for digital preservation, since its responsibility is for digital information within the UK Further and Higher Education sectors only.

Other suggestions made included the UK Data Centres and data archives such as the AHDS, which are both funded by JISC. As one interviewee said:

> *In Britain, in the academic sector, anyway, there's … already a strong, established principle that national services will provide preservation on a disciplinary basis.*

These services, while they are already heavily involved in preservation, only exist in some subject areas and are mainly responsible for primary research output such as datasets, not for published materials.

Interviewees also mentioned several regional, national and international organisations in the library sector. For example, Resource, the Council for Museums Archives and Libraries, could have an advisory role. The regional Museums, Libraries and Archives Councils could also be involved; these councils should meet one interviewee's concern that any such body should also serve other sectors. Other suggestions made were OCLC, JSTOR, the Public Record Office in the UK, and libraries with national importance, such as the National Library of Medicine in the USA.

One interviewee suggested that creators of digital materials ought to take some responsibility for the long-term preservation of their works:

> *It would help if they took responsibility for … keeping it and ensuring that it gets to some place that can look after it.*

In fact, six of the 13 authors surveyed agreed with this. However, a library questionnaire respondent suggested that 'individual authors simply do not have the resources to preserve digital materials'.

7.1.8 Need for discussion

A further pressing need identified is for the different stakeholders, particularly libraries and publishers, to work together to overcome their existing conflict and mistrust. This needs to happen soon if we are to ensure that digital publications are preserved for the long-term and to avoid losing significant amounts of data. A librarian interviewed said that said that libraries need to 'strike up relationships with the relevant agencies who are doing the digital preservation', and that they need to do this 'fairly soon' to avoid 'a black hole as far as the 90s are concerned'. His view was that it is essential to build relationships between the different stakeholders as a precursor to decisions about how to manage digital preservation. As another interviewee commented, however, we are still in the 'very early days' of doing this.

Librarians and publishers interviewed explained that the traditional roles of libraries and publishers have changed in the digital environment, and that this is currently causing conflict. As one interviewee explained:

> *I think it's part of the struggle that … all the people who add value between the writer and the end-user have. Everybody's struggling to redefine their roles, and everybody's sort of cannibalising the other roles.*

Another interviewee commented that these roles will continue to develop, and that we are still in the 'transition phase' where there is a 'hybrid library mix'. These changes are happening as a direct result of the digital environment, which an interviewee from a publishing organisation described as a 'dream come true' for libraries, but a 'nightmare' for publishers. This environment enables both libraries and publishers to make publications more widely available than before, and some publishers view libraries' dissemination activities as a threat, feeling that these activities may 'change the environment for publishing'. Some also feel that they no longer need libraries to

help with disseminating their publications: a publisher representative suggested that libraries in the future would just help people find information, and not provide or preserve it themselves:

> *I would find that a perfectly legitimate role for librarians, who have to say, 'we don't preserve anything, we've got a library which is nothing but a series of … servers and meeting rooms and computer terminals, and what we do is help people get to the information which is suitable for them in the context that they need it'.*

Given that library budgets are under increasing pressure, some publishers feel that libraries are using such activities to try to 'find extra-budgetary ways of accessing' the materials they need. For example, they feel that libraries are using the British Library Document Supply Centre (BLDSC) instead of subscribing to journals themselves. This represents a misunderstanding of the role of the BLDSC, but is one cause of the tension between the two groups.

In addition, while it is traditionally libraries that have provided access to publications that were no longer in print, it is now becoming cheaper and simpler for publishers to keep publications in print using short-run digital printing. One publisher was anxious that libraries should not be able to provide access to digital publications that publishers have ceased to support: he described this as 'republishing'. A related issue is that publishers are now starting to sell digital access to their older materials, where they would in the past just have sold current, print materials. As a publisher representative explained:

> *the value of [the publisher's] own archive has increased, and that has become more part of the particular product that publishers provide themselves.*

Again, this causes conflict with libraries, since in the past, users would have approached libraries not publishers to access these. Some interviewees were positive about the prospect of working together on this issue; one said, 'I see this not as an area of conflict. I think this is an area of cooperation'. A representative of a rights

holder organisation welcomed the prospect of discussion on these issues, since their experience has shown that people are generally more understanding of its viewpoints when it has the opportunity to explain them properly. However, not all the interviewees were so positive and optimistic about libraries' and publishers' ability to work together. There is currently felt to be 'stereotypical prejudice about what each party would like to do' and an 'us and them attitude'. One interviewee, recognising that librarians and publishers are not very good at understanding each other's points of view, suggested that discussions might be more successful at the representative body level, since representatives tend to be more reasonable than individuals. She also suggested that the government might be able, or might be needed to help mediate between the different parties. Discussion with other stakeholders such as authors and creators will also need to take place.

More than one interviewee expressed frustration about the slow progress of work towards a solution to the digital preservation problem. One librarian complained that digital preservation activities have yet to get past the 'pilot project' stage to large-scale activity, and that progress has been very slow to date:

> *I think we're getting there, but it's going more slowly than I would like to have thought, say two years ago. I would have liked to have thought we would have had a Digital Curation Centre by now, and that institutions and projects would have been set up to do some real, practical preservation work. Whereas at the moment, projects are still writing feasibility studies and that kind of thing. Which is helpful and useful to get people thinking about issues and to clarify the issues, but I think we need some more practical stuff going on.*

Another interviewee attributed this to a lack of organisation and coordination, adding that the consequences of this could be serious:

> *when the great preservation crash comes in twenty or thirty years time, everyone will look back and say, "oh, we could have got it, but we just didn't organise things.*

Another comment made was that: 'if you get the managerial stuff in place, then half the problems are suppressed'. Technical aspects such as 'metadata', 'formats' and 'workflows' were also said to need further work.

One interviewee explained that the negative views of preservation within the e-print community might result from this slowness to act. He said that some who advocate the preservation of e-prints are still talking about how to do this and 'haven't done anything at the moment'. Preservation is thus viewed as a 'distraction'. This interviewee therefore advocated a 'middle way', which involves investigating and piloting preservation activities 'in the background, in parallel' while continuing to raise awareness of e-prints.

Interviewees made several general points relating to the need to talk and work together on digital preservation. One of the librarians interviewed asked that publishers should:

> *Make public statements about what they're doing and put those in their licences, and … come to some sort of agreement.*

However, another librarian added that we need to ensure that 'publishers' interests are preserved'. Another recommendation is that we need 'a general recognition of the repository role of the libraries in this new environment'. A comment from one interviewee summarised many of these concerns: 'it feels like a very big task'.

7.2 Stakeholder attitudes towards the rights issues of digital preservation

7.2.1 Stakeholder concerns about copying of digital materials

Interviewees felt that copyright law is not yet properly adapted to the digital environment. Several of them commented that UK copyright law was clearly written for the print environment and does not adequately reflect 'what people both want to do and can do in the new technical environment'. The interviewees felt that 'everybody' needs to understand the 'differences with digital material compared to

the print world' better than at present, and also to acknowledge that 'it is not possible to map digital to paper clearly'. An underlying issue with this is that the whole concept of 'copying' is different in the digital environment. One interviewee suggested that it is 'crazy' to speak of 'copying' in the digital environment, since computers have to make a temporary copy of a Website to allow it to be viewed; the law does acknowledge this particular situation, and allow this to happen.

These comments reflect another issue that is of great concern to the rights holders and publishers interviewed: the ease of copying in the digital environment. New technologies now allow 'added convenience' for copying and 'unlimited exploitation' of published materials. File sharing was mentioned as a particular concern in this respect:

> *Given that we're in the world of Napster … publishers are obviously a bit paranoid about allowing anyone to get one copy available, because they can take that copy, sign up to Gnutella or something like that, and suddenly … five million people can access that.*

It is also simple to alter publications, which can mean that:

> *an image could be manipulated, its appearance changed, modified, cropped, details taken out, even just things being inversed or a red being not quite red or the scale being wrong. All of those sorts of things that the digital environment allows you to do at the touch of a button.*

The interviewees gave examples that suggest that their concerns are, at least to an extent, justified. A representative of a rights holder organisation said that:

> *every day there is a story about one of our members works is pirated, or they appeared here and they hadn't agreed.*

However, this has not been as widespread as had been feared.

This is an issue for publishers because they want and need to control the use that is made of their own intellectual property, and that of the creators and other publishers whose material they publish. In the digital environment, it is impossible to do this. As a legal expert explained:

> *Publishers, rights holders in general, often are most concerned about losing contact with what is happening rather than necessarily what is happening. They don't want things to happen beyond a certain amount, but what they also don't like is to not know that things might be happening.*

Rights holders are unhappy about this, having been used to greater levels of protection for their print publications. As one interviewee explained, they:

> *can't transport the print world onto the electronic world and maintain the same kind of protection levels that they enjoyed.*

A further problem is that some publishers also do not trust libraries to control the use of their digital materials. One commented:

> *How could we guarantee that the library was not allowing that work to be used in a way that they might in all innocence think that they had the right to do?*

One interviewee did suggest that rights holders are now becoming 'less concerned' and 'more comfortable' about the digital environment and that there is now 'less head-in-the-sand hysteria'. They suggested that this has mainly happened because they have seen that digitisation has not been 'the end of the world as people predicted a couple of years ago'. They are also now more used to licensing for digital publications. However, it is not clear that this view is representative of all stakeholders.

7.2.2 Stakeholder views on what could be done to help libraries undertake preservation copying

The library respondents were asked their opinions on what could help libraries to preserve digital resources and were given several options. One hundred and sixty respondents replied.

	No. of Responses	**% of Responses**	**% of Respondents**
Extension of legal deposit	108	35.6	67.5
Change in library privileges	56	18.5	35.0
Collective licensing	52	17.2	32.5
Provision of digital preservation rights metadata by publishers	39	12.9	24.4
Don't know	32	10.6	20.0
Individual licenses	10	3.3	6.3
Other	6	2.0	3.8
Total	**303**	**100**	**189.4**

Table 7.5 What would help libraries to preserve digital resources

Some respondents selected more than one option. The most frequent by far was the extension of legal deposit to cover digital materials. Among the "other" responses was less bureaucracy over copyright and licensing.

The comments made by librarians about these issues largely relate to non-legal issues. One asked for:

> *a source of good information for libraries intending to do this [digital preservation]- legal, technical, networking. I find the current resources either provider led, too disparate or too non-specific.*

Seminar delegates agreed that the organisational, managerial and administrative processes for digital preservation need to be standardised and made as simple and transparent as possible, so that everyone knows what to do. Another respondent thought that more information is needed:

> *There does not seem to be a central source of information to help libraries which wish to collect digital resources. Such a source could include publishers, Websites, copyright advice, technical advice. There is a definite need to extend legal deposit to digital resources since so much valuable information is nowadays only published on the internet or on CD-Rom. The British Library could archive whole sections of the World Wide Web, since so many sites exist for short periods and are then removed by their creators. There is a need for a UK-wide policy on digital preservation, linked to the People's Network.*

A comment from a publisher revealed a similar view:

> *Digital publication long-term storage is a constant concern. The sooner a coherent set of national guidelines are established, digital publication will remain a short term answer to long term needs.*

These comments reflect the need for raising awareness of digital preservation.

7.2.3 Stakeholder views on the type of preservation copying that libraries should be allowed to undertake

When asked what preservation actions they would allow libraries to undertake, 26 publishers responded. Making backups was the most frequent response, but a small number of publishers said they would allow refreshing and migration of material. No respondents selected emulation as a response. However, some respondents either would not allow any preservation copying or were unsure.

	No. of responses	% of responses	% of respondents
Backup	36	40.4	47.4
Don't know	18	20.2	23.7
None	14	15.7	18.4
Refreshing	10	11.2	13.2
Migration	7	7.9	9.2
Other	4	4.5	5.3
Emulation	0	0	0
Total	**89**	**100**	**117.1**

Table 7.6 Preservation actions that publishers would allow libraries to undertake

One respondent said that they would allow preservation copying of CD-ROMs but not online material, but did not explain why. In addition, another respondent said that they would 'very occasionally' allow copying 'on a case by case basis if the title is unavailable' or if they could be 'assured that the material will be protected and/or be used for a limited time'. Again, this reflects concerns about rights issues.

Authors were also asked whether they would permit libraries to copy their digital material for preservation purposes, and which preservation methods they would permit them to use. No respondent said that they would not allow any such copying, although two respondents commented that they would only permit this under licence. Another respondent said that:

> *I leave that to the E-journal to handle on my behalf. They (Internet Archaeology) use a trusted repository … and I would permit them to undertake whatever steps they thought necessary.*

Roughly similar numbers of respondents said that they would permit each of the methods mentioned (between six and eight respondents for each). Five respondents said that they would be happy with all three methods, but the rest selected one response only. Given that two of the respondents who were happy with all three options used the comments box to express this rather than selecting more than one option, it may be that the way the question was phrased confused respondents.

Method allowed	No. of responses	% responses	% respondents
Yes – migration	8	34.8	61.5
Yes – refresh	7	30.4	53.8
Yes – backup copies	6	26.1	46.2
Other	2	8.7	15.4
No	0	0	0
Total	**23**	**100**	**176.9**

Table 7.7 Whether authors would permit preservation copying by libraries, and methods that they would permit

7.3 Legal solutions for digital preservation

7.3.1 General comments about legal solutions

The interviews revealed a general lack of awareness of copyright law, particularly how it relates to preservation. The library questionnaire asked respondents whether they thought current legislation provides for the preservation needs of libraries. Forty-one per cent of the 161 respondents thought that it does not, only 6.8% thought it does, but more than half of respondents said that they did not know. Interviewees were similarly unsure about this. A digital preservation expert interviewed had encountered this lack of awareness in his work, and commented that there is:

> *a real confusion about what to do with copyright and how to deal with copyright with print and digital resources.*

This lack of awareness was said to be widespread; for example, one academic librarian involved in digitisation commented that academic staff 'don't understand the law'. One interviewee attributed this to a 'real lack of information', while another said that copyright law is viewed as 'impossibly difficult'. Interviewees agreed that UK copyright law is 'confusing' and contains 'a lot of subtleties', for example with the period of copyright protection. There is also a lack of relevant case law to clarify the uncertainties, particularly in the digital environment.

Some interviewees felt that the rights issues of digital preservation can only be addressed by legislation. One librarian explained that, 'legislation is the only thing that will do any good', since 'cooperation alone will not work'. However, a majority of interviewees, who represented all stakeholder groups, felt that legal solutions would not be ideal. One publisher argued that cooperation on digital preservation would be sufficient, since 'voluntary participation is always much better than compulsory. More effective.' A major reason given as to why legislation would be inadequate is that the legislative process is very slow. For example, in the case of the Legal Deposit Libraries Act 2003, a legal expert advised that it would take at least a year to develop the necessary regulations. As a librarian explained, this is too slow to be useful:

> *we need to have separate instruments for each different type of material, which doesn't sound highly workable to me, because they'll take longer to get through than it'll take the technology to advance!*

Publishers interviewed also agreed that the law would be slow. For example, one suggested that 'claims and counter-claims' would slow the process down, and another commented that it would not be satisfactory to 'resort to judges who have to then interpret … the law.' Interviewees also felt that laws would need to be able to evolve over time, because:

> *the requirements and … things libraries would want to do to preserve and to allow access over time will change.*

Again, they felt that the law would be too slow to achieve this satisfactorily. Another concern expressed was that laws would not be able to clarify details of agreements precisely enough. For example, a legal expert suggested that the Legal Deposit Libraries Act is only of limited usefulness, since it 'regulates some aspects of the relationship, but … remains silent on others'. It is not clear what aspects were being referred to here. She therefore suggested that even legal deposit libraries might want or need to have licensing agreements in place 'on top of' the law, for the sake of clarity.' It is not clear how this would be achieved in practice.

7.3.2 Legal deposit law and digital preservation

The Legal Deposit Libraries Act 2003 was passing through the House of Lords at the time of the interviews. The British Library representatives interviewed thought that this Act will have 'major implications', and will facilitate digital preservation, though only for legal deposit libraries. The Act will help the British Library to acquire comprehensive holdings of digital publications, although it will only apply to UK publications and will not apply retrospectively. Some of the publishers interviewed were unhappy about this Act; even in the last stages of its passage through parliament, one still felt that it was 'an awful mess', and had been 'rushed through in somewhat imperfect form'. He suggested that various things have been omitted, and that DCMS had not taken all of publishers' concerns into account. Publishers' views of Legal Deposit have been described in Section 5.4.

The Legal Deposit Libraries Act should provide a solution to the rights issues of digital preservation for legal deposit libraries. This is positive, because it shows that the government has recognised that:

> *there's not much point preserving a national archive of non-print material if in ten years time no one can look at it.*

This Act will allow legal deposit libraries to carry out prescribed preservation activities without infringing copyright. However, it is not yet clear precisely what these activities will be, since these will only be defined in the regulations. The Act therefore goes further than Section 42 of the CDPA since it takes into account the fact that:

> *for non-print material that we do need to preserve for posterity, the preservation activities may well need to be done at a very early stage. There's no point waiting to do your preservation until your machinery is obsolete.*

Legal deposit libraries will therefore be allowed to take action to preserve digital materials early in their life cycles. The Legal Deposit Libraries Act will also allow

legal deposit libraries to harvest Websites 'as an alternative to deposit for publishers of material on the Web.' Again, this will not infringe copyright.

7.3.3 Amending the law to allow other libraries to copy for preservation

The Legal Deposit Libraries Act will not allow libraries other than legal deposit libraries to carry out preservation copying. Further changes to the law may therefore be called for to allow this. More than a third of the libraries that responded to the library questionnaire wanted library privileges to be changed in this way. The respondents to the rights holders' questionnaires also agreed with this. Sixty percent of publisher respondents also felt that libraries should be given the legal right to make copies for preservation, with just over 19% saying that the law should not be changed, and nearly 22% not sure. Only one author respondent felt that libraries should not have this right, although two thought that preservation copying should only be allowed under licence. One author respondent commented that this would be a 'safeguard, in case publishers are unable or unwilling to do so'.

However, interviewees seemed much less in favour of changing the law in this way. Few interviewees thought that this needs to happen. Some of the legal experts interviewed felt that it would be very difficult to make any further legislation affecting existing copyright law. They suggested that any such changes could not be made lightly, since it would be unrealistic to expect publishers to trust all libraries with their content:

> *to trust every library, they wouldn't do anything else that they weren't permitted to do with that downloaded copy, I think would be going quite a long way further than would be necessary to achieve the objective.*

They also thought that 'it is unlikely that preservation provisions will be seen as high priority.' One legal expert did, however, advise that the EC Copyright Directive is not 'prescriptive in the detail,' meaning that further changes are theoretically possible. One digital preservation expert called on librarians and others to be more forceful in pushing back against rights holders for the right to preserve and read.

7.4 Licensing solutions for digital preservation

7.4.1 General comments about licensing as a solution

It has already been seen that many interviewees feel that legal solutions to the rights issues of digital preservation would be inadequate. In general, they believed that some sort of licensing solution, whether collective or individual, would be better for both libraries and publishers. A publisher representative suggested that the best solution would be for the stakeholders rather than the government 'to agree on standards, and then get them developed.' A legal expert also advised that licensing would enable libraries to get 'something more generous … than statute would ever be able to give them'. This interviewee added that publishers would prefer licensing since it would enable them to retain more control over what is done with their works:

> *they often are prepared to go further and permit more if they are licensing it, because at least they know it's happening, than if something is statutory … They don't like that at all, because they don't know exactly what's happening out there.*

Licensing was also viewed as more flexible than legislation. As a legal expert explained, licensing arrangements would be able to 'take account of individual circumstances, which the law cannot predict and incorporate'.

In view of these concerns, seminar delegates recommended that libraries should insist that their licences contain suitable clauses, and should refuse to sign unsuitable clauses, if user needs and demands allow this.

7.4.2 Individual licences for preservation

Both questionnaire respondents and interviewees generally felt that individual, transactional licensing would not be an ideal solution to the rights issues of digital preservation. Some library interviewees reported negative experiences of obtaining copyright clearance from publishers for digitisation or preservation purposes. The main complaint from librarians was that publishers and licensing societies were often

very slow to respond to their requests, although one did add that, 'generally, the ones that do reply are very positive'. Interviewees were also concerned that transactional licences for the use of copyright digital material are generally only granted for a limited period of time. This means that licences may have to be renewed regularly, which again takes time. However, a representative of a reproduction rights organisation said that they had stopped granting short licences because they had recognised that this was not useful 'in the long term.'

7.4.3 Collective licensing for preservation

In general, interviewees thought that there should be a blanket, collective license for preservation copying, like those that currently exist for photocopying and scanning. They felt that having a 'clearing house' that 'deals with everybody' would save time and energy for everyone involved. The different stakeholders also agreed that collective licensing would offer 'the best chance of a consistent approach.' Several interviewees suggested that a 'standard' or 'model' licence should be developed, including a licence or 'code of conduct' for legal deposit.

However, several interviewees felt that this type of licensing would be difficult to implement. Developing a standard licence would be a slow process, since:

> *people have had different views of what it should be. We can't … impose that philosophy*

Another interviewee agreed, explaining that 'you have to sometimes start with a less-than-ideal to get there'; she therefore proposed that stakeholders should:

> *start with individual licenses to get people to understand that they can trust this sort of approach.*

A representative of a rights holder organisation thought that rights holders might need to become more comfortable and familiar with other blanket licenses for digital materials before they could accept a licence for preservation copying.

A representative of a rights holder organisation suggested that the existing licensing societies are best placed to administer such licences since they already have extensive experience of transactional and collective licensing. They may also already be aware of stakeholders' views. For example, the organisation this interviewee represents has already consulted its members about the inclusion of their materials in digital publications, and is also involved in ongoing discussions with other stakeholders, consumers, and related organisations abroad. A legal expert interviewed agreed that negotiations should perhaps start with representative bodies, although she suggested that the impetus might equally come from individual stakeholders:

> *it could either start because there's one particular publisher who's happy to talk about licensing and then that could be expanded out, or it can start at the top level, you've got a representative body that's happy to talk about this and then try and get third parties to sign in to it when they've developed it.*

This interviewee felt that representative bodies were likely to be more reasonable than individuals. She suggested that if these methods failed, stakeholders could be encouraged to negotiate by someone from outside, for example a government representative.

7.4.4 Hybrid licences for preservation

Because of the perceived difficulties of developing and implementing a standard, blanket licence for digital preservation, more than one interviewee argued for a 'hybrid licence' combining elements of collective and individual licences:

> *It might be that it has to be a hybrid of the two things. … It maybe has to have a core of … very prescribed transactional uses, and then some collective elements that shoot off from that, to deal with low-risk things.*

One interviewee who has been involved in granting licences for digital materials thought that hybrid licences would work best because users often want licences for a

whole ‘suite of uses.’ Some of these uses are viewed by reproduction rights agencies as ‘primary uses’ for which ‘transactional’ licences would be more appropriate. Hybrid licences could contain ‘a number of optional clauses that could be evoked or not as the case would be.’ These clauses could, for example, cover embargoes on access.

7.5 Other solutions to the rights issues of digital preservation

7.5.1 Metadata

All three questionnaires asked respondents about rights metadata, and respondents were generally positive about using it. Almost a quarter of libraries thought that the provision of rights metadata by publishers would help them to preserve digital material. No publishers said they would refuse to provide metadata ‘detailing what copying [they] will allow for digital preservation purposes’. However, only twenty-seven publishers responded to this question, perhaps because they did know much about the topic. Several publishers interviewed already produce metadata for their publications. It is not clear what type of metadata this is; one publisher said that metadata is most important when publishers give their content to others to host, since it facilitates searching and distribution. Similarly, only two author respondents currently generate rights metadata for their digital materials, but nine said that they would be willing to do this. Only two felt that this was someone else’s responsibility. It would be interesting to know whether they are aware of what is involved in producing metadata, and what would motivate them to start doing this.

Interviewees also believed that metadata has a role to play in rights management. A legal expert interviewed observed that rights metadata is not yet widely enough used to judge its usefulness. However, this interviewee believed that it would be useful, perhaps in conjunction with licensing solutions, since it could ensure that:

> *[some] activities could be permitted, but … other ones that aren't desirable could be more readily policed and stopped.*

This would work as long as it were respected by both rights holders and users, and not used unreasonably:

> *I suppose that's most likely to be the case if the different parties understood each other's needs and desires, and being reasonable about it. Be not used merely as a mechanism for stopping everything, more as a mechanism for enabling reasonable things.*

Interviewees suggested different ways in which rights metadata should be used. One publisher suggested that digital publications should contain copyright statements similar to those in printed books, or perhaps giving more sophisticated rights information. This publisher also thought that:

> *One of the issues we have … would be in terms of paying our authors and monitoring use of material. So the use of metadata would be important in tagging material in terms of rights and also agreements that we have in place with our authors.*

A digital preservation expert thought that recording licence agreements in metadata would facilitate the copying of particular types of material for preservation purposes:

> *This would help practically, as people would be able to see what had been done in the past with particular licenses or legal cases, and see what to do in the future.*

Metadata could therefore help inform digital preservation agencies about rights issues.

Several interviewees also mentioned preservation metadata. Some of the libraries and preservation agencies interviewed currently assign preservation metadata to the digital files that they receive. This includes basic metadata that describe the content of a work and is usually produced by its creator, and more detailed metadata that are usually produced by the preservation agency when the material is received there. This second type of metadata may include information about 'the original format, modification history, current format, file size,' that will facilitate preservation.

Several interviewees emphasised the importance of metadata in helping them to preserve digital materials for the long-term. Deborah Woodyard of the British Library explained that an institution needs to know:

> *what you've got - and where it is … from the point of view of what media it's on, what format it's in, what technology it needs to be used with…*

Preservation metadata can help to record such information. The British Library is therefore intending to include such metadata in its Digital Object Management System, when this is introduced.

One of the barriers to the implementation of metadata is that it is costly to do this. A small publisher interviewed explained that: 'we're not of a size where it's economic yet to go through the steps to set that up.' Metadata was also said to be very time-consuming to produce. Deborah Woodyard therefore suggested that libraries and preservation agencies could be helped by:

> *ways of being able to automatically record it, or get the information from the publishers*

Automatic assignment of metadata is now becoming simpler because of the development of software to do this. Such software:

> *will look and identify text in the header page or the title page of a book, and then insert that metadata into your fields and into your indexes.*

One interviewee suggested that another major barrier to the widespread implementation of preservation metadata is that there is currently no standard metadata scheme. Interviewees mentioned a wide variety of schemes, including company schemes and the Dublin Core and Open Archives Initiative schemes. Specialist preservation metadata schemes are being developed in a number of places, including by OCLC and RLG, the National Library of Australia and in the

Netherlands. In the UK, work has also been done on this by the Cedars Project and the AHDS. One preservation expert commented that:

> *The only well-defined set of technical metadata that would be usable for preservation is the NISO Still Images, but the problem with that is that it's comprehensive, there's about 300 fields and that's about 290 fields more than you're ever going to have. There isn't really anything much for other types of material.*

One interviewee, who is a member of an international working group on preservation metadata, felt that it would not be easy to standardise metadata schemes:

> *we haven't done it yet, so I guess it's pretty hard … because we've been talking about it for a long time*

A suitable scheme will also need to be able to record detailed information about file formats and system requirements that will be particularly complicated to achieve for certain materials, such as interactive CD-ROMs. A further issue is that the existing schemes are all free-text. One interviewee felt that this would not be 'manageable' in the long term, partly because it is time-consuming to create.

However, another preservation expert was positive about the emergence of a standard metadata scheme:

> *I think a valid standard will emerge. … If you … compare them all, everybody's fairly much aware of what sort of things they want to record, so I think it'll be possible to map between them, even if you don't have a single standard.*

In fact, OCLC/RLG have now brought the different schemes together, and created the basis of a standard preservation metadata scheme. However, the metadata schemes used by publishers still have to be integrated with those used by libraries.

Digital Object Identifiers (DOIs) may also be able to help with preservation. Nick Dempsey explained that DOIs are primarily being used by the CrossRef movement of STM publishers, as part of a set of metadata that will eventually allow users to cross search journals from different publishers. The main object of DOIs is to enable users to locate documents over the long-term, even if they are moved from one Website to another, and this may help with digital preservation. DOIs can do this because:

> *a DOI has an identifier for its content, can then link to all kinds of other information about that content*

This information may include rights metadata that would make DOIs very useful for preservation.

DOIs have not yet widely been adopted by the publishers interviewed. In general, it is larger publishers who are aware of DOIs, and are already using them or considering using them in the future. A smaller publisher interviewed explained why they do not yet use DOIs:

> *With the volume of material we have, it's not economic to try and implement something like that.*

Another publisher said that they would only be likely to implement DOIs if so many other publishers had done so that they felt they would lose out by not doing so. By contract, one publisher that was already using DOIs commented that:

> *if there's something new that comes up and that's considered to be necessary, we'll do it too.*

This publisher is apparently more forward-looking than many of the publishers interviewed, although this may just be because it has greater resources. Nick Dempsey explained that while publishers were initially enthusiastic about DOIs, they are currently more sceptical 'about where CrossRef is going'. However, DOIs are becoming more widely used outside the STM area. For example, The Stationery

Office is a DOI registry, and is assigning DOIs to government documents. Nick Dempsey therefore thought that:

> *identifiers will be very important in the future. Particularly as … futuristic things like the semantic Web get a bit more foothold, and metadata surrounding the documents on the Web becomes more important.*

7.5.2 Digital Rights Management Systems

Several interviewees mentioned digital rights management systems as a possible solution to the legal issues surrounding digital preservation. Publishers are in favour of these as they can use them to control access to and usage of digital publications, and to calculate royalties. Another advantage is that technological protection measures have been given legal protection under the 2003 Copyright Regulations. A legal expert interviewed explained that the legislation contains provisions enabling circumvention of such measures by those who have a legitimate exception allowing them to copy a protected item. However, she explained that it is not yet clear whether this procedure will work well in practice.

Digital rights management systems are apparently little used at present. Only one publisher interviewed currently uses such a system. This is a digital-only publisher that protects its standalone CD-ROMs to prevent them from being copied. Another publisher interviewed is currently investigating sealed media as a way of protecting its PDFs. He explained that sealed media is:

> *a proprietary piece of software for sealing PDFs so that … the content and the licence are kept separately. The licence is held on a main server and the publisher can then determine exactly what form of licence they give with their content, whether it's to open to read only, read for three minutes, read for four weeks, pass on to four colleagues, whatever it is.*

This publisher felt this was a good system, but explained that they would not adopt it until the readers needed to use it are more widely available. They can already be downloaded simply and at no cost, but until they are more widely used the publisher's materials would be inaccessible to many users.

Nick Dempsey of EPS Ltd explained that while 'there are digital rights management systems out there at the moment', they have not been frequent since 'they're all a bit clunky and a bit hard to use.' He said that there is a drive towards using such systems, coming from the technology sector, but felt that they would not be widely adopted until they became more user-friendly. He was unsure how long this would take, but estimated that it could take five years. A publisher interviewed agreed that such systems still need much investment.

7.6 Solutions to the preservation of licensed materials

As has been seen, an important issue for libraries that license digital publications is whether they lose all access to materials if they cancel their subscription. One of them explained that:

> *I think … the whole question of perpetual access is becoming … more visible in the library and information world. And … more and more libraries are getting concerned about this.*

Interviewees felt that discussion between libraries and publishers is needed to resolve this issue that they variously described as a 'question which is open to debate' and as 'a battleground'. As a representative of a rights holders' organisation explained:

> *I just think there's no agreement yet on the philosophy, and whether … we should maintain the virtual world exactly the same way as we maintain … the paper world*

Two publishers interviewed understood libraries' concerns; they agreed in principle that libraries should continue to have access to journals electronically, as they would

have done with print. One publisher stated that they saw themselves as having 'an obligation' to provide access in perpetuity. However, interviewees clearly feared that a satisfactory solution would not be found for some time.

The interviewees expressed different preferences on whether publishers should provide access to their materials to libraries that have cancelled their subscriptions. One publisher thought it would be best if libraries continued to access materials via the publisher's Website:

> *from our point of view it would be quite nice if they would carry on using a section of the site because it would keep them aware that perhaps they ought to re-subscribe.*

An academic librarian agreed with this, saying:

> *I'd prefer not to get them on a bunch of CD-ROMs, I'd prefer access as we had been getting it, via the publisher's site or via another site, because managing your own CD-ROMs is a pain, CD-ROMs are not easily networkable, and so on*

A publisher agreed that giving libraries materials to host themselves might be difficult, since few libraries have the storage capacity to host large amounts of data themselves. However, the British Library representatives interviewed indicated that they prefer to be given files to host themselves, in keeping with their preservation role. It is not clear whether libraries would actually have the right to copy such files for preservation. Another library interviewee was concerned about what might happen if publishers gave physical copies of materials to libraries:

> *we've got to be careful they're not just giving us old … technologically challenged material that they just can't provide access to any more, we need it while we can still use it, and can still migrate it to something else, maybe.*

One interviewee suggested that an alternative solution might be for publishers to deposit their digital materials with a 'trusted repository' of some description; another interviewee added that depositing them with third parties 'whose mission is to preserve access to material' would also provide 'safeguards'. This should also help to alleviate libraries' concerns about whether publishers can be trusted to preserve their own materials. It would also mean that libraries would no longer be dependent on publishers keeping their servers active.

Not all publishers are clear about how they will provide such access. One librarian said that some licences provided options, while others 'might be a little bit vague about it'. This may be for different reasons: one publisher said this was because 'we're quite flexible' while another said 'it's basically unclear what would be preferable'. One publisher that has not yet made a decision about this said:

> *fortunately we haven't had anybody cancel a site licence that insists on getting the data.*

This reflects comments from the librarians interviewed, that they had not yet had to cancel a subscription, so did not know how their agreements would work in practice.

Another issue with licensed materials is how they will be preserved if their publisher merges or ceases to publish. Seminar delegates recommended that publishers should deposit copies of their materials with a national library or with subscribing libraries. This would best be done on publication, with an embargo on access, but could also be done when a publisher goes out of business. An agreement about preservation should be made at the time of depositing. This may not be overridden thereafter, even if the publisher is taken over by a company with a different policy about preservation.

8 CONCLUSIONS

The aim of this research was to investigate whether copyright and licensing issues are affecting the abilities of libraries to preserve digital content. The fundamental issue is whether libraries will be preserving digital material. There are two aspects of this question: whether libraries are or are planning to preserve, and whether they should be. If the answer to these questions is yes, then we can consider the copyright and licensing issues. It is important to take this approach, because one of the clear findings of this research is that the issue of responsibility for preservation has to be dealt with before any progress can be made on the legal issues.

8.1 The need to preserve digital library collections

The only really clear responsibility for preservation of digital material lies with the legal deposit libraries. The main aim of legal deposit is preservation and the legal deposit libraries are charged with maintaining and providing access to an archive of the nation's published output. While there is a lot of work to be done to implement deposit of digital publications in the UK, the new UK legal deposit legislation at least makes provision for exceptions to copyright to allow the legal deposit libraries to acquire and preserve future digital legal deposit collections.

This leaves the question of whether other types of library have digital collections that they wish to preserve. If the findings of this research are at all representative of UK libraries, then many libraries in all library sectors have growing digital collections, acquired through different means. Some of these collections will have no print equivalent, so someone somewhere has to preserve material of lasting value. This research provides no conclusive data on whether the proportion of born digital material in library collections will grow. However, comments from project participants suggest that there will be increasing divergence between so-called parallel-published material, that is, the digital version will become increasingly different from the print version of a work. Comments from participants and in the literature suggest that online publishing will become the first choice of dissemination channel, although some libraries may take physical delivery of material to mount on their own servers. Libraries are also investing in creating digital content through

making digital copies of their existing collections, and the preservation of these collections needs to be considered.

So libraries do have digital collections, but the question is whether they will be preserving them. The explosion in digital preservation research and development work and the inclusion of libraries in cooperative work in digital preservation suggest that at least some libraries other than national libraries will want to preserve their collections. While legal deposit libraries have led a lot of the work, academic libraries have also been taking the lead in this issue. Prominent examples considered in this research include the Cedars project and the Digital Preservation Coalition in the UK, Stanford University in the United States and the Research Libraries Group. More than half of the respondents to our library survey will be taking responsibility for preserving their collections now or in the near future, although few currently have established policies.

Taking responsibility for preservation does not necessarily mean that libraries will actually be carrying out preservation themselves. This is especially true for material that libraries do not physically own. Indeed, depending on licensing arrangements, they may not even be able to preserve material they do physically own.

8.2 Copyright legislation and digital preservation

One of the questions this research aimed to answer is whether copyright legislation will allow the sorts of activities required to implement digital preservation strategies. A lack of case law makes it impossible to provide a clear answer to this question. Evidence from the literature and from interviews with preservation and legal experts indicates that digital preservation strategies will involve activities that could infringe copyright law in several ways, and that the existing preservation exception will not allow all these activities to be carried out lawfully. The activities include making multiple copies for backup purposes, frequent replication to overcome media deterioration, reformatting material, saving software and running emulators to access material. In saying this, the legal experts were not unanimous in their interpretations of the law. While some activities were considered to adhere to the spirit of the law, some were considered to go beyond it. A particularly interesting point was that while

actions to preserve may be considered to adhere to the spirit of the law, subsequent access to preserved material would require permission from rights holders. Another rights issue that would affect the ability of libraries to preserve is a lack of publicly available documentation on file formats as rights owners seek to protect their commercial interests.

A survey of the law in other countries did not find much of use for reframing the UK preservation exception for libraries. The two laws that looked most useful required material to be already obsolete before preservation strategies could be implemented, which is not helpful. Issues to consider would include allowing for replicating the original material more than once, and allowing emulated material to be accessed even if only in a very limited way. When preservation can be carried out also has to be considered; waiting until material is not available commercially or is already obsolete will not help preservation of digital materials. The issue that only material held in permanent collections may be copied under the existing copyright exception was mentioned as problematic in the literature and by project participants because of the trend toward remote access. Whether this should also be changed is a moot point because it has to be considered along with the provisions of licence agreements.

A point made by participants in this research has also been made in other discussions of copyright in the digital environment. The point is whether the concept of "copying" is still useful. In the print environment, collection management and provision of access did not involve making copies. In the digital environment, pretty much all activities do. The question is whether these activities really threaten the ability of rights holders to benefit commercially from their intellectual property. Copyright law is supposed to provide a balance between allowing rights owners to benefit from their property and allowing access to it. Perhaps the type of activities that infringe intellectual property rights should be reconsidered.

This research has found no clear answer to the question of whether changing copyright law would help solve the problems described above. New legal deposit law in the UK allows for changes in copyright law for legal deposit libraries and some other countries have also taken similar measures. Changing copyright law, whether in a limited or more fundamental way, takes time and may be difficult. In the meantime,

if libraries want to carry out preservation themselves, the most practical way to do so is to ask for permission.

As for moral rights, although participants in the project admitted there might be cases where the result of digital preservation actions may result in changes to digital material, this was unlikely to be a big problem. As far as creators are concerned, the issues seem to be that words should not be altered and attributions should be retained. Whether changes to look and feel and functionality would result in infringement of moral rights would depend on the nature of the work and the degree and nature of the changes. The publishers who participated in this research expressed different opinions on what they thought about changes and their typographical rights. Again, there seemed to be some agreement that the basic content should not be changed, but a range of opinions on whether they could accept changes to other aspects of digital material such as interfaces and functionality. Commercial issues were mentioned here; there was some concern that libraries would improve material through preservation actions and potentially benefit in a financial way from making it accessible in its new form. There is little evidence to suggest that this would be the case.

8.3 Licensing and digital preservation

There are several aspects to the issue of licensing and digital preservation. One such issue is the preservation of remotely accessed licensed content and licences for preserving physically owned content. Some of the library respondents to the questionnaire made comments that seemed to refer to dissatisfaction with provision of current or ongoing access to online digital material. This issue was outwith the scope of this research project. However, the inclusion of older material within subscriptions is a preservation issue for libraries who would like to discard paper copies of older material, but still require access to it. There is also the issue of whether libraries retain access to material they paid for during their subscriptions when subscriptions come to an end.

The research indicated that there is little standardisation in licence agreements for remotely accessed content. Although some model licences exist and are used by some

publishers, there is no collective licensing in this area and different publishers do different things. While access to older material may be possible, it may have to be paid for and the back files may not be extensive. The practice of using a "rolling wall" of access to back files does not seem to be frequent. Commercial interests can play a role in provision of access to older material. It may not be economic to convert older material to digital form and publishers may want to benefit commercially from their older material. Comments form librarians suggest that they are not happy to pay "extra" for older material, either during a subscription or after it ends.

The findings of the research were not encouraging as far as the ability of publishers to preserve their own material over time. Since legal deposit will not include all material of possible longer-term interest to other libraries, this is a concern. While some publishers consider they have a duty to keep material available long-term, others are motivated by commercial concerns and are therefore not able to give long-term guarantees to libraries. There is also an indication from the publisher survey that publishers have not thought through arrangements for longer-term care of their material. There is little indication of concern about this in either library or publisher responses. This may be because arrangements have not yet been tested. A point raised here was whether the digital environment should mirror the print environment. Should access to digital content follow the same rules as print, as librarians seem to want? Or can publishers change the existing norms to fully exploit their digital intellectual property.

As far as licensing for preservation goes, we did not get enough data to comment on the views of creators on preservation of their work or of how they currently assign and licence rights to publishers. The publisher survey suggested that publishers' arrangements with original rights holders could introduce complications as a far as licensing for preservation is concerned. This would be particularly true for software, where the agreement with the software provider is very narrow.

The research indicated that libraries would prefer collective licensing of digital preservation, based on prior experience of licensing other types of digital activity. This would save time and resources. The academic sector does benefit from the use of a model licence for access with preservation provisions, but this is silent on

preservation by libraries of content provided by publishers. However, the ease of implementing collective licensing was questioned and a gradual move towards this and the use of hybrid standard/individual licences was suggested. These suggestions imply the need for cooperative working between libraries and publishers to develop such licences, and also for reproduction rights organisations to get actively involved.

The research did indicate some enthusiasm for the use of rights metadata to facilitate digital preservation, but comments from interviewees confirmed that this area needs further development both in terms of metadata schemes and finding economic ways of gathering this metadata.

8.4 Responsibility for digital preservation

So far this chapter has focused on the difficulties of preserving digital information when libraries do not physically hold content or do not have the rights to preserve. There is a question of how libraries can preserve material if they do not trust publishers to do it. There was no clear consensus from either publishers or libraries on who should be responsible for digital preservation, although the majority of librarians and publishers selected legal deposit libraries as an option. This is not just a rights issue; respondents to the library questionnaire were not actively preserving material other than using good security practices. There was a comment in the interviews that even if publishers handed over content to libraries, the libraries may not have the capacity to mount it on their own servers. There was also an indication of lack of knowledge to carry out digital preservation among library respondents. This is hardly surprising since even the most advanced libraries in the world are still working out how to do this. The use of trusted repositories was suggested. There are precedents for this in other countries, for example Elsevier's arrangements with libraries in the Netherlands and the United States.

8.5 Lack of awareness

A major finding of this research is a lack of awareness of digital preservation and related copyright and licensing issues. This does not apply to all participants, but there was evidence in the questionnaire responses and comments from interviewees. This

lack of awareness was not confined to a particular stakeholder group; librarians, publishers and authors all showed a lack of awareness of certain issues. The authors showed a general lack of interest in the project. However, a creator representative organisation did take part in the interviews. Respondents were unsure of what copyright law allows and of preservation provisions of rights agreements in particular. Only specialists such as legal deposit libraries, preservation and legal experts showed a real awareness of the issues. The publisher survey respondents who include preservation provisions in their agreements with customers did not seem to have thought through how they would implement guarantees.

There was also some evidence of a lack of trust between publishers and libraries, quite possibly based on a lack of awareness, particularly on the part of publishers. The surveys and the interviews indicated some caution on the part of publishers in what they would allow libraries to do in terms of digital preservation. This caution has surfaced in the literature, particularly around discussions of extension of legal deposit law in the UK. It seems that at least some publishers are concerned that their commercial interests could be damaged by loss of control of their material, and by libraries using their intellectual property to compete with them.

9 RECOMMENDATIONS

9.1 Need to raise awareness

It is hard to see how any progress can be made in addressing the legal issues arising from preservation of digital information without raising awareness among the interested parties. It is not just a case of raising awareness of the rights and licensing issues but of digital preservation issues generally. This is not the first research project to come to this conclusion, and action is being taken to raise awareness through various means, including the UK Digital Preservation Coalition.

This does not only apply to legal deposit and research libraries, as libraries in different sectors also acquire and create digital content. Publishers and creators also need to be more aware of the issues. Each of these groups needs to know more about what digital preservation involves, what the rights issues are and what the perspectives, concerns and priorities of other groups are. A greater awareness of these issues should help ease the lack of trust between rights holders and libraries and pave the way towards rights regimes that will meet the needs of all interested parties.

There are already awareness raising activities in progress in the UK and other countries too. The legal deposit libraries have been working with publishers groups, the UK higher and further and education sector has also been working with publishers. Prominent projects in this area, including the Cedars project, played a significant role in investigating preservation issues and disseminating their findings. The UK Digital Preservation Coalition is working on awareness raising in this country and also has contacts in other countries, for example in Australia. However, there is still a lot of work to be done. There is a role for organisations representing different stakeholder groups, for example the Museums, Libraries and Archives Council (formerly known as Resource: The Council for Museums, Archives and Libraries) and CILIP, the Chartered Institute of Library and Information Professionals in the library sector, the Publishers Association and other specialist publishing associations, creator representatives and reproduction rights organisations such as the Authors Licensing and Copyright Society, the Publishers Licensing Society, the Design and Artists Copyright Society, the Copyright Licensing Agency and the Newspaper Licensing

Agency. These organisations can cascade information to their members and work cooperatively to raise awareness.

9.2 Examine issue of responsibility for preservation

This and other research in this area has found a greater need for cooperation in the preservation of digital information. The current situation is very fragmented, with the British Library in particular taking the lead, but as yet no coordinated policy on who should be responsible for digital preservation. The rights issues differ, depending on who is carrying out preservation. There is a need to seriously engage with this issue and to examine the possibility of specialist repositories responsible for preserving digital information. This could take the pressure off rights holders and libraries. There is no denying that this would be a complex undertaking and there would be a lot of issues to resolve, including access agreements, commercial confidentiality, security and funding.

Organisations such as the national and other large libraries could become trusted repositories for publishers and other libraries, or new organisations could be set up. This is an issue that has to be discussed by representatives of all library sectors, publishers, including open access publishers and institutional repositories, the UK Government. It may be that arrangements can only be implemented piecemeal with individual publishers making deals with preservation organisations. On the other hand, there may be scope for international cooperation on this issue. Publishers could, for example deposit material in their home country and access to preserved material could be provided through negotiated agreements. However, this would be complex to set up and manage, particularly for material that is still commercially valuable.

Centralised repositories have to be funded in some way. If they are housed in existing institutions, these institutions will require additional funding to cover additional activities. Existing candidates in the UK would be legal deposit and large research libraries and data centres. The UK Government is the most obvious source of funding, through central and devolved Government departments and education funding councils. However, the Government will need to be convinced that this issue is sufficiently important to be worth supporting on a sustainable basis. The UK

Government has shown a commitment to extending legal deposit to cover non-print material, but reluctance to provide additional funding. A concerted lobbying effort would be required to have any chance of success. This would have to include rights holder and user representatives as well as libraries. The Digital Preservation Coalition, library organisations such as CILIP, the Consortium of University and Research Libraries and the Association of Learned and Professional Society Publishers could have a role in this. Open access publishers may also be able to play a role.

A lot of work on digital preservation has focused on technical issues and has only looked at management issues at the level of individual institutions. There is some movement towards taking a more strategic approach to digital preservation. The Library of Congress, for example, is leading the National Digital Information Infrastructure and Preservation Program (NDIIP) in the United States. There is an urgent need to look at models for preservation that transcend individual institutions and how these can be funded. If individual institutions cannot afford to preserve without additional resources and Governments are unable or unwilling to take on the financial burden of preservation, it is crucial that other possibilities need to be explored. The situation is very difficult because digital preservation cannot be implemented until sustainable models are found. However, researching possible models is difficult because the questions to be asked are still hypothetical.

9.3 Cooperative discussions on licensing

Establishing responsibility for digital preservation will make it easier to resolve licensing issues for carrying out preservation. However, there is also a need to look at the agreements that libraries have for access to remotely accessed material. Publishers need to be fully aware of the concerns of their customers, particularly about access to back files during and after termination of subscriptions. At the same time, libraries that wish to preserve material themselves need to reassure publishers that they are not planning to compete with them.

The starting point for such discussions could be the group set up to work on developing regulations for the implementation of legal deposit for non-print

publications. The higher and further education sector already has some infrastructure in place for joint working. The public and special library sectors also need to be involved. There are some model licences around, in particular the NESLI electronic journal licence that could be used as a starting point for discussions. Discussions could examine areas that could be standardised and other areas where options or customised clauses would be more appropriate. Discussions on licensing also need to include creators. Creators need to be aware of preservation needs and to make their own views on preservation and use of their work clear.

The licensing and reproduction rights organisations have a role here, as do organisations representing different library sectors, for example the Joint Information Systems Committee, the MLA and CILIP. Libraries need to actively communicate their requirements to their suppliers, either in their representative groupings and/or through JISC. They also need to communicate with the reproduction rights agencies, who can then talk to their members about the possibility of developing collective licensing for preservation.

9.4 Copyright legislation

One of the original objectives of this research was to make recommendations on licensing and changes to legislation if appropriate. It is not clear from the findings of this research whether it is appropriate to change current legislation, apart from to recommend that the provisions in the Legal Deposit Libraries Act 2003 are implemented.

As for other libraries, the scope of the preservation exception to current UK copyright does still need to be clarified. However, it is likely that libraries at least will avoid becoming case law in this area. The question is whether the rights issues can be solved through licensing and cooperation between stakeholder groups. If not, and the future of valuable content is endangered, then maybe the law should be changed. This would not be easy because technology and the electronic publishing industry are constantly evolving and the exception would have to be made as future proof as

possible. While librarians may prefer this option, it is not clear that rights holders would, even though respondents to our publisher survey seemed amenable.

In the long-term, a reassessment of the whole concept of copyright in the digital environment is probably needed and this should take place not at the national, but at the European or even world level. There is some scope for research in this area, perhaps through a Delphi exercise in the first instance, to gather views on this issue. There would then have to be debate at the international level on whether and how the traditional balance of interests represented in copyright should be addressed.

10 BIBLIOGRAPHY

Abbott, D., 2003. Overcoming the dangers of technical obsolescence: rescuing the BBC Domesday Project. *DigiCULT.info: a newsletter on Digital Culture* [online], 4. <http://www.digicult.info/downloads/digicult_newsletter_issue4_highres.pdf>, [accessed 11.02.04].

AIIM International, 2003. *PDF Archive*. <http://www.aiim.org/standards.asp?ID=25013>, [accessed 11.02.04].

AOP and British Library Discuss Archiving the Web, 2003. Managing Information [online], 18th February 2004. <http://www.managinginformation.com/news/content_show_full.php?id=2409>, [accessed 23.02.04].

Australia, 1968. *Copyright Act 1968 (Act No. 63 of June 7, 1968 as amended in 2002).* <http://clea.wipo.int/>, [accessed 11.02.04].

Bearman, D., 1999. Reality and chimeras in the preservation of electronic records. *D-Lib Magazine* [online], 5(4). <http://mirrored.ukoln.ac.uk/lis-journals/dlib/dlib/dlib/april99/bearman/04bearman.html>, [accessed 11.02.04].

BioMed archive deal, 2003. *Library and Information Update*, **2**(12), 13.

British Library, 2003. *Policies: legal deposit in the British Library*. <http://www.bl.uk/about/policies/legaldeposit.html>, [accessed 23.02.04].

Byers, F.R., 2003. *Care and handling of CDs and DVDs: a guide for librarians and archivists*. Washington, D.C.: Council on Library and Information Resources; Gaithersburg, MD: National Institute of Standards and Technology.

Canada, 1997. *Statutes Of Canada, 1997. C. 24 An Act to amend the Copyright Act*, *Second Session, Thirty-fifth Parliament, 45 & 46 Elizabeth II.* <http://laws.justice.gc.ca/en/C-42/>, [accessed 11.02.04].

CEDARS Project, 2001. *The Cedars Project report: April 1998-March 2001*. <http://www.leeds.ac.uk/cedars/admin/CedarsProjectReportToMar01.pdf>, [accessed 11.02.04].

CEDARS Project, 2002. *Cedars guide to intellectual property rights*. <http://www.leeds.ac.uk/cedars/guideto/ipr/guidetoipr.pdf>, [accessed 11.02.04].

Chalcraft, A., Prytherch, R., and Willis, S., 1998**.** *Walford's guide to reference material. Vol. 3, Generalia, language and literature, the arts,* 7th ed. London: Library Association.

Charlesworth, A., 2002. *Legal issues arising from the work aiming to preserve elements of the interactive multimedia work entitled "The BBC Domesday Project"*. <http://www.si.umich.edu/CAMILEON/reports/IPRreport.doc>, [accessed 11.02.04].

Creative Commons, [n.d.], <http://creativecommons.org/>, [accessed 11.02.04].

Consultative Committee for Space Data Systems, 2002. *Reference model for an Open Archival Information System (OAIS). Blue Book, Issue 1 CCSDS 650.0-B-1 January 2002*. <http://www.ccsds.org/docu/dscgi/ds.py/Get/File-143/650x0b1.pdf>, [accessed 17.02.04].

Darlington, J., 2003. PRONOM - A Practical Online Compendium of File Formats. *RLG DigiNews* [online], 7(5). <http://www.rlg.org/preserv/diginews/diginews5-4.html>, [accessed 11.02.04].

Darlington, J., A. Finney, & A. Pearce, 2003. Domesday Redux: the rescue of the BBC Domesday Project vidodiscs. *Ariadne* [online], 36. <http://www.ariadne.ac.uk/issue36/tna/>, [accessed 04.09.03].

Denmark, 1997. *Act on Copyright Deposit of Published Works*. <http://www.kb.dk/kb/dept/nbo/da/pligtafl/pligt-en.htm>, [accessed 11.02.04].

Digital Object Identifier System, 2004. *Welcome to the Digital Object Identifier System.* <http://www.doi.org/>, [accessed 26.02.04].

Digital Preservation Coalition, 2002, <http://www.dpconline.org/graphics/>, [accessed 4.12.03].

Donoghue, A., 2003. *XML: Extremely critical or exhaustingly complex?* <http://insight.zdnet.co.uk/0,39020415,39115499,00.htm>, [accessed 11.02.04].

Entire French Web to be archived, 2001, <http://www.europemedia.net/shownews.asp?ArticleID=4075>, [accessed 18.11.02].

European Parliament and Council of the European Communities, 1993. *Council Directive 93/98/EEC of 29 October 1993 harmonizing the term of protection of copyright and certain related rights.* <http://www.ebu.ch/departments/legal/pdf/leg_ref_ec_directive_copyright_duree_protection_291093.pdf>, [accessed 25.02.04].

European Parliament and Council of the European Communities, 1996. *Council Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases.* <http://europa.eu.int/smartapi/cgi/sga_doc?smartapi!celexapi!prod!CELEXnumdoc&lg=en&numdoc=31996L0009&model=guichett>, [accessed 16.02.04].

European Parliament and Council of the European Communities, 2001. *Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society.* <http://europa.eu.int/smartapi/cgi/sga_doc?smartapi!celexapi!prod!CELEXnumdoc&lg=EN&numdoc=32001L0029&model=guichett>, [accessed 16.02.04].

Feeney, M., 1999a. *Digital culture: maximising the nation's investment.* London: National Preservation Office.

Feeney, M., 1999b**.** Towards a national strategy for archiving digital materials. *Alexandria*, **11**(2), 107-121.

Gadd, E., C. Oppenheim, & S. Probets, 2003. *RoMEO Studies 6: Rights metadata for open archiving.* <http://www.lboro.ac.uk/departments/ls/disresearch/romeo/index.html>, [accessed 11.02.04]. This article has been accepted for publication in Program 38(1) 2004.

Gilheany, S., 1998. *Preserving information forever and a call for emulators: presented at the Digital Libraries Conference and Exhibition.* <http://www.archivebuilders.com/aba010.html>, [accessed 16.02.04].

Google, 2004, <http://www.google.com>, [accessed 11.02.04].

Granger, S., 2001. *Digital preservation & emulation: from theory to practice.* <http://www.leeds.ac.uk/cedars/pubconf/papers/ichim01SG.html>, [accessed 11.02.04].

Granger, S., 2002. Digital preservation and deep infrastructure. *D-Lib Magazine*, [online], 8(2). <http://www.dlib.org/dlib/february02/granger/02granger.html>, [accessed 11.02.04].

Great Britain, 1988. *Copyright, designs and patents Act 1988, (c. 48).* <http://www.hmso.gov.uk/acts/acts1988/Ukpga_19880048_en_1.htm>, [accessed 16.02.04].

Great Britain, 1989. *Copyright: the copyright (librarians and archivists) (copying of copyright material) regulations 1989.* <http://www.legislation.hmso.gov.uk/si/si1989/Uksi_19891212_en_1.htm>, [accessed 17.02.04].

Great Britain, 1995. *The duration of copyright and rights in performances regulations 1995.*

<*http://www.legislation.hmso.gov.uk/si/si1995/Uksi_19953297_en_1.htm*>, [accessed 17.02.04].

Great Britain, 1997. *Copyright rights in databases: the copyright and rights in databases regulations 1997.* <*http://www.hmso.gov.uk/si/si1997/1973032.htm*>, [accessed 17.02.04].

Great Britain, 2003a. *The copyright and related rights regulations 2003.* <http://www.legislation.hmso.gov.uk/si/si2003/20032498.htm>, [accessed 11.02.04].

Great Britain, 2003b. *Legal Deposit Libraries Act 2003.* <http://www.hmso.gov.uk/acts/acts2003/30028--a.htm#1>, [accessed 16.02.04].

Hendley, T., 1998. *Comparison and methods & costs of digital preservation.* British Library Research and Innovation report 106. London: British Library Research and Innovation Centre.

Hirtle, P, 2003. *Digital Preservation and Copyright.* <http://fairuse.stanford.edu/commentary_and_analysis/2003_11_hirtle.html>, [accessed 11.02.04].

Holdsworth, D. & P Wheatley, 2001. Emulation, preservation, and abstraction. *RLG DigiNews* [online], 5(4). <http://www.rlg.org/preserv/diginews/diginews5-4.html#feature2>, [accessed 11.02.04].

Hong Kong, 1997. *Copyright Ordinance, Chapter 528.* <http://clea.wipo.int/>, [accessed 11.02.04].

Institute for Information Law, 1998. *Copyright aspects of the preservation of electronic publications.* IViR Reports 7. University of Amsterdam. <www.ivir.nl/Publicaties/koelman/KBeng2.doc>, [accessed 11.02.04].

Internet Archive, 2001, <http://www.archive.org>, [accessed 17.02.04].

James, H. *et al*., 2003. *Feasibility and Requirements Study on Preservation of E-Prints.* <http://www.jisc.ac.uk/uploaded_documents/e-prints_report_final.pdf>, [accessed 11.02.04].

Joint Information Systems Committee, 2003. *JISC Circular 6/03 (Revised). An invitation for expressions of interest to establish a new Digital Curation Centre for research into and support of the curation and preservation of digital data and publications.* <http://www.jisc.ac.uk/index.cfm?name=funding_digcentre>, [accessed 11.02.04].

Jones, M., 2003. Archiving e-journals consultancy – final report. <http://www.jisc.ac.uk/uploaded_documents/ejournalsfinal.pdf>, [accessed 11.02.04].

JSTOR, 2004. *JSTOR – the scholarly journal archive*. <http://www.jstor.org/>, [accessed 11.02.04].

Kranch, D. A., 1998. Beyond migration: preserving electronic documents with digital tablets. *Information Technology and Libraries*, **17**(3), 138-145.

Larivière, J., 2000. *Guidelines for legal deposit legislation,* rev., enl. and updated ed. Paris: Unesco. <http://www.ifla.org/VII/s1/gnl/legaldep1.htm>, [accessed 11.02.04].

Lee, K.-H. *et al.*, 2002. The State of the art and practice in digital preservation. *Journal of Research of the National Institute of Standards and Technology,* **107**(1), 93-106.

Lehmann, K.-D., 1996. Making the transitory permanent: the intellectual heritage in a digitized world of knowledge. *Daedalus*, **125**(4), 307-329.

LexisNexis (TM) Professional, 2004. http://Web.lexis-nexis.com/professional/ [accessed 11.02.04].

Licensingmodels, 2001. *Academic: single institution licence: Version 2.0 14/4/00.* <http://www.licensingmodels.com/academic.htm>, [accessed 26.02.04].

Liblicense, 2001. *Standard License Agreement.* <http://www.library.yale.edu/~llicense/standlicagree.html>, [accessed 16.02.04].

Library and Information Science Abstracts. Online via Cambridge Scientific Abstracts. <http://www.csa1.co.uk>, [accessed 11.02.04].

LOCKSS Web site, [n.d.], <http://lockss.stanford.edu/>, [accessed 11.02.04].

Lorie, R., 2001. A Project on preservation of digital data. *RLG DigiNews* [online], 5(3). <http://www.rlg.org/preserv/diginews/diginews5-3.html#feature2>, [accessed 11.02.04].

Lorie, R., 2002. *The UVC: a method for preserving digital documents: proof of concept*. IBM/KB Long-term Preservation Study Report Series 4. Amsterdam: IBM Netherlands; The Hague: Koninklijke Bibliotheek.

Lynch, C., 1999. Canonicalization: a fundamental tool to facilitate preservation and management of information. *D-Lib Magazine* [online], 5(9). <http://www.dlib.org/dlib/september99/09lynch.html>, [accessed 11.02.04].

Martin, D., 2001. *ONIX International Guidelines for Publishers Level 1*. London: Book Industry Communication. <http://www.editeur.org/onixfiles1.2.1/ONIX%20Guide%20Level%201%20R1.2.1.PDF>, [accessed 11.02.04].

Mauritzen, I., & S. Solbakk, 2000**.** *A study on copyright and legal deposit of online documents*. Helsinki: Edita.

Mayfield, K., 2001. Wayback goes way back on Web. *The Wired News* [online], 29 October 2001. <http://www.wired.com/news/culture/0,1284,47894,00.html>, [accessed 11.02.04].

Mellor, P., P. Wheatley & D. Sergeant, 2002. *Migration on request, a practical technique for preservation.* <http://www.si.umich.edu/CAMILEON/reports/migreq.pdf>, [accessed 11.02.04].

Muir, 2004. *Digital legal deposit*. PhD thesis, Department of Information Science, Loughborough University.

Nasjonalbiblioteket, [n.d.]. *Paradigma: preservation of digital material.* <http://www.nb.no/paradigma/eng_utskriftsvennlig.html>, [accessed 11.02.04].

NESLi, 2002. *NESLi2 Licence for journals.* <http://www.nesli2.ac.uk/nesli2_lic_010903.htm>, [accessed 11.02.04].

New decree for Kulturarw3, 2002, <http://www.kb.se/Info/Pressmed/Arkiv/2002/020605_eng.htm>, [accessed 11.02.04].

New e-preservation deals, 2002. *Library and Information Update*, **1**(7), 10.

New Zealand, 1994. *Copyright Act 1994 No. 143.* <http://www.legislation.govt.nz/browse_vw.asp?content-set=pal_statutes>, [accessed 11.02.04].

Norway, 1989. *Act No. 32 of 9 June 1989 relating to the legal deposit of generally available documents*. <http://www.pliktavlevering.no/html/legal_deposit.html>, [accessed 11.02.04].

OCLC FirstSearch (1992-2004). Available online from OCLC First Search

OCLC/RLG Working Group on Preservation Metadata, 2002. *Preservation metadata and the OAIS information model: a metadata framework to support the preservation of digital objects.* <http://www.oclc.org/research/pmwg/pm_framework.pdf>, [accessed 11.02.04].

Open Archives Initiative, [n.d.]. *Organization.* <http://www.openarchives.org/organization/index.html>, [accessed 11.02.04].

Pinfield, S., & H. James, 2003. The Digital Preservation of e-Prints. *D-Lib Magazine* [online], 9(9). <http://mirrored.ukoln.ac.uk/lis-journals/dlib/dlib/dlib/september03/pinfield/09pinfield.html>, [accessed 11.02.04].

Project RoMEO, [n.d.], <http://www.lboro.ac.uk/departments/ls/disresearch/romeo/index.html>, [accessed 11.02.04].

Representation and Rendering Project, 2003. *Survey and assessment of sources of information on file formats and software documentation : Final Report.* <http://www.jisc.ac.uk/uploaded_documents/FileFormatsreport.pdf>, [accessed 11.02.04].

République française, 2001. *Project de loi sur la société de l'information*. (NOR: ECOX0100052L/B1), p. 6. <http://www.lsi.industrie.gouv.fr/observat/innov/lsi/exp.pdf>, [accessed 21.10.2003].

Research Libraries Group, 2002. *Preserving Digital Information: Final Report and Recommendations.* <http://www.rlg.org/ArchTF/>, [accessed 11.02.04].

Robertson, S. B., 1996**.** *Digital Rosetta Stone: a conceptual model for maintaining long-term access to digital documents.* MSc dissertation, Air Force Institute of Technology, Air University.

Rothenberg, J., 1995. Ensuring the longevity of digital documents. *Scientific American*, **272**(1), 24-29.

Rothenberg, J., 1999a. *Ensuring the longevity of digital information.* <http://www.clir.org/pubs/archives/ensuring.pdf>, [accessed 11.02.04].

Rothenberg, J., 1999b. *Avoiding technological quicksand: finding a viable technical foundation for digital preservation: a report to the Council on Library and Information Resources.* Washington: Council on Library and Information Resources. <http://www.clir.org/pubs/reports/rothenberg/pub77.pdf>, [accessed 11.02.04] .

Rothenberg, J., 2000**.** *An Experiment in using emulation to preserve digital publications*. Den Haag: Koninklijke Bibliotheek.

Russell, K., 1999. *Digital preservation: ensuring access to digital materials into the future*. <http://www.leeds.ac.uk/cedars/Chapter.htm>, [accessed 11.02.04].

Shepard, T., & D. MacCarn, 1998. *The Universal Preservation Format: background and fundamentals*. Sixth DELOS Workshop : Preservation of Digital Information. <http://www.ercim.org/publication/ws-proceedings/DELOS6/upf.pdf >, [accessed 17.02.04].

South Africa, 1997. *Legal deposit act*. <http://www.nlsa.ac.za/docs/legaldep.pdf>, [accessed 11.02.04].

Thibodeau, K. 2002. Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years. In: *The State of Digital Preservation: An International Perspective: conference proceedings, Documentation abstracts, Inc. Institutes for Information Science, Washington, D.C., April 24-25, 2002*. Washington, D.C.: CLIR. <http://www.clir.org/pubs/reports/pub107/contents.html>, [accessed 11.02.04].

Tizard, J., 2003. *Digital Technology and the Copyright Act 199 : Policy Recommendations: Cabinet Paper.* <http://www.med.govt.nz/buslt/int_prop/digital/cabinet/index.html>, [accessed 11.02.04].

United Kingdom Parliament, 2003. House of Commons Hansard for 4 Jul 2003 (pt 22). http://www.parliament.the-stationery-

office.co.uk/pa/cm200203/cmhansrd/cm030704/debtext/30704-22.htm, [accessed 23.02.04].

United States, 1998. *The digital millennium copyright act of 1998: U.S. Copyright Office summary.* <http://www.copyright.gov/legislation/dmca.pdf>, [accessed 17.02.04].

UPF Home, 2000, <http://info.wgbh.org/upf/>, [accessed 11.02.04].

Van Bogart, John W.C., 1995. *Magnetic tape storage and handling: a guide for libraries and archives*. <http://www.clir.org/pubs/reports/pub54/4life_expectancy.html>, [accessed 11.02.04].

Waters, D. & J. Garrett, 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*. <ftp://ftp.rlg.org/pub/archtf/final-report.pdf>, [accessed 11.02.04].

Wheatley, P., 2001. Migration: a CAMiLEON discussion paper. *Ariadne* [online], 29. <http://www.ariadne.ac.uk/issue29/camileon/>, [accessed 11.02.04].

World Intellectual Property Organisation, [n.d.]. *Collection of laws for Electronic Access.* <http://clea.wipo.int/clea/lpext.dll?f=templates&fn=main-h.htm&2.0>, [accessed 11.02.04].

APPENDIX 1

Copyright and Licensing for Digital Preservation

A research project funded by

Arts and Humanities Research Board

Library Questionnaire

All information will be treated in confidence.
Please tick box(es) as appropriate.

A. DIGITAL COLLECTIONS

1. Is your library:

Academic ☐

Public ☐

Special ☐

National ☐

Other (please specify) _____

2. Does your library have digital resources in its collections?

Yes ☐ *Go to question 5*

No ☐ Go to question 3

3. Do you expect to acquire digital material in the next five years?

Yes ☐ *Go to question 5*

No ☐ *Go to question 4*

Don't know ☐

4. What are your main reasons for <u>not</u> acquiring digital material?

_____

_____ *Go to question 14*

5. What are your main reasons for acquiring digital material ?

---

6. Approximately how much of the digital material in your library is:

Unique (only exists in digital form) _____%

Duplicate of another format (e.g. print) _____%

Don't know ☐

7. How was this digital material acquired?

Donation _____%

Purchased outright _____%

Licensed from a vendor _____%

Created in-house _____%

Voluntary deposit _____%

Harvested from Web _____%

Other (please specify) _____

B. DIGITAL PRESERVATION POLICIES

Digital preservation: *Storage, maintenance and access to digital objects/materials over the long-term. This may involve one or more digital preservation strategies including technology preservation, technology emulation or digital information migration.*

8. Do you assume responsibility for the preservation of any digital material in your collections?

Yes	☐	*Go to question 9*
No, but will in future	☐	*Go to question 11*
No	☐	*Give reasons if you wish, then go to question 14*

---

9. Which digital material have you assumed preservation responsibility for?

Material deposited by publishers ☐

Donated material ☐

Material purchased outright ☐

Licensed material ☐

Other (please specify) _____

10. Does your library currently have a formal digital preservation policy?

Yes ☐

No ☐

Don't know ☐

If you have a formal policy and you would be willing to let us to see it, please attach it to your returned questionnaire

11. Will you be developing a digital preservation policy in the next 12 months?

Yes □

No □

Don't know □

Further comments

Refresh: *to copy digital information from one long-term storage medium to another.*

Technology Preservation: *digital data are stored at bit streams on a stable digital medium (and refreshed to new media as required) and associated with that object are preserved copies of the original application software, the operating system that this would normally run under and the relevant hardware platform.*

Emulation: *digital materials are stored in their original format as a bit stream and software and hardware emulators are employed to mimic the behaviour of obsolete hardware platforms and emulate the relevant operating system to allow for access.*

Migration: *a set of organised tasks designed to achieve the periodic transfer of digital materials from one hardware/software configuration to another, or from one generation of computer technology to a subsequent generation.*

12. Do you use any of the following preservation strategies?

Refreshing □

Technology preservation □

Migration □

Emulation □

Don't know □

None □

Other (please specify) _____

_____

13. Is digital preservation carried out:

In house ☐

Externally ☐

N/A ☐

Don't know ☐

Further comments _____

C. COPYRIGHT AND LICENSING ISSUES FOR DIGITAL PRESERVATION

14. Do you think that current UK copyright and database legislation provides for the digital preservation needs of libraries?

Yes ☐

No ☐

Don't know / not sure ☐

Further comments _____

_____

_____

15. If you have negotiated permission to digitise copyright analogue materials, did you also get permission to copy the resulting digitised material for preservation purposes?

Yes ☐

No ☐

Don't know/not applicable ☐

Further comments _____

_____

_____

16. If you copy born digital resources that you physically own for preservation purposes, do you have to seek permission from the copyright holder(s) to do so?

Yes ☐ *Go to question 17*

No ☐ *Go to question 18*

Don't know/not applicable ☐ *Go to question 18*

Further comments _____

_____

_____

17. Approximately how much time in man-days is spent annually on rights clearance for digital preservation purposes in your library?

_____

_____

18. If you have licensed digital content in your library, which of the following licensing options do you use?

(*Please tick all that apply*)

Shrink-wrap licence ☐

Individual licence agreements with publisher/ aggregator ☐

Licence agreements with publisher/ aggregator based on a model licence ☐

Not applicable ☐ *Go to question 22*

Other (please specify) _____

Further comments _____

_____

_____

19. Is there provision in the licence agreement for access to back files?

(*Please tick all that apply*)

Yes - during subscription period, to all material	□	*Go to question 20*
Yes - during subscription period, to material published during the subscription period	□	*Go to question 20*
Paid access only once subscription ends	□	*Go to question 20*
No access once subscription ends	□	*Go to question 22*
No access to back files	□	*Go to question 22*

Other (please specify) _____

20. How do publishers ensure this access to back files?

(*Please tick all that apply*)

Publisher undertakes to provided remote access	□
Publisher relies on a third party to provide remote access	□
Publisher provides copies of material to the library	□
Don't know	□

Other (please specify) _____

21. Is this access provided satisfactorily?

Yes	□ *Go to question 22*
No	□ *Please give details*

_____

_____

_____

D. SOLUTIONS TO COPYRIGHT AND LICENSING ISSUES IN DIGITAL PRESERVATION

22. Who should be responsible for preserving digital publications?

Authors □

Publishers □

Legal deposit libraries □

Libraries in general □

E-print archives □
Don't know □

Other (please specify) _____

_____

_____

23. Which of the following do you think would help libraries to preserve digital material?

(*Please tick all that apply*)

Changes in library privilege exemptions to copyright legislation □

Provision of digital preservation rights metadata by publishers □

Collective licensing of digital preservation copying □

Individual licences negotiated with publishers aggregators □

Extension of legal deposit to cover digital material □

Don't know □

Other (please specify) _____

_____

_____

You are invited to make any other additional comments here.

_____

_____

_____

_____

_____

_____

Thank you very much for the time you have taken to fill in this questionnaire. Please return the questionnaire in the pre-paid envelope provided by 31st March 2003 to:

Margaret-Mary O'Mahony,
Department of Information Science,
Loughborough University, Loughborough, LE11 3TU, UK.

If you would be willing to be interviewed to discuss these issues further, please contact:

Margaret-Mary O'Mahony
Tel. 01509 223053
Email m.m.omahony@lboro.ac.uk

For further information about the CLDP project please see our Website http://www.lboro.ac.uk/departments/dis/disresearch/CLDP/index.htm

APPENDIX 2

Copyright and Licensing for Digital Preservation

A Research Project funded by the

Arts and Humanities Research Board

Publisher Questionnaire

All information will be treated in confidence.

Please tick the appropriate box(es).

A. PUBLISHING ACTIVITIES

1. What are the main subject areas you publish in?

(*Please tick all that apply*)

STM ☐

Professional ☐

Trade ☐

Education ☐

Other (please specify) _____

_____

_____

2. Do you publish material in digital formats?

Yes ☐ *Go to question 4*

No ☐ *Go to question 3*

3. If **NO**, are you planning to publish in digital formats within the next 12 months?

Yes ☐ *Go to question 5*

No ☐ *Go to question 28*

Don't know ☐ *Go to question 28*

Further comments _____

_____

_____

4. How long have you been publishing digitally?

Less than 12 months ☐

1-3 years ☐

3-5 years ☐

5-10 years ☐

5. How do/will you distribute your digital products?
(*Please tick all that apply*)

Offline, e.g. on CD ROM ☐

Online (direct) ☐

Online (intermediary/aggregator) ☐

Other (*please specify*) _____

_____

6. What types of materials do/will you publish digitally?

(*Please tick all that apply*)

Serials □

Monographs □

Popular magazines □

Textbooks □

Other (*please specify*) _____

_____

_____

7. What format or encoding standards do/will you use ?

Open source standards □
(e.g., plain text files, HTML, XML)

Proprietary standards □
(MS Word, MS Access, Adobe Acrobat)

Both open and proprietary standards □

Other (*please specify*) _____

_____

_____

B. DIGITAL PRESERVATION

***Digital preservation**: Storage, maintenance and access to digital materials over the long-term This may involve one or more digital preservation strategy including technology preservation, technology emulation or digital information migration.*

8. What percentage of the preserved digital copies also have print equivalents?

Less than 5% □

5% - 25% □

25% - 50% □

50% - 75% □

75% - 100% □

9. Which of the following do you carry out?

(*Please tick all that apply*)

Short-term archiving of digital publications □

Long-term preservation of digital publications □

Neither □

Don't know □

10. Do you have a formal long-term digital preservation policy or strategy?

Yes □ *Go to question 12*

No □ *Go to question 11*

If **NO**, is your organisation planning to develop a preservation policy in the next 12 months?

Yes □ *Go to question 12*

No □ *Go to question 16*

Don't know □ *Go to question 16*

Other (*please specify*) _____

11. How long do you/will you preserve material for?

As long as it is commercially valuable □

As long as possible □

Other (*please specify*) _____

Refresh: to copy digital information from one long-term storage medium to another.

Technology Preservation: digital data is stored as a bit stream on a stable digital medium (and refreshed to new media as required) and associated with that object are preserved copies of the original application software, the operating system that this would normally run under and the relevant hardware platform.

Emulation: digital materials are stored in their original format as a bit stream and software and hardware emulators are employed to mimic the behaviour of obsolete hardware platforms and emulate the relevant operating system to allow for access.

Migration: a set of organised tasks designed to achieve the periodic transfer of digital materials from one hardware/software configuration to another, or from one generation of computer technology to a subsequent generation.

12. Do you use any of the following preservation methods?

(*Please tick all that apply*)

Refreshing □

Technology preservation □

Migration □

Emulation □
Other (*please specify*) _____

If you are willing to let us see your preservation strategy or policy statement, please attach it to this questionnaire.

If you ticked any boxes in question 13 , please go to question 14, otherwise go to question 15.

13. How frequently do carry out this/these activity/ies?

Every 12 months or less □

Between 1-5 years □

Every 5 years or more □

14. Can you still access your earliest digital material?

Yes □

No □

C. RIGHTS AND LICENSING ISSUES

15. Do your authors/content creators assign:

All rights □

Publication only rights □

Publication and preservation rights □

Other (*please specify*) _____

_____

_____

16. If you license content from a third party for inclusion in your material, please indicate which of the following you acquire:

Software □

Text □

Images □

Multimedia □

Don't know □

Other (*please specify*) _____

_____

_____

18. Do you licence access to your digital material?

Yes □

No □

Don't know □

19. If you licence access to your digital material, do you use a model licence agreement when licensing your publications?

JISC Model Licence (based on the NESLI licence and the PA/JISC model) □

Liblicense standard agreement □

Don't know □

Other model licence e.g. John Cox Associates (*please specify*) _____

_____

_____

20. If you licence access to your digital material, do you provide access to "back files"?

Yes - during subscription period, to all material □ *Go to question 22*

Yes - during subscription period, to material published during the subscription period □ *Go to question 22*

Paid access only once subscription ends □ *Go to question 22*

No access once subscription ends □ *Go to question 21*

No access to back files □ *Go to question 21*

Other (*please specify*) _____

_____

_____

21. If **No,** do you intend to provide access to "back files" within the next 12 months?

Yes □ *Go to question 22*

No □ *Give reasons if you wish, then go to question 28*

Don't know □ *Go to question 28*

_____

_____

22. How do you / are you planning to, achieve this provision of access?

Will take responsibility for preservation □

Will use a third party for preservation □

Will provide physical "copy" of material to customer □

Don't know □

23. How do you plan to provide access should you merge with another company or cease to publish altogether?

_____

_____

24. Are there any digital titles which you do not hold the **full** rights to make an archival copy?

Yes ☐ *Go to question 25*

No ☐ *Go to question 26*

25. If **YES,** approximately what percentage? _____%

***Digital rights metadata**: a language for expressing the rights, conditions and fees for using digital works*

26. Do you/do you plan to generate rights metadata for your digital materials?

Yes ☐

No ☐

Don't know ☐

27. Would you be willing to provide metadata detailing what copying you will allow for digital preservation purposes by preservation institutions?

Yes ☐

No ☐ *Give reasons if you wish*

Don't know ☐

_____

_____

D. SOLUTIONS TO COPYRIGHT AND LICENSING ISSUES IN DIGITAL PRESERVATION

28. Who should be responsible for the short-term archiving of digital publications?

Authors ☐

Publishers ☐

Legal deposit libraries ☐

Libraries in general ☐

E-print archives ☐

Don't know ☐

Other (*please specify*) _____

_____

_____

29. Who should be responsible for the long-term preservation of digital publications?

Authors ☐

Publishers ☐

Legal deposit libraries ☐

Libraries in general ☐

E-print archives ☐

Don't know ☐

Other (*please specify*) _____

_____

_____

30. Do you participate in the British Library's voluntary deposit scheme for digital publications in the UK?

Yes □ *Go to question 31*

No □ *Give reasons if you wish, then go to question 32*

Don't know □

_____

_____

Web harvesting: *where Web pages are harvested, indexed, and are made available to end-users.*

31. If **YES**, how do you submit your publications?

On a physical storage medium □

Digital transfer (e.g. email attachment, FTP) □

Allow publications to be harvested from the Web □

Other (*please specify*) _____

32. Do you/ would you permit libraries to copy digital material for preservation purposes?

Yes - backup copies □

Yes – refreshment □

Yes – migration □

No □

Don't know □

Other (*please specify*) _____

33. Do you think libraries should be given the legal right (e.g. under copyright legislation) to copy digital work for preservation purposes?

Yes □

No □ *Give reasons if you wish*

Don't know □

_____

_____

_____

You are invited to make any other additional comments here.

_____

_____

_____

_____

_____

_____

Thank you very much for the time you have taken to fill in this questionnaire.

Please return the questionnaire in the pre-paid envelope provided by the 31st March 2003.

If you would be willing to be interviewed to discuss these issues further, please contact:

Margaret-Mary O'Mahony
Tel. 01509 223053
Email m.m.omahony@lboro.ac.uk

For further information about the CLDP project please see our Website
http://www.lboro.ac.uk/departments/dis/disresearch/CLDP/index.htm

APPENDIX 3

Arts and Humanities Research Board Copyright and Licensing for Digital Preservation Research Project

Authors Questionnaire

All information will be treated in confidence.

Please tick the appropriate box(es).

A. GENERAL

1. Is your work published in :

Academic / scholarly publications ☐

Non-academic/scholarly publications ☐

Official/government publications ☐

Other (please specify) _____

2. What subject areas do you publish in? [Please give details]

_____

_____

3. Is any of your work published electronically?

Yes ☐ *Go to question 4*

No – but may in the future ☐ *Go to question 12*

No – no plans to ☐ *Please go to end of questionnaire for details on how to return it.*

4. If **YES**, which of the following best describe the content:

Text only [alpha numeric etc.] □

Text and images □

Multimedia □

Other (please specify) _____

_____

_____

5. Have you ever published your own work electronically?

Yes □ *Go to question 6*

No □ *Go to question 7*

Open source standards permit copies of the original software to be distributed enabling anyone to use or adapt it. ***Proprietary standards*** limit the rights allowing the use or distribution of the software.

6. Do you adhere to any specific format or encoding standards such as:

Please tick the appropriate box(es).

Open source standards □
(e.g., plain text files, HTML, XML)

Proprietary standards □
(e.g. MS Word, Adobe Acrobat, MS Access,
MS Excel, Adobe Photoshop)

Both open and proprietary standards □

Don't know □

Other (please specify) _____

_____

_____

B. ELECTRONIC PUBLICATIONS

7. In what type of e-publication is your work published?

Please tick the appropriate box(es).

e-journals / e-zines □

online databases □

e-books □

personal Websites □

Weblogs □

open archives □

Other (please specify) _____

8. Is your work published:

Please tick the appropriate box(es).

Simultaneously in print and electronically □ Please give % per year _____ *Please go to Q9*

Electronic only □ Please give % per year _____ *Please go to Q10*

Electronic later than print □ Please give % per year _____ *Please go to Q9*

9. If your work is published both in print and electronically, please indicate if the content of the electronic version differs:

Please tick the appropriate box(es).

Not at all from the print version □

A little from the print version □

Considerably from the print version □

Please give details_____

_____

10. For how long has your work been published in electronic/digital format?

Please tick the appropriate box.

Less than 1 year □

1 – 3 years □

3 – 5 years □

More than 5 years □

11. What proportion of your electronic/digital work is still accessible?

Please tick the appropriate box.

Less than 5% □

5% - 25% □

25% - 50% □

50% - 75% □

75% - 100% □

Don't know □

C. PRESERVATION (both published and self published work)

Digital preservation: Storage, maintenance and access to digital objects/materials beyond the limits of media failure or technological change. This may involve one or more digital preservation strategies including technology preservation, technology emulation or digital information migration.

12. How long do you want your work to be available in electronic format?

Please tick the appropriate box.

As long as commercially valuable □

As long as anyone wants access □

As long as possible □

Other (please specify) _____

_____

_____

13. Is a publisher's policy on preservation an important criterion for you when choosing a publisher?

Please tick the appropriate box.

Yes □

No □

N/A □

Further comments _____

_____

14. If you self publish, do you:

Please tick the appropriate box.

Delete earlier versions of your work
and save only most recent version □

Preserve all versions of your work □

Preserve certain versions only □

N/A □

15. Do you notify users if you change URLs, internal or external links or the structure of your electronic/digital material?

Yes □ *Go to Q16*

No □ *Go to Q18*

N/A □ *Go to Q18*

16. If **YES**, how often?

Every 12 months or less □

Between 1-5 years □

Every 5 years or more □

17. How do you notify users of these changes ? _____

_____

_____

Voluntary legal deposit scheme for electronic/digital publications: Under current legislation, printed material published and distributed in the United Kingdom and in the Republic of Ireland must be deposited in the six legal libraries and archives. There is currently no legislation with regard to the legal deposit of electronic/digital materials, however a code of practice exists in the United Kingdom for the voluntary deposit of electronic publications.

18. Who backs up or preserves your electronic/digital work?

Please tick the appropriate box(es).

Yourself □

Employer □

E-print archive □

Library □

Legal deposit library □

Other voluntary deposit scheme □

Publisher □

Other (please specify) _____

_____

Further Comments _____

_____

_____

_____

D. COPYRIGHT & MORAL RIGHTS

Copyright *is the exclusive right to reproduce a work, and also to prevent third parties from copying the work, without prior permission.*

Moral right *is the right of an individual author of a work to be acknowledged as the author or creator. It is the right not to have his or her work subjected to derogatory treatment and the right to refuse to be associated with something he or she did not create.*

19. Do you ever have to ask permission of a third party to have their material included in your work?

Please tick the appropriate box.

Yes ☐ *Go to question 20*

No ☐ *Go to question 21*

20. If **YES,** what permissions do you usually request?

Please tick the appropriate box.

Permission to reproduce material for inclusion in one particular edition of a publication ☐

Permission to reproduce material for inclusion in several editions of a publication ☐

Permission to reproduce material in perpetuity ☐

Permission to reproduce material in the context of the preservation of overall publication ☐

Other(please specify) _____

_____

_____

***Assign**: The rights are sold or given away for good.*

***Licence**: The licensee is granted the right to do certain restricted acts over a specified period of time.*

***Exclusive licence:** the licence is granted to just one person/organisation*

***Non-exclusive licence:** the licence might be granted to several people/organisations*

21. What preservation rights do you assign or licence when you assign or licence rights to your own work?

Please tick the appropriate box(es).

I don't assign or licence any rights at all ☐

Non-exclusive licence *including* a clause which allows copying for preservation ☐

Non-exclusive licence *without* a clause which allows copying for preservation ☐

Exclusive licence *including* a clause which allows copying for preservation ☐

Exclusive licence *without* a clause which allows copying for preservation ☐

Assign some rights including right to copy for preservation ☐

Assign some rights *not including* right to copy for preservation ☐

Assign rights for copying for preservation only ☐

Assign all rights ☐

Don't know ☐

Other(please specify) _____

_____

22. If you have not hitherto licensed/assigned the right to copy for preservation purposes, would you be willing to do so in the future?

Please tick the appropriate box.

Yes □

No □

Don't know □

Further comments _____

_____

23. How important to you is it that your preserved works remain identical to the original electronic versions?

Please tick the appropriate box.

Very important □

Important □

Not important □

Not sure/don't know □

24. What aspects are most important to you?

Please tick the appropriate box, where 5 is most important and 1 is least important.

Look and feel □

Functionality □

Content □

Further comments _____

E. SOLUTIONS TO COPYRIGHT AND LICENSING ISSUES IN DIGITAL PRESERVATION

25. Who do you think should be responsible for preserving electronic/digital materials?

Please tick the appropriate box(es).

Authors ☐

Publishers ☐

Libraries (general) ☐

Legal deposit libraries only ☐

Don't know ☐

Other (please specify) _____

_____

Web harvesting: where Web pages are harvested, indexed, and are made available to end-users.

26. If you self publish, how would you prefer to submit your publications for preservation?

Please tick the appropriate box.

On a physical storage medium ☐

Electronic transfer (e.g. email attachment, FTP) ☐

Allow publications to be harvested from the Web ☐

Other (please specify) _____

_____

_____

***Refresh*:** to copy digital information from one long-term storage medium to another.

***Technology Preservation*:** digital data is stored as a bit stream on a stable digital medium (and refreshed to new media as required) and associated with that object are preserved copies of the original application software, the operating system that this would normally run under and the relevant hardware platform.

***Emulation*:** digital materials are stored in their original format as a bit stream and software and hardware emulators are employed to mimic the behaviour of obsolete hardware platforms and emulate the relevant operating system to allow for access.

***Migration*:** a set of organised tasks designed to achieve the periodic transfer of digital materials from one hardware/software configuration to another, or from one generation of computer technology to a subsequent generation.

27. Do you/would you permit libraries to copy your electronic/digital material for preservation purposes?

Please tick the appropriate box.

Yes - backup copies □

Yes – refresh □

Yes – migration □

No □

Other (please specify)_____

_____

_____

28. Do you think libraries should be given the legal right (e.g. under copyright legislation) to copy electronic/digital work for preservation purposes?

Please tick the appropriate box.

Yes □

No □ *Give reasons if you wish*

_____

_____

Don't know □

Digital rights metadata: is information expressing the rights, conditions and fees for using electronic/digital works.

29. Do you generate rights metadata for your electronic/digital materials?

Please tick the appropriate box.

Yes □

Not at present, but am willing to □

No □ *Give reasons if you wish*

_____

_____

_____

This is the publisher's / someone else's responsibility □

Don't know □

Thank you very much for the time you have taken to fill in this questionnaire.

If you would like to be interviewed to discuss these issues further please contact:

Margaret-Mary O'Mahony
Tel. 01509 223053
Email m.m.omahony@lboro.ac.uk

For further information about the CLDP project please see our Website
http://www.lboro.ac.uk/departments/dis/disresearch/CLDP/index.htm

APPENDIX 4

COVERING LETTER - LIBRARIES

August 10, 2005

Continued access to and preservation of digital content – rights issues

Librarians increasingly need to manage and preserve digital resources. These may be created by digitisation programmes, purchased or licensed from publishers or other content providers.

An issue of concern to librarians and publishers is intellectual property rights. We are investigating whether current copyright legislation and licensed access to digital content threaten the ability of libraries to provide long-term access to that content and to suggest ways in which the problems can be overcome.

We are surveying the views of relevant stakeholders in this issue, including:

- information professionals from all LIS sectors
- publishers and information aggregator
- authors
- reproduction rights organisations.

We would love to hear about your digital preservation requirements and your views on the impact of rights issues on your ability to ensure access to your collections for your users.

We have enclosed a questionnaire and would be grateful if you could return it in the reply paid envelope by 31 March 2003. The questionnaire will also be available via our Web site http://www.lboro.ac.uk/departments/dis/disresearch/CLDP/index.htm . Your comments will be treated in confidence and if later published will be in an anonymised and aggregated form.

The Copyright and Licensing for Digital Preservation project is funded by the Arts and Humanities Research Board. If you would like any more information please contact us at the address above.

Yours sincerely,

Adrienne Muir
Project Director

APPENDIX 5

COVERING LETTER – PUBLISHERS

Direct Line: +44(0)1509 223064
Fax: +44(0)1509 223053
E-mail: a.muir@lboro.ac.uk
url:http://info.lboro.ac.uk/departments/dils/index.htm

August 10, 2005

Continued access to and preservation of digital content – rights issues

Publishers are increasingly making their content available in new and innovative ways by using digital technologies. Libraries are increasingly managing and providing access to digital resources. The preservation of digital resources is an increasingly prominent issue for both publishers and librarians. There are concerns about how to preserve digital resources and who should take responsibility for preservation. An issue of great concern to librarians and publishers is the intellectual property right issues associated with ensuring continued access to digital resources.

We are investigating whether current copyright legislation and licensing of rights in digital content threaten the ability of publishers and libraries to provide long-term access to that content and to suggest ways in which the problems can be overcome.

We are surveying the views of relevant stakeholders in this issue, including:

- authors and creators
- publishers and infomediaries
- all types of library and information services
- reproduction rights organisations.

We want to find out how creators/publishers want their digital content to be preserved, the rights issues involved and how these issues can be managed to the benefit of all stakeholders.

We have enclosed a questionnaire and would be grateful if you could return it in the reply paid envelope by 31 March 2003. The questionnaire will also be available via our Web site http://www.lboro.ac.uk/departments/dis/disresearch/CLDP/index.htm .We appreciate you have many calls on your time, but this is an important issue that will have to be resolved. Your comments will be treated in confidence and if later published will be in an anonymised and aggregated form.

The Copyright and Licensing for Digital Preservation project is funded by the Arts and Humanities Research Board. If you would like any more information please contact us at the address above.

Yours sincerely,

Adrienne Muir
Project Director

APPENDIX 6

INTERVIEW SCHEDULES FOR LIBRARIES

Section 1: Holdings

1. What types of digital materials does the library hold?

2. What were the main reasons for acquiring them?

3. How did you acquire them? (e.g. bought, licensed, donated, deposited)

4. What proportion of them are unique (i.e. not also print copies of them).

Section 2: Preservation

5. Do you have responsibility for preserving any of your digital materials?
   - If so, what type?
   - If so, how long for?

6. Do you have a formal policy for preserving digital materials / preservation strategy?
   - If yes, how was this developed? What does the policy contain?
   - If not, why not? (Not necessary? Not able to? Planning to in future?)

7. Do you currently carry out any preservation activities?

8. Are these done in-house or externally? (Why?)

9. What preservation methods are currently used (e.g. migration, emulation, technology preservation)?
10. If you copy digital resources for preservation purposes, do you have to get permission from copyright holder(s) to do this?

11. Is the right to carry out preservation activities already included in your licences?

12. How difficult / time-consuming is it to arrange permission?

Section 3: Licensing

13. What type of licences do you use for digital materials (e.g. model, individual, shrink-wrapped).

14. Do your licences make provision for perpetual access to back files? If yes, how is this arranged (e.g. files given, still access via publisher Website).

Section 4: Responsibility for preservation

15. Who do you think should have responsibility for preservation (libraries, publishers, other organisations)? Why?

16. How can libraries take responsibility for material that they do not physically own (i.e. material which is licensed)

17. What changes could be made (e.g. in the law, by publishers) which would make it easier for libraries to preserve electronic materials?

18. What else could / should people be doing now to facilitate preservation later (e.g. rights metadata, closer co-operation with publishers, collective licensing).

Section 5: Copyright issues of preservation

19. Do you think that the law will allow libraries to preserve digital publications?

20. What do you think are the intellectual property rights issues of preservation copying (e.g. making multiple copies for redundancy, changing look and feel by migration)

APPENDIX 7

INTERVIEW SCHEDULE FOR PUBLISHERS/VENDORS

Section 1: Publications

The aim of this section is to investigate whether the way the publisher makes their publications available may have implications for preservation in libraries. For example, will libraries have equivalent parallel print versions that can be preserved, how easy would it be for libraries identify and to preserve content, would libraries deal directly with the publisher if access was not available for some reason?

1. Ask about company and publications.

2. Do you have any digital only publications?

3. Where there are parallel print and digital versions of publications, are there significant differences between the two versions?

4. How do you provide access to your digital publications?

5. Do you provide physical copies of any publications and how are these delivered?

6. Where you provide online access to your publications, is this always directly from your own site or do you use an intermediary?

7. Do you provide online only access to publications, or only bundled print and online access?

8. What formats do you use for your content (e.g. use of SGML/XML, PDF, image formats, sound files)?

9. What identification schemes do you use for your content (e.g. ISBN, ISSN, DOI)?

10. Do you use metadata to describe your content?
    - If so, what scheme do you use? (e.g. internal scheme, ONIX)?

Section 2: Licensing for libraries

The aim of this section is to investigate whether digital access models are different from print to assess whether there are any potential preservation issues that might deter libraries from moving to electronic only collections.

11. Do libraries have access to archived material as part of their digital subscription?

12. Do the digital archives cover all material published (e.g. all issues of a journal title)?
    - If not, would you change your policy

13. Do libraries lose all access to material when a subscription is terminated?
    - If yes, what is the thinking behind this – differs from print model?
    - If yes, would you change your policy in response to demand from your customers?
    - If not, what archiving/preservation provisions are included in licences?

14. If you use intermediaries, how much value do they add to your content? (e.g. they may add functionality)

Section 3: Publishers' archiving and preservation activities

The aim of this section is to investigate the publisher's preservation activities. Are publishers carrying out short-term archival activities or are they planning to preserve their material long-term?

Digital preservation: Storage, maintenance and access to digital/electronic materials over the long-term This may involve one or more digital preservation strategies including technology preservation, technology emulation or digital information migration.

15. How long have you been publishing in digital form?

16. How long do you retain all the material published in digital form? (e.g. forever, as long as possible, as long as valuable, etc)?
    - What are the reasons for this?

17. Is all your digital material still accessible?

18. Have you had to migrate earlier material?

19. Do you have a plan for keeping material accessible?

20. Do your agreements with authors allow you to carry out preservation actions, such as migrating material to new formats?

21. If your licence agreements commit you to guaranteeing access to material, how do you do this?
    - Do it yourself – how?
    - Use a third party – details?
    - Give copies of material to libraries? Willing to consider an agreement similar to that of Elsevier Science and the Royal Library of the Netherlands?

Section 4: Views on legal and preservation issues

Digital preservation activities will involve copying and possibly changing material in some way – refreshing, media migration, migration, emulation. The aim of this section is to gauge publisher views on whether copyright and related law allows preservation by libraries and whether it should.

22. Who do you think should be responsible for the long-term preservation of digital publications? (Publishers, libraries, other organisations)
    - If libraries, which – only legal deposit, all libraries
    - If publishers – do you think they are in a position to do this, do you think they will be willing to do this
    - If other organisations, which?

23. Do you participate in the current UK voluntary deposit scheme for digital publications?

24. Do you think UK legal deposit should be extended to cover digital publications?
    - If yes, do you think your digital publications should be included?
    - If yes, would you allow legal deposit libraries to harvest your material from the Web?
    - If no, why not?

Under current UK libraries can make a copy of material for preservation purposes under very limited circumstances.

25. Do you think current law will allow libraries to carry out the copying that will be needed for digital preservation? (e.g. harvesting material, backups, multiple copying to refresh data and media, multiple migrations, creating adaptations through emulation)
    - If not, do you think that libraries should be given the rights to carry out this copying?
    - If no, why not?

- If no, would you be willing and able to grant permission for libraries to copy for preservation purposes?Do you think there is a role for preservation rights metadata?

26. Are you aware of the LOCKSS and JSTOR initiatives and are you/would you be willing to participate in them?
    - If no, why not?

APPENDIX 8

INTERVIEW SCHEDULE FOR LEGAL EXPERTS, PRESERVATION EXPERTS AND REPRESENTATIVES OF REPRODUCTION RIGHTS ORGANISATIONS

Section 1: Technical preservation strategies

1. It is necessary to remove digital publications from their original storage media for preservation purposes? (As recommended by various projects)
   - If yes, what are the practical difficulties likely to be? (E.g. copyright protection measures, encryption, etc)

2. What copying will be required for the maintenance of bitstreams? (E.g. replication, repackaging/frequency)

3. What are the different likely migration options and what are the copyright implications (E.g. changing look and feel, moral rights, recreating interfaces, recording screen shots and operation)

4. What are the different likely emulation options and what are the copyright implications (E.g. refreshing bitstreams, saving third party software, adapting software)

5. What are the differences between the migration on demand and the UCV data preservation approaches?

6. What are the differences between the Rothenberg, CAMiLEON and UVC emulation approaches? (At what level should emulation take place, when should emulators be developed)

7. Are there any other technical approaches to digital preservation?

Section 2: Copyright issues

8. Given the copying implications of the above, will current provisions for preservation copying allow the sort of copying that will be required? (E.g, multiple acts of copying over time, multiple copies for redundancy, "adaptation" of works so they will run in new environments).
    - If not, should copyright legislation be changed and how?
    - Should it only cover material in permanent collections?
    - Should it include harvesting material from the Web?
    - Should all libraries be included or should it only be legal deposit libraries? (prescribed libraries, position of commercial libraries)

9. Do any other provisions of current and future UK rights law impact on preservation copying activities? (E.g. database right, circumvention of technological measures to protect copyright material)?
    - If so, what should be done about this?

Section 3: General and metadata issues

10. Who should be responsible for the long-term preservation of digital information? (E.g. legal deposit libraries, other libraries, publishers)

11. If copyright law does prevent all or some preservation copying, what are the alternatives? (e.g. individual licensing, collective licensing – any roles for reproduction rights organisations or other intermediaries)

12. Is there a role for rights metadata?
    - Is any work being doing in this area?
    - How could provision/creation of rights metadata be managed?

APPENDIX 9

Interviewees

Name	Company / Organisation
Richard Balkwill	PLS
Jens Bammel	PLS
Joanna Cave	DACS
Andrew Charlesworth	Bristol University
Nick Dempsey	EPS Ltd
Chris Dodd	UCE
Alastair Dunning	AHDS
Iain Ferguson	OpenDemocracy
Dick Fletcher	New Media
Steven Hall	Proquest
David Hoole	Nature
Hamish James	AHDS
Stephen Jeffrey	Pearson Education Ltd
Simon Lake	Pearson Education Ltd
Neela Mann	Scholastic Press
Edward Milford	James & James Ltd
Jenny Pickles	Emerald
Stephen Pinfield	University of Nottingham
Judith Sullivan	Patent Office
John Sweeney	British Library
Kathryn Toledano	Emerald
Jan Velterop	BioMed Central
Paul Wheatley	Leeds University
Deborah Woodyard	British Library

APPENDIX 10

Briefing document for seminar, 19th November 2003

Project Background

The Copyright and Licensing for Digital Preservation project is investigating whether and how copyright legislation and licensed access to digital content affect the ability of libraries to provide long-term access to that content, and aims to suggest ways in which any problems can be overcome. It is an eighteen month project, due to finish in March 2004, and is funded by the Arts and Humanities Research Board.

Aims and Objectives

The objectives of the project are to:

- Assess whether UK copyright legislation meets the digital preservation needs of UK national, academic, public and special libraries
- Identify if and how copyright legislation in other countries addresses this issue
- Investigate to what extent licensed access to digital material in libraries takes account of preservation needs and identify examples of best practice in the UK and in other countries
- Investigate how publishers and information providers are planning to achieve the provision of perpetual access to digital material for libraries, including how access to third party information and software is being dealt with
- Make recommendations for amendments to UK legislation if appropriate
- Develop model licences for long-term access if appropriate
- Make recommendations on how legislators, information providers and libraries can work together to ensure long-term access to digital information

Methods

These objectives have been addressed using several methods.

Literature Review

A comprehensive review of the library, law and related literatures was carried out to investigate relevant areas including different approaches to digital preservation and their copying implications, preservation provisions of copyright legislation in the UK and overseas, and issues related to licensed access to digital information.

Questionnaire Surveys

Questionnaires were distributed by post to 1,000 libraries and 600 publishers and also mounted on the project Website. One hundred and sixty-eight libraries (16.8%) and 111 publishers (18.5%) responded. A questionnaire for authors was also mounted on the project Web site, but only ten responses were received.

The questionnaire surveys aimed to obtain an overview of stakeholders' perceptions of and involvement in digital preservation issues. These included current and future digital publications and digital holdings, experience of problems with long-term access to digital material, experience of licensed access, awareness and involvement in digital preservation, and awareness of copyright issues.

Interviews

Twenty in-depth interviews were carried out with librarians, publishers, legal experts, representatives of rights holder organisations and people involved in digital preservation projects.

Summary of findings

The digital preservation problem

Electronic storage media have shorter life expectancies than traditional media. However, the biggest threat to digital information is the obsolescence of the hardware and software needed to use and view it.

42% of the publishers surveyed are already publishing digitally, and they are producing a wide range of materials. Digital publications are also becoming increasingly important to libraries, with 72% of the libraries surveyed currently holding some form of digital content. Some of these publications do not have print equivalents: 20% of publishers surveyed stated that the vast majority of their material is only available electronically, and 52.5% of the libraries surveyed hold some publications in digital format only. The need for digital preservation is perhaps most acute for these publications.

The questionnaire surveys suggest that both librarians and publishers lack awareness of the need for digital preservation. Only 41.8% of the libraries surveyed are currently taking responsibility for the preservation of their digital material, and only 5.8% of libraries and 30.2% of publishers surveyed said that they had a digital preservation policy. 57.3% of libraries are not currently carrying out any preservation activity, although 54.8% of publishers are doing so.

Preservation activities

A number of technical methods for digital preservation are being developed. The following are the main methods currently suggested:

- **Technology preservation** preserves the information in its original form, together with the software and hardware needed to view it. This is seen by experts as a short-term measure at best.
- **Migration** converts information from one format to another, for example from an older version of software to a newer one, or to a standard format. It can happen either periodically or on request.
- **Emulation** uses software to enable a new technology platform to mimic an older one. It aims to preserve the functionality and look and feel of digital objects.
- There are also methods which combine elements of migration and emulation, such as the so-called **Universal Virtual Computer** approach.

The questionnaires and interviews suggest that many libraries and publishers who are currently carrying out preservation activities are only partially addressing the preservation problem, since they are following good practice in storage, redundancy and making backups, but are not addressing digital obsolescence. Another approach used by many publishers is to rely on the use of standard formats, which will probably only delay the problem of digital obsolescence. However, the literature and interviews have shown that some, mainly large, publishers are seeking to set up preservation agreements with National Libraries.

Respondents gave various reasons for not carrying out digital preservation, with both publishers and libraries mentioning the costs involved. Some libraries see technical preservation methods as too difficult, while many publishers do not see a need to preserve their publications when they are no longer commercially valuable. This lack of preservation activity could also result from the lack of awareness of this problem.

Rights issues for digital preservation

All the methods of digital preservation mentioned above would involve activities that are restricted under copyright law, including straight replication to migrate information to new storage media, reformatting, and the development and use of

software to allow material to be accessed on new technological platforms. Awareness of these is generally quite low; people are unsure about them because there is a lack of case law, and people involved in digital preservation research have been concerned not to become test cases.

The **Copyright Designs and Patents Act** (CDPA) of 1988 restricts the reproduction and making of adaptations of copyright works, which include literary, and artistic works and computer software. This protection varies from medium to medium, but typically lasts for 70 years from the creator's death. In addition, the typography and layout of literary, dramatic and musical works are protected for 25 years from publication. New regulations which came into force at the end of October 2003 include a new provision prohibiting the circumvention of technological measures which protect copyright materials. However, libraries will be allowed to circumvent these mechanisms if they copy under an exception for preservation copying

Section 42 of the CDPA allows librarians and archivists to make a copy of any item in their permanent collection in order to preserve or replace that item, or to replace an item in the permanent collection of another library or archive, but only if the item is damaged or lost. In principle, this exception applies to electronic resources. This exception might not apply if a library were copying early in an object's life cycle, because it only applies where it would not be reasonably practicable to purchase another copy. Section 42 only allows a single copy to be made, that is, networked access would not be permitted.

The Copyright and Rights in Databases Regulations 1997, which amended the CDPA, may also affect preservation copying. This applies 'where there has been a substantial investment in obtaining, verifying or presenting the contents of the database' and prevents unauthorised extraction and re-utilisation of such material. This database right only lasts for 15 years, although it can be extended in certain circumstances, and offers different exceptions from copyright law.

It is also not clear whether making multiple copies for the purposes of redundancy or over a period of time would be allowed. One backup copy of software may currently be made, but this would not suffice for long-term preservation.

Moral rights are included in the CDPA. These include the right to be acknowledged as author or creator of a work, and the right not to have work subjected to 'derogatory' treatment. Several interviewees suggested that there would be moral rights issues if the look and feel of an object were changed during preservation. However, a copyright expert advised that these concerns are unfounded, because copying for preservation could hardly be considered to be 'derogatory treatment'.

Of the thirty-three publishers who answered a question on the licensing of intellectual property from third parties, around half license text and images and just over 21% license software. This may affect their ability to grant permissions to libraries for preservation copying, but this is unclear and needs further exploration. Publisher agreements with authors could also affect this, depending on which rights authors assign or license to publishers.

Preservation of licensed material accessed remotely

Just over fifty-two per cent of libraries surveyed have licensed access to some of their digital resources and 48.9% of publishers surveyed license access to some of their material. Libraries have particular concerns about preserving this type of material; they do not physically own the materials, and this affects their ability to preserve them.

One major issue with licensed material is whether libraries have access to back files during their subscriptions. From the publisher questionnaire, 23.5% of publishers provide no access to back files and 14.7% only provide access to these during the subscription period, which is unsatisfactory for many libraries. Equally, some publishers use a 'rolling wall' approach to their back files, which means that older content eventually stops being available, even while a subscription continues.

Libraries also feel that they should still be able to access material to which they have subscribed if they cancel their subscription, as happens with print subscriptions. However, 44% of library respondents have agreements which give them no access to material once their subscriptions have ended. Once they have cancelled their

subscriptions, research has shown that they would like to be given copies of the material to host, though not on CD-ROM; again, only 18.5% of libraries surveyed said that publishers do this, with 64.2% saying that publishers continue to provide remote access. Even if publishers do say they will give libraries their material, there are still preservation issues with this: it is not clear whether they would have the right to copy this material for preservation, and few libraries have yet had experience of whether publisher guarantees are actually met.

Responsibility for long-term preservation

The issue of who should be responsible for long-term preservation is seen as a pressing one. More than one interviewee felt that digital materials would be lost primarily because of the failure to put organisational structures in place, rather than because of the failure to develop technological solutions.

The project has found that there is no consensus about who should be responsible for long-term digital preservation. The most frequent options in the questionnaires were legal deposit libraries (53.8% of libraries and 42.5% of publishers), followed by publishers (31.3% of libraries and 38.8% of publishers), although 28% of libraries and 13.8% of publishers did not know who should be responsible. Some interviewees thought that other organisations should be involved in preservation, mentioning the forthcoming Digital Curation Centre and bodies such as Resource; one interviewee mentioned the need for cross-sectoral working. Several interviewees felt that this issue would be determined by the availability of willingness, funding and technical skills to do so. The new legal deposit bill (see later) may well mean that legal deposit libraries take primary responsibility for preservation; centralising preservation would avoid duplicating effort.

Publishers and libraries will need to work together on this issue, but they will have to overcome their mutual mistrust to do so. Libraries often do not trust publishers with preservation, since many have traditionally not seen preservation as a priority; some publishers do not see the need to preserve their materials long-term. This problem is particularly acute with licensed materials, since libraries do not own copies and have

to rely on publishers for preservation. They are particularly concerned about what would happen if publishers were to merge or cease operation: 15.8% of publishers who answered a question on this stated that no access would be provided and over half could give no firm answer. Equally, many publishers do not want libraries to preserve their material, since they do not want to lose control over it, and they fear losing revenue if legal deposit libraries provide access to their material. However, some publishers think that libraries should preserve their material since they have the enthusiasm and expertise to do so. Despite these concerns, libraries and publishers are already working together with initiatives such as LOCKSS and JSTOR.

Need to raise awareness

There is a widespread lack of awareness about digital obsolescence. Libraries and publishers involved in the project were generally unaware of the problem or had not thought through its implications; some of them assumed the project was about digitisation, since this is something which they have already encountered. Other rights holders such as authors, artists and illustrators may also be concerned about the control of their copyright material, but the poor response from authors suggests that they, too, are largely unaware of the problem.

Awareness of the technical methods of digital preservation was also very low; this is perhaps inevitable since these are still being developed. Even if these methods will only be used by a few specialists, awareness needs to be raised to prevent libraries and publishers relying on inadequate methods of preservation.

Raising awareness of the issues should be a priority, but it is not clear who should be responsible for this, nor how this should be done. There may well be a role in this for the Digital Preservation Coalition or for other national bodies.

How to deal with the rights issues for preservation

Many libraries want to be able to preserve the digital materials which they hold, but if they are to do this, the rights issues will first need to be resolved. Various solutions to this problem have been proposed:

Legal solutions

The project has shown that it is unlikely that current UK law will allow the copying necessary for digital preservation. The new UK Legal Deposit Libraries Act will allow for the extension of legal deposit to electronic publications and for copyright law to be changed to allow legal deposit libraries to copy these for preservation purposes. However, implementation of the Act will require further regulations to be developed. In addition, it would not allow any other libraries to copy for preservation purposes, and library exceptions in copyright law would need to be extended to enable this to happen. Thirty-five per cent of libraries surveyed thought that this would make preservation easier, although more than half of respondents did not know. There was also support for this from publishers: 60% of those surveyed agreed that libraries should be given the legal right to make copies for preservation, with only 19% saying that the law should not be changed.

Licensing

Some interviewees thought that legal solutions to digital preservation would be inadequate and too slow to be useful. They felt that some form of licensing would bring benefits for both libraries and publishers, since licences could state clearly what was and was not allowed, and whether any charges would be made. **Collective licensing** was generally seen as preferable to individual licences, since these were felt to be too time-consuming to administer. However, it was not clear whether this should form part of existing blanket licences, or be administered by a separate body. Some interviewees felt that **individual licensing** would be needed, at least initially, as rights holders would want to know exactly what was being done. However, many respondents felt that this would be too time-consuming to administer. It was not clear

whether publishers would be willing to grant such permission, and some interviewees felt that publishers would not see such requests as a priority.

Other solutions

Other solutions to the rights issues have been proposed. **Metadata** is seen by some as being extremely important: 24.4% of the libraries surveyed thought that the provision of digital preservation rights metadata by publishers would help facilitate preservation, and no publishers said they would refuse to do this. Likewise, one interviewee involved with publishers felt that **digital rights management systems** would provide a solution. However, both of these still need further work: there is currently no standard metadata scheme, and digital rights management systems are still in their infancy.

Issues for discussion groups

LONG-TERM PRESERVATION

How should long-term preservation be carried out and what barriers are currently preventing this?

Who should be responsible for the long-term preservation of digital publications?

- Should it be legal deposit libraries only, should other types of library be involved? Should publishers be involved in long-term preservation?
- Should authors be involved?
- Should other types of organisation? Which?

What are the constraints that would prevent each of these stakeholders from undertaking long-term preservation of digital publications?

Stakeholders:

- Legal deposit libraries
- Other libraries
-
- Publishers
- Authors
- Others?

Constraints, for example:

- Lack of awareness of the need for long-term preservation
- Do not physically own publications
- Lack of resources, and open-ended financial commitments. Lack of expertise
- Do not have the rights to carry out preservation strategies

Any others?

What could be done to remove or minimise constraints on the stakeholders that should undertake long-term preservation?

- How can awareness of the issues be raised further among the stakeholders?
- Is there a role for individual or collective licensing of permissions to preserve?
- How can stakeholders work together?
- Is there a role for a central body to undertake preservation?
- What are the roles of organisations such as the Digital Preservation Coalition, trade and professional bodies? Is there is a role for government and legislators?

What types of material should be prioritised for long-term preservation?

- Can material be categorised and prioritised for preservation. If so, according to which criteria?

- Who should decide which categories of material are worth preserving?

Should long-term preservation be undertaken on a national or even international level?

- Should preservation decisions be taken on a national level, rather than by individual organisations and institutions?
- How could stakeholders work together?
- Is cooperation on an international level desirable or feasible?

CONTINUING ACCESS TO LICENSED CONTENT IN LIBRARIES

How can libraries ensure their users have ongoing access to remotely accessed digital publications?

Should licence agreements include provision for continuing access to remotely accessed material?

- Should licence agreements include provisions for material available during subscriptions to continue to be available when subscriptions end (as in the print environment)?

- Should there be provisions in licence agreements for ongoing access to material subscribed to if publishers merge or fail?

How can guarantees of continuing access to remotely accessed material be implemented?

- What are the implications of “moving walls” of provision of older material for libraries, i.e. if publishers provide access to current material and material from the last, say, ten years only?

- Should libraries pay for extra for access to material predating subscriptions?

- How can publishers implement guarantees of continued access? Should publishers and libraries work together on this.

APPENDIX 11

Seminar Report

Introduction

The Copyright and Licensing for Digital Preservation Project Seminar was held on Wednesday 19th November 2003, at the Policy Studies Institute in London. It was attended by 25 invited delegates, who represented the following stakeholder groups:

- Publishers (including academic, directory, educational, electronic and STM publishers)
- Librarians (including academic and legal deposit libraries)
- Digital preservation experts
- Legal experts
- Reproduction Rights Organisations

The seminar was chaired by Mark Bide of Rightscom.

Opening session

After welcoming delegates to the seminar, Mark Bide made some introductory comments about the issues to be covered:

- It is fundamental that we resolve the issues surrounding giving access to electronic materials which are being preserved. Preservation without access is a sterile activity.

- Digital preservation is like sending messages into the future, to an unknown recipient. We need to ensure that the information is sent in a form that can be received and understood.

- Digital preservation is at the difficult intersection of business, technology and law. We also need to consider the organisational and managerial issues of digital preservation, which are a greater problem than the technology.

- We need to work out who should carry out digital preservation, rather than how they will do it.

- The Legal Deposit Libraries Act 2003 will not bring an end to the legal issues surrounding digital preservation, but will just be the beginning of a solution.

- Copying is essential to preservation. Traditionally, multiple copies were produced and stored, in the hope that some of them would survive. We need to determine what sort of copying is reasonable in the digital environment, i.e. how many copies can be made and in what format.

- Many publishers' business models are now about providing access, not about selling copies. They are therefore unhappy about libraries providing access to the materials which they are preserving, since they would then be competing with libraries to provide access to their materials. Any solution to the issues of digital preservation therefore needs both to preserve the cultural heritage and to protect publishers' business models.

- We need to overcome the mutual lack of trust between the different stakeholders.

Presentation

The main presentation was given by Adrienne Muir (Project Manager) and Charles Oppenheim (Professor of Information Science, Loughborough University).

Adrienne Muir began with a brief overview of the Copyright and Licensing for Digital Preservation Project. The project, which is funded by the Arts and Humanities Research Board, began in September 2002 and is due to end in March 2004. It aims

to '*investigate the impact of copyright legislation and licensed access to digital content on the ability of libraries to provide long-term access to that content, and to suggest ways in which any problems can be overcome'*. The project has used a literature review, questionnaire surveys and interviews to gather data. This seminar is intended to disseminate interim results, and to discuss and recommend solutions to the issues identified. This section of the presentation ended with an outline of the copying requirements of the different digital preservation methods that have been proposed:

Strategies	**"Copying" requirements**
Refreshing bits **Media migration**	• Periodic copying of bit streams from one physical medium to another
Migration	• Content format conversion • Recording and saving information about original software environment • Recreating software environment
Emulation	• Encapsulation of content, original software, specifications, etc • Developing new software to allow original software to be run. Reverse engineering of original software • Development and use of emulation software
Re-creation	• Recreating content and software

Charles Oppenheim then outlined the relevant clauses in UK Copyright law:

- The Copyright, Designs and Patents Act 1988 defines 'restricted acts' and the length of the period of copyright.
- Database right may be relevant to preservation.
- Section 42 of the CDPA is important for preservation, since it allows prescribed libraries and archives to make copies of copyright works for preservation purposes, subject to certain conditions. This clause has not been affected by the new copyright regulations implemented on 31st October 2003.

- Under the new copyright regulations, a lawful user is allowed to bypass "technical measures" in order to enjoy an exception to copyright; this includes the preservation exception in Section 42. However, as explained later, there is a complicated process to go through to do this.
- Moral rights may not be overridden when copies are made under Section 42. This means that libraries may not subject a work to 'derogatory treatment', by making changes which impugn the reputation of the creator.

Adrienne Muir then summarised the implications of copyright law for digital preservation:

- Technical methods for digital preservation will involve copying and will therefore potentially infringe rights under copyright law.
- Section 42 permits some copying for preservation purposes, but it is too limited to be useful for digital preservation purposes

There are particular issues with preserving electronic information which libraries access remotely, under licence. Libraries want to be able to access this for as long as it is useful for their users, but since they do not own physical copies of the information, they are dependent on publishers for ongoing access and long-term preservation.

Three possible solutions to the legal restrictions on preservation copying have been identified:

Legal deposit: The Legal Deposit Libraries Act 2003 will enable legal deposit to be extended to cover non-print publications, excluding film and sound material. However, regulations need to be drafted before this law is implemented and it will be implemented in phases. Access to preserved publications will be limited, and these regulations will only apply to material "published" in the UK.

Changing copyright law: The current exceptions to copyright law could be changed to facilitate preservation copying. Most countries' laws are similar to the UK's in this

respect, or do not contain any specific provision for preservation copying. However, the US and Canadian laws, and the proposed changes to the New Zealand laws contain provisions which could be useful. However, it would probably be slow and difficult to amend the UK law to allow preservation copying by all libraries. Such changes would not help libraries to preserve items which they license and do not own physically.

Licensing: It could be very onerous for an individual library to have to seek a licence for each work it wants to preserve. Collective licensing would be simpler, but there are currently no such schemes in existence. Clauses could also be included in existing licence agreements to provide for preservation; some model licences already contain such clauses.

The Project has investigated stakeholders' views of preservation and current preservation activities. The stakeholders identified are libraries and their users, rights owners and their representatives (publishers, authors/creators, reproduction rights organisations), preservation experts/agencies and legal experts and policy makers.

- A significant proportion of libraries have digital collections. Some libraries are taking responsibility for preserving these, but few have preservation policies. Libraries are generally not clear whether their digitisation agreements allow them to preserve materials that they have digitised.

- It is not clear whether publishers will preserve their materials for the long term. The publisher survey indicates that they are not keen to allow libraries to carry out preservation copying, but that they are not entirely opposed to the law being amended to allow this.

- Both libraries and publishers lack awareness of the legal issues surrounding preservation copying. The preservation activities they are carrying out are not addressing all the issues; some are relying on good storage, which will not protect against technological obsolescence.

- Licence agreements do not always guarantee long-term access to materials. Where licences do contain relevant clauses, few libraries have yet had to test whether publisher guarantees are implemented successfully.

- There is no consensus about who should take responsibility for long-term preservation. The changing roles of libraries and publishers are contributing to tension between these two groups.

Adrienne Muir concluded that:

- *We need to decide what should be preserved and for how long, who should preserve, and how to preserve before legal issues can be considered.*

- *Rights owners and libraries need to work together and work out how to meet each others' concerns.*

Questions and discussion

Delegates asked a number of questions during and after the presentation. The presentation was also followed by a general discussion about the issues raised. The main points made are summarised below, arranged by category.

Legal issues

- Publications to which libraries license access are not considered to be in their 'permanent collections' for the purposes of Section 42, since their licence agreement, and thus their access, could end at some point.

- Technological protection measures can be bypassed in order to 'enjoy' exceptions.
  - A user must first apply to the publisher for permission to do this. If permission is not granted, they can then appeal to Secretary of State. This is an elaborate procedure.

- Theoretically, someone with a legitimate reason for doing so can bypass this procedure by hacking.
- People are technically allowed to produce tools to bypass technical measures, as long as they are only for use by those with a legitimate exception. However, it would be very difficult to prove what these tools were designed and intended for.

- Moral rights only apply in the UK to works produced since 1st August 1999. There is no case law in the UK relating to moral rights with literary works; the cases which have been brought tend to be about artistic works and films.

- The Digital Millennium Copyright Act allows US libraries to make copies of works and lend these to other libraries. These could not be lent to UK libraries, since the copies would infringe UK law.

- Canadian law allows libraries to make a preservation copy in another format, if the format of the original is obsolete. It is not clear whether any preservation action can be taken before a work has actually become obsolete, nor how obsolescence is defined. Ultimately, as with so many copyright related issues, this will only be decided by a judge.

- Preservation is in the public interest so there needs to be framework legislation to ensure it happens. The current Section 42 is too limited. Legislation and / or a code of conduct is needed to set out what publishers must do.

- The government is unlikely to legislate if people can develop licence agreements, because of the work involved.

Licensing issues

- Some publishers have opted out of the archiving and long-term access clauses in the JISC/NESLI model licence.

Preservation issues

- Lifecycle management is important with digital materials. The preservation agency needs to think about how to preserve these at the time of acquisition.

- At present, legal deposit libraries are only accepting electronic-only publications for deposit and preservation. This may change in the future, since print and electronic versions of so-called parallel publications are gradually diverging.

- It is important to decide why something is being preserved, whether as part of the cultural record, or to enable continued access.

- It is difficult to know which of today's publications will still be seen as valuable in a hundred years time. This has always been an issue, but it is more significant for electronic publications, since fewer of these are likely to be preserved because of resource and other issues.

- While a central facility for preservation would be desirable, the UK market is too small for this.

- There will need to be different types of arrangements for preserving different types of materials.

- Preserving dynamic, e-only publications such as databases is particularly difficult, and more modelling is needed of how this could be done. One possibility would be to preserve the underlying dataset. Another possibility would be to preserve snapshots of sample, user-defined outputs, although it is questionable how important these would be to the cultural record.

- Some people feel that computer games should be preserved, since they are an important part of cultural history. However, the British Library has never kept absolutely everything published, and would not see computer games as falling

within its remit. They would be difficult and expensive to preserve, and some might object to public money being spent on this.

Publishers and preservation

- In the project questionnaires, a larger percentage of publishers than libraries claimed to have preservation policies. Publishers who claim to have policies may not actually be carrying out any preservation activities. They may just have thought about this, or may have policies which merely state that 'we are trying not to lose anything'.

- Results from the publisher questionnaire suggest that most publishers can still access all their electronic publications. However, many of them have not been publishing electronically for very long, so problems may only arise later.

- Some publishers apparently do want to preserve their materials, but do not know how to do this. They are not deliberately setting out to remove access to their materials after a fixed period.

- Publishers have varying policies on preservation. This may lead to difficulties if a publisher is taken over.

- Libraries are unwilling to trust publishers with preservation. One publisher was said to have lost a whole year's worth of files from a particular journal when it had taken over another publisher.

- The digitisation of back files and the use of printing on demand mean that old publications may become commercially valuable again, and that they may never technically go out of print. This is leading to new patterns of user behaviour. While libraries traditionally carried out long-term preservation, these changes mean that the traditional value equations have also changed. There is therefore some confusion between preserving the cultural heritage, and preserving things with commercial value. We are currently in a period of transition, and it is not clear how this will develop.

- Similarly, the new business models that are emerging in the publishing industry are about access. Publishers may licence publications for very short periods. This is a legitimate business model, but it is unpopular with users.

Libraries and preservation

- The library questionnaire found that 29% of academic libraries claimed not to have any digital holdings. This seems highly unlikely, and it may be that the person answering the questionnaire either did not know the correct answer or misunderstood the question.

- The Royal Library in the Netherlands was said to be looking to earn money from the digital preservation activities it is carrying out.

Authors and preservation

- If an author assigns all his or her rights to a publisher, it is then the publisher, not the author, who will determine whether an article is still valuable. The author will have no say over this, and may disagree with the publisher's decision.

Mark Bide concluded the morning session by explaining the purpose of the afternoon's breakout groups. Changes to the law were the last thing we could expect to achieve. He requested the groups to focus on actions which could be taken, and to report on what can be done, how, and by whom. They should seek to find a consensus.

Breakout groups

In the afternoon, there were three breakout groups. These were chaired by Maggie Jones, Charles Oppenheim and Anthony Watkinson, who are on the Project's

Advisory Board. The groups all discussed the same questions, which had been sent to delegates in advance (see Appendix).

Final discussion

After the breakout groups, each chair person reported their group's conclusions and recommendations to the other groups. This was followed by general discussion of the points made. Mark Bide concluded the seminar by commenting on some of the recommendations that had been made.

Recommendations

Responsibility for long-term preservation

- Legal deposit libraries should take the primary responsibility for digital preservation. They are well placed to do this since they have traditionally preserved the materials they hold. However, they will not be able to do this on their own, since they have never preserved everything which has been published.

- Other libraries, particularly specialist libraries may play a more limited role in preservation. They may wish to preserve rare or unique materials in their collections, or special collections which they have built up. It would not be sensible for each individual library to carry out its own preservation activities. Consortial arrangements might be beneficial, and would bring economies of scale.

- Publishers should not be expected to take primary responsibility for digital preservation. Preservation is generally low on publishers' agendas, and they do not want to pay for preservation. Libraries do not trust publishers to preserve their own materials, and they fear what may happen if a publisher goes out of business or is taken over.

- Publishers could contribute to preservation in some way. They could make a financial contribution, publish or deposit their materials in forms which are simple to preserve, or supply metadata when depositing their publications.

- Other organisations may have a role to play in preservation. These may include JISC, the Digital Preservation Coalition and the forthcoming Digital Curation Centre, although the exact function of this is not yet certain. The government may have some involvement, although some may oppose government 'interference'. There may also be a role for existing intermediaries such as OCLC and JSTOR, or for new intermediaries which may emerge.

- The technical side of preservation could be outsourced to specialist companies, since libraries and publishers may not have the skills and resources to do this themselves.

- Ultimately, responsibility for digital preservation may be determined by who is willing or able to pay for it. Costing models need to be developed, to discover whether preservation is affordable.

- There needs to be a central body to coordinate preservation. The Digital Preservation Coalition is a suitable candidate to do this, but needs additional resources for this task. It also currently lacks involvement from publishers, and needs the involvement of trade and professional bodies to encourage their members to participate.

Constraints that are currently preventing stakeholders from undertaking long-term preservation of digital publications.

- **Technical issues**: Long term preservation is technically difficult, and reliable technical preservation methods are still being developed. The necessary tools and skills are currently scarce. Preservation activity has been almost entirely project based until now, and larger scale, ongoing activities need to be set up.

- **The scale of the task**: A great number of publications need preserving. These are in a wide range of formats, some of which (e.g. dynamic databases) are difficult to preserve.

- **Legal issues**: The Legal Deposit Libraries Act 2003 will allow legal deposit libraries to carry out preservation copying, but only of UK publications. This law does not apply to other libraries, so these will need licences if they are to copy for preservation.

- **Funding issues**: It is currently not clear who will pay for digital preservation. Stakeholders are either unable or unwilling to pay for preservation and government or other public funding is likely to be inadequate.

- **Lack of will to preserve**: some stakeholders do not currently believe that digital preservation is necessary and worthwhile.

- **Changes in the publishing industry**: For example, publishers are increasingly selling access rather than selling copies; this causes conflict when libraries want to provide access to materials which they are preserving.

- **Changes to the ways libraries operate**: For example, libraries are moving from a 'just in case' set up to a 'just in time' set up. This means that fewer copies are being kept, and that these are being accessed by more people, than with print materials. In the print environment, several copies of each work were kept to ensure that at least one would survive. Electronic publications could be at greater risk if fewer copies are kept.

Recommendations as to how these constraints can be removed or minimised.

- **Legal solutions** are not seen as desirable as they take a long time to be implemented. The regulations relating to the Legal Deposit Libraries Act 2003

are not expected for another 18 months to 2 years and legal deposit libraries want to be able to take action sooner than this.

- **Collective licences** are favoured as a solution to the rights issues of digital preservation, since these are simpler to administer than individual licences. Standard contractual terms would be desirable, but would be hard to achieve, partly because of competition law. Collecting societies could administer these.

- **Funding for digital preservation** could be acquired in a number of ways:
  - Publishers could contribute to the cost of preservation by paying a lump sum, or a fixed percentage of the cost. They could charge lower subscription fees to libraries which are preserving their materials, in return for the services which the library is carrying out on their behalf.
  - Users could be charged for accessing preserved publications on a 'pay-per-view basis. Libraries might feel that charging users would go against their nature, and users might be unhappy at having to pay for access, as it has always been free at the point of use. This could be integrated into license agreements. The charges paid could be divided between the library (to pay for preservation activities), the publisher (as an incentive to deposit its materials for preservation) and the author. It would need to be clear who has to pay, since there could be legal difficulties if libraries charge those who are allowed to view a publication under an exception to copyright law. Good administrative procedures would be needed to enable libraries to operate this system successfully.
  - If a pay-per-view system is adopted for access to preserved publications, some initial investment will still be needed to set up preservation activities.
  - Business models need to be developed for licensed materials so that preservation and continuing access will be funded adequately, even if a library stops subscribing to a publication, or the publisher stops publishing it.

Prioritisation of materials for long-term preservation

- Decisions about which materials should be prioritised should ultimately be made by preserving libraries, and these may reflect collection development policies. However, there should also be input from other stakeholders.

- The following categories should be prioritised:
  - Publications which are in particular demand from users.
  - Publications which are only available in electronic form. Electronic publications with print equivalents should be prioritised if they differ substantially from the print version, or if the original is rare or fragile.
  - Publications which are simple to preserve.
  - Handheld materials rather than online. This is probably because libraries own handheld materials, and only have access to online ones.
  - Libraries do not want to have 'dark archives', although publishers are generally happy with these. Libraries might be reluctant to preserve materials if the publisher did not want them to provide any access to these, although initial embargoes on access are likely.

Preservation on a national or international scale

- Preservation needs to be organised on an international level, so that users can continue to have access to works published in other countries. Access to international publications could be organised by creating a designated centre for these. Alternatively, national libraries could give other libraries access to the materials they hold; the foreign library would need to licence access to these.
- Preservation should be organised on a national or even institutional level initially, since this may be quicker than waiting for international solutions.

Provision of continuing access to remotely accessed digital materials

- Licence agreements should include provision for continuing access to remotely accessed material. Libraries should have perpetual access to the volumes to which they have actually subscribed.
- Libraries should insist that their licences contain such clauses. They should refuse to sign licences which do not contain such clauses, if user needs and demands allow this.

- Publishers need to offer libraries guarantees about preservation. Licences should contain clauses stating what will happen in a 'disaster', or if the publisher ceases to make a publication available itself.
- Publishers should deposit copies of their materials with a national library or with subscribing libraries. This would best be done on publication, with an embargo on access, but could also be done when a publisher goes out of business. An agreement about preservation should be made at the time of depositing. This may not be overridden thereafter, even if the publisher is taken over by a company with a different policy about preservation.

Other recommendations

- Awareness of the issues needs to be raised. The Digital Preservation Coalition should play a role in this, and there need to be publications and presentations.
- The different stakeholders, including authors, need to work together to find solutions to the problems identified. They need to learn to trust one another, and to reconcile their different aims. There need to be committees or groups which include all stakeholders; solutions need to be developed by all groups since a decision developed and imposed by one group will not be accepted by the others.
- The organisational, managerial and administrative processes for digital preservation need to be standardised and made as simple and transparent as possible, so that everyone knows what to do.
- Access to preserved materials needs to be clarified. Preservation agencies see no point in preserving materials if they cannot provide long-term access to them,

while publishers do not want others to provide access to their materials while they are still commercially valuable. This could be resolved by publishers depositing works for long-term preservation, but providing access under licence as at present.

- The authenticity of preserved copies must be ensured; this may be important for authors and publishers. Version control and changes made need to be recorded in metadata, and there need to be standard, transparent processes for this.

APPENDIX 12

Dissemination activities undertaken

ARTICLES

Muir, A., 2003. Copyright and licensing for digital preservation. *Library & Information Update,* 2(6), 34-36.

Muir, A., 2003. Digital preservation : how long will digital content in libraries be accessible? *IMI Insights*, September 2003, 1-3.

Muir, A., 2004. *Journal of Information Science,* 30(1), 69-88.

Ayre, C. & A. Muir, 2004. The right to preserve. *D-Lib Magazine*, March 2004.

Ayre, C. & A. Muir, 2004. Ensuring long-term access to digital information in libraries. *IMI Insights.* (In preparation).

CONFERENCE PAPERS

Muir. A., 2003. Copyright and licensing issues for digital preservation and possible solutions. *In*: *Proceedings of the ICCC/IFIP Seventh International Conference on Electronic Publishing, 25-28 June 2003, Guimarães, Portugal. Universidade do Minho*.

OTHER

Dissemination of project papers on project Web site.

Invitation-only seminar held at the Policy Studies Institute in London on 19th November 2003.

Dissemination of executive summaries of project findings to stakeholder organisations and interested individuals.