



This item was submitted to Loughborough's Institutional Repository (<https://dspace.lboro.ac.uk/>) by the author and is made available under the following Creative Commons Licence conditions.



CC creative commons  
COMMONS DEED

**Attribution-NonCommercial-NoDerivs 2.5**

**You are free:**

- to copy, distribute, display, and perform the work

**Under the following conditions:**

 **Attribution.** You must attribute the work in the manner specified by the author or licensor.

 **Noncommercial.** You may not use this work for commercial purposes.

 **No Derivative Works.** You may not alter, transform, or build upon this work.

- For any reuse or distribution, you must make clear to others the license terms of this work.
- Any of these conditions can be waived if you get permission from the copyright holder.

**Your fair use and other rights are in no way affected by the above.**

This is a human-readable summary of the [Legal Code \(the full license\)](#).

[Disclaimer](#) 

For the full text of this licence, please go to:  
<http://creativecommons.org/licenses/by-nc-nd/2.5/>

**Legal deposit of digital publications**

**by**

**Adrienne Muir**

**A Doctoral Thesis**

**Submitted in partial fulfilment of the requirements for the award of**

**PhD of Loughborough University**

**January 2005**

**© Adrienne Muir 2005**

## **ABSTRACT**

Legal deposit is an obligation to deposit publications with specified depositories. The purpose of legal deposit is to preserve and provide long-term access to the national intellectual heritage. Extending legal deposit to digital publications presents many challenges for the framing of legislation, selection and acquisition of material, preservation and provision of access. The aim of this research was to: explore the potential issues related to the extension of UK legal deposit to digital publications and identify the implications for existing legal deposit arrangements. The research was based on Soft Systems Methodology. Data was gathered through two rounds of interviews with stakeholder groups, including legal deposit libraries, publishers, government and technical experts. Preservation is accepted as the main purpose of legal deposit, but there is some tension and lack of trust between publishers and legal deposit libraries on access to legal deposit collections. The new UK legal deposit law is enabling only; implementation will require further regulations that will be subject to detailed consultation and negotiation. While there has been a voluntary scheme in place for some time in the UK, the research found a lack of readiness amongst the UK legal deposit libraries. They still had to develop cooperative arrangements between themselves and publishers; policies, procedures, especially for online publications; and did not have all the necessary technical infrastructure in place. The deposit of digital publications is an extra role for legal deposit libraries and it is not clear that they will receive extra funding, as is the case in some other countries. There is currently no full-scale fully functional digital legal deposit system in the world. However, there are lessons to be learned from other legal deposit libraries and research and development work is providing partial solutions. The key issues are the need for communication and collaboration between UK legal deposit libraries and cooperation and trust between legal deposit libraries and publishers. Developments since the research was carried out demonstrate some progress in this. Without these, digital legal deposit cannot be successfully implemented in the UK. There is also a need to look at what the users require in terms to material collected and how it is preserved.

# TABLE OF CONTENTS

Abstract .....	ii
Table of Contents.....	iii
Acknowledgements .....	xi
List of Tables .....	xiii
List of Figures .....	xiv
List of Acronyms.....	xv
1. INTRODUCTION.....	1
1.1. UK Legal Deposit Libraries.....	1
1.2. Scope of Legal Deposit in the UK.....	3
1.3. The Advent of Digital Publishing.....	4
1.4. The Push Towards Extension of Legal Deposit .....	4
1.5. The Research Problem .....	5
1.6. The Structure of the Thesis .....	7
1.7. The Significance of the Research .....	8
2. THE HISTORY AND DEVELOPMENT OF LEGAL DEPOSIT .....	9
2.1. Beginnings .....	9
2.2. Purpose: Present and Past .....	9
2.3. Overview of Legal Deposit Systems.....	12
2.4. History of Legal Deposit in the United Kingdom .....	13
2.4.1. Origins in the United Kingdom.....	15
2.4.2. The 1911 Copyright Act .....	17
2.5. Extension of Legal Deposit in the UK.....	23
2.5.1. The BL Proposal and the Government’s Response.....	23
2.5.2. The Kenny Working Party.....	26
2.5.3. The Voluntary Code of Practice for Non-Print Publications .....	29

2.5.4.	Waiting for Legislation .....	31
2.5.5.	The New Legislation.....	32
2.6.	Scope of Legal Deposit in Other Countries .....	44
2.7.	Guidelines for Legal Deposit Legislation .....	46
2.8.	Summary.....	49
3.	Practical Issues .....	51
3.1.	Definitions: Documents, Publications and Publishing .....	51
3.2.	Discovery, Selection, Acquisition .....	56
3.2.1.	Discovery.....	56
3.2.2.	Selection .....	59
3.2.3.	Acquisition .....	61
3.3.	Accession, Processing and Storage .....	63
3.3.1.	Accession and Cataloguing .....	63
3.3.2.	Storage .....	66
3.3.3.	Preservation Metadata.....	67
3.4.	Preservation .....	68
3.4.1.	Media Stability .....	69
3.4.2.	Technological Obsolescence .....	69
3.4.3.	What to Preserve .....	70
3.4.4.	Authenticity .....	72
3.4.5.	Preservation Strategies.....	74
3.4.6.	Disasters and Rescue of Digital Information .....	83
3.5.	Access .....	84
3.5.1.	Restricting Access.....	84
3.5.2.	Use of Legal Deposit Material.....	85
3.5.3.	Networked Access.....	86
3.6.	Intellectual Property Rights and other Legal Issues .....	86

3.6.1.	Intellectual Property Rights .....	86
3.6.2.	Other Legal Issues .....	89
3.7.	Costs of Legal Deposit and Digital Preservation .....	90
3.7.1.	The 1995 British Library Cost Estimates.....	90
3.7.2.	The DNH Draft Cost Compliance Assessment .....	93
3.7.3.	The 2002 Cost Estimates .....	95
3.7.4.	Cost of Digital Preservation.....	99
3.8.	Summary.....	103
4.	MANAGING DIGITAL LEGAL DEPOSIT COLLECTIONS .....	104
4.1.	Digital Preservation Policies.....	104
4.1.1.	Digital Information Life Cycles .....	104
4.1.2.	Categorisation for Deposit and Preservation .....	105
4.1.3.	Risk Management of Digital Material.....	107
4.1.4.	Assessment of Preservation Strategies.....	111
4.1.5.	Assessment of Costs.....	113
4.2.	Open Archival Information System and its Implementation .....	113
4.2.1.	OAIS Components.....	114
4.2.2.	Towards Implementing the OAIS .....	118
4.2.3.	“Persistent” Archives .....	119
4.3.	Preservation Metadata.....	119
4.3.1.	Preservation Metadata Schema .....	120
4.3.2.	Provision and Management of Preservation Metadata.....	121
4.3.3.	The Role of Persistent Identifiers .....	125
4.4.	Infrastructure.....	127
4.4.1.	Repositories for Metadata and Documentation .....	128
4.4.2.	“Trusted” Digital Archives.....	131
4.4.3.	Stakeholders and Roles.....	132

4.4.4.	Incentives to Preserve .....	134
4.5.	Summary.....	135
5.	RESEARCH, DEVELOPMENT AND INITIATIVES .....	137
5.1.	United States.....	137
5.1.1.	Task Force on Digital Archiving.....	137
5.1.2.	Research Libraries Group and OCLC Joint Working .....	138
5.1.3.	National Digital Information Infrastructure and Preservation Program.....	139
5.2.	The United Kingdom .....	143
5.2.1.	JISC/NPO Studies and the Warwick Meetings .....	143
5.2.2.	The CURL Exemplars in Digital Archives Project .....	147
5.2.3.	The Digital Preservation Coalition.....	148
5.2.4.	Digital Curation Centre.....	149
5.2.5.	United Kingdom Web Archiving Consortium (UKWAC).....	150
5.3.	Australia.....	150
5.3.1.	National Leadership .....	151
5.3.2.	PADI (Preserving Access to Digital Information) .....	151
5.3.3.	Safekeeping Project.....	152
5.3.4.	International Participation.....	153
5.4.	Pan-European Initiatives .....	153
5.4.1.	BIBLINK Project .....	153
5.4.2.	Electronic Deposit (ELDEP) Study .....	156
5.4.3.	Networked Electronic Deposit Library (NEDLIB).....	158
5.4.4.	The European Library .....	161
5.4.5.	The Proposed PATCH Project .....	162
5.4.6.	PRESTO Space .....	162
5.5.	International Web Archiving Activities .....	163
5.5.1.	Internet Archive.....	163

5.5.2.	International Internet Preservation Consortium.....	164
5.5.3.	Nordic Web Archive .....	164
5.6.	Initiatives in Individual National Libraries .....	165
5.6.1.	Koninklijke Bibliotheek, Netherlands.....	165
5.6.2.	National Library of Australia.....	171
5.6.3.	National Library of Canada .....	176
5.6.4.	Library of Congress .....	177
5.6.5.	Nasjonalbiblioteket, Norway .....	177
5.7.	Summary.....	178
6.	METHODS .....	180
6.1.	Aim and Objectives.....	180
6.2.	Literature Review .....	181
6.2.1.	Summary of Search Tools Used.....	181
6.2.2.	Searches .....	182
6.3.	Development of the Research.....	183
6.3.1.	Research Problem .....	183
6.3.2.	Methodology .....	186
6.4.	Application of Soft Systems Methodology.....	189
6.4.1.	Limitations of the Research .....	190
6.4.2.	Interviews and Selection of Subjects.....	191
6.4.3.	Cultural Analysis.....	192
6.4.4.	Logical Analysis.....	193
6.4.5.	Analysis of the Intervention.....	194
6.5.	Data Analysis.....	196
6.6.	Summary.....	197
7.	CULTURAL ANALYSIS .....	198
7.1.	Definition and Purpose of Legal Deposit.....	198



7.2.	Alternative Mechanisms for Building a National Archive.....	201
7.3.	Stakeholders.....	203
7.3.1.	Legal Deposit Libraries.....	203
7.3.2.	Publishers and Authors .....	205
7.3.3.	Users and Other Libraries .....	207
7.3.4.	Government.....	208
7.4.	Relationships Between Stakeholders.....	211
7.4.1.	Deposit Libraries and their Sponsoring Departments .....	211
7.4.2.	Between the Legal Deposit Libraries .....	213
7.4.3.	Relationships Between Legal Deposit Libraries and Publishers .....	220
7.4.4.	Joint Committee on Voluntary Deposit .....	221
7.4.5.	Legal Deposit Libraries, the Research Community and Publishers ..	225
7.5.	Extending Legal Deposit Legislation.....	225
7.6.	New Players.....	228
7.7.	Summary.....	228
8.	LOGICAL ANALYSIS .....	230
8.1.	Legal Deposit – the Real World .....	230
8.2.	Developing the Conceptual Model.....	230
8.3.	Create Publications.....	235
8.3.1.	Print Publishing .....	235
8.3.2.	Definitions .....	235
8.3.3.	Lack of Standards and Technological Dynamism.....	237
8.3.4.	Formats and Distribution Channels.....	238
8.3.5.	Create Metadata.....	241
8.4.	Discover Publishers and Publications .....	243
8.4.1.	Identifiers.....	246
8.5.	Select Publications .....	248

8.5.1.	Selection Criteria.....	250
8.6.	Assemble Publications .....	252
8.6.1.	Number of Copies Deposited and Shared Acquisition .....	252
8.6.2.	Who Should Deposit What.....	255
8.6.3.	Lack of Established Processes and Procedures .....	257
8.6.4.	Provision of Publications .....	258
8.6.5.	Provision of Metadata .....	260
8.6.6.	Processing Publications .....	261
8.7.	Store Publications.....	264
8.8.	Preserve publications.....	265
8.8.1.	Redundancy .....	268
8.8.2.	What to Preserve .....	269
8.8.3.	How to Preserve Publications.....	271
8.8.4.	Rights Issues .....	272
8.8.5.	Identifiers and Preservation Metadata.....	274
8.8.6.	International Responsibility for Digital Preservation .....	276
8.9.	Provide Access .....	278
8.9.1.	Networked Access.....	278
8.9.2.	Current Access to Material .....	279
8.9.3.	Access Restrictions .....	280
8.10.	Looking Forward.....	283
8.10.1.	Skills .....	283
8.10.2.	Preparing for Digital Legal Deposit .....	284
8.10.3.	Costs and Financial Resources .....	285
8.11.	Summary.....	287
9.	DISCUSSION.....	288
9.1.	The Extension of Legal Deposit to Digital Publications .....	288

9.2.	Readiness for Digital Legal Deposit.....	291
9.3.	Organising the Digital Legal Deposit System .....	293
9.4.	Managing Digital Legal Deposit Collections .....	295
9.4.1.	Identification of Digital Publications .....	296
9.4.2.	Selection .....	296
9.4.3.	Preservation .....	297
9.4.4.	Metadata .....	298
9.4.5.	Access .....	299
9.4.6.	Costs.....	300
9.4.7.	Legal Deposit in the UK: the Wider Context .....	301
9.5.	Overview.....	302
10.	CONCLUSIONS AND RECOMMENDATIONS.....	304
10.1.	Aims and Impact of the Research.....	304
10.2.	Major Issues and Recommendations.....	305
10.2.1.	Scope of Legal Deposit.....	306
10.2.2.	Deposit Libraries and Publishers: Cooperation and Communication.....	307
10.2.3.	Communication and Preparedness in Deposit Libraries .....	308
10.2.4.	Increased Funding .....	309
10.3.	Recommendations for Future Research .....	309
10.3.1.	Review of Legal Deposit of Print Publications .....	310
10.3.2.	Review of Digital Legal Deposit.....	310
10.3.3.	Sustainable and Integrated Research on Preservation .....	310
10.3.4.	Sustainability, Cost and Funding Models .....	311
10.3.5.	Creation and Provision of Preservation Metadata .....	312
10.3.6.	Legal Issues .....	312
10.4.	Concluding Remarks.....	313
	References.....	314

Other Works Consulted.....	353
Appendices.....	362

## ACKNOWLEDGEMENTS

I am extremely grateful to everyone who participated in this research for their time, hospitality and for making the interviews so enjoyable. Eric Davies, my supervisor, kept me motivated and did everything he could to make life easier for me. LISU colleagues also looked after me, kindly providing me with tea and biscuits whilst Eric was doing his stuff. I am also grateful to colleagues in the Department of Information Science of Loughborough University for practical and moral support, in particular Cliff McKnight and Shirley Horner. I am also grateful to my heads of department for giving me time and space. The research was initially funded by a studentship from the Faculty of Science of Loughborough University, then by the Department of Information Science. Thanks to my friends for being so patient for so long, particularly John and the Edinburgh girls. Charles inspired me to do a PhD and put up with me and fed me while I was doing it. This thesis is dedicated to the Muir girls, with love. What would I have done without the tea, cakes, retail therapy and running?

## LIST OF TABLES

Table 2-1 Main Developments in UK Legal Deposit Law Until 1842.....	14
Table 2-2 Development of Legal Deposit in the UK, 1911-2003 .....	18
Table 2-3 New UK Legal Deposit Legislation: Parliamentary Process .....	33
Table 3-1 Physical, Logical and Conceptual Digital Objects.....	70
Table 3-2 Copying implications of digital preservation strategies.....	87
Table 3-3 Estimated costs to UK-based legal deposit libraries of deposit of CD-ROMs...	94
Table 3-4 Estimated cost of compliance for deposit of CD-ROMs .....	94
Table 4-1 Digital resource lifecycle .....	105
Table 4-2 Human/toolbox interaction in VRC organisational stages .....	110

## LIST OF FIGURES

Figure 4-1 The Open Archival Information System Reference Model .....	115
Figure 5-1 BIBLINK Work Space .....	155
Figure 6-1 Research Model.....	185
Figure 6-2 Soft Systems Methodology .....	188
Figure 8-1 Model of UK Legal Deposit System .....	233
Figure 8-2 Human Activity Model of a National Archive of Digital Publications.....	234

## LIST OF ACRONYMS

AIHT	Archive Ingest and Handling Test
AIP	Archival Information Package
ALPSP	Association of Learned and Professional Society Publishers
API	Application Programming Interface
ARK	Archival Resource Key
ASCII	American Standard Code for Information Interchange
BFI	British Film Institute
BIC	Book Industry Communication
BICI	Book Item and Contribution Identifier
BL	The British Library
BNB	British National Bibliography
BNF	Bibliothèque nationale de France
CAMiLEON	Creative Archiving at Michigan and Leeds Emulating the Old on the New
CCLRC	Council for the Central Laboratory of the Research Councils
CDNL	Conference of Directors of National Libraries
CD-R	CD-Recordable
CD-ROM	Compact Disc-Read Only Memory
CEDARS	CURL Exemplars in Digital Archives
CENL	Conference of European National Libraries
CILIP	Chartered Institute of Library and Information Professionals
CLA	Copyright Libraries Agency This should not be confused with the Copyright Licensing Agency
CLIR	Council on Library and Information Resources
CNRI	Corporation for National Research Initiatives
CoBRA	Computerised Bibliographic Records Action
CPA	Commission on Preservation and Access
CUL	Cambridge University Library
CURL	Consortium of Research Libraries
DAWG	Digital Archiving Working Group
DCC	Digital Curation Centre
DCF	Digital Content Forum
DCMA	Digital Millennium Copyright Act
DCMI	Dublin Core Metadata Initiative
DCMS	Department for Culture, Media and Sport
DFES	Department for Education and Skills
DIAS	Digital Information Archiving System
DIP	Dissemination Information Package



DNEP	Deposit of Netherlands Electronic Publications
DNH	Department of National Heritage
DOI	Digital Object Identifier
DOM	Digital Object Management system
DPA	Directory Publishers Association
DPC	Digital Preservation Coalition
DSEP	Deposit System for Electronic Publications
DTD	Document Type Definition
DTI	Department of Trade and Industry
EAN	European Article Numbering
EBLIDA	European Bureau of Library, Information and Documentation Associations
ELDEP	Electronic Deposit
eLib	Electronic Libraries Programme
EPPP	Electronic Publications Pilot Project
EPS	Electronic Publishing Services Ltd
EPSRC	Engineering and Physical Sciences Research Council
ERPANET	Electronic Resource Preservation and Access Network
FEP	Federation of European Publishers
FTP	File Transfer Protocol
HEFCE	Higher Education Funding Council for England
HMSO	Her Majesty's Stationery Office
ICSTI	International Council for Scientific and Technical Information
IETF	Internet Engineering Task Force
IFLA	International Federation of Library Associations and Institutions
IIPC	International Internet Preservation Consortium
INDECS	Interoperability of Data in E-Commerce Systems
ISBN	International Standard Book Number
ISO	International Standardization Organisation
ISSN	International Standard Serial Number
JCLD	Joint Committee on Legal Deposit
JCVD	Joint Committee on Voluntary Deposit
JISC	Joint Information Systems Committee
KB	Koninklijke Bibliotheek, Netherlands
LISA	Library and Information Abstracts
MARC	Machine Readable Cataloguing
METS	Metadata Encoding & Transmission Standard
MINERVA	Mapping the INternet Electronic Resources Virtual Archive
NAA	Name Assigning Authority
NAAN	Name Assigning Authority Number
NARA	National Archives and Records Administration

NBN	National Bibliography Number
NCBI	National Center for Biotechnology Information
NCL	National Central Library
NDIIP	National Digital Information Infrastructure and Preservation Program
NEDLIB	Networked European Deposit LIBrary
NIST	National Institute of Standards and Technology
NLA	National Library of Australia
NLC	National Library of Canada
NLLST	National Lending Library for Science and Technology
NLM	National Library of Medicine
NLS	National Library of Scotland
NLW	National Library of Wales
NMAH	Name Mapping Authority Hostport
NORDINFO	Nordic Council for Scientific Information
NPO	National Preservation Office
NSF	National Science Foundation
NWA	Nordic Web Archive
OAI	Open Archives Initiative
OAIS	Open Archival Information System
OCLC	Online Computer Library Center
ONIX	Online Information eXchange
OSTI	Office of Scientific and Technical Information
PA	Publishers Association
PADI	Preserving Access to Digital Information
PALS	Publisher and Library/Learning Solutions
PANDAS	PANDORA Archiving System
PANDORA	Preserving and Accessing Networked Documentary Resources of Australia
PARADIGMA	Preservation of Digital Material
PATCH	Permanent Access Toolbox for Digital Cultural Heritage
PDF	Portable Document Format
PDI	Preservation Description Information
PREMIS	PREservation Metadata Implementation Strategies
PRESTO	Preservation Technology for European Broadcast Archives
PRISM	Preservation Reliability Interoperability Security Metadata
PURL	Persistent Uniform Resource Locator
RLG	Research Libraries Group
SGML	Standard Generalized Markup Language
SICI	Serials Item and Contribution Identifier
SIMONE	Simple ONIX Editing Tool
SIP	Submission Information Package

SSM	Soft Systems Methodology
TCD	Trinity College Dublin
TEL	The Electronic Library
TOM	Typed Object Model
UK	United Kingdom of Great Britain and Northern Ireland
UKOLN	UK Office of Library and Information Networking
UKWAC	United Kingdom Web Archiving Consortium
UKOP	U K Official Publications
UNESCO	United Nations Educational, Scientific and Cultural Organization
UPC	Unique Product Code
UPF	Universal Preservation Format
URI	Uniform Resource Identifier
URL	Uniform (previously Universal) Resource Locator
URN	Uniform Resource Name
UVC	Universal Virtual Computer
XML	EXtensible Markup Language

# 1. INTRODUCTION

Legal deposit is a statutory obligation, usually on publishers, to deliver copies of all material published in a country to designated depositories, such as national libraries. The legal depositories then act as a national archive of the nation's published output, preserving and providing limited access to this material. A sixteenth century French King is widely credited with initiating legal deposit through a wish to enrich his library while exercising control over the written word. The origin of legal deposit in the United Kingdom of Great Britain and Northern Ireland (UK), is a seventeenth century private agreement between a publisher body and an English university. Legal deposit proper in the UK started in the early eighteenth century and was part of the first copyright system in the world.

While legal deposit may have had its origins in the control of intellectual output, there was also a notion that the intrinsic value of this output makes it worth collecting and keeping. The private agreement in seventeenth century England involved a publisher organisation and an academic library and the purpose was to provide access to the collection for scholars. While the main purpose of legal deposit today is "the preservation of the nation's cultural heritage." (Jasion 1991, p.7), the purpose of preservation is to ensure that it is available for use for posterity.

Currently, seven libraries benefit from UK legal deposit provisions: three national libraries, three university libraries and a private professional library.

## 1.1. UK Legal Deposit Libraries

The Bodleian Library (Bodley's Library) of the University of Oxford opened in 1602 and is the University's main research library. The Bodleian has benefited from deposit of publications since 1610. Sir Thomas Bodley, the Library's founder, negotiated with the Company of Stationers in London to send a copy of every new book printed by its members to be added to the Library's collections.

A library has probably existed in the University of Cambridge since the mid-fourteenth century. Now Cambridge University Library (CUL) is centrally administered by the

University, although it is made up of five separate libraries. The CUL became a beneficiary of statutory deposit in 1662, when an act designed to regulate the press in England stipulated that copies of all publication be deposited in three libraries: the Bodleian Library, the library of Cambridge University and the Royal Library.

The British Library (BL), as its title suggests, is the national library of the United Kingdom. The Library was formed in 1973 and it brought together several existing library collections and services, including legal deposit collections inherited from one of its predecessor bodies. The constituent parts of the new Library included the library collections of the British Museum, the Patent Office Library, the National Central Library (NCL) and the National Lending Library for Science and Technology (NLLST). Later additions were:

- 1974 - the British National Bibliography (BNB) and the Office of Scientific and Technical Information (OSTI)
- 1982 - the Foreign and Commonwealth Office's India Office Library and Records
- 1983 - the British Institute of Recorded Sound.

Until recently, the BL had premises scattered around London and also in the north east of England. The long-awaited new St Pancras building consolidated the operation to just three locations: St Pancras and the Colindale newspaper library in London and the former NLLST complex in Boston Spa in Yorkshire. The reference collections, reading rooms and some supporting functions are based in London and the national bibliography, document supply and acquisition and processing functions are based in Boston Spa.

The National Library of Scotland (NLS) was formed in 1926. At that time, the Library of the Faculty of Advocates passed over its non-legal collections to the nation. The 'Advocates Library' had benefited from legal deposit since 1709, shortly after the Union of the Scottish and English parliaments in 1707. This entitlement also passed over to the NLS, although the Advocates Library shares this entitlement with the NLS for some legal material. Although located in the Republic of Ireland, Trinity College Dublin (TCD) can claim copies of publications under UK legal deposit law. This arrangement is a historical one: it has been in place since the union of Great Britain and Ireland 1801 and has survived

the secession of the Irish Free State in the early twentieth century. This arrangement is a reciprocal one with UK legal deposit libraries receiving publications under Irish legal deposit regulations. While Northern Irish publications are subject to UK legal deposit, there is no legal deposit library in Northern Ireland.

The National Library of Wales (NLW) became a legal deposit library in 1911, shortly after its foundation in 1907. At first, its entitlement was limited compared to the other libraries, but this has since been rectified. Whilst the BL is entitled to automatically receive a copy of all new UK publications, the other libraries are entitled to claim them.

## **1.2. Scope of Legal Deposit in the UK**

The UK has lagged behind some other countries in updating its legal deposit law to take into account new developments in the publishing industry. At the time of writing, UK legal deposit law only applies to print material, although a new law *enabling* future expansion of the law to non-print material has been in force since the beginning of 2004. For the first time in three hundred years, legal deposit law has been separated from copyright law. The new law replaces legislation that was in force for over ninety years. During that time, extension of the law to categories of material other than print have been considered, and rejected several times.

UK publishers have always had mixed attitudes to legal deposit. There have always been complaints that the provision of free copies of material to libraries is commercially burdensome. Legal deposit has been described as a tax on knowledge and publishers have questioned why their industry is penalised in this way. The same arguments against legal deposit have been used over the centuries and compliance with the law was patchy until the nineteenth century. The counter arguments are that legal deposit is a public good. The building and maintenance of a national archive supports research and scholarship, the output of which becomes the raw material of the publishing industry. The burden of legal deposit is not borne solely by the publishers, the State also contributes through the funding of preservation of and provision of organised access to recorded knowledge by legal deposit libraries.

### **1.3. The Advent of Digital Publishing**

Legal deposit originally applied to print on paper publications. However, the publishing industry has evolved over time and so-called ‘non-print’ publications have become an increasingly prominent part of recorded intellectual output. This non-print material includes:

- *microforms*, particularly microfilm and microfiche, and
- *audiovisual material*, for example cine film and video cassettes

The mid-twentieth century saw the advent of electronic or digital publishing. This began with the use of information and computer technology to store and manipulate large amounts of data, including bibliographic data, and to make information accessible across computer networks. Initially digital publishing was limited to textual and numeric data, but increasingly it has incorporated sound and still and moving images, so-called ‘multimedia’ publications. New recording media, such as optical discs, have increased storage capacities. Software has improved retrieval, presentation and ability to interact with information. The development of browser software and the World Wide Web has caused a publishing explosion. A recent trend has been the Open Access movement that has the potential to change profoundly the nature of scholarly publishing.

### **1.4. The Push Towards Extension of Legal Deposit**

Real progress in extending the scope of UK legal deposit to non-print publications has only really been made in the last ten years. In January 1996, the British Library, supported by the UK legal deposit libraries, Trinity College Dublin and the British Film Institute, put forward proposals to the then Department of National Heritage (DNH) for the legal deposit of non-print publications, including digital material. The Library argued that this extension was needed to build a comprehensive national archive of publications. The rationale was that non-print material also forms part of the nation’s intellectual heritage and the UK’s published output is increasingly made available in formats other than print. The UK Government responded by setting up a Working Group to examine the case for legislation. The Group confirmed that deposit of non-print material should be on a statutory basis. The

Government then confirmed its commitment to legislation and asked the Group to develop an interim voluntary scheme. This has now been operating for nearly five years. The new legislation that came into force at the beginning of 2004 is merely a first step towards implementation of legal deposit for non-print material. Implementation will be in stages, most likely starting with 'easier' categories of material, such as microforms followed by hand-held digital publications, with online material coming later. The new legislation completely excludes film and sound material altogether.

## **1.5. The Research Problem**

Legal deposit was originally concerned with the physical deposit of copies of mass-produced discrete units of static information that were made available within individual nations. While this model may still work reasonably well for some non-print material, such as microforms, it starts to break down when it comes to digital information for several reasons. Concepts such as 'publication', 'publisher', 'published in the UK', 'deposit' and 'copy' can change or lose their traditional meanings in the digital environment. This then causes difficulties for framing legal deposit law. Digital publications, because of their relative novelty, are in many respects an unknown quantity. The deposit, long-term management and provision of access to digital material are also problematic. They pose especially complex challenges such as:

- *technical and preservation issues* – given the variety of formats and media used, there are questions about the form in which material should be deposited, how online material can be 'deposited', how the obsolescence of hardware and software platforms is to be managed and how continuing access is to be achieved over time
- *legal issues* - digital and networking technology raise questions of rights and rights management. Acquiring, preserving and providing access are likely to involve actions that are prohibited under copyright law. The publishing industry has traditionally acted as a filter for low-quality, illegal or improper material and this filtering mechanism is increasingly breaking down as creators can by-pass the traditional system.
- *economic issues* – it is not clear how much it will cost to maintain a UK digital archive or how it will be funded. There is no sign of a slowdown in the print publishing industry in the UK, so there will be an additional cost. Given the potential ease with



which digital information can be copied and manipulated, the commercial interests of rights holders may need protection

- *management issues* – how can publishers and libraries organise themselves for the legal deposit of digital material and is there still a role for legal deposit in the digital environment?

When this research idea was conceived in the summer of 1999, there was already some work on identifying the potential issues related to the deposit of digital publications, but there were few answers to the questions raised. Some legal deposit libraries in other countries were experimenting with digital publications, but there were (and still are) no full-scale, fully-functional digital legal deposit systems in place. However, the UK legal deposit libraries were actively campaigning for extension of legal deposit law, so it was important that the practical implications for the legal deposit libraries and the publishers were addressed. It was already clear that libraries and publishers would have to work more closely together than ever before and that publishers were concerned about the commercial implications of digital legal deposit. There was a clear need to carry out research in this area and so the research was timely.

The aim of the research was to explore the potential issues that would arise from the extension of legal deposit in the United Kingdom to digital publications in order to identify the implications for the existing legal deposit arrangements. The specific objectives were:

- To review the development of legal and voluntary deposit systems, in the UK and other countries
- To examine the current voluntary and legal deposit arrangements for printed material in the UK
- To assess proposals made for the extension of legal deposit to digital publications and progress made on their implementation
- To identify countries that already had, or were planning to, introduce digital voluntary or legal deposit and investigate their plans and arrangements
- To identify and explore the potential issues arising from the extension of legal deposit to digital publications

- To identify all the stakeholders who will be affected by digital legal deposit in the UK and evaluate their interests and concerns
- To assess preparedness and attitudes of stakeholders who will be affected by digital legal deposit
- To identify the management implications for legal deposit of digital publications
- To suggest changes to the current legal deposit arrangements where necessary.

In the end, the timing of the research fitted very neatly with the move towards extending UK legal deposit legislation. The research was carried out as the libraries and publishers were actively thinking through the issues and exploring practical issues through a voluntary scheme. Worldwide research and development activities in legal deposit and digital preservation became increasingly intense over the course of the research. The new, enabling, legislation was passed as this thesis was being written up.

## **1.6. The Structure of the Thesis**

The research involved a review of the literature to provide background information on legal deposit, practical issues arising from digital legal deposit and relevant research and activities taking place in the UK and overseas. The findings of the literature are reported in chapters 2 to 5. Chapter 2 traces the evolution of legal deposit in the UK and elsewhere and highlights some of the issues, particularly the concerns and attitudes of publishers. Chapter 3 brings together the literature dealing with the practical issues of digital deposit. This chapter has expanded over the course of the research as more has become known about the challenges, particularly in the area of digital preservation. This is the core concern of legal deposit libraries. Chapter 4 follows up this analysis by identifying the various developments that could potentially be used to implement and manage digital legal deposit. The final literature-based chapter approaches the topic from a different perspective and provides an overview of the most prominent and influential initiatives and developments in the area over the last few years. It puts the specific issues highlighted in the previous chapters into perspective and hopefully provides some insight into the development of digital legal deposit and digital preservation at national and international levels.

Primary data was gathered through two rounds of interviews with identified stakeholder groups, including UK and overseas legal deposit libraries, publishers, relevant representative organisations, other digital archives and researchers and experts in the area. The research was designed in order to take into account the different perspectives and concerns of the different stakeholder groups and the methodology is described in chapter 6. The results of the primary research are described in chapters 7 and 8. Chapter 7 examines the stakeholders, their roles, perceptions, relationships and concerns. Chapter 8 focuses on how the UK legal deposit system functions at the moment, how well or how badly it will adapt to digital publications, what the problem areas are and what potential solutions might be, if any. Chapter 9 brings together the results of the literature review and primary research and discusses the key issues and findings. Chapter 10 draws conclusions and makes recommendations.

### **1.7. The Significance of the Research**

This research provides a timely synthesis of sources of relevant information that will be of use not only in the implementation of legal deposit but also for the digital preservation community in general. It also presents an independent view of the readiness of the UK legal deposit libraries for digital deposit that takes account of the political and cultural issues that will have an impact on how successfully participants in the UK legal deposit system will be able to implement digital legal deposit.

## 2. THE HISTORY AND DEVELOPMENT OF LEGAL DEPOSIT

### 2.1. Beginnings

France is widely credited as the first country to introduce legal deposit. King François I's *Ordonnance de Montpellier* of 28 December 1537 obliged all "imprimeurs et librairies" (printers and booksellers) to deposit "toutes les œuvres dignes d'être vues", at the Royal Library (quoted in Partridge 1938, pp. 227-228). There was no exception to this obligation; punishment for non-compliance was confiscation of the entire edition and an arbitrary fine.

However, the idea of building up a library through the compulsory deposit of books may well have originated at the ancient library at Alexandria. One acquisition method apparently involved searching all ships docking in Alexandria harbour, then confiscating and copying all books discovered. The copies were then returned, with some compensation, whilst the originals were retained for the Library (Blum 1991, p.103, Canfora 1989, p.20, El-Abbadi 1990, p.27, Parsons 1952, p.163).

Other countries also have long-standing legal deposit systems. Jasion (1991) compiled a survey of legal deposit around the world that is a useful source of dates for original legislation. For example, Belgium's earliest legislation dates from 1594, while other European countries introduced legislation in the 17<sup>th</sup> century, including Sweden (1661), England (1662), Poland (1645) and Denmark (1697).

### 2.2. Purpose: Present and Past

The *Ordonnance de Montpellier* gave the reason for the deposit of works as:

... pour avoir recours aux dits livres, sit de fortune ils étoient cy après perdus de la mémoire des hommes, ou acunement immués, ou variés de leur vraye et première publication. (quoted in Partridge 1938)

The Royal Library was to act as an archive of the printed intellectual and cultural heritage of France in order to ensure that these works were not lost to the nation and also to allow verification of the integrity and authenticity of individual works.

However, the aim was not just to preserve the country's cultural heritage; it was also censorship. Anyone planning to sell a foreign work in France first had to check with the Royal Library that it was “...tolérable destre vu, afin d'obvier aux méchantes œuvres et erreurs”. (quoted in Partridge 1938)

The law was a means of acquiring all published material in order to examine it for sedition before allowing wider dissemination.

Jasion (1991, p. 7) summarises the evolution of legal deposit. Currently, the main purpose is “the preservation of the nation's cultural heritage.” Other purposes are or have been: copyright protection; censorship; provision of exchange material; bibliographic control; provision of statistical information; provision of library stock; provision of government and legal publications and provision of universal access to published information. Although he also mentions censorship, none of the countries he surveyed explicitly mentioned censorship in their legal deposit laws (Jasion 1991, p. 8).

In his UNESCO (United Nations Educational, Scientific and Cultural Organization) guidelines Larivière (2000, p. 4) set out various objectives for legal deposit. Legal deposit ensures “the acquisition; the recording; the preservation and the availability of a nation’s published heritage.” The role of the legal deposit system is the development of a national collection of published material, the compilation and publication of a national bibliography and the provision of access to national and overseas researchers. Legal deposit is the “foundation of a national policy of freedom of expression and access to information”. Related to the concept of universal access to publications, is the concept of freedom of information. Lor (1995, p. 96) comments “Without free access to what has been already published by either the government or private citizens, freedom of information is limited and incomplete”.

Of course, legal deposit only contributes to freedom of information; other legal systems, such as the deposit of public records have a role to play.

The historical connection of legal deposit with copyright protection is still evident in the legislation of many countries. For example, legal deposit is part of copyright law in the USA (1976), Ireland (2000) and, until recently, the UK (1911). Vitiello (1993, p.49) stated that deposit is only currently a requirement for copyright registration in a few countries, including the USA. This is no longer the case in the USA, although works need to be registered before legal cases for infringement can be brought (United States Copyright Office 2004). The UNESCO guidelines advocate the separation of legal deposit and copyright legislation, guidelines from the Council of Europe and EBLIDA (European Bureau of Library, Information and Documentation Associations) recommend that the two pieces of legislation should be coordinated to allow reasonable access to users and compensation for copyright holders (Council of Europe & EBLIDA 1999).

Jasion (1991, p. 3) identifies four requirements for legal deposit. These are:

- exhaustiveness
- preservation
- information about deposited material should be made known, usually through listing in the national bibliography
- access.

Exhaustiveness implies all material, regardless of quality or format, should be deposited. As will become clear in subsequent chapters, this is a requirement that is being increasingly questioned. While legal deposit material is accessible, this access is usually restricted in some way. Legal deposit libraries are often styled 'last resort'; users usually must have a research need that cannot be fulfilled elsewhere.

Vitiello says that there has been a shift in emphasis in deposit libraries over time. The preservation function has become more important. Deposit libraries are no longer focussed solely on the comprehensive collection of material, but are also concerned with its use and long-term care (Vitiello 1993, p. 42).

### **2.3. Overview of Legal Deposit Systems**

Jasion lists the different bases for legal deposit used around the world. These include: parliamentary acts or laws; cabinet decrees and orders; ministerial regulations and directives; government departmental orders; regulations, circulars, rules, policy statements; library regulations and statutes and municipal ordinances (Jasion 1991, p.7). In some countries, legal deposit is carried out on a national level. In others legal deposit is on both a national and regional level. For example, Canada has national and provincial deposit laws and Germany has national and federal laws. Switzerland has no national legal deposit law; the national library relies on voluntary deposit and purchasing to fill gaps and the cantons have their own laws. In the Netherlands there is no statutory deposit system either; all deposit of print and non-print material is carried out on a voluntary contractual basis. The system appears to work very well and the depository institutions seem to enjoy good relations with publishers (Noordermeer 1997).

Publishers usually deliver material, although sometimes the obligation is on printers. In France, both the publishers and printers are still required to deposit material. The purpose of this seems to be a double check on whether all eligible material has been delivered (Smith 1994, p.143). The depositor is usually responsible for the full cost of deposit. However, in some countries, depositors are reimbursed for their material or can send material post-free. The number of copies to be deposited differs from one to several (Jasion 1991). If a single library receives legal deposit material, then all material is usually sent directly to that library. In the UK, one library receives material directly and the others mainly claim material through an agent acting on their behalf. This agent then forwards material. In other countries with distributed systems, material is sent to one library and then forwarded to other libraries. Finland uses this type of arrangement (Helsinki University Library 1999). The time limits for deposit range from before publication to once a year. Penalties for non-compliance vary from small fines to prison sentences.

There is great variety in the types of material collected through legal deposit. The requirement is usually material available to the public whether for sale, hire or for free. Printed publications are really the only common factor. Other types of material collected include sound recordings, audiovisual material and software.

Often the depositories are national libraries. Parliamentary libraries also act as depositories, for example, the US Library of Congress and the Japanese National Diet Library. In the UK, national and university libraries act as depositories. National archives are often used for non-print material, including film and sound archives. In some countries, including Finland and the former Yugoslav Republic of Macedonia, a university library also acts as the national library. In the UK, legal deposit is extended to a foreign country, i.e. the Republic of Ireland.

Rugaas (1990, p.157) states that in many countries, legal deposit arrangements do not, in practice, promote the aim of safeguarding the cultural and intellectual heritage of a country. Legal deposit is often “developed without any overall plan in mind to see this development as different parts of an integrated solution for the preservation of the intellectual and cultural heritage”. He refers to the frequent pattern of different institutions taking responsibility for different information carriers, and a lack of integrated systems of bibliographic tools. Rugaas does not advocate that one institution should be responsible for all legal deposit material, he just wants integrated legal deposit systems.

#### **2.4. History of Legal Deposit in the United Kingdom**

The standard work on the history of legal deposit in the UK is by Partridge (1938). This deals with the legal deposit in the whole of the former British Empire. The first part of the work is devoted to the situation in Great Britain and Ireland between 1610 and 1932. Bell acknowledges that he has drawn heavily on Partridge in his two-part article on the history of legal deposit, although he seems to have done more than this by copying passages verbatim from the earlier work (Bell 1977a; Bell 1977b). Feather devotes a chapter to legal deposit in his work on the history of copyright in Britain (Feather 1994, pp. 97-121). This deals with the legislative process and the politics of the development of legal deposit. Feather brings into sharp focus the often vehement opposition of the publishing trade to the whole principle of legal deposit, and the trade's history of reluctant or non-compliance with the law. Between them, these three authors provide a good overview of how legal deposit has developed in the UK up to the early 20<sup>th</sup> century. The following section therefore mainly focuses on developments since then. The history of legal deposit in the UK up to the end of the 19<sup>th</sup> century is summarised in Table 2-1.



Year	Development	Details
1610	Agreement between Sir Thomas Bodley and the Stationers' Company	The Stationers' Company passed a copy of every book deposited at Stationers Hall to Bodley's Library
1637	Star Chamber decree (Court of Star Chamber 1637)	Confirmed private arrangement above.
1662-1679 1685-1695	Press Licensing Acts (Great Britain 1662, pp.23-25, 31-32, 41, 69, 293) (Great Britain 1664, pp.26, 293-4) (Great Britain 1665)	1662: Three deposit libraries: <ul style="list-style-type: none"> <li>• Bodleian Library, Oxford University</li> <li>• University Library, Cambridge</li> <li>• Royal Library</li> </ul>
1709	Copyright Act (United Kingdom 1710)	Added six deposit libraries (total 9): <ul style="list-style-type: none"> <li>• Faculty of Advocates Library, Edinburgh</li> <li>• University of Edinburgh</li> <li>• University of Glasgow</li> <li>• University of St Andrews</li> <li>• Kings and Marischal Colleges, Aberdeen (1 copy shared)</li> <li>• Sion College, London</li> </ul>
1801	Copyright Act (United Kingdom 1801, s. 6, p.689)	Added a further two deposit libraries (total 11): <ul style="list-style-type: none"> <li>• Trinity College Dublin</li> <li>• King's Inns Dublin</li> </ul>
1836	Copyright Act (United Kingdom 1836)	Reduced deposit libraries to five: <ul style="list-style-type: none"> <li>• Bodleian Library, Oxford University</li> <li>• University Library, Cambridge</li> <li>• British Museum (was Royal Library)</li> <li>• Faculty of Advocates Library, Edinburgh</li> <li>• Trinity College Dublin</li> </ul> Plus £3,000 per annum compensation to libraries losing their privileges.
1842	Copyright Act (United Kingdom 1842, pp. 579-581)	Extended scope of legal deposit to publications of the entire British Empire

**Table 2-1 Main Developments in UK Legal Deposit Law Until 1842**

#### 2.4.1. *Origins in the United Kingdom*

On 12 December 1610, the University of Oxford and the Company of Stationers in London signed a private agreement. The Company agreed to send to Oxford a copy of every new book deposited by its members at the Company of Stationers Hall. The Company of Stationers kept registers of all books printed by its members (Arber 1950); registering a work gave its owner the exclusive right to print it. Apart from the presses at the Universities of Oxford and Cambridge, printers could only print publications if they were members of the Company. Although printers were required to register their works, printers were not penalised for not depositing their books. A 1637 decree (Court of Star Chamber 1637) attempted to strengthen the Bodley agreement by threatening printers who failed to deposit new books with imprisonment and other punishments.

However, the real legal basis for legal deposit in the United Kingdom was control of publishing through the mid-17<sup>th</sup> century press licensing acts (Great Britain 1662, pp.23-25, 31-32, 41, 69, 293). This aim was stated explicitly in the 1664 renewal of the *Press Licensing Act* of 1662:

Bee it enacted ... That an Act ... that now is entituled An Act for preventing Abuses in printing seditious treasonable and unlicensed Bookes and Pamphlets and for regulateing of Printing and Printing Presses shall be continued and remaine in force until the end of the next Session of Parlyament.  
(Great Britain 1664, pp.26, 293-4)

The 1662 Act required deposit of three copies of works at Stationer's Hall before any sales were made. The intention was that these copies would then to be forwarded to the Royal Library, and the libraries of the Universities of Oxford and Cambridge. The *Press Licensing Act* was renewed again in 1665 (Great Britain 1665). The intention was made explicit in the 1665 Act. The *Press Licensing Act* lapsed and was renewed over the course of the next few years and finally expired in 1679.

The UK *Copyright Act* of 1709 addressed the issue of piracy of works. The Act linked deposit with the granting of copyright and gave deposit privileges to several Scottish libraries in recognition of the union of the Scottish and English parliaments in 1707. There was some argument in the 18<sup>th</sup> century about what books should be deposited. The printers argued that only works registered at Stationers' Hall were eligible for deposit. The (*University*) *Copyright Act* (United Kingdom 1775) made it clear that copyright protection was contingent on deposit of works at the time of registration at Stationers' Hall.

The *Copyright Act 1801* marked the union of Great Britain and Ireland and aimed to prevent piracy of British works in Ireland. The Act gave two Irish libraries legal deposit privileges (United Kingdom 1801, s. 6, p.689). The first half of the nineteenth century was very turbulent as far as copyright legislation was concerned. Feather (1994, pp. 101-119) describes the intense activity that resulted in the Copyright Acts of 1814, 1836 and 1842. The issue of which books should be deposited resurfaced after the 1801 Act. While it was now clear that all books registered with Stationers' Hall were to be deposited, the 1709 and 1801 Acts were ambiguous on whether books that were not registered were also eligible for deposit. The 1814 Act removed this ambiguity (United Kingdom 1814).

Feather (1994, p. 113) says "in the immediate aftermath of the 1814 Act we can see the remote origins both of the employment of a London Agent by the out-of-town libraries, and of their regular meetings to discuss matters of common concern". The first part of this statement contradicts Bell's (1997a, p. 7) assertion that the use of an agent started much earlier in 1726.

In 1836, six libraries lost their legal deposit rights. Instead they were entitled to an annual compensatory payment of £3,000 to be shared between them (United Kingdom 1836). Feather (1994, p. 119) implies that the link between deposit and copyright was broken at this point. It is not clear why he says this because the need to register appears to continue in the *Copyright Act 1842* (United Kingdom 1842, pp. 579-581). However, the British Museum's copy was now to be delivered direct rather than through Stationers' Hall, so perhaps the direct link between deposit and copyright *registration* was broken.

At the beginning of the 19th century, the amount of material being deposited was unsatisfactory for the deposit libraries. This situation improved somewhat through the

determination of Antonio, later Sir Anthony, Panizzi at the British Museum. Panizzi is credited, through his energetic enforcement of legislation, for establishing what may be regarded as 'proper' legal deposit in the UK (Altick 1957, p.215, Harris 1991). The supply of legal deposit material to the other libraries also apparently improved as a consequence of the actions of the British Museum.

#### **2.4.2.            *The 1911 Copyright Act***

At the beginning of the 20<sup>th</sup> century, there were five legal deposit libraries. The Copyright Act of 1911 included provisions for the legal deposit of 'books'. While this provision was amended over time, it remained in force for more than ninety years. In 2003, new legislation was passed that not only separates legal deposit law from copyright law, but also paves the way for the extension of legal deposit to non-print material. While this creates the potential for legal deposit to catch up with changes in publishing over the last century, the fact that sound and films are excluded under the new law means that there will be no statutory underpinning for deposit of the recorded intellectual heritage for the foreseeable future. The main steps in the process of extension of legal deposit in the UK are summarised in Table 2-2 below and are discussed in this and the following sections.

<b>Year</b>	<b>Development</b>	<b>Details</b>
1911	<i>Copyright Act 1911</i>	Extended copyright libraries to six:  Bodleian Library, Oxford University  University Library, Cambridge
	<i>British Library Act 1972</i>	British Museum (later British Library)
	<i>National Library of Scotland Act 1925</i>	Faculty of Advocates Library, Edinburgh (later National Library of Scotland) National Library of Wales Trinity College Dublin
	<i>Copyright (British Museum) Act 1915</i>	Exclusion of: advertisements, voters rolls, design specifications, timetables, calendars, blank forms and wall sheets
	<i>Theatres Act 1968</i>	Inclusion of: published scripts of theatre productions
1969	<i>Film (Statutory Deposit) Bill</i>	Failed attempt to extend legal deposit to films
1996	British Library proposal (British Library 1996)	To extend legal deposit to non-print material. Supported by the other legal deposit libraries and the British Film Institute
1997	Government Consultation Paper	Confirmed commitment to legal deposit, consultation on scope of legislation
1998	Kenny Working Party Report (Working Party on Legal Deposit 1998)	Recommendations for legislation and principles for an interim voluntary scheme
1999	Voluntary Deposit Code of Practice (see Appendix 1)	For voluntary deposit of microform and digital publications distributed on physical media. Voluntary scheme commences January 2000.
2002-2003	<i>Legal Deposit Libraries Bill</i>	Draft legislation makes its way through Parliament
2003	<i>Legal Deposit Libraries Act</i>	Retains regulations for printed material and makes provision for the development of regulations to extend legal deposit to non-print material.

**Table 2-2 Development of Legal Deposit in the UK, 1911-2003**

For the purposes of the Act, 'books' included "every part or division of a book, pamphlet, sheet of letterpress, sheet of music, map, plan, chart or table separately published, but shall not include any second or subsequent edition of a book unless such edition contains additions or alterations either in the letterpress or in the maps, prints, or other engravings belonging thereto" (United Kingdom 1911, p. 190.) The 1911 Act required publishers to send a copy of all printed books published in the country to the British Museum within a month of publication. In addition, five other libraries had the right to claim copies of material they required for their collections. Feather (1994, p. 120) states that the deposit libraries had six months from publication to make claims; the Act specified twelve months (United Kingdom 1911, S. 15(2), p. 189). Publishers sent requested copies to the Copyright Libraries Agent, who acts on behalf of the other libraries. These provisions transferred unchanged to the new law of 2003 (United Kingdom 2003d).

Some parts of the relevant section of the 1911 Act were amended. Some of the institutions receiving the deposit privilege changed. On the creation of the National Library of Scotland (NLS), the Advocates' Library passed over its non-legal collections to the nation (United Kingdom 1925). As Feather (1994, p. 121) states, the Advocates' Library still receives legal works through legal deposit. These items appear in the catalogues of both the Advocates' Library and the NLS. They are available to readers in the NLS, but they are in fact housed in the Advocates' Library for the use of the Faculty members. The National Library of Scotland cares for non-legal material and legal material not wanted by the Advocates' Library. The National Library of Wales (NLW) became a deposit library, but had narrower rights compared to the other libraries. The NLW was entitled to claim Welsh, Welsh language and Celtic material. The British Library (BL) inherited the British Museum's right to receive legal deposit material automatically on its creation in 1973 (United Kingdom 1972).

The Government set up a copyright committee to review the 1911 Act. The 'Gregory Committee' included legal deposit in its deliberations. The Committee supported the continued existence of legal deposit (Board of Trade 1952, S. 58, p. 24), but felt that it should not be extended to films and recorded sound. The Committee accepted that such an extension was logical, but since copyright was no longer dependent on deposit the Committee felt that the issue should not be considered within the scope of the review of the Copyright Act (Board of Trade 1952, S. 72, pp. 28-29). In fact, there was an

unsuccessful attempt to introduce legislation for the deposit of films in 1969 (United Kingdom 1969). The Committee did not say that legal deposit for print material should be taken out of copyright legislation, which would be a more consistent position. The Committee did recommend that works *first published* in "micro-film" should be deposited because these works were the equivalent of books (Board of Trade 1952, S. 67, p. 27).

An interesting point raised in the Committee's report is that the provisions regarding the delivery of books to the British Museum and the Copyright Agent in the 1911 Act were inconsistent (Board of Trade 1952, S. 65, p. 27). The Act stated explicitly that copies to be delivered to the Museum should be at the publisher's expense. This was not so for deliveries to the Copyright Agent. The Committee recommended that the terms should be the same for all the deposit libraries. Despite these recommendations, the new *Copyright Act* (United Kingdom 1956) did not amend any of the legal deposit provisions of the 1911 Act. The next Committee set up to review copyright did not consider the issue of who should bear the cost of delivery of material to the Copyright Agent. There is no evidence of any other measure dealing with the ambiguity, but in practice publishers do deliver material at their own expense.

In 1973, the Government announced that it was setting up a committee to investigate whether any changes needed to be made to the *Copyright Act 1956* and the *Design Copyright Act 1968*. The 'Whitford Report' (Board of Trade 1977) included a summary of submitted evidence that illustrated the contradictory views of legal deposit libraries and publishers as to whether legal deposit arrangements should be maintained or reduced (Board of Trade 1977, SS. 809-819, pp. 205-207). The major objection from the publishers' perspective was the financial burden placed on the trade. The report points out that moral objections to the "confiscatory nature" of the deposit requirement is weakened by the willingness of publishers to provide one or two copies rather than six (S. 810). Other points included whether an additional legal deposit library should be introduced into the system (S. 826), whether Channel Islands and Isle of Man publications should be included (S. 826) and whether legal deposit should be extended to non-print material (S. 830).

There the need for clarification of the definition of publication was highlighted. The reason for this was to ensure that future legal deposit provisions include "items distributed

free of charge and items paid for by subscription in advance of publication and to establish a criterion of publication for microform material produced in single copies on demand" (S. 817).

The discussion on the types of non-print material that could be included in legal deposit included microfilm, sound and audiovisual material, such as films (S. 830). Digital material is not mentioned. Although online databases were established at this time, there was little else in the way of electronic publishing. However, the point on the need to develop a criterion of publication for on demand microform material is relevant in the digital environment. The assumption that the deposit libraries would not be able to cope with sound and audiovisual material and that separate archives should be considered (S. 830) is interesting for three reasons. First, the National Sound Archive that was set up is now part of the BL. Second, the NLW has always collected Welsh material in various media. Third, the convergence between the different media that has since occurred means that the deposit libraries could potentially deal with audiovisual material that forms part of digital publications.

The report argues for the retention of legal deposit on three grounds:

- the value to scholarship of legal deposit depends to a great extent on the continuing comprehensiveness of the legal deposit collections
- there is value in the distributed nature of the collections
- the legal obligation on publishers to deposit newly published material is a simple and effective mechanism ensuring continuing comprehensiveness of collections (S. 821)

According to the report, the work of scholars is aided by having immediate access to the whole literature on their subject (S. 821). Several advantages to the geographic dispersal of collections were identified (S. 822). The National Libraries of Wales and Scotland have regional collection responsibilities and their collections are enhanced by the inclusion of all UK literature. Geographical dispersion of multiple copies also serves a preservation function, especially in the case of disaster. The existence of several libraries means that the pressure of access is shared and there are savings in publicly funded travel and subsistence for researchers. According to the report, legal deposit is a more reliable way of



ensuring continuing comprehensiveness of the national archive than relying to public funds.

The report said that there was no need for legal deposit to continue to be part of copyright law as the connection between ownership rights and deposit had long ceased to exist (S. 832).

The Committee had no formal recommendations to make, but the report summarises the views of the Committee. These were:

- The question of relieving publishers from the financial burden involved in legal deposit, possibly through fiscal concession, should be considered.
- All deposit libraries should be on the same footing.
- The John Rylands University Library of Manchester should be accepted as a deposit library, at least in principle.
- The question of books published in the Channel Islands and the Isle of Man should be considered.
- The financial penalties for non-compliance with the legal deposit obligation should be revised.
- The question of microform publications should be considered.
- The revised provisions for deposit should be the subjects of legislation separate from the law of copyright.
- The Committee was in favour of the establishment of national archives of films and published visual and sound recordings, but considered that further investigation of the desirability and feasibility of such action was required (S. 833).

The views of the Committee on legal deposit were generally not taken up, even though a later (Department of Trade 1981, pp. 56-57) consultative document on copyright reform agreed with some of them. This review agreed that all the legal deposit libraries should be on the same footing. This referred to the entitlement of the NLW, which had been further restricted by legislation subsequent to the 1911 Act (Department of Trade 1981, Ch. 17, Paras. 1 & 4). The various anomalies were addressed over time, and in 1987 the last of the restrictions was removed (United Kingdom 1987a, 1987b). The need to review penalties

for non-compliance with legal deposit was endorsed. The deposit of microforms was deemed to require further debate (Para. 5). The suggestion of increasing the number of legal deposit libraries was rejected, as was providing further financial concessions to publishers (Para. 2). A Prime Minister's Working Party on the future of the British film industry recommended that there should be statutory deposit of films in the National Film Archive (Prime Minister's Working Party 1976, Para. 106, p. 26).

This recommendation was repeated in a report on proposals for setting up a British Film Authority (Interim Action Committee on the Film Industry 1978, Para. 25, p. 9). However, nothing came of the proposal. Indeed, the 1981 consultation document (Department of Trade 1981, Ch. 17, Para. 6) stated that the case had not yet been made for the creation of film and audiovisual archives, citing technical problems of storing this material and the need for more space and staff to take on the extra responsibilities. Deposit in film and audiovisual archives was to remain on a voluntary basis. The issues of books published in the Isle of Man and the Channel Islands does not seem to have been considered in 1981 and legal deposit remained part of the 1911 Copyright Act after the new Copyright Designs and Patents Act (United Kingdom 1988) appeared.

## **2.5. Extension of Legal Deposit in the UK**

### **2.5.1. *The BL Proposal and the Government's Response***

The BL commissioned several studies on potential issues arising from the extension of legal deposit to non-print publications (Electronic Publishing Services Ltd 1996, Hendley 1996, Martin 1996). The findings from the studies fed into a proposal the BL (1996), supported by the other legal deposit libraries and the British Film Institute, submitted to the then Department of National Heritage (now the Department for Culture, Media and Sport) in January 1996.

The UK Government issued a consultation paper in response in February 1997 (Department of National Heritage *et al.* 1997). The Government confirmed its commitment to maintaining a national published archive, and that legislation would be

required to extend legal deposit to digital (or electronic) publications in physical form. Views were solicited on whether online publications should be included at some point in the future (S. 3.28). Any changes should not result in overburdening publishers. Views were also sought on whether deposit libraries should be given statutory powers to copy for preservation purposes and to give access to material (S. 4.5).

The Government prepared a summary of responses to its consultation document (Department for Culture, Media and Sport 1998). Respondents mainly supported a system of legal deposit. Unsurprisingly, what opposition there was came from publishers. Generally, they were worried about lost sales and the cost and viability of extending the system.

According to Davies (1998) the publishers' preferred option was deposit in a single repository. The Publishers Association (PA) (1997, Para. 3.2) response to the consultation document also suggested that a single copy of a digital work should be deposited with a designated repository. For CD-ROMs, a single copy with a possible back-up should be deposited at a single repository. Most respondents to the Government consultation document thought that deposit should be extended to online publications (Department for Culture, Media and Sport 1998, Para. 9), but agreed that there were substantial difficulties to be overcome. The PA (1997, Para 1.7.5) did not think that it was acceptable for legal depositories to have full access to all online publications, and pointed out that some have strict use restrictions and are expensive to produce. Therefore, the PA suggested that a single library should be able to take a 'snapshot' of a particular online database on an annual basis (Para. 3.7). However, the deposit of dynamic databases would be problematic: it would not be feasible to take a copy of every updated version of a particular database.

The PA (1997, Para. 1.7.4) stated that it is difficult to separate so-called multimedia publications into different categories, such as text, sound or moving images. This classification would have been necessary under the system mooted in the Government's consultation document because different deposit conditions would apply to different types of publication. The PA suggested that publishers, rather than deposit libraries, should be able to decide categories to minimise the burden of having to comply with different systems for different types of publications.

Publishers were concerned that access to their material might result in commercial loss (Department for Culture, Media and Sport 1998, Ss. 5 & 11). This would be either because the depositories would be effectively acting as competition or because of illegal copying by users. Publishers wanted to restrict access as much as possible and were opposed to networked access.

The maintenance of appropriate hardware for mounting and viewing material was an issue for the legal deposit libraries, no doubt because of the rapid cycle of obsolescence (Department for Culture, Media and Sport 1998, S. 20). The libraries preferred to apply the existing deposit arrangements to CD-ROMs (S. 5). The BL said it would prefer multiple deposit of CD-ROMs because of the restricted form of access proposed by publishers (British Library [n.d.]). Depository libraries were in favour of providing access to material deposited, but were sensitive to the concerns of publishers and the need for negotiation for the system to work (British Library [n.d.]). Many publishers thought that the deposit of commercially sensitive or high value material should be delayed. The main concern of libraries here was that if material is not deposited at the time of publication, it may be lost if the publisher goes out of business or for some reason cannot provide the material at a later stage (Department for Culture, Media and Sport 1998, S. 14). A possible compromise here would be deposit at the time of publication, together with an embargo on access as agreed with the depositor.

While there was agreement about the right of depositories to make copies for preservation purposes, publishers were concerned that the right to copy should not be used for making further copies available to users (S. 12). They had concerns about depositories making copies in new formats, presumably because this would change their original product. This view would be in conflict with the ability of depositories to provide long-term access if it turns out that migration to new formats becomes a widely used preservation strategy (see section 3.4.4).

## 2.5.2. *The Kenny Working Party*

The Government set up a working group comprising representatives of the various interested parties, under the chairmanship of Sir Anthony Kenny, then chairman of the BL Board. Its remit was to:

- advise on how to achieve a comprehensive national archive of non-print material, taking into account the concerns of publishers and exploring the potential for the use of secure networking for access;
- draw up and agree a voluntary code of practice for digital and microform material, to be used until Government introduced legislation;
- ensure compatibility between the provisions of this code of practice and existing voluntary arrangements for film and sound material (Working Party on Legal Deposit 1998, S. 1.8)

The Working Party reported in July 1998. In its opinion, extension of legal deposit legislation was the only way to achieve a comprehensive national archive; voluntary arrangements would not work (S. 2.2). The Working Party recommended that legislation should allow the appropriate Secretary of State to declare from time to time, after consultation, that 'publications in specific new media' should be subject to legal deposit (S. 2.5(1)). In its general principles for an interim voluntary system (S. 3 and Appendices), the Working Party advocated a distributed archive that would include at least the current legal deposit libraries (S. 4.1). Rights owners should allow repositories to hold items for both archiving and access purposes.

According to the Working Party, if an item is published in more than one medium, the repository should be able to choose which should be deposited (Summary, S. 3).

Publishers queried this point. In its response to the report, the Newspaper Association pointed out that letting the deposit libraries choose the medium of the deposit copy would create another administrative tier. This is because the publisher would have to tell the libraries about the existence of different media<sup>1</sup>. According to the Newspaper Society, publishers may find it easier and cheaper to submit the hard copy version of their papers.

---

<sup>1</sup> C. Courtney letter to Sir Anthony Kenny, 10 June 1998.

Echoing the point made by Rugaas about the coordination of preservation activities mentioned earlier, the Newspaper Society also questioned how well the NEWSPLAN newspaper microfilming initiative would “dovetail with the proposals for a national archive of non-print material”<sup>2</sup>.

The Working Party (1998, Summary, S. 4) recommended that authorised users of a repository should have access to items, unless there is some special temporary restriction, for example for very valuable material (Summary, S. 6). Dissemination of material beyond the 'national published archive' should be allowed only with the agreement of copyright holders or on expiry of copyright and might include royalty payments (S. 2.5).

Certain categories of material may be exempted from legal deposit obligations if deposit would place an unreasonable burden on publishers (S. 6). There should be a standing committee, responsible to the Secretary of State, made up of representatives of repositories, their users and publishers (S. 4.2). This committee would deal with any applications for exemption or embargo, and disputes.

The report defined several different publication types and publication media for 'electronic publications' (Appendix A). It also defined place of publication. These definitions were based on work carried out for the European Commission (Mackenzie Owen & Walle 1996, see section 3.1).

The Working Party (1998, Summary, S. 4) recommended that publishers should deposit CD-ROMs after twelve sales in the UK. The report does not specify how this mechanism would be enforced. Indeed, if a particular title did not achieve twelve sales, a copy might never be deposited. The recommendation on access was a compromise between publishers and legal deposit libraries: at least one, but preferably up to six, copies should be deposited and access should be restricted to a single user at a time. The publishers on the Working Party preferred access in one physical location only (Appendix B, Ss. 5-7).

Legal deposit library representatives agreed that access should be limited, but preferred the option that after more than 50 sales in the UK, there should be access in each of the

---

<sup>2</sup> NEWSPLAN is a cooperative microfilming programme for local newspapers in the UK and Ireland.

existing legal deposit libraries. Again, no details were given on how the libraries would know when this level of sales was achieved. Wider access could be achieved either through multiple deposit or by single deposit with networked access to other libraries. A standard set of terms and conditions should be drawn up between repositories governing wider access and use. Repositories should indemnify publishers against 'any breach of the terms and conditions by users which results in a publisher incurring a legal liability' (Appendix B, S. 9).

The report recommends that 'Legislation should be drafted in such a way that on-line publications are brought within the scope of legal deposit, when this is the primary form of publication, though only some forms of such publication, in particular monographs and journals, would be included from the introduction of deposit arrangements' (Working Party on Legal Deposit 1998, p. 14 Appendix B, S. 8).

The Government's response in December 1998 (Smith 1998) to the Working Party's report was to ask Sir Anthony Kenny to lead a group drawn from libraries and publishers in developing a code of practice for the voluntary deposit of non-print publications. The Government recognised that this would be a purely interim measure and accepted the Working Party's recommendation that the only long-term solution was through legislation. The Government also asked the group to carry out more work on definitions and the impact on publishers of an extension to legislation. The Government's 1997 consultation document included a draft compliance cost assessment (see section 3.7.2), but this was criticised by the PA (1997, S. 1.13). According to the Publishers Association, publishers felt that they had not been adequately consulted and did not agree with what was said (Lakeman Fraser 2001). There is no mention in the literature of the outcome of the work on definitions. The BL agreed to commission a Regulatory Impact Assessment study, to be submitted to the Government on 31 March 2000 (Working Party on Legal Deposit 1999, S. 1.3). Again it is not clear from the literature whether this was done. However, a cost and impact study was finally carried out in 2002 (Electronic Publishing Services 2002), the results of which are discussed in Chapter 3. This work fed into the Regulatory Impact Assessment required for the new legislation (Department for Culture, Media and Sport 2003).

### 2.5.3. *The Voluntary Code of Practice for Non-Print Publications*

The voluntary code of practice, endorsed by the PA, the Association of Learned and Professional Society Publishers (ALPSP) and the Periodical Publishers Association (PPA), came into operation on 4 January 2000 (see Appendix 1). However, publishers were encouraged to also deposit publications made available before that date. The code sets out the purpose of legal deposit and of the voluntary scheme itself. There is a discretionary element in the code; this is something that publishers saw as important (Working Party on Legal Deposit 1999).

The purpose of legal deposit is defined as: 'to ensure that the nation's published output (and therefore its intellectual record and future published heritage) is collected as systematically and as comprehensively as possible, both in order to make it available to current researchers within the libraries of the legal deposit system and to preserve the material for the use of future generations of researchers.' (Working Party on Legal Deposit 1999, S. 2). This reflects the view accepted by libraries, publishers and Government of the need for a comprehensive national archive. However, the PA (1997) in its response to the 1997 Government consultation document did make the point that because of the problems associated with deposit of online publications, comprehensiveness is not possible and is therefore not an appropriate goal. In practice, legal deposit of even print material is not comprehensive. In the UK, a high proportion of material is collected by the deposit libraries, but some material does fall through the net, especially 'local publications', or publications circulated only in a particular locality (Eden *et al* 1999).

The issues covered in the code include definition of the place of publication; types of publication, medium and format to be deposited; dealing with publications appearing on more than one medium; exclusions from deposit; numbers of copies to be deposited; access and restrictions on access, printing, copying and downloading.

The code focuses on offline digital publications, but there are some recommendations for so-called static online publications and the online element of hybrid physical format/online publications. Dynamic databases are completely excluded from the code. Neither does the code cover the types of publications that are already subject to voluntary codes in the UK:



film, sound and Ordnance Survey digital mapping products. The problem of defining which category of publication digital items belong to is addressed in the code insofar as text-based, or items intended for the purposes of information rather than entertainment, are covered.

The requirements for the formats and timings for deposit are those recommended by the Working Party. Offline publications should be deposited in the form in which they were issued, with all their supporting software and documentation (Working Party on Legal Deposit 1999, S. 9.2). They should be deposited after twelve copies have been sold in the UK (S. 10.2). Publishers are asked to deposit one copy; the other deposit libraries may request further deposit copies (S. 12).

When the scheme became operational, publishers were asked at the time of deposit to state on a standard form what level of access they would allow (S. 13.2). The levels are:

- Single-user access at a standalone workstation in the holding library
- Single-user access at a time, in the holding library, via an intranet to designated terminals
- Networked access across a secure network between deposit libraries (when the publisher is depositing a single copy only), at designated terminals, to either a single user at a time on the network, or to a single user at a time in each library.

While the first option presents the least risk for publishers, this may well present preservation problems in the form of increased handling and theft. A feasibility study on the design and implementation of a secure network between legal deposit libraries was carried out by the libraries and is discussed later.

As far as use of digital material is concerned, downloading is not permitted at all (S. 16), and printing is to be limited according the 'fair dealing' principles agreed by the Joint Information Systems Committee (JISC) of the UK Higher (and now Further) Education Councils and the PA (1998).

The voluntary code advocates an 'opt out' system for allowing the copying of deposit copies for preservation purposes (S. 17). Unless explicitly forbidden, libraries will assume

that they can make copies 'to other media' for preservation purposes. However, this is subject to the “preservation of the individual publication's identity and integrity” (S. 17). This may be a problem in the long-term, because preserving integrity, over time, is difficult if certain preservation strategies are used (see section 3.4).

#### **2.5.4.            *Waiting for Legislation***

The Department for Culture, Media and Sport (DCMS) put forward an unsuccessful proposal for legal deposit to be included in the legislative programme at this time, but there was not enough parliamentary time for this to go forward (Lakeman Fraser 2001, p.27).

While the publisher representatives involved in looking at legal deposit and developing the voluntary code seemed to have accepted the continuing need for legal deposit, there was some evidence of dissent in the publishing industry. In an article aimed at the book trade, Whitaker (2001b) objects to the right of the six deposit libraries to receive books at no cost to themselves. He mentions that the Copyright Libraries Agency makes blanket requests for books for these libraries. Whitaker's view is that legal deposit should end or at the least that the proposal in the draft 1836 copyright act that only the BL receive a copy of all material at cost price should be enacted. He also questions the need of the libraries of Oxford and Cambridge for non-academic material. Whitaker's tone is contentious, using phrases such as the “seedy history” [censorship] of this “tax” on publishers and the deposit libraries putting a “spin” on their “shameful past”. He cites the great cost of legal deposit to publishers. Replies to this article included comments that the burden on publishers is probably not as onerous as Whitaker makes out and that legal deposit ensures at least one copy of a work will survive (Adkins & Adkins 2001). A publisher agreed with Whitaker's remarks, but suggested that the number of free copies be reduced to one (Adams 2001).

The Library Association published another article by Whitaker three months later (Whitaker 2001a). Not surprisingly, given the target audience, the tone is less contentious. One could make the case that Whitaker is implying that the libraries do not even keep all legal deposit material in the first article. In the second article he acknowledges that the libraries do not dispose of this material. Whitaker also admits that legal deposit, or rather keeping a record of the published output, is a national good. He just does not agree that

publishers should be providing the books for free and repeats that legal deposit is an anomalous tax in kind on the publishing industry. Rather than saying that legal deposit should be abolished, he is now looking for an “equitable outcome” and he is looking to Her Majesty’s Treasury for a solution. In other words, if having a national archive is a public good, then the public should pay for it. The fact that the public pays for the processing, storage and long-term care for legal deposit collections is ignored. In the one reply to this article, the burden on publishers is questioned (Dellar 2001). Deposit obligations are factored into publishing businesses as review copies are and the legal deposit collection is also for the benefit of publishers.

Another article in *The Bookseller* sums up the activities and thinking that had been going on since the start of the voluntary scheme for digital material and while legislation was awaited. Lakeman Fraser quotes a representative of the Publishers Association who said that he would prefer legislation not to come too quickly to give time to develop digital deposit mechanisms (Lakeman Fraser 2001, p.27). This article also states that the deposit libraries would prefer the deposit of one copy of a digital publication with restricted network access. This is a shift in opinion by the libraries from wanting the option of multiple deposit. The article reported that the BL was working on a digital store and metadata system for digital publications. However, the article also reported that a bid for external funding for the secure network was unsuccessful.

### **2.5.5.            *The New Legislation***

In 2002, Chris Mole MP introduced a so-called ‘hand out bill’ (see Appendix 2), a Government-sponsored Private Members Bill, to the UK Parliament. Bills pass through several stages, including three ‘readings’ in the House of Commons, detailed discussion in a House of Commons committee and further discussion in the House of Lords. The *Legal Deposit Libraries Bill* went through the following process.

<b>Date</b>	<b>Stage</b>
11 December 2002	First Reading of <i>Legal Deposit Libraries Bill</i> , House of Commons
14 March 2003	Second Reading, House of Commons
4 June 2003	Standing Committee, House of Commons
4 July 2003	Third Reading, House of Commons
7 July 2003	First Reading, House of Lords
12 September 2003	Second Reading, House of Lords
6 October 2003	Committee stage waived, House of Lords
13 October 2003	Third Reading, House of Lords
30 October 2003	Royal Assent
1 February 2004	<i>Legal Deposit Library Act 2003</i> enters into force

**Table 2-3 New UK Legal Deposit Legislation: Parliamentary Process**

The concerns of publishers were raised in the House of Commons at the Second Reading and Committee stages. These concerns included that they were not sufficiently consulted on the drafting of the Bill (Moss 2003a, col. 581), the scope was potentially very wide and could prove unacceptably burdensome to at least some publishers. Dodd (2003) also reported the initial reaction to the Bill from the Digital Content Forum (DCF). Dodd reports a much stronger reaction than reported in parliamentary proceedings, with a representative of the DCF calling for the proposed legislation and possibly legal deposit itself to be scrapped altogether. Todd (2003, Column 8) also referred to the perceived haste with which legislation was drafted. He would have preferred a fundamental reassessment of legal deposit, including its purpose and how it is implemented at the moment. There was also concern over the implications of a clause on preservation and access (Mole 2003a, col. 580) and that publishers would not be able to decide the versions to be deposited (Moss 2003a, col. 583). However, the Members of Parliament who mentioned publishers' concerns also stressed that publishers were agreed on the need for legislation.

Mole (2003a, Col. 580, United Kingdom 2003a, S. 8(8)) emphasised that regulations would be developed in consultation with stakeholders and that the draft regulations were subject to parliamentary approval. Dodd (2003) reports that the DCMS commented that details for each non-print medium will be worked out in consultation with stakeholders.

Mole and the relevant Government Minister repeated this commitment at the Standing Committee stage.

Publisher anger at a perceived lack of consultation in the drafting of the Bill is perhaps surprising because publisher and library representatives had been working together for some time on the voluntary deposit scheme. In a briefing document to its members, the PA sheds some light on the strength of the reaction. The concerns were raised by members of the DCF, some of whom had not been involved in previous joint working with the legal deposit libraries. The PA (2003) apparently managed to persuade them “that to object to the Bill would be to lose it”.

The initial draft of the Bill allowed for the transfer of existing legal deposit provisions unchanged, the continuation of the voluntary schemes for sound and film materials and extension of legal deposit to all other non-print publications (United Kingdom 2003a). It is not clear why sound and film material were excluded from legal deposit. However, the Handout Bill process apparently works best with items that are not controversial and will have a low impact on the industry; it is Government-sponsored and Government is unlikely to support controversial bills. Bringing sound and film into legal deposit would involve creating a new legal deposit institution – the British Film Institute. The 1996 proposal to Government included a submission from the BFI and there was therefore an implied expectation that films would be included. However, the sound and film industries may not be happy about changing to a mandatory system. The exclusion of film and sound material remained after the committee stage. It does seem like a wasted opportunity to introduce comprehensive legal deposit, especially when other countries have been able to do this.

There was some discussion along the lines advocated by Todd (2003, Column 8) at the Committee stage, in that roles of the existing legal deposit libraries for print material were questioned (House of Commons, Standing Committee F 2003, Cols. 28-32). The need for three deposit libraries in England was questioned and there was a request for assurance that the Bodleian Library and Cambridge University Library do not make legal deposit material more widely available to their own academic communities rather than purchasing material. The issue of why there is still a UK legal deposit library in the Republic of Ireland, but none in Northern Ireland was also raised. Members of the Committee were

not arguing against the need for several deposit libraries, but were highlighting the anomalies. Arguments for the retention of several libraries were on the basis of maintaining redundancy of copies for preservation purposes. The point that being entitled to copies did not mean that every library needed to claim them, especially if a secure network is in place, was raised. Another point was that where the provision of several copies of a publication would be financially onerous on a publisher, alternative arrangements could be made in such cases.

The Bill contained no explicit statement of the purpose of legal deposit, but the long-title included “to make provision about the use and preservation of material deposited”. Mole did, however, state that

The purpose of legal deposit is to ensure the nation’s published output, and thereby its intellectual record and future published heritage, is collected systematically and as comprehensively as possible. We do this to make material available to current researchers in the libraries of the legal deposit system, and to preserve it for the use of future generations of researchers. Both purposes are important. (Mole 2003a, Column 577)

The first section of the first draft of the Bill dealt with the duty to deposit (United Kingdom 2003a, S. 1(1)). For non-print material, the copy delivered has to “in the medium in which it is published”. The address to which the copy has to be delivered is “an address in the United Kingdom or an electronic address” (United Kingdom 2003a, Ss. 1(3, 6-7), 4 (3, 6)). This section remained unchanged, although there was much discussion surrounding definitions and scope, with some members of the Committee expressing concern on behalf of publishers (House of Commons, Standing Committee F 2003, Cols. 3-16). The gist of the concerns is that publishers did not know what they would be signing up for and some were worried about the potential impact on their business. Todd (2003, Col. 7) put the risk into perspective

Sometimes, there is confusion about the purpose of legal deposit. It is not about making information available to people freely because that is something that they should have as of right. It has a specific, narrow purpose. In that context, we have to recognise the specialist section of information providers who are deeply impacted by this legislation.

While the aim of introducing enabling legislation was to allow flexibility for future developments, the Bill (and subsequently the Act) did not incorporate any of the technology neutral definitions used in legislation in other countries, such as Norway (see 3.1). The Government Minister stated that definitions will be set out in regulations (Howells 2003, Col. 14).

Another concern that was raised was whether the UK legal deposit libraries would want to collect material available in the UK from publishers based overseas, in effect extending UK jurisdiction in this area (Moss 2003c, Col. 5).

The question of what happens with alternative versions of works was dealt with. The Bill used the formulation of the Kenny Working Group and does not apply to a work which is “substantially the same as one already published in the same medium in the United Kingdom” (United Kingdom 2003a, S. 2(1)). For works published in more than one medium, the “substantially the same work” rule applies and the work should only be deposited in one medium (United Kingdom 2003a, S. 2(2)). No definition was given for “substantially the same” in either the Bill or accompanying explanatory notes. Decisions on this are to be taken by the Secretary of State at a later date (United Kingdom 2003, S. 2(3)). The definition for ‘medium’ is “any medium of publication, including in particular any form of on line or off line publication” (United Kingdom 2003a, S. 9). This clause remained unchanged after the Committee stage. An amendment to allow publishers to determine the medium of deposit was proposed, then withdrawn. The motivation for the amendment was a concern for the potential costs to the publisher and that the deposit libraries should be able to take whatever medium the publisher gives them, rather than imposing delivery formats on the publishers. The seems to have been confusion as to the purpose of this clause, which is more to do with choosing between alternative versions of the same material rather than imposing format requirements on publishers (House of Commons, Standing Committee F 2003, Cols. 16-20).

An interesting change in the enforcement of legal deposit was introduced. In the 1911 Act non-compliance could result in criminal proceedings. The provisions in the Bill refer to county and sheriff (Scotland) court orders to comply or for publishers to pay the library “an amount which is not more than the cost of making good the failure to comply” (United Kingdom 2003a, S. 3). This clause remained substantially unchanged, but there was some

discussion of its provisions (House of Commons, Standing Committee F 2003, Cols. 21-24). There was a question over what the “cost of making good the failure to comply” means and why the enforcement provision of the 1911 Act could not be retained. There was a concern that for publishers of low volume, high value material deposit would be financially ruinous. The reply was that the impact on these publishers would be considered when regulations were made. The issue of an arbitration mechanism was also raised. This possibility was not ruled out (Howells 2003, Col. 26).

The enabling provisions for non-print material indicated that the print model will not necessarily be followed because the Secretary of State will make regulations on what is to be deposited and how, when and by whom (United Kingdom 2003a, S. 6). This clause was amended at the Committee stage; the Secretary of State may not make regulations unless they have considered the impact on publishers who would be affected (United Kingdom 2003b, S. 6(3)). Another amendment was proposed obliging the deposit libraries to maintain adequate measures to ensure the “secure and confidential storage of the data and content to be deposited.” (House of Commons, Standing Committee F 2003, Col. 33). The proposer referred to the protection of material of a sensitive nature (Moss 2003c, Col. 33). The term “sensitive” was questioned because deposited material is published and therefore not confidential. It seems that the real concern was immediate use or unauthorised copying of material that is still commercially valuable. Assurances were given that this issue would be taken into account in regulations (Mole 2003, Col. 35).

A member of the Committee (Allan 2003, Col. 37) queried the requirement for the provision of any software, information or accompanying documentation (United Kingdom 2003b, S. 6(2)(b)). He made the point that the publisher may not have the rights to deposit such software. The Bill’s proposer replied that this should only occur rarely and that the deposit libraries may well purchase the software in these cases (Mole 2003c, Col. 38). This remark would apply to software required to view material, such as readers, but it may well not be possible to purchase retrieval software or the publishing platform used for delivering online content. Indeed, where this software is proprietary, it may not be wise from a preservation point of view because it could be difficult to keep proprietary software working over time. Given these issues, another Committee member sensibly pointed out the need for the libraries to develop retrieval systems for deposited material and the need for this work to be funded (White 2003, Col. 37). The point was made that the



Government department sponsoring the Bill, the DCMS, only funds one of the deposit libraries; the other libraries are funded by various means (Thomas 2003). This requires cooperation between the libraries and leadership from the DCMS (Allan 2003, Thomas 2003). There would also need to be closer cooperation between libraries and publishers in the deposit of non-print material.

The section of the Bill on use and preservation of non-print publications turned out, unsurprisingly, to be the most controversial because it dealt with rights issues. To briefly summarise, preservation of and access to digital information is likely to require copying (see section 3.6). This may also be the case for acquisition of material, for example if it is harvested from the Internet. The original clause acknowledged this and provided for amendments to rights legislation so that “copyright, publication right, database right and any similar right” is not infringed by deposit libraries who need to copy for preservation or the provision of access or users making temporary copies for access (United Kingdom 2003a, S. 7). The reason given for proposing these amendments is that the original clause was thought to be too permissive (Mole 2003c, Col. 40). It is true that the original clause did not say exactly how rights legislation would be altered.

The sub-clause on copyright issues relating to use was replaced by a new clause (United Kingdom 2003b, S. 8) specifying amendments to copyright legislation, that is, the *Copyright Designs and Patents Act 1988* (United Kingdom 1998) and the *Copyright and Rights in Databases Regulations 1997* (United Kingdom 1997). These amendments provide that actions carried out under the proposed *Legal Deposit Libraries Act 2003* will not infringe the provisions of copyright legislation. However, as a Committee member pointed out, whereas the original clause explicitly made the preservation of non-print publications a legitimate reason for copying material, the new clauses do not (Thomas 2003, Col. 40). The concern was whether publishers would object to the possibility of delivering several copies of a publication, which could then be copied without infringing rights legislation, unless it is made clear that this copying is for the legitimate purpose of preservation (Thomas 2003, Col. 41).

It is understandable that publishers would not want to lose the protection of rights legislation. Again, it was confirmed that regulations made that allowed for copying would be subject to affirmative resolution (Howells 2003, Col. 43). The Minister also confirmed

that the Government department responsible for rights legislation, the Department of Trade and Industry, was “fully on board” and that the two departments were working closely together. Explicit provisions for Web harvesting were added at the third reading stage of the Bill (United Kingdom 2003c, S. 8(1)(1)).

The sub-clause relating to people and acts agreed at the Committee stage gave slightly more specific details of what will be covered by regulations: purposes for which material can be used; time or circumstances of use; definitions of who may use material and limitations on the number of simultaneous users (United Kingdom 2003b, S. 7(4)). There was also provision for redress for publishers for contravention of any regulations governing use of material (United Kingdom 2003b, S. 7(5)). This provision was developed further at the third reading stage. The clause was renamed “Restrictions on activities in relation to non-print publications (United Kingdom 2003c, S. 7). The amended clause sets out activities that cannot be carried out unless allowed by regulations. These are: using deposited material; copying material, adapting material comprising or containing a computer programme or database; lending the material to a third party other than a reader on library premises; transferring material to a third party and disposing of material (United Kingdom 2003c, S. 7(2)). There are no references to any of these activities in the provisions for the deposit of print material, although most are applicable.

Copying and adapting deposited material could cover preservation actions, for example making backups, migration or emulation (see section 3.4). The inclusion of lending may have been in response to publisher fears that legal deposit libraries lend deposited material. Disposing of material could apply to discarding the original medium of offline material if the content is transferring to an archival environment or deleting older versions of migrated content. During discussion at the third reading stage, the concern that the Bill does not specifically provide for libraries copying material for preservation purposes was repeated. The reply was that the DCMS lawyers think the provisions do cover this. Other amendments in this section recognise the need for copying by deposit libraries, because “relevant material”, i.e. deposited material, includes “a copy (at any remove)” (United Kingdom 2003c, S. 7).

While the implementation of the proposed legislation was to be left to secondary instruments, or regulations, the Bill included provision for different arrangements for

different classes of non-print material and different deposit privileges for different libraries (United Kingdom 2003, S. 8(1)). In addition there was provision for regulations on removing legal deposit privileges and conferring privileges on libraries (United Kingdom 2003, S. 6(1)). So the Bill was not only flexible in terms of future types of publication, but also deposit institutions. However, any proposed regulations that will create differences in provisions and privileges that affect the National Libraries of Scotland and Wales will have to be agreed by the devolved government in Scotland and the devolved Assembly in Wales (United Kingdom 2003, S. 8). This, however, does not apply to electronic publications where a means for accessing these publications is provided. Presumably this means that when drawing up the regulations, the decision could be taken that only one library would actually own a copy of the material. This raises preservation questions, for example could this one library be relied upon to carry out all the necessary precautions to guard against the impact of disaster. There is also the question of whether the national libraries of Scotland and Wales could end up in the position of not owning Scottish and Welsh material and what the implications of this would be. A sub-clause relating to Trinity College Dublin was added in the amended Bill. This sub-clause deals with ensuring that any entitlements would require the Secretary of State to be satisfied that “the restrictions on use of relevant materials imposed by the laws of Ireland are not substantially less restrictive than those imposed by the laws of any part of the United Kingdom.” (United Kingdom 2003a, S. 9(9)).

A quite detailed amendment to the original Clause 8 was proposed at the Committee stage but withdrawn after discussion. This involved requiring the Secretary of State to consult with a Technical Advisory Committee on the likely impact of any regulations. This Committee would also oversee the deposit of publications. The Minister had already mentioned that he favoured some sort of advisory committee, but stated that he did not think this should be specified in detail in primary legislation because it would reduce the flexibility of the Committee and the Secretary of State. The proposer agreed to withdraw the amendment after being assured that a committee really would be set up and the issues of “proportionality”, or the burden on publishers, and copyright protection would be given due importance (Moss 2003c, Col. 48, Howells 2003, Col. 48).

Concerns about potential liability for infringement of intellectual property rights or breaches of contract arising from the deposit of and access to materials raised at the

Committee stage were addressed at the Third Reading stage through the introduction of two Exemption of Liability clauses: deposit of publications etc. and activities in relation to publications (United Kingdom 2003c, Ss. 9 & 10). The first clause related to accompanying material deposited in order to make publications accessible. There was some debate on this since Mole (2003b, Col. 673) confirmed that it could only apply to contracts governed by English law. Presumably, he meant UK law, since the Bill applies to UK publishers. Mole admitted that this provision would not protect publishers from actions in other countries, but said that it was unlikely that foreign courts would act since the publishers would be complying with their national law. However, the point was made that overseas third parties might terminate their contracts with UK publishers if they were required to deposit licensed material and this may cause commercial damage to publishers (Moss 2003b, Col. 675). A proposed amendment to waive the duty to deposit in such circumstances was rejected on the grounds that it would create a potential loophole with publishers entering into contracts with overseas third parties to avoid depositing material (Morris 2003, Col. 677). It was pointed out there has never been a problem with liability with printed material, but while print material may contain content licensed from third parties it does not require expensive software to make it work.

The new clause on activities relating to publications deals with the liability of libraries for making defamatory material available. No doubt the deposit libraries will welcome this in their Web harvesting activities because they could easily unknowingly acquire such material through this method. The fact that these two clauses only cover liability for defamatory material was questioned and not other potential offences such as “negligent mis-statement, contempt of court, privacy, racism, obscenity and misleading advertising” (Moss 2003b, Col. 676).

At the Third Reading stage there was a lengthy discussion of a proposed amendment setting out what “first published from the United Kingdom” means. The aim of this was to limit application of the proposed legislation because the Bill did not do this. The focus of this amendment was online publications. Some of the participants in the discussion showed a complete lack of understanding of the nature of the World Wide Web. Mole (2003b, Col. 686) made an interesting comment “Issues to do with the location of web material are irrelevant ... because the Bill treats the role of the individual in publications as critical. The key point is who is responsible for the publication and not where it physically

sits”. One interpretation of this remark could be that a Briton resident in another country, making material publicly available from that country would be required to deposit the material in the UK. At the moment, the legal deposit libraries have to purchase print material by British authors that is published abroad. The amendment failed and this issue is to be dealt with in regulations.

This debate continued with points about the volume of material that would be potentially eligible for deposit. Another proposed amendment set out some selection criteria: works published online should “contribute to the national intellectual, cultural or scientific record.” (Lord 2003, Col. 680). The debate centred on how this could be implemented. While the mechanism of legal deposit collection policies tempered by financial constraints did not entirely satisfy participants in the debate, this amendment was dropped. There was some concern at the Third Reading stage that the proposed legislation would apply retrospectively.

The amended Bill (see Appendix 3) did give some more detail on how the enabling legislation will be enacted. This will be through statutory instrument laid before and approved by resolution by both Houses of Parliament (United Kingdom 2003b, S. 9(9), United Kingdom 2003c, S. 11(6)). The Secretary of State will have to consult both legal deposit libraries and publishers (United Kingdom 2003b, S. 9(7), United Kingdom 2003c, S. 11(2)), the Scottish Ministers and the National Assembly of Wales (United Kingdom 2003b, S. 9(2, 4-6), United Kingdom 2003c, S. 12). The Bill, as amended at the Committee Stage, provided that the Secretary of State also has to consider the impact of regulations on the publishers likely to be affected (United Kingdom 2003b, S. 6(4)). This was further elaborated at the Third Reading stage with the addition of the need to consider the costs versus the benefits of the result of regulations (United Kingdom 2003b, S. 11(4)) and whether they would “unreasonably prejudice the interests of persons who publish works” (United Kingdom 2003b, S. 11(5)). The Government minister on the Committee gave an assurance that there would be consultation before regulations were drawn up and that these would be subject to impact assessments (Howells 2003, Columns 11 & 12). The Minister also agreed with the suggestion that a way forward would be to initially concentrate on the uncontroversial offline media, whilst carrying out extensive consultation on online information (Howells 2003, Column 10).

The Government intends to create an independent advisory panel to help with the drafting of regulations. The proposed membership of this panel is “deposit library and publishing industry representatives and other stakeholders and experts” (Howells 2003, Col. 14).

The Bill made its way through the House of Lords without further amendment. However, a Lord commented that the DCMS told the House of Lords’ Delegated Powers Select Committee that the Bill would not be allocated further time in the House of Commons if the House of Lords amended it. This Committee made no recommendations for amendments, but did draw the House of Lords’ attention to two issues. The Committee suggested that the Lords ask the Government for more information on its intentions for taking the development of regulations forward. The other issue was to request that the Government also explain the need for and use of powers in clause 10(8) of the Bill. This provides for the extension of exemption from liability to categories other than defamation. A Lord stated that in practice this would mean that the Bill would be defeated and that in view of its importance and the support it enjoyed that the Lords should not put the Bill at risk (Prys-Davies 2003, Col. 573).

So the Lords passed the Bill unamended, but raised some concerns and requested various assurances from the Government that were duly given. The concerns were the same as those raised earlier in the Bill’s passage through Parliament. Another assurance given was that it was unlikely that “on-line aggregation services” providing access to content that is published elsewhere would be subject to legal deposit (Evans, 2003, Cols. 576-580). Lord Evans, representing the Government, commented that regulations for off-line material would be the first to be implemented and these were likely to be heavily influenced by the existing voluntary arrangements (Evans, 2003, Col., 577).

The Legal Deposit Libraries Act 2003 (Appendix 4) came into effect on 1 February 2004 (United Kingdom 2004). The BL press release announcing the legislation indicates that the publishers were reassured by the Government’s responses to their concerns during the passage of the legislation through Parliament.

After some difficult moments during the passage of the Bill, publishers have negotiated key reassurances from Government and crucial changes to the wording of the legislation itself. I am pleased to say this establishes a basis for publishers and

libraries to work together constructively to develop practical ways to capture electronic as well as print publications. (Mills Wade 2003)

At the time of writing, a Joint Committee on Legal Deposit had been set up (Fell 2004). Representatives of the BL and the DCF jointly chair the JCLD. All of the legal deposit libraries are represented, unlike the Joint Committee on Voluntary Deposit that had one representative for all the so-called Agency libraries. Publisher representation has also been broadened to take in the Newspaper Publishers' Association, the Association of Online Publishers, STM (science, technical and medical publishers) and Reuters. The PA, Directory Publishers Association (DPA), Periodical Publishers Association (PPA) and ALPSP continue to be represented. So this new committee does seem to encompass the range of both library and publisher interests. The Committee will apparently meet regularly and discuss all aspects of legal deposit, not just deposit of non-print publications. This will no doubt please publishers. However, the major role of the Committee is to advise the proposed Government Advisory Panel for "ongoing legislation on legal deposit" (Fell 2004).

## **2.6. Scope of Legal Deposit in Other Countries**

There is no comprehensive up-to-date international survey on the scope of legal deposit legislation. The National Library of Canada carried out a survey of national libraries in 1989, to which it received 46 replies (McCormick & Williamson 1990). Jason's (1991) international survey of legal deposit summarises the provisions for different publication types. However, this survey is now dated and digital publications are not specifically mentioned. Hoare (1996) carried out a limited international survey for the BL and updated this with an article on the situation in eleven western European countries (Hoare 1997). Martin (2001) carried out a limited survey of legal deposit in a selection of countries in 1999 and revised this in 2001. The National Library of Australia's Preserving Access to Digital Information (PADI) Web site (<http://www.nla.gov.au/padi/>) also covers a small selection of countries.

Information on the legal deposit situation in some countries is often hard to find. Even if there is easily accessible information, it can be ambiguous and different sources can be contradictory. Descriptions of what types of publication are covered by legal deposit are

often too general either to deduce whether digital publications are included at all, or what types of digital publication are covered. The situation in former Soviet Union and Eastern Bloc countries is not clear from the literature. Information on many African countries is almost impossible to find. Appendix 5 contains a table that briefly summarises what it has been possible to establish from a variety of sources. An earlier version of this summary was published in 2000 (Muir & Davies 2000). The contents of the table graphically reflect the problems of finding clear information. Since the history of legal deposit provisions in different countries is not a primary objective of this research, mostly secondary sources have been used and gaps in information have not been followed up. Where evidence of coverage of digital publications has been found, this has been followed up as far as possible.

Some countries have extended their legislation, mostly initially to cover only tangible digital publications (for example, magnetic tape and diskettes and optical discs). Legislation in some countries, for example Norway, theoretically extends to all digital publications, but in practice only certain types are collected or are collected on a limited basis. Other countries are currently taking a hybrid approach: legislation covers some types of material, with voluntary arrangements for other types. Australia and the UK have been taking this approach. Canada and Sweden have legal deposit for physical format digital material, but have been experimenting with collecting online material for some time. In the Netherlands all deposit of print and non-print material, including digital material, is carried out on a voluntary basis. This began in 1974 and the scope of voluntary arrangements has been expanded over time.

Legal deposit law in Denmark (1997) could be useful as an example in the framing of UK Regulations. This law applies to hand-held digital publications and online publications whose content does not change. While technical documents “necessary for making the work available when the technical equipment used is no longer on the general market” have to be deposited on demand, the depositor can specify that this documentation is not to be made available to any third parties. This may be a useful provision because the legal deposit libraries have information to help them preserve material, but commercial organisations do not risk valuable proprietary information becoming publicly available. For databases, the producers can inform the legal deposit library that the database is available while providing a password to allow the library to retrieve and copy information



for its legal deposit collection. This provision may be helpful in that it takes the burden of deposit off the producer. On the other hand, producers may feel nervous about doing this.

## **2.7. Guidelines for Legal Deposit Legislation**

International organisations have provided guidelines for countries planning to amend or introduce legislation. The Conference of Directors of National Libraries (1996) issued a document on the legal deposit of electronic publications. Its target audience was countries thinking of preparing a case for introducing legislation. There is also practical advice on handling electronic publications. The report contains some definitions of electronic publications. Case studies from Canada, France and the USA are included.

The Council of Europe and EBLIDA (1998) produced guidelines on library legislation and policy in Europe that included legal deposit. These state that the objectives of legal deposit should be to accumulate a national collection in order to preserve it and pass on the national culture to future generations, to compile and publish a national bibliography and to provide access to stored publications. There seems to be some overlap in these objectives in that transmitting the national cultural to future generations must surely include the provision of access. On the other hand, there is an implication that access should be current as well as for future generations.

The guidelines include suggestions on the scope of legal deposit regulations, including limiting the number of copies of print material. While legislation should cover all types of publication, different policies for different categories of material should be avoided. This is an important point for the UK, because it looks like future regulations will not only exclude sound and audiovisual material but may also include differing provisions for different categories of material. The guidelines also cover the roles and responsibilities of the different parties affected by legal deposit. National libraries and/or bibliographic agencies should provide efficient bibliographic services and adequate access to deposited material and they should cooperate with other organisations to “better serve the entire library community, as well as society at large.” (Council of Europe & EBLIDA 1998). Non-compliance with legal deposit should be sanctioned with “financial or other

measures". There is also a specific guideline on the need for research into the legal, technical and financial aspects of the legal deposit of digital publications.

The Council of Europe/EBLIDA guidelines are based on the UNESCO legal deposit guidelines of 1981 (Lunn 1981). These have now been superseded by a revised and updated edition (Larivière 2000). The Council of Europe/EBLIDA guidelines are limited to very brief statements and lack detail and it is hard to see how helpful they would be to countries setting up or amending legal deposit systems. The current UNESCO legal deposit guidelines provide a much more comprehensive framework for legal deposit and are more up to date than both the Council of Europe/EBLIDA and CDNL guidelines. The guidelines discuss the nature, role and history of legal deposit. The elements of a legal deposit scheme are set out as are the different types of material to be included. There is a separate chapter for electronic publications, although they are also included in the general chapter on types of material. This all comes together in a legal framework for legal deposit. The document discusses the various issues and difficulties of implementing legal deposit for digital material.

The guidelines give the following definition of legal deposit:

... a statutory obligation which requires that any organization, commercial or public, and any individual producing any type of documentation in multiple copies, be obliged to deposit one or more copies with a recognized institution. (Larivière 2000, p.3)

The reference to multiple copies immediately throws up a potential problem of interpretation for online material. The question is whether a commercial online product counts as 'multiple copies' because this phrase implies physical objects. Since the guidelines later refer to the need to include online material, it has to be assumed that the answer is yes.

The case for legislation is that without it, not all publishers will comply; legislation has to be enforceable (Larivière 2000, p.4). The guidelines suggest a penalty structure, with reasonable levels of fines, and that fines should be a last resort. There is no guidance on how these fines are to be extracted from publishers. While Whitaker (2001b) asks if there are alternative funding mechanisms to publishers having to provide their publications for

free, the UNESCO guidelines state that national institutions should not be dependent upon “discretionary assignment of public funds for purchasing the deposited items” (Larivière 2000, p.10).

The UNESCO guidelines also recommend that the objectives of legal deposit should be clearly stated in legislation. The reason given is to avoid national institutions prioritising one objective over another or ignoring one of them (Larivière 2000, p.10). This is relevant to the current UK situation where the aim of legal deposit is not stated explicitly. Another reason for making objectives explicit would be to avoid the situation in the UK, where only the act of deposit is specified. The custom and practice of preservation of and provision of access to deposited material has developed over time, but the lack of a statutory basis for this is a source of disagreement between publishers and deposit libraries.

One area where the Council of Europe/EBLIDA and the UNESCO guidelines are at odds is compensation for depositors. The former says this should be considered and the latter recommends that there should be no compensation.

Larivière recommends that legislation should not be retroactive and the gaps in collections should be filled through voluntary deposit or purchase. Legislation should be as broad as possible. The national library or similar institution should be the depository, although the legal deposit system can be decentralised with coordination between the depositories. The recommended minimum deposit is two copies, one for preservation and one for use, with exceptions for valuable or low volume material. For digital publications the number of concurrent users replaces the number of copies with a recommendation of a minimum of one user at a time. There are no recommendations on maximum numbers of copies or simultaneous users.

The UNESCO guidelines contain some potentially controversial points. For example, there is the recommendation that “The legislation should include a commitment by the depository that all possible and reasonable efforts will be made to permanently keep all materials deposited, but the legislation should also include a right of disposal.” (Larivière 2000, p.15). The guidelines do state that disposals should only be under certain conditions, but given the recent uproar (Brindley 2000, Ezard 2000, Hensher 2000, McCue 2000, Reynolds 2000), caused by the decision of the BL to dispose of some newspaper titles (in

fact purchased rather than deposited), any proposal to include disposal in any future legislation may well anger publishers and revive protests against the existence of legal deposit.

Another point is that of material that is “forbidden” in some way. The examples given in the guidelines include pornographic and hate material. The guidelines recommend that even though there are laws in a country to prevent the creation and distribution of such material, where it exists, it should be subject to legal deposit (Larivière 2000, p.15). The case made for this is that the value of the historical record may be compromised if material is excluded under earlier, less tolerant standards. So the material should be deposited but access should comply with legal requirements.

The various guidelines discussed above advocate the extension of legal deposit to digital material. It takes time to draft and enact legislation and while this is happening digital publications are not being collected. Various countries are operating voluntary schemes to overcome this problem.

The Conference of European National Libraries (CENL) and the Federation of European Publishers (FEP) made a declaration in 2001, advocating the immediate implementation of voluntary schemes (Conference of European National Libraries & Federation of European Publishers 2001). This declaration includes a model voluntary code for local adaptation. The model code is based on the UK scheme and should include offline and online publications. The declaration recommends that there should be a transnational steering group for these initiatives. It is not clear whether such a group has been set up, but a joint committee of the two organisations has continued to be involved in this issue by acting as an advisory body for a study on deposit of digital publications by publishers that formed part of The Electronic Library (TEL) project (Oskamp 2003).

## **2.8. Summary**

In the UK, as in some other countries, legal deposit has its origins in the control of intellectual property. However, legal deposit is now established as a means of collecting the recorded heritage so that it can be preserved and made available for posterity. Since the

recorded heritage is increasingly available in forms other than print, there has been a worldwide move towards extension of legal deposit to include ‘non-print’ material. While Norway’s early (1990) legislation used technology neutral definitions so that any future type of publication would be covered, most early adopters restricted their legislation mainly to material published on discrete physical media. In the mid to late 1990s, library and publisher organisations developed guidelines for legal deposit legislation and systems. The UK has not been a leader in extending the scope of its legal deposit regulations. Despite several reviews of legal deposit provisions during the 20<sup>th</sup> century, the scope of legal deposit remained limited to print material. Since its inception in the 17<sup>th</sup> century, there has been some resistance to legal deposit amongst UK publishers. Publisher concerns – especially the obligation to supply several free copies of material and the provision of access to that material - have not diminished over time. Despite this, new legislation was passed at the end of 2003. The new legislation has benefited from the experience of the early adopters in that it deals with legal issues and recognises the need for flexibility as new types of publication become available. Despite publisher concerns, the publishers and legal deposit libraries have worked together first to develop a voluntary deposit scheme and then on the development of legislation. The new legislation is enabling and will require further detailed negotiations for its implementation. These negotiations are likely to be difficult and will require that the Government, legal deposit libraries and publishers find ways to address the many issues that are discussed in Chapter 3.

### **3. PRACTICAL ISSUES**

The literature identifies many issues that arise from the legal deposit of digital publications. The legal deposit of digital publications presents challenges at every stage of the deposit system, from the framing of legislation to the provision of long-term access to the publications. In fact, each stage of the legal deposit process throws up legal, economic, technical and managerial/organisational issues. The issues are discussed within the framework of different stages in the legal deposit process. Many well-established concepts either do not apply in the digital environment or need redefinition. This issue is so fundamental that it is discussed separately in section 3.1. Legal and cost issues are also dealt with separately in section 3.6 as it is difficult to separate them out for each part of the legal deposit process.

#### **3.1. Definitions: Documents, Publications and Publishing**

This issue was raised during the passage of new UK legal deposit legislation and runs right through the literature. Terms that are well understood in the print environment are no longer relevant, become ambiguous or have new meanings. Examples include terms relating to documents or publishing such as ‘publication’, ‘edition’, ‘authenticity’, ‘place of publication’ and ‘publisher’. Another problem is that the same words have different meanings for different communities. For example, ‘archives’ and ‘metadata’ have different meanings for different professional communities (Russell 1999) and this may hamper cooperative working. Researchers working in this area have developed glossaries of terms. Unfortunately, these are of limited use because they have been developed specifically for the use of individual projects. Examples include the CEDARS (1999) and NEDLIB (Borbinha *et al.* 2000, Clavel-Merrin 2000) glossaries.

For legal deposit, the main issues are what constitutes a publication and how ‘published in the UK’ can be defined in a global publishing environment. Many of the problems of definition associated with legal deposit stem from the fact that the concept was originally based on mainly textual information first made available in individual nations via a physical carrier.

Some of the traditional concepts still apply to digital information on physical media. However, telecommunications and global networking have changed the nature of information dissemination. Many online publications are frequently, if not continuously, updated and they are globally available. Some 'publications' are only generated on demand. Initiatives such as the Open Archives Initiative (OAI) are influencing the nature of scholarly communication. E-print archives provide alternative access routes to formally published research papers (post prints) and rapid access to draft material (pre-prints). There is the question of whether pre-prints, in particular, are 'publications' and should be collected under legal deposit. The Internet has spawned other new types of communication, such as email, mailing lists, chat rooms, personal World Wide Web home pages and dynamic Web pages that are generated 'on the fly' from databases. How much of this information can be called a 'publication' in the traditional sense is debatable.

In 1996, the BL commissioned a study on the definition of terms to aid the structuring and wording of any future UK legislation. Existing sources of definitions were given in appendices to the study report. The study found that, at the time, existing definitions were not helpful because they did not deal well with new types of material, including digital material (Martin 1996, p. 1). One point made was that definitions should be format or medium-independent to make them 'future proof'.

As far as the concepts of documents and publishing were concerned, Martin (1996) did not suggest definitions, but provided "an overall framework of analysis within which to work". Martin defined a document as:

- (a) A combination of a work or compilation of works,  
The medium on which the work or compilation is stored, and  
Any access technology which is specific to the document;
- or (b) Any one of a number of copies of such a combination

This definition seems to be largely based on Norwegian legal deposit legislation, which has been used by other countries in their laws (Sjøvoll 1996). The intention of the Norwegian legislation is that it will apply to information in all forms, current and future. This is done through careful wording and definitions. The set of definitions used includes medium, document, publisher, producer and importer.

- Medium – a means of storing information
- Document – one or more identical copies of a medium, by which information is stored for subsequent reading, listening, showing, or transmission
- Publisher – any person who at his own expense produces or arranges for the production of a document to make it available to the public
- Producer – an person who produces copies of a document for a publisher
- Importer – any person who at his own expense brings into the country documents published abroad in order to make them generally available

In Norway, the main criterion for eligibility for legal deposit is not “the method of presentation or the size of the issue, but the availability of the information contained in it” (Rugaas 1990, p. 161). Information has to be available outwith a limited defined group. Norwegian legislation deals with editions by relating to copies of documents. New editions that are not identical to the original document are different documents. If the same information is released in more than format, each format constitutes a different document. So, unlike British legislation, copies of each format would have to be deposited. Where information is stored in a database from which paper documents are generated on demand, the database itself is not eligible for deposit, but the paper copies will be because they will be available, rather than the database.

Martin identified four classes of documents according to method of publication:

- Documents published by replication on a material support
- Documents published by telecommunication to a receiving device
- Documents published by performance before an audience
- Hybrid documents (information content is disseminated through a combination of two or more of the other classes)

Martin also defines ‘published within the United Kingdom’. This is “the public to which it is offered or broadcast or made available or before which it is performed includes a part of the United Kingdom ... and the publisher or an importer or distributor or an agent of any of the aforementioned is domiciled in the United Kingdom” (Martin 1996, p. 8). Martin



admits that this definition creates a potential loophole for publishers whose entire operation is outside of the UK, but whose offering is aimed directly at a UK audience. He also rightly points out that it would be difficult to enforce UK law in this situation.

The report also advised against extending legal deposit to online digital publications at the outset because of their complexity (Martin 1996, p. 11). This advice was based on the findings of another BL sponsored study, on the deposit of online databases (Electronic Publishing Services 1996). This study spells out the potential problems associated with depositing online material. Information available via the Internet was not explicitly covered. However, many of the problems identified apply equally to both traditional online and Internet-based services. Given EPS Ltd's (2002) findings in a later study of the costs of digital legal deposit – that the use of offline media for digital publishing is in decline – legal deposit libraries cannot afford to ignore online publications.

The 1996 EPS report provides definitions for different types of database. However, a major point is that the traditional concept of publishing is not applicable in the digital environment. The publication process is not the same in the print and online environments and different entities are involved. In the online world, no single entity has overall control of the process, and intellectual property rights are normally created at several points. The entity that owns the rights to the data may be different from the entity hosting it. The entity owning the rights to the retrieval software may be different from the data owner and/or the host entity. This raises the question of which of these entities is responsible for deposit. UNESCO suggests that for online databases the “owner” or “distributor” are responsible. (Larivière, 2000, p. 21)

The EPS report flags up the term ‘in the UK’ as problematic and predicts that attempts to differentiate between UK and overseas based entities may cause both database owners and hosts to move their operations offshore. Like Martin, this report warns that enforcement would be difficult in this case.

The Conference of Directors of National Libraries (CDNL) has defined what they call "electronic publications" (Conference of Directors of National Libraries 1996). Electronic publishing is "the use of electronic means of communication to make information available to the public". Electronic publications are "stored in computers and may be

displayed for viewing either on a computer screen or as a print-out". There seems to be an incompatibility between "stored in a computer" and the fact that the definition is supposed to apply to publications stored on physical media, such as CD-ROMs.

UNESCO guidelines for legal deposit legislation state that "In general, any type of library material, regardless of the format, as long as it is made available to the public and produced in multiple copies, should be an object of legal deposit." (Larivière 2000, p. 19). In this case, the 'size of the issue' does matter. The UNESCO guidelines use the South African definition of material to be deposited (South Africa 1997). This in turn was based on the Norwegian legislation (Larivière, 2000, p. 39), but unlike the Norwegian legislation, the production of multiple copies is a criterion for legal deposit. This could then exclude on demand publishing. It also raises a question about information that is made widely accessible online. This is whether the term 'multiple copies' applies to the offer of online access to databases held on a remote server. This point is debated in the guidelines and the arguments against are rejected because, as a matter of principle, "all electronic publications, both off- and on-line should be subject to legal deposit." (Larivière 2000, p. 19).

A working group of the International Association of STM publishers has made a helpful contribution to the definition of 'publication' in the digital environment (Frankel *et al* 2000). This work is timely given the rise in open access publishing and self-archiving. Since being 'published' usually means being available to the public, much material held in e-print repositories could fall within the scope of legal deposit. However, much of this material will not be the final version of material and it may be changed on an ongoing basis. The National Library of Australia (NLA) excludes this sort of material from its collecting activities (National Library of Australia 2003a). The international STM working group focused on the peer-reviewed article as the primary vehicle for the dissemination of research results. They have identified the "Definitive Publication" as the final published version of an article after peer review (Frankel *et al* 2000).

The definitive publication should:

- Be publicly available

- Be known to the relevant community
- Be accessible through a system for long-term archiving, preservation, retrieval and access
- Not be changed
- Not be removed
- Be unambiguously identified
- Have a bibliographic record

This could be helpful in the selection of research material, but will be of limited use for other types of digital material.

It is interesting that the draft UK legal deposit legislation does not use any of the definitions suggested above. However, the issue of definitions will have to be addressed when the detailed regulations for implementing the legislation are drafted. Despite the work that has been carried out in the area since the 1990s, there is still a need to decide what ‘published in the UK’ means. This was a finding of a survey carried out in 2002 (Electronic Publishing Services 2002, p. 40), and is also a task of the new Committee on Legal Deposit (see chapter 2).

## **3.2. Discovery, Selection, Acquisition**

### **3.2.1. *Discovery***

The 1996 ELDEP study reported that the amount of material published only in digital form was quite small compared with the volume of traditional publishing output (Mackenzie Owen & Walle 1996, p. 17). This view was repeated in 1999 in another study (Bide *et al.* 1999). These studies are now dated, but they both stated that the proportion of published output released only in digital form was likely to increase over time.

In order to acquire all the information they are entitled to, depositories have to discover its existence. Bibliographic control is well developed in the print environment, less well-developed for non-print formats and, it seems, virtually non-existent for digital

publications. One area of concern is that of unique identification of digital publications. Current identification mechanisms include the International Standard Book Number (ISBN), International Standard Serial Number (ISSN), Book Item and Contribution Identifier (BICI) or Serials Item and Contribution Identifier (SICI) systems. While identification of offline digital material such as CD-ROMs may be reasonably straightforward using identifiers such as ISBNs, Unique Product Codes (UPC) or European Article Number (EAN) barcodes, online material presents problems (Bide *et al.* 1999, p. 21).

One problem is that there may be different manifestations, or versions, of the same content. The question is whether each manifestation should have a different identifier, or whether there should be one identifier for the underlying work. This raises the further question of how to identify each manifestation and relate this to the underlying work. Another problem in the online environment is the lack of persistence of identifiers. Universal Resource Locators (URLs) identify World Wide Web documents. Unlike traditional identifiers that identify items, the URL indicates the location of the item. URLs change over time as Web content is rearranged; they lack persistence. This causes problems when documents are linked because the links will break. It also causes problems with finding documents over time. There are several initiatives exploring the development of persistent identifiers.

Identifiers used for Web material are called Uniform Resource Identifiers (URI). URIs can be locators, or names, or both. URLs are a type of URI. The Internet Engineering Task Force (IETF) Uniform Resource Names (URN) are also URIs, but these identify resources by name or attribute rather than by access mechanism. URLs are widely used, whereas the development of URNs has been slow. This is because there has been a lack of agreement on how the URN idea should be formalised. Sollins and Masinter (1995) set out the minimum requirements of the URN. The new types of identifier being developed include the Digital Object Identifier (DOI), an implementation of the Uniform Resource Name (URN) approach. A DOI consists of a prefix and a suffix divided by a forward slash. The prefix is assigned to the organisation that wants to register the DOI in a similar way to ISBNs. In fact, an ISBN can be used as the prefix of a DOI. The prefix is followed by a unique suffix. The DOI can be assigned to any item of intellectual property, but this must be clearly specified. This is done through the use of structured metadata and a metadata

dictionary is part of the system. Once assigned, the DOI is not altered even through successive ownership changes. In order to ensure persistence the DOI system uses a resolution system, the Handle system (Corporation for National Research Initiatives, 2002). Rather than relying on the original URL, the DOI resolves to current information such as the current URL (International DOI Foundation 2002).

Persistent Uniform Resource Locators (PURL) are URLs that do not point directly to the item, but to a resolution service. The resolution service then points the PURL to a URL. OCLC (Online Computer Library Center) developed the PURL system and assigns PURLS to resources catalogued in its Internet Cataloguing Project (OCLC 2002). The part of the URL before the first forward slash (the domain name) is for the resolution service rather than the actual location of the object, for example [www.purl.oclc.org/](http://www.purl.oclc.org/). While the URL may change, PURLS cannot be changed, so are persistent. OCLC considers PURLS to be an intermediate step in the move towards URNs.

While work is being carried out to address the problem of unique identification and persistence of identifiers, until coherent systems are implemented there is still the problem of knowing what exists. One suggestion here is that legal deposit could require all publishers to register their publications (Webb 1997, p. 11). The existence of publications would then be known, even if they were not all collected. This idea is a good one in theory, but in practice, it may be impossible to enforce precisely because of the large number of publications involved and the possible ignorance of many Internet ‘publishers’ of legal deposit systems.

Despite the problems with identification, the British Library (BL) estimated that in 2002 it was receiving around 75% of all material published on physical carriers and between 45-50% of “electronically delivered publications (mostly e-serials)” (Electronic Publishing Services Ltd 2002, p. 18). However, the BL does not know how much of this material is significantly different from printed counterparts. This was suggested as an area for research before digital legal deposit legislation comes into force (Electronic Publishing Services 2002, p. 21).

### 3.2.2. *Selection*

Mackenzie Owen and van de Walle (1996, p. 21) point out that legal deposit laws are often selective in their coverage: some types of material are included and some are not. For example, current UK deposit law excludes types of printed matter. As was shown in chapter 2, different countries take different approaches. Mackenzie Owen and van de Walle recommend that all digital publications should be collected, including those published in parallel with print equivalents. However, some exceptions are specified:

- Items that cannot be physically acquired and stored under the full control of the deposit library
- Very frequent updates – snapshots should be taken
- Items distributed by private individuals
- Items distributed by publishers other than trade publishers if insufficient resources and expertise are available
- Highly distributed, open networked publications
- Public communications
- Informal publications and pre-prints
- Parallel digital versions
- Publications that cannot be handled and preserved within the technical environment of the library (Mackenzie Owen & Walle 1996, p. 23)

Some writers, such as Rugaas (1990) have questioned whether comprehensive collection of material is still a feasible proposition. The main reason behind this is the widely acknowledged information explosion and the inability of libraries to cope with increased national outputs. Rugaas describes this debate as “the end of all and forever”. Legal deposit libraries will have to accept that they may never be able to collect all digital publications. There will be too many publications, too many publishers, and the rate of technological change is too fast.

Under the UK voluntary deposit scheme, the BL has not been actively selecting material; it has been passively receiving it.

However, the other legal deposit libraries, in particular the national libraries of Scotland and Wales, have apparently been proactively looking for digital material relevant to their collections (Electronic Publishing Services Ltd 2002, p. 19). While the BL has always aimed to be the most comprehensive in its legal deposit collections, the volume of material potentially eligible for digital legal deposit is likely to require some filtering so that material of value is top priority. This will require cooperation between the acquisitions and curatorial staff (Electronic Publishing Services Ltd 2002, p. 19). This will have to be well managed since the processing staff is based in Yorkshire, whilst the curatorial staff is based on London.

A study of the cost implications of extending UK legal deposit to non-print publications suggested some selection priorities for the BL. These are:

1. A high proportion of material that only exists in digital form
2. Material with significant additional content and functionality

Material that is more or less a digital facsimile of printed material should be of lower priority. Two areas of staff responsibility, which really apply to digital serials, are identified: identification of what exists and dialogue with publishers to ascertain which digital publications have additional content and functionality (Electronic Publishing Services 2002, p. 19).

Selection problems also apply to Internet publications. The National Library of Australia has perhaps the most developed selection guidelines for online material. The selection criteria are wider than that for traditional legal deposit in that they focus on material of Australian interest no matter where the information is located. On the other hand, the selection criteria address many of the questions of definition, such as the boundaries of a digital publication, what will be considered publications and therefore considered for collection. Internet material that will not be collected include Web logs, datasets, games, news sites, organisational records, digitised material that is mounted on the Web but is not accompanied by any interpretation or other intellectual input and sites that are compilations of material from other sources. Print versions are usually preferred to digital versions, except where there is a significant difference and online versions are preferred to offline versions because online is apparently easier to preserve.

This is interesting since early work digital legal deposit in the UK assumed that it would be more difficult to preserve online material. The National Library of Australia admits that it has not solved the problem of how to archive Web content that is in the form of a database (National Library of Australia 2003a).

### **3.2.3.            *Acquisition***

The acquisition of dynamic digital information is particularly problematic. UK publishers have commented on the impracticality, or even impossibility, of capturing every version of databases that are amended very frequently or in real-time (Publishers Association 1997, (Para. 3.7). The acquisition method here would be samples or snap shots. However, Mandel (1996, p. 45) stated in 1996 that there is no commonly accepted practice and little practical experience of sampling techniques and the deposit of databases was still a concern for publishers in 2002. An albeit very small survey of publishers by Electronic Publishing Services Ltd (2002, p. 40) found that publishers questioned how very large databases that change every day could be deposited. Hyper linked documents present problems of deciding where the boundaries of the documents are. Decisions have to be taken on which linked sources should also be selected.

With traditional publications, deposit usually means that some responsible entity sends physical objects to depositories. The situation is more complex in the digital environment. Online publications are not available in physical form and they cannot just be sent through the post. At present, there are three main options for acquiring online information. Publishers can transfer the information onto a physical medium and send that to the depositories. Publishers can arrange to transfer, or ‘push’, information to depositories via networks. Spivey (2002) described the experience of Emerald, a journal publisher involved in experimenting in the deposit of online journals with the BL. The first question arising was what constituted an ‘online’ journal, and therefore what should be deposited. Spivey (2003, p. 213) identified sixteen elements, including full-text article and non-article content, such as editorials, book reviews and conference reports. In addition to this material there was also information about the journal, such as details of editors, editorial boards, notes for authors and subscriptions.



Finally, there were elements contributed by the organisation providing the publishing platform for Emerald, including the search and retrieval functions and linking all the component parts of the journals together. While Emerald could provide content and content related material, it is stored in different parts of Emerald's system and would require changes to this system. Otherwise, Emerald would have to ask its service provider to deposit material, and this may incur charges (Spivey 2003, p. 215).

Alternatively, libraries can 'pull' from publishers' sites themselves. A variant of this activity is 'harvesting'. The depositories use software to identify and pull information from sites. Another suggestion is that there would be a link from the legal deposit library to the Web version (Electronic Publishing Services 2002, p. 40). However, then there is the question of whether this will have to be paid for, and if not, how to control this access in order not to harm the publisher commercially. This would be a particular problem for publishers whose core customers are libraries (Electronic Publishing Services 2002, p. 40).

Several depositories are working with harvesting software to acquire publications, including the National Libraries of Australia, Denmark, France, Finland, the Netherlands and Sweden. Harvesting information is problematic in that existing tools do not entirely meet the needs of the depositories at the moment. There are also legal issues that arise. These include copyright because pulling information from web sites involves copying. These issues are discussed in section 3.6. Different legal deposit libraries are taking different approaches. Generally, smaller countries have been experimenting with comprehensive collection of material, whereas larger countries started with a selective approach. However, the National Libraries of Australia and Canada both started experimenting with selective approaches to acquiring online information in the mid-1990s (Gatenby 2002). The Kungliga Biblioteket in Sweden has been running the Kulturarw<sup>3</sup> project since 1996. A snapshot of the "Swedish Web" is taken twice a year.

Gatenby (2002) has discussed the advantages and disadvantages of selective archiving. Selective archiving allows quality-controlled selection of resources according to selection criteria. Technical problems and access arrangements are dealt with on a case-by-case basis. However, this approach is resource intensive and the amount of information collected may be a very small proportion of all that is available.

In addition “the selective approach takes resources out of the Web context in which they reside and consequently they can lose some of the meaning that the context provides” (Gatenby 2002). While comprehensive harvesting is more automatic, it is not without its problems. Currently technology cannot deal with the so-called ‘deep Web’ and can only gain access to publicly available sites, the quality of material gathered will also be variable and there are legal issues involved with copying and providing access to material (Gatenby 2002). These issues have been acknowledged in the new UK legislation. Gatenby (2002) comments that only the libraries that started experimenting in the mid-1990s have routine procedures in place for collecting online resources. She also points out that national libraries are moving towards a hybrid approach to gathering online resources to off-set the disadvantages of each approach. In fact, the Bibliothèque nationale de France has experimented with different approaches to Web archiving and has proposed a mixed approach or “integrated functional model” (Giannattasio 2002).

### **3.3. Accession, Processing and Storage**

#### **3.3.1. *Accession and Cataloguing***

CEDARS sets out the activities required to add digital material to collections (CEDARS 2002a, p. 14). These include:

- Receipt of a physical object
- Verification of the object and its integrity
- Take decisions on the long-term role of material and assessment of its Significant properties for preservation purposes
- Cataloguing, including assigning unique identifiers and preservation metadata
- Classification
- Provision of discovery aids
- Application of appropriate access controls and security
- Consideration of technical preservation strategies

Most of these will apply to legal deposit libraries, although decisions on the long-term role of legal deposit material may be less necessary. In 1996, the BL stated that initial processing of non-print material would be similar to that of print (see section 3.7.1).

Activities involved include: logging of receipt; checking exterior packaging; assigning accession numbers and claiming procedure. However, some additional activities would be required, including: checking the correct number of copies have been sent; ensuring that material received is in the preferred format (e.g., for PCs rather than Apple computers); all application software normally sold with the publication is present; log receipt and check all user and technical documentation and signing licence agreements where appropriate.

According to Steenbakkers (2000, p. 17), registration of off-line material can be carried out using existing systems in legal deposit libraries. Mackenzie Owen and Walle (1996, p. 29) recommend that quality checks and functional tests should be carried out for all items received.

The purpose of such procedures is to check that the item is:

- The correct version
- In the required medium and format
- Complete
- Undamaged
- Error free and fully functional
- Not copy protected

Off-line material has to be installed in order to carry out checks, which is why the existence of copy protection mechanisms is problematic. The NEDLIB project recommended that online material is loaded into a temporary storage space for checking purposes before it is moved to archival storage. Offline material should be copied from its original storage medium to the institution's archival storage (Steenbakkers 2000). This recommendation is also made by the CEDARS preservation project (CEDARS 2001). It is a departure from traditional legal deposit procedures, where material is stored largely in the form it is published (book jackets may discarded or stored separately), unless it starts to deteriorate. The transfer from the original publication environment may be a concern to

publishers, because there are then issues about how widely material can be made available via networks.

In 1996, the BL outlined how it thought publications should be tested – the “first handling” stage (British Library 1996, p. 8). CD-ROMs would be loaded and both the publication and the associated software would be checked to ensure the following:

- Supplied software contains no viruses
- Material is not copy protected
- Material is in working order

However, apparently only limited checks would be carried out as more rigorous testing would be carried out on a sample basis. While the BL was not proposing to separate the content from its original medium as advocated by the NEDLIB project, the BL had been advised to experiment with downloading some data from CD-ROMs for preservation purposes (British Library 1996, p. 9)

Ensuring the authenticity of documents could be more difficult for digital material (Mandel 1996, p. 456). While the checks mentioned above will address this issue, there are other techniques such as time stamping and digital signatures. Since legal deposit libraries are providing a record of the intellectual output of the country, authenticity is something they will have to consider both in terms of the material acquired and how it is managed over time. This issue is discussed in sections 3.4.4 and 3.4.5.4.

The BL has been struggling to cope with the amount of digital material it has been receiving under the voluntary scheme. It seems that electronic monographs are not a cause for concern. The BL received 332 in both 2000/01 and 2001/02 (Electronic Publishing Services Ltd 2002). More recent figures were not publicly available at the time of writing. Processing of this material has been slower than for print. This is both because of the need to load material and there have been password and encryption problems. This has at times required members of staff to discuss solutions with publishers (Electronic Publishing Services Ltd 2002, p. 19). Despite also being more difficult to catalogue than print material, these items have been processed without building up a substantial backlog.

The volume of digital serials to be dealt with is the main area of concern “both in terms of absolute volumes and rate of growth” (Electronic Publishing Services Ltd 2002, p. 16). The majority of the digital serials being deposited fall outwith the scope of the voluntary scheme, i.e., they are not distributed on physical carriers. During the 2001/02 period, the number of serials on CD-ROMs deposited was around 75 per month, whereas 20,000 items were received by email. The volume of material deposited increased over the course of the year.

A study of the situation of other UK legal deposit libraries in 2002 highlighted various issues relating to the processing of material (Electronic Publishing Services Ltd 2002, pp. 32-36). Despite taking on a new member of staff, the Bodleian has also been unable to deal with the volume of digital serials received. Material received via email is not being catalogued, except for one title – the *BBC Monitoring Service Global Newslines*.

The majority of material received by Cambridge University Library is via the Copyright Libraries Agent. At the time of the survey, none of the other libraries was cataloguing digital serials. The NLS is apparently receiving fewer digital serials than the other legal deposit libraries. Some of the libraries were recruiting staff to deal with deposited material. The NLS has recruited two new staff to catalogue digital material and a more senior post covering selection and accession. The National Library of Wales has recruited one new member of staff in acquisitions and two cataloguers. The NLW has also reorganised its systems section to deal with digital deposit.

### **3.3.2.            Storage**

The storage issues for digital and non-digital material will differ to some extent. While physical space may be less of a problem for digital material, there are still issues about how material should be stored and the storage capacity required. There are various alternatives for long-term storage, including magnetic disc, magnetic tape and optical disc technology. There are trade-offs between speed of access and costs (Steenbakkens 2000, p. 15). Different types of material take up more space, for example audio and images take up more space than text-based material. While audio and moving image material is not

included in forthcoming UK legislation, the legal deposit libraries may find that they have to store so-called multimedia material. Another storage issue arising from proposed UK legislation is that the libraries may also be storing accompanying software and documentation in digital and non-digital formats.

Since digital material issued on physical carriers usually has some protective packaging, it is being stored in a similar way to print material in the BL (Electronic Publishing Services Ltd 2002, p. 18). Only an accession record is being created for digital serials, then the material is burned to CD and stored (Electronic Publishing Services Ltd 2002, p. 16). The material is not being catalogued and is therefore unavailable to users.

However, staff at the BL think that it may take less time to process electronically delivered material because there are fewer loading problems than with hand-held publications (Electronic Publishing Services Ltd 2002, p. 20). On the other hand, the volume of material to be processed is the more significant problem.

In 1996, the BL envisaged using jukebox systems for high volumes of data, but aiming for an integrated system to store all digital material, including material digitised by the Library (British Library 1996, p. 12). The BL (1999) issued a briefing document in 1999 on the Digital Library System it wanted to procure. The procurement process never came to fruition. The BL is currently in the process of establishing an interim Digital Object Management (DOM) system. This will provide storage for all deposited material (Electronic Publishing Services 2002, p. 22). In 2002, CUL and the national libraries of Scotland and Wales were also in the process of acquiring storage systems for digital material. CUL has been implementing the Massachusetts Institute of Technology DSpace digital repository software (University of Cambridge 2003). While this software is to be used to store born digital material produced in the University and also material digitised by the University Library, it could also perhaps be used for legal deposit material.

### **3.3.3.            *Preservation Metadata***

The amount and types of metadata gathered at the accession stage will affect preservation of and long-term access to digital publications (Bide *et al* p. 22). On the other hand, decisions on how material is to be preserved will have an impact on the amount of

metadata that will need to be stored to ensure material remains accessible (CEDARS 2002a, p. 15). Deposit libraries will need more and better from publishers at the point of accession than is necessary for printed information. There is also a need for some standardisation in this metadata (Bide et al, p. 22). The NEDLIB project found that publishers cannot always provide all the required information and a lack of standardisation in the metadata that was provided (Steenbakkers 2000, p. 17). The ELDEP study particularly focused on the bibliographic control of digital legal deposit collections (Mackenzie Owen & Walle 1996, pp. 59-92).

The BL has recognised the need for technical and administrative metadata to facilitate the management and preservation of digital publications. The BL has been represented in international work on preservation metadata (see chapter 4). The BL also produced a set of metadata to meet its own requirements (Clayphan 2001). Whether this initial metadata set will be the one used when digital legal deposit is implemented is not clear. Since there are likely to be as many technical and administrative metadata elements as traditional descriptive elements, there will be extra work to do, particularly since the BL assumes that only 20% of this metadata can be captured automatically (Electronic Publishing Services Ltd 2002, p. 23) in the short-term.

### **3.4. Preservation**

Digital preservation involves "the storage, maintenance, and accessibility of a digital object over time" (Russell 1999). Preserving information artefacts ensures the preservation of information they contain. This is unlikely to be the case in the digital environment because there is an intermediary technology layer between the digital object and its reader. Of course, this is also the case for some other types of non-print material, including analogue sound material. Even if a tape or a disc can be kept in good working order, the technological environment required to read it may no longer exist, making the information stored on the medium effectively inaccessible. This section briefly describes the problems of preserving digital information and then discusses preservation strategies.

### **3.4.1.            *Media Stability***

The literature reflects a sharp increase in interest in digital preservation since the mid-1990s. However, the earliest references in the literature come from the mid-1980s, with the introduction of CD-ROM technology into libraries. Early studies in the BL and the Library of Congress either ignored or only mentioned the potential of the new technology as a preservation medium (Clarke 1986, Price 1985), although the Library of Congress has since carried out research on the preservation qualities of optical media.

Estimates of the likely life expectancy of various storage media vary from around 1 to more than 100 years (Rothenberg 1999a, p. 3, National Media Laboratory 1995). Even the high-end estimates do not compare well with established archival media such as permanent paper or preservation microfilm. Like other media, the physical carriers used for digital information react to environmental factors (Kranich 1998, p. 138, Feeney 1999b). Van Bogart produced a report on the storage and handling of magnetic tape (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.

### **3.4.2.            *Technological Obsolescence***

However, media instability is not the main problem for the preservation of digital information; it is technological obsolescence of the hardware and software used to create and use it. This is not a new problem; earlier examples of this include Readex Microcards. Recognition of technological obsolescence as the main threat to the long-term survival of digital information became 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, thereby



requiring customers to constantly upgrade technology and perhaps locking them into proprietary technology.

### 3.4.3. *What to Preserve*

According to Thibodeau (2002, p. 5), “it is the essential character of the information object, not the way it happens to be encoded digitally, that must be preserved.” The question is how to identify, define and develop strategies to retain this “essential character”. Thibodeau continues (2002, p. 6),

... the properties of any digital object are inherited from three classes. Every digital object is a physical object, a logical object, and a conceptual object, and its properties at each of these levels can be significantly different.

Table 3-1 briefly explains each of these classes of objects.

Object Class	Explanation	Examples
Physical object	Inscription of signs on a medium	Magnetic tape - +ve & -ve charged particles CD-ROMs - pits and lands
Logical object	Object that software can recognise and process – encoding format	ASCII, PDF, MS Word JPEG, TIFF WAV, MP3
Conceptual object	Object recognised and understood by humans	Journal article, book, Web page, map

**Table 3-1 Physical, Logical and Conceptual Digital Objects**

The relationship between levels can be simple or more complex (Thibodeau 2002, p. 11). There may well be a one-to-one relationship, for example in the case of a simple text file. Alternatively, the relationship may be one-to-many or many-to-many. For example a Web page may include text, images, sound and links to other Web pages.

Physical and logical objects can be changed without changing the conceptual object. For example, a file can be copied from a magnetic hard disc to an optical disc, or text and images can be encoded in different formats, and the conceptual object will still look the same. Preservation purists could argue that any change is unacceptable, but in order to

keep digital information accessible and meaningful over the long-term some changes will have to be made. According to Rothenberg (2000a, p. 56), “we must choose what to lose”. This is not a new concept in preservation. While a microfilm version of a printed newspaper will retain much of the information content of the original, other features, such as the experience of holding and reading a broadsheet, will be lost. Digital information cannot be preserved as a physical object in the digital environment. We can only preserve the “ability to reproduce the document” (Thibodeau 2002, p. 16).

Assessing how the information will be used in future is one way of making this decision. Rothenberg (2000a, p. 57) gives some examples of how different uses would impact on what is preserved. An extreme example is that the digital material should “retain as much as possible of the exact function, form, appearance, look, and feel that the [sic] it presented to its author or publisher ... to enable future researchers to evaluate the range of alternatives that were available to the author”. Another example is that the function, form, appearance, look, and feel as presented to its original audience should be retained so that future users could use it in the same ways.

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?

It is unlikely that staff in legal deposit libraries can predict all likely future uses of digital material. Wheatley has introduced two further issues to be considered: time and cost. Neither of these are new in preservation management. Thibodeau (2002, p. 15) suggests four criteria to apply to selecting preservation methods. These are that methods should be feasible, sustainable, practicable and appropriate.

The CEDARS project investigated the idea of Significant Properties. Significant Properties refer to the level of content and functionality to be retained (CEDARS 2001, p. 14) and equate to Thibodeau's "essential characteristics". The Significant Properties "determine the underlying technical components that need to be documented and supported to ensure preservation of those Significant Properties" - or the "Underlying Abstract Form" (CEDARS 2001, p. 14). Thibodeau's criterion of appropriateness reflects the CEDARS (2001, p. 14) suggestion that 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. In order to ensure that the significant aspects of the conceptual object are preserved, the underlying physical and logical objects and their relationships have to be known (Thibodeau 2002, p. 11).

#### **3.4.4.            *Authenticity***

This acceptance that preservation will involve change, raises questions of authenticity of preserved digital material. Authenticity refers to "the degree of confidence a user can have that the object is the same as that expected based on a prior reference or that it is what it purports to be" (National Library of Australia 2002b). One way of determining authenticity of a publication is to compare it against a reference copy (Levy 2000, p. 30). While it would be relatively simple to provide reference copies of print material, this would be problematic in the digital environment. If there are no stable digital objects, then we have to find another way of determining authenticity (Levy 2000, p. 30). Levy suggests keeping audit trails of changes leading back to the creation of the document or to an authentic version. Mandel states context also has to be preserved and that the concepts involved in preserving context are the same in digital and older formats (Mandel 1996, p. 458). These concepts are fixity and provenance. For Mandel, fixity relates to the reasons why a particular version was selected, and how this version relates to other versions, which may not be extant. This term has a different meaning in the preservation metadata schemes that are currently being developed; the term context is used for this purposes (see section 4.3.1). Provenance relates to the origins and custody chain of the information.

Rothenberg (2000a, p. 57) discusses a strategy for defining authenticity in the digital environment that focuses on the originality of the information or whether it has been

altered from its original state. This strategy does not involve audit trails but “intrinsic properties” and provides criteria for assessing whether each of these properties is present and unaltered. Rothenberg points out that using audit trails means relying on the authenticity of the documentation about an object rather than these intrinsic properties. So we are back to the CEDARS idea of Significant Properties and how to define these. Rothenberg (2000b, p. 56) pointed out that we have no definitive taxonomies for either the attributes of “digital informational entities” or the uses that will be made of them in the future. Thibodeau has suggested that it is easier to describe the physical and logical objects than the significant properties of the conceptual object. According to Rothenberg (2000b, p. 57), we cannot yet reliably and consistently define attributes or uses. However, he does try and set out what a ‘suitability strategy’ would involve, “tactics to specify and test whether an informational entity fulfils [sic] a given range of purposes or uses.”

Rothenberg anticipates that criteria for assessing authenticity could differ between disciplines and that they may not be able to anticipate future needs. He suggests that “authenticity principles” should be developed. These would encapsulate “the overall intent of authentic preservation from a given legal, ethical, historical, artistic or other perspective” (Rothenberg 2000b, p. 59). It would be desirable to develop a small set of alternative authenticity principles that would cover most perspectives (Rothenberg 2000b, p. 60), so the range of expected uses will have to be described.

Different uses could result in alternative principles, for example:

- Same to all extents and purposes
- Same functionality and relationships to other information objects
- Same “look and feel”
- Same content
- Same description (metadata, the object could be discarded) (Rothenberg 2000b, p. 62)

A set of authenticity criteria could be derived from each principle. Rothenberg (2000b, p. 66) discusses the concept of a “digital-original” as “any representation of a digital informational entity that has the maximum possible likelihood of retaining all meaningful

and relevant aspects of the entity.” There is no single, unique original; originals may exist in different collections and in different contexts. According to Rothenberg, if digital-originals can be created, then a single preservation strategy will be effective across disciplines, however they define authenticity. This notion was at odds with most other writers in the field at the time.

### 3.4.5. *Preservation Strategies*

There are a number of potential preservation strategies that address different preservation requirements and timeframes. Various works provide snapshots of the state of the art at a particular period in time, including Fresko & Toombs (1998), Feeney (1999), Yakei (2001), Lee *et al* (2002), the Council on Library and Information Resources (2002) and CLIR with the Library of Congress (2002). The main possible preservation strategies at the time of writing are migration and emulation or some mix of the two. The focus of migration is on changing the information to make it readable in a new technological environment, the aim of emulation is to preserve something of the original technological environment. True technology preservation involves no change to the digital information other than refreshing it and preserving the original hardware and software to access it. While technology preservation is largely discounted as nothing more than a stopgap measure, there has been ongoing work on migration and emulation over the last few years.

#### 3.4.5.1. Migration

Migration has been defined 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 in 1999. Simple strategies included "refreshing" information and media migration to combat deterioration and obsolescence in storage media respectively. Conversion strategies to combat software obsolescence relied on backward compatibility of new application software (Feeney 1999b, p. 44) and interoperability between rival softwares. The first of these would require regular migration of material because backwards compatibility may only last for a few generations, and the second could result in loss of information or formatting.

A final migration strategy discussed by Feeney (1999b, p. 45) involved conversion to standard formats to address the problem of 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. Alternatively collection managers only accept publications in one of the preferred formats. Aschenbrenner (2004) argues that a small number of standard formats will not meet all the requirements of individual software applications, so they may hinder active use. He does concede that they may be useful for preservation purposes. 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”.

Suggestions for potential standard formats that would apply to legal deposit libraries include an archival version of Adobe’s Portable Document Format (PDF/A) and XML (eXtensible Markup Language). Work is ongoing to “specify the use of PDF in a manner that is specifically geared to long-term management and use” (LeFurgy 2003). This format could be useful for documents whose content and appearance should not change. The PDF/A format was going through the ISO approval process at the time of writing. XML is text-based and therefore easily readable by people; XML is also easily processed by computers (Digitale Bewaring Testbed 2001, p. 8).

However, Aschenbrenner (2004) argues that the human readability that would be useful for long-term preservation is not *inherent* in the XML format and that there are other data formats that are more easily understandable. It is difficult or impossible to use XML for images, audio, video and 3D simulation models (Aschenbrenner 2004). Apparently using XML to encode documents may not even retain the appearance of documents (LeFurgy 2003).

While the use of standard formats 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. In addition, legal deposit libraries may not be in a position to dictate which formats they are willing to receive. Rothenberg (1999a) also criticised migration. He likened 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). 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 200a, p. 1).

Wheatley (2001) considers the migration strategies described by Feeney (1999a, 1999b) as "traditional" and attempted to further clarify the meaning of migration. He mentions the categories of migration included on the Open Archival Information System (OAIS) Reference Model (see section 4.2). 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. Only transformation is real migration because the bit stream is actually transformed. The OAIS model also specifies that migration can be reversible or non-reversible. The reversibility of preservation actions is part of the ethical code of conservators. 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) calls the basic activity of preserving the bit streams minimum preservation. His migration categories are: minimum migration; preservation migration; human conversion migration; automatic conversion migration and recreation. In fact, Wheatley sees recreation as a separate, but related, process because recreation creates a wholly new digital object. This represents the significant properties of the original, but does not incorporate any elements of the original digital object. It may not be appropriate to do this in the context of legal deposit, except in very exceptional circumstances.

Minimum migration involves a minimal amount of technical work to make material viewable. It looks like 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 key processes involved in information use. These actions could be accompanied by textual descriptions of the look and feel. 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 by Feeney. The CEDARS migration on demand approach has been further developed through the CAMiLEON project emulation work, described below. In theory, the CAMiLEON migration on demand tool gets over some of the disadvantages identified by Rothenberg and others (Mellor, Wheatley & Sergeant 2002). The original bit streams 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 evolves. So, the problem of introducing errors that are then compounded through successive migrations is removed. The coding of the input module never has to be rewritten, saving resources. However, there is still the possibility of loss of information if a poor migration tool is developed.

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. 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.

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 changes, the essential properties stay the same. A respectful conversion has occurred when the input and output of a conversion “cannot be distinguished when viewed through the interface of that type” (Ockerbloom 2002). Ockerbloom (2002) describes the TOM idea as a

... data model for describing a wide variety of data types and formats. In a broader sense, the term “TOM” is also used to describe a supporting architecture built



around mediator agents called “type brokers”, that receive and maintain descriptions of data formats, describe them to clients, and contact servers that interpret and translate data in those formats.

It sounds like this approach is similar to the migration on demand approach and a way of implementing Lynch’s canonicalisation idea (see section 3.4.5.4). TOM is apparently flexible in that it allows users to define new types, encodings and conversions without having to go through a centralised standards body as long as they follow the TOM rules for type definition.

### 3.4.5.2. Emulation

The aim of emulation is to retain the look, feel and functionality of digital information through the use of software that allows new technological platforms to mimic the behaviour of older platforms. Granger has briefly described different approaches to emulation (2001).

Rothenberg (1995) has been a prominent supporter of emulation. Emulation can be carried out at different levels, including application software used to create and view material, operating system or hardware. 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. Rothenberg initially thought that this would need the development of emulators, or software programmes 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 (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 (1999b) set out the requirements for implementing emulation of hardware. 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 the “ancillary” issue of intellectual property rights that would arise from the use of emulation. The legal issues of digital preservation are discussed in section 3.6. All that needs to be said here is intellectual property rights is not just an ancillary issue.

Rothenberg’s emulation approach has its critics. Bearman (1999) 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.

Bearman may well be right in his claim that emulation may be overkill in certain situations. However, in making the comments about records and evidence, Bearman is perhaps unfair in dismissing preservation of functionality as a legitimate aim of preservation strategies for digital publications that are increasingly rich in functionality.

Rothenberg further developed his approach through an experiment in emulation with the KB, as part of the NEDLIB project. The aim was that the experiment would go through several iterations between 1999 and 2001. At the time, the NEDLIB project was building a demonstrator digital deposit system and publishers participating in the project were to providing material 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 test bed. 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. Layers of emulation would be required (Rothenberg 2000a, pp. 7-13); the so-called "emulation virtual machines" that would run on a future platform would have to be emulated when 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. This sounds a very complex approach.

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 2000a, p. 83). This claim seems somewhat inflated since the experiment actually involved running Windows 95 publications on an Apple Mac using Connectix VirtualPC software as the emulator. The most Rothenberg should claim is that this particular software does what it says it does. The NEDLIB project has now ended and it is not clear whether later stages of the experiment were completed. The project leader, the KB, is now working with IBM on the use of the Universal Virtual Computer approach to long-term preservation (see section 3.4.5.3) and Rothenberg has some involvement in this.

Holdsworth and Wheatley (2000) of the CAMiLEON project were nervous about Rothenberg's reliance on the production of specifications for emulators to be built in the future. Even if carefully written, there is no guarantee that a specification could be used to build a future emulator and no way of testing the success of the emulation because it will not be possible to compare it with the obsolete original technological environment. Holdsworth and Wheatley suggested 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 later generations of computers. They have also considered the level at which emulation should be carried out. Unlike Rothenberg, Holdsworth and Wheatley recommend emulation at the Application Programming Interface (API) 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. Thibodeau (2002, p. 19) considers that emulation of application software would be more difficult than operating systems because there are more of them.

Neither the Rothenberg nor CAMiLEON approaches have been tested by practical experience over the long-term. However, the CAMiLEON team have at least successfully emulated an obsolete system (the BBC 1986 Domesday Project), unlike Rothenberg who only demonstrated emulation on contemporary but non-compatible platforms.

### 3.4.5.3. 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 similar to the CEDARS/CAMiLEON migration on request approach in 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 who then decides whether to import it to a current format. 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 key could be changed to accommodate new technologies. IBM has carried out a 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 working on this approach it must be assumed that Thibodeau misunderstands the UVC approach because he says "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).

#### 3.4.5.4. Strategies for Ensuring Authenticity

Technical strategies for asserting the authenticity of digital resources include unique document identifiers, documenting changes in metadata, hashing, digital stamping, encapsulation techniques, digital watermarks and digital signatures. The Koninklijke Bibliotheek in the Netherlands has expressed interest in 'stamping' migrated material to indicate to users that the information has been altered and to reassure publishers (Bide 2000).

Lynch (1999) points out the problems with using computational techniques over time. Digital signatures are bound to a specific representation of an object and this may change. An author may be able to sign the reformatted object, but this would be unfeasible for a legal deposit library given the number of authors involved and the long-term nature of their remit. The alternative is for the preserving institution to sign metadata linking the new object and the original signature. Lynch (1999) discusses the role of "canonical" formats and "canonicalization" in digital preservation. This assumes that a canonical form capturing the "essential characteristics" for a class of digital objects can be defined. This sounds similar to Rothenberg and Thibodeau's thinking. 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 for given types of object to the canonical form. This will not include irrelevant data included in some data formats, so a

reverse translation may not result in an identical object at the bit level. 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 migrated object match.

Since Lynch's canonicalisation idea combines abstract descriptions of the "essence" of an information object with computational algorithms, there is a need for precision in descriptions of "essence" and consensus about the right algorithms for different contexts (Lynch 2000, p. 38). In 2002, a meeting of preservation experts agreed that there was a need for a more consistent approach to authenticity across different heritage domains and for more communication and understanding of the perspectives of the different domains. Another issue discussed was that of the role of trust in authenticity and the limited understanding of what drives trust decisions (Ross 2002, p. 7)

#### **3.4.6. *Disasters and Rescue of Digital Information***

Ross and Gow (1999) investigated approaches taken to access digital information when it has already become inaccessible. Possible techniques for data recovery include heat treatments, chemical treatments, searching binary structures to identify recurring patterns and reverse engineering of content. Heat treatments can help make media affected by moisture readable. Chemical treatments can be used to clean soiled media. One of the findings of the study is that there is a distinction between data recovery and data intelligibility. While it may be possible to recover data through searching binary structures, technological developments mean it will become harder to read the recovered data. A future possibility is the use of magnetic force microscopy to read damaged media. Another is cryptography to help in the interpretation of recovered data.

The report includes a number of recommendations for further research, documentation for hardware and software and the potential of magnetic force microscopy and cryptography for aiding digital preservation. Ross and Gow recommend that a media quality index should be developed. They suggest various criteria that should be taken into account. All storage media should be marked with a quality rating. Ross and Gow have no suggestions as to how a quality rating system would work.

### 3.5. Access

The provision of access to digital deposit publications is a source of tension between deposit libraries and publishers. This tension is demonstrated in library and publishers' responses to the UK Government's consultation document and the drafting of the voluntary code for deposit (see section 2.3). The main issues in this context are legal and commercial. Publishers do not want libraries to act as competitors by providing widespread access to deposited material, especially when this material required high development costs and commands a high market price.

#### 3.5.1. *Restricting Access*

Mackenzie Owen and Walle (1996, pp. 41-42) also set out a number of access modes for digital deposit material. These were:

- *On-site consultation* - providing access in a controlled environment, this approach does not take advantages of networking technology
- *Controlled remote access* - online access for registered users, use is monitored, similar to campus site licences
- *Time windowing* - limited access for a certain period of time, then less restricted access later on

The concerns of rights holders will mean that the full potential of new technology cannot be harnessed for providing wider access to legal deposit material. There is nothing in the literature that directly reports the views of users, so it is not clear whether they want remote access to digital publications or whether they would object to limited access rights, say limited or no printing or downloading of material.

UK publishers have suggested various solutions to the access issues including restricting the number of simultaneous users and employing embargoes on access for high, “fast-changing dynamic data” or “time-sensitive market data” (Electronic Publishing Services

2002, p. 44). However, it seems that the more restricted the access, the happier some publishers will be because they believe any access will be detrimental to their businesses.

### 3.5.2. *Use of Legal Deposit Material*

Another legal issue is that of security. It is not really possible to alter printed material other than to deface or remove it. Users can potentially manipulate digital information and both libraries and publishers want to eliminate this possibility. Authentication of users and the setting up of access rights are security measures that have implications for users and libraries.

New technology provides the means of closely monitoring information use. The purpose may be to police user behaviour to ensure access agreements with publishers are not breached. Williamson (1994, p. 101) points out that logging use may be burdensome for deposit libraries. There is also the potential for other uses of data, for example providing feedback to publishers, which has data protection implications.

While stating that libraries accept that "the legitimate interests of publishers require that access is limited and controlled", Mackenzie Owen and Walle (1996, p. 41) suggest that access should be encouraged to facilitate preservation of digital publications. They reason that if electronic publications are not used for some time, they may be found to no longer work. On the other hand, a high level of access will "check the operability of electronic publications and ... identify and remedy any access problems that may occur" (Mackenzie Owen & Walle 1996, p. 41). The truth of this statement would depend on the preservation and storage policies adopted. It may be that a systematic checking process is set up for all publications, regardless of frequency of use.

Before using digital information, users have to be able to find it. Bibliographic records traditionally provide details of shelf marks of physical information objects deposited in the library. In the digital environment, bibliographic records can link directly to the material. However, as Mackenzie Owen and Walle (1996, p. 44) point out, there can be several pointers, including to the original and the archival storage locations. Both of these may be



physically located in the library. In the case of online publications, the original location may be a network address.

The BL has a number of reader work stations in both its Yorkshire and London reading rooms. In 2002, there were 100 workstations at St Pancras, with a further nine at Colindale in London and Boston Spa in Yorkshire. At this time, there was 80% occupancy of the workstations. Despite there being some scope to cope with increased demand, the BL estimates that additional places will be required. A study by Electronic Publishing Services suggested (2002, p. 23) that by 2005, the number of places should be increased by 50%.

### **3.5.3.            *Networked Access***

The development of a secure network between the legal deposit libraries will also be required, particularly if digital publications are not deposited in all the legal deposit libraries. The publisher representatives who have worked with the legal deposit libraries through the Joint Committee on Voluntary Deposit (JCVD) have been convinced that a secure network is viable by the results of the proof of concept trials run by the libraries (Electronic Publishing Services 2002, p. 42). However, some publishers are “extremely anxious about the risk of lost sales as a result of insufficiently strict access arrangements or ‘leakage’” (Electronic Publishing Services 2002, p. 42). Some publishers are also worried that their material could be used for inter-library loans and document delivery, despite the fact that this is not normally the case for print material (Electronic Publishing Services 2002, p. 42).

## **3.6.            Intellectual Property Rights and other Legal Issues**

### **3.6.1.            *Intellectual Property Rights***

NORDINFO supported a study on the copyright questions related to the legal deposit of online material. This study concentrated on the European, and especially Nordic, legal environments. The study found various copyright questions arising from the acquisition, storage and retrieval of online material. The study report concludes that there is a gap

between copyright provisions and legal deposit objectives and that legal deposit and copyright laws should be harmonised “in order to secure the objectives of legal deposit ...” (Mauritzen & Solbakk 2000, p. 14).

The CEDARS project explored intellectual property rights issues in digital preservation and concluded that the two primary issues are the need to copy and to modify the digital resource (CEDARS, 2002, p. 10). The need to copy Internet material to get it into a digital depository has been discussed above. Projects such as CEDARS and NEDLIB have also advocated copying digital information from their original publication environment to a secure archival store (CEDARS 2002b, p. 10; Steenbakkers 2002, p. 17). Depending on the technical strategy used, the result of a preservation strategy may be that the information looks different (CEDARS 2002b, p. 10) and/or there is some data loss.

Table 2-3 below (Muir 2003) summarises the copying likely to be required for digital preservation based on current recommendations for preservation strategies.

<b>STRATEGY</b>	<b>COPYING ACTIONS REQUIRED</b>
Media refreshment and change	Copying from an old medium to a new one of the same type Copying from a digital medium to paper or microform Copying from an one type of digital medium to another
Migration	Media refreshment and migration of content Producing a new version of the original information or a migration tool in a new format Acquisition and preservation of format specifications
Emulation and UVC	Media refreshment and migration of content, application software, operating software, hardware specifications, software specifications Reverse engineering of software

**Table 3-2 Copying implications of digital preservation strategies**

Another problem with digital preservation is that it is not just the basic content that may need to be copied. Digital preservation will involve acquiring and preserving applications, retrieval and operating system software, and documentation, such as format, software and

hardware specifications. The software and documentation may be readily available, but copying it from Web sites has copyright implications. Preservation may also require reverse engineering of software in order to recreate interfaces, or write emulators. 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. Sections 37-42 of current UK copyright law deal specifically with library privileges. One library privilege of interest permits copying for purposes of preservation or replacement (United Kingdom 1988, s. 42).

This exception is very narrowly defined; it permits a librarian or archivist of a *prescribed library*, including school, university, further education, public and government libraries (United Kingdom 1989), 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 (United Kingdom 1988, s. 42). One of the prescribed conditions is that copying is only permitted where it is not reasonably practicable to purchase a further copy of the item. This exception only applies to literary, dramatic and musical works, not artistic works. Software is protected under UK copyright legislation as a literary work. The Act, however, allows for one backup copy of software to be made only. Another question is whether changes to the look and feel of material resulting from preservation strategies would infringe publishers' rights in the typographical layout of publications.

While legal deposit libraries (with the likely exception of the Faculty of Advocates Library) qualify as prescribed libraries, the copying that is likely to be required for preservation may not be possible under the exception. There are also copyright issues related to acquisitions, for example with regard to digital archiving, especially web archiving, the "legislation as currently worded is not terribly helpful." (Charlesworth 2003, p. 8). Charlesworth states 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).

A fuller explanation of the copyright issues associated with digital preservation in all library sectors is available in Ayre and Muir (2004). The important point is the new UK legal deposit legislation recognises the issues and makes provisions to address some of these issues for legal deposit libraries (see section 2.5).

Other countries have come to similar conclusions regarding copyright and digital legal deposit and are taking steps to address the issue. For example, the Norwegian National Library has embarked on a project that will, amongst other things, explore the legal issues of collecting online material. The Swedish government issued a decree relating to the Kulturarw<sup>3</sup> Heritage Project at the Royal Library. The Decree authorises the Royal Library to collect Swedish Web sites and allow the public access to material within the library premises (New decree for Kulturarw<sup>3</sup> 2002). It was reported that the French government was to adopt a law requiring every French web site to be archived (Entire French web to be archived 2001), but it is not clear whether this happened.

Another point on copyright made in the literature is that publishers may not have the right to deposit material. Spivey (2003) works for a journal publisher and states that rights have not been assigned for all of the journal content. Spivey (2003, p. 215) contends that “publishers can only deposit content for which they either own the copyright or hold an exclusive licence from the author”. It is not clear why this should be the case, as publishers deposit their print material without getting any additional permissions from authors.

### **3.6.2.            *Other Legal Issues***

The JISC and Wellcome Trust commissioned Charlesworth to study the legal issues relating to web archiving. He covers defamation, content liability and data protection in his report as well as copyright and legal deposit (Charlesworth 2003). The concerns raised by Charlesworth on defamation and content liability were also voiced during the passage of the new UK legal deposit law and were duly acknowledged in the wording of the *Legal Deposit Libraries Act 2003*. However the issue of data protection (mentioned in section 3.6) is not explicitly acknowledged in the Act.

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. Moral rights are distinct and separate from intellectual property rights and in the UK 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. (United Kingdom 1988, s. 77)

The issue of "derogatory" treatment may become important depending on the results of preservation action. It is not clear what authors' views are on this issue. Neither is there any provision in legal deposit legislation to give protect UK legal deposit libraries from the consequences of infringing moral rights.

### **3.7. Costs of Legal Deposit and Digital Preservation**

There have been three exercises in estimating the costs of extending legal deposit to non-print materials, in 1995, 1997 and 2002. The cost of legal deposit of digital publications is made up of several elements. Perhaps the most difficult to quantify is preservation, since there are no full-scale digital deposit systems in existence that include preservation. Until full-scale operational deposit and preservation systems have been running for some time, the nature and extent of costs cannot be known for certain. However, there have been various attempts to identify the types of costs associated with long-term care of digital information and the likely costs of digital legal deposit, at least in the short-term.

#### **3.7.1. *The 1995 British Library Cost Estimates***

As part of its 1996 proposal for the extension of UK legal deposit to non-print material, the BL included an account of how it proposed to acquire, process and store digital publications, and attached cost estimates to these activities (British Library 1996, pp. 21-26). The BL estimated that it would cost more to deal with digital publications than print

publications. This exercise was limited in its scope in that it only included CD-ROMs, or at best, digital material published on physical media. It also only included costs for the BL operation; submissions from the other deposit libraries included in the BL proposal did not have any costs attached. The BL document did not consider the costs incurred by publishers in depositing digital publications.

The cost models were based on an estimated intake of publications. The BL document points out the difficulties of estimating future output of CD-ROMs (British Library 1996, p.17 of British Library submission), but presents minimum, expected and maximum estimated numbers for published CD-ROMs for the five financial years from 1997/98 to 2001/02. The minimum estimates were 1,870 up to 3,190 and the maximum estimates were 8,400 up to 13,975. The expected figures, an average of the minimum and maximum figures, were from 5,134 to 8,582 items. The cost models are, however, based on the maximum expected intake. The BL proposed a “minimal” and “realistic” model of costs. Minimal provision referred to a single, stand-alone workstation, with discs and documentation stored on shelves or in cabinets. Realistic provision referred to discs stored in “towers”, jukeboxes and shelves according to demand. Access would be via five networked workstations and two stand-alone workstations. It is not clear how the numbers of machines for the realistic model were derived.

The cost models are broken down by activity and include staff and equipment costs. The activities include the following:

- Accessioning, first handling, record creation and a 10% management charge
- Storage and access

The BL did not include selection costs in the model. This is perhaps because “It is proposed that new legislation to cover non-print publications should also enable automatic deposit” (British Library 1996, p. 3 of British Library submission).

The first set of activities include some that are common to print and non-print material, but also include some additional activities. The models assume that, at least initially, it would take longer to process digital publications. CD-ROMs would have to be loaded and

accessed via computers before they could be tested and catalogued (British Library 1996, p. 8 of British Library submission). Assumptions about bibliographic control included that the basic cataloguing principles apply, additional metadata would be needed, metadata would need to be updated and added to over time and cataloguing time may be reduced in the long-term (British Library 1996, p.10 of British Library submission).

It was assumed that refreshment, media migration and migration might be used but that it was not possible to estimate the cost of migration. While the BL submission does estimate a cost for media migration, this activity is not included in the cost models (British Library 1996, p. 25 of British Library submission) because this activity would not occur in the first five years of digital legal deposit.

The technology required to store and provide access is a major cost element. The costs that are similar to that for print include labelling, shelving and retrieving shelved items and documentation. The possibility of payment of access fees for unrestricted access to deposited item is mentioned, but it is assumed that access, even networked access, within the BL reading rooms would not attract a fee. The BL did not consider the possibility that publishers might require recompense for access beyond a single, stand-alone workstation, even if it is confined to the BL reading rooms. However, the BL acknowledged that the costs would influence the access model used. The costs of managing access and user agreements have been included in the cost models. Other additional access costs would be the requirement for staff to advise and support users in accessing publications.

The estimated additional costs to the BL of receiving CD-ROMs through legal deposit is summarised in Appendix 6. The estimated staff costs for accessioning, first handling and record creation far outweigh the equipment costs for these activities (see Appendix 7). Not surprisingly, the equipment costs for the realistic model for storage and access costs are much more substantial than those for accessioning and they rise steadily over the five years. The staff costs associated with storage and access also greatly outweigh those for accessioning from 1998/99 onwards.

Some estimates are given for the unit costs of activities that make interesting comparisons. For example, the BL estimated the unit cost of processing print items at £1.13 and CD-ROMs at £3.39, assuming that it takes three times as long to do this for CD-ROMs. First

handling would introduce a new cost of £4 per item. For cataloguing, the estimate for print is £7.06 per item whereas CD-ROM is £14.12 because it would take twice as long. This would reduce over time, so that by 2000 it would only take 1.5 times as long and therefore cost £10.59 per item. No comparisons of preservation costs are given.

The most striking aspect of the models is cost of staff time and that this cost could be reduced over time if ways were found to speed up the process of processing and providing access to CD-ROMs. If the model had considered online material, the models would have looked completely different.

### **3.7.2.                    *The DNH Draft Cost Compliance Assessment***

The UK Government's 1997 consultation paper included a draft Cost Compliance Assessment for extending legal deposit to non-print publications (Department of National Heritage, *et al*, 1997, pp. 45-61). The BL commissioned this survey of costs from Deloitte and Touche on behalf of the Department of National Heritage (Department of National Heritage *et al* 1997, p. 38).

Like the earlier BL exercise, the only digital publications included were those published on physical carriers, concentrating on CD-ROM publications. The estimated industry output figures used were the minimal figures estimated by the BL in 1995. The assessment assumed that CD-ROM title output would continue to increase in the future at a similar rate.

The cost estimates are not explained in detail. Cost elements mentioned are: sorting; cataloguing; storage and provision of access. Again, preservation costs were not included because "the processes necessary for preservation of electronic material are not yet clear" (Department of National Heritage *et al*, 1997, p. 41). The cost estimates given are for all five of the UK-based legal deposit libraries. What is specified is that the costs include one off costs of acquisition and initial processing of publications plus the recurring costs of managing and providing access to publications. The estimated costs are shown in Table 3-3.



Deposit Options <sup>3</sup>	£k				
	1997/98	1998/99	1999/2000	2000/01	2001/02
Option 1	242.3	151.7	189.6	352.2	303.6
Option 2	484.8	303.4	379.2	704.3	607.3
Option 3	607.4	622.1	795.1	1,050.4	1,147.6

**Table 3-3 Estimated costs to UK-based legal deposit libraries of deposit of CD-ROMs**  
(Department of National Heritage *et al* 1997, p. 42)

The costs for each option are based on a 50% deposit rate either to the BL alone; shared amongst all the libraries, with the BL taking on 67%; or with all the libraries taking 50% of the UK output. It was made clear that the Government had no intention of providing additional funding to legal deposit libraries to take on these additional responsibilities (Department of National Heritage *et al* 1997, p. 43).

The cost to publishers of compliance with legal deposit of CD-ROMs was also estimated. These costs are included in Table 3-4.

	1997/98	1998/99	1999/2000	2000/01	2001/02
Lowest/least onerous	£19,000	£22,000	£25,000	£29,000	£32,000
Highest/most onerous	£1,268,000	£1,492,000	£1,715,000	£1,939,000	£2,163,000

**Table 3-4 Estimated cost of compliance for deposit of CD-ROMs**

The lowest figures are based on Option 1 in Table 3-3 (50% compliance with up to two copies deposited, and no legal fees). The highest are based on Option 3 (100% compliance with deposit in all the legal deposit libraries and 50% of publishers incurring one-off legal fees). The elements included in compliance costs were the cost of publications and the cost of providing these publications (recurrent costs). The unit cost (£10-13) of supplying a publication was multiplied by the number of publications to be deposited.

---

<sup>3</sup> Option 1 - Voluntary deposit with one or more of the deposit libraries; Option 2 - Legal deposit of one or two copies of publications with a single legal deposit library; Option 3 - Legal deposit of one copy with the British Library and up to five copies with the other legal deposit libraries on request (Department of National Heritage *et al* 1997).

The one-off legal costs referred to legal advice for publishers on whether they can deposit material. However, the new legislation may obviate the need for this. Costs for compliance for individual publishers were estimated at between £30 and £2,034, depending on how many copies are deposited. This does not seem to be particularly costly.

It is not clear exactly how many digital publishers provided data on costs, but it is clear that the number was very low. Only fifty-six organisations participated in the survey and only twenty-nine of these provided data on costs. Thirteen out of an estimated 321 UK CD-ROM publishers were approached and ten responded to the survey. It is not clear if this was a random sample or how many of the respondents provided cost data.

The fact that the cost of potential lost sales is not included in the costs of compliance is made explicit. The potential of increased sales due to inclusion in the national bibliography is mentioned. The potential for disproportionate commercial damage to publishers of certain types of material is also acknowledged and the possibility of access restrictions and exemptions to the obligation to deposit is raised. The potential for disproportionate commercial impact on small publishers is raised, but not taken seriously. Criticism of this costing exercise were reported in section 2.5.2. It seems likely that these criticisms were due to the limited nature of the exercise and the methods used to derive cost estimates. Apparently, rather than trying to identify characteristics of “typical” and “small” publishers, the costs are based on an “average” publisher and so may not be representative. This problem is also explicitly recognised.

### **3.7.3.                    *The 2002 Cost Estimates***

This exercise benefited from the experience of the voluntary deposit scheme, in operation since the beginning of 2000. The JCVD commissioned Electronic Publishing Services Ltd (EPS) to carry out this study. The study had two aims:

- To gather data on costs and “quantifiable impacts” of legal deposit of non-print publications for both publishers and legal deposit libraries
- To develop a model for calculating and “illustrating” these costs and impacts that could be “used against variable assumptions as to the types of material to be

deposited, and over time as the types, amount and value of material published changes” (Electronic Publishing Services Ltd 2002, p. 5)

The aim of tackling “quantifiable impacts” makes this assessment different from its predecessor in that it is not merely acknowledges that there will be costs other than those incurred in supplying publications. The second aim reflects the nature of the draft legislation being planned at the time in that a model that could be used to look at the impact of extending legal deposit to new publication types over time was to be developed.

The study was carried out very quickly, between June and September 2002. It was much broader in its scope than previous work, that is, not restricted to CD-ROM publishing. By the time of this study, it was clear that the use of CD-ROM as a publishing medium was in decline (Electronic Publishing Services Limited 2002, p. 1).

The study involved estimating industry output between 2002 and 2005 and surveying publishers and legal deposit libraries on their costs. Methods used for data gathering included desk research, postal questionnaires and interviews. Thirteen interviews were used to gather data from the legal deposit libraries and the Copyright Libraries Agent. The cost model for the libraries was based on processes and procedures in the BL. The study report recognises that the other libraries “have different throughput rates and different cost structures”, but says that the model will be valid for them if they input their own variables (Electronic Publishing Services Limited 2002, p. 2). The study also included a postal questionnaire of all the publishers participating in the voluntary deposit scheme. The researchers received contact details for 350 publishers from the BL and received twenty-one “valid and usable” responses. In order to improve the representativeness of the response to the research, the researchers supplemented the questionnaire returns with nineteen telephone interviews.

Like the previous cost compliance exercise, the research was based on a very small sample of the digital publishing industry, and ideally it would have been better to use some other sampling frame and a larger sample. However, given the short timescale, it was probably the most practical approach in the circumstances and at least the publishers surveyed had some experience upon which to base their cost estimates and comments.

Seven years on from the BL's costing study, the lack of reliable figures for the output of the digital publishing industry was still an issue. This study uses a wider range of sources for publishing output than previous studies, including the products of bibliographic agencies such as Book Data and Whitaker and subscription agents for serials. The forecasts made for this study are based on current trends and the study team's knowledge of publisher strategies. Electronic Publishing Services Ltd were well placed to carry out this sort of exercise because of their close connections with the digital publishing industry. It is made clear that the forecasts are open to debate, but state that the general trends are not disputed by the industry.

The study gives figures for the output of new CD-ROM titles from 1998 to 2001 and estimates output from 2002 to 2005. The actual output figures are not directly comparable with previous estimates because these were based on financial years whereas the figures for this study are for calendar years. Nevertheless, the estimates from the previous studies are higher than the actual output. This study estimates that the number of new CD-ROM titles issued between 2002 and 2005 will drop steadily from 1,150 titles in 2002 to 125 titles in 2005. Conversely the number of new digital only serials titles is 270 for 2002 and is estimated to rise to 2,250 by 2005 (Electronic Publishing Services Ltd 2002, p. 12). The focus of the study was electronic monographs and serials published on physical carriers or in "pure" digital form and did not cover any other sorts of digital publishing, such as Web sites. The unit for deposit for serials was also taken to be the issue, ignoring the possibility that the volume/issue model may not survive in the digital environment.

EPS Ltd estimated that the BL was receiving 75% of digital monographs published in the UK, but only 45-50% of electronic serials through the voluntary scheme. The cost models developed for the BL were for three years and included two scenarios for serials: 60% or 80% deposit of digital only serials by year three. The model for monographs assumes 80% deposit in 2003, 85% in 2004 and 90% in 2005.

The cost models include processing, equipment (including storage) and telecommunications costs. It also includes staffing estimates to allow space considerations to be costed. The cost of running the secure network is also included. Estimates for storage and the secure network are based on existing estimates. Unlike the BL model of 1995,

overheads are not included. Total estimated costs to the BL between 2003 and 2005 are shown in Appendix 6.

The network is the biggest cost element. The study report makes clear that this is based on a cost estimate from 2000 and that it represents the worst case. It is not clear whether the cost of networked access to material is less than the costs for each library to acquire, process and provide access to their own copies of material.

The storage costs are initially high as this would involve setting up a storage solution, but drop to a lower level in 2004 and 2005. The biggest processing element in this model is the cost of the technical and administrative metadata. This is not a task that can be dispensed with given the preservation role of legal deposit libraries. However, even within three years, the cost of this task is predicted to fall.

Apart from monographs on CD-ROMs the other legal deposit libraries are not receiving anything like the volume of digital publications as the BL. If each library had to deal with a similar volume, the overall cost of digital legal deposit to the libraries would be significant. The benefit of the model to the other libraries is that they could use it for entering data on their own operations in order to identify what proportion of the overall work and costs they could each bear. It is interesting to note that discussion with the researchers and the other libraries indicated that processing costs were cheaper than those of the BL.

The study identified two categories of cost for publishers: the cost of delivering up to six copies of material and the “consequential costs” of material being available via the legal deposit libraries or of other obligations associated with the deposit process (Electronic Publishing Services 2002, p. 40). Publishers thought that the cost of delivery would be marginal. The cost of duplication of material is low as are the costs of adding legal deposit libraries to distribution or mailing lists. Even transmission costs were thought to be insignificant. There was more concern about potential consequential costs. These costs were categorised as use of their publications in legal deposit libraries and providing the extra metadata not normally supplied to customers. The costs of metadata to publishers would depend on the “depth” of metadata required by the libraries. The study assumes that the publishers have all the data that would be required, the costs would arise from bringing

all of it together, in a suitable format, and attaching it to deposited publications. The study estimates that there would be initial costs in setting up a system, but after that the costs would be low.

The publishers felt that there would be a burden in terms of the need for liaison between the publishers and the libraries and the need to supply documentation with publications (Electronic Publishing Services 2002, p. 42). However, discussions were underway at the time of this study under the auspices of Book Industry Communication (BIC) to find a procedure that would help the libraries, but pose the least burden on publishers (Electronic Publishing Services 2002, p. 45).

The study provides a range of costs to publishers in its cost model. The minimum (£1) and maximum (£5) costs per unit given are not fully explained. Presumably the £1 cost is the bare minimum cost of producing extra copies and delivery whereas the £5 cost includes metadata provision. Estimates are provided for both one and six copy delivery. The EPS study found that it was not possible to quantify potential loss of sales.

This study of costs seemed to be as thorough as possible, given the unknowns and short timescale. Publishers probably will not have to deposit six copies of their material, so their costs are likely to be nearer to the one copy deposit scenario. Ideally the libraries would want as high a rate of compliance as possible. So costs to publishers are likely to be up to £1.5 million pounds by 2005. However, since the libraries cannot cope with the volume of material they are receiving under the voluntary scheme, they may not actively pursue comprehensiveness, particularly if the Government keeps its word and does not produce additional funding for the libraries.

#### **3.7.4. *Cost of Digital Preservation***

While there are no fully functional digital preservation systems, there has been some work on identifying cost elements of digital preservation. Hendley's study of different preservation methods and associated costs took into account diversity in digital materials (Hendley 1996). Hendley adopted the framework developed by Greenstein and Beagrie (1998) to "define and agree the context within which digital preservation is being

addressed” (see section 4.1.1). Hendley does not mention exploring alternatives to this framework and there is also no critical discussion of the value of the framework. This is perhaps because the framework used was part of the same programme of work as the Hendley study.

The approach taken to achieve the study aims was:

- Define the key tasks in digital preservation
- Review of the three preservation strategies – migration, emulation, technology preservation – given in the study remit
- Define all digital information resources and data types covered by the study
- Develop a decision model to assess categories of digital resource and select the most appropriate preservation strategy
- Develop a cost model to assess costs according to category and preservation strategy and also to allocate costs to the stages in the management process in the framework developed by Beagrie and Greenstein

Hendley does not consider costs of provision of access to material through discovery services and use of material as preservation costs. What Hendley provides is a detailed analysis of cost elements based on knowledge and practical experience available at the time.

The CEDARS project also investigated costs of long-term preservation. The project found that digital preservation requires an ongoing financial commitment, but that the level of costs will depend, at least in part, on the archiving model used (CEDARS 2001, p. 67). Like Hendley, CEDARS work refers to the lifecycle of a digital resource, concluding that the earlier in the lifecycle that preservation decisions are taken, the less the costs are likely to be (CEDARS 2001, p. 67). The need to assess costs against benefits is also identified, “an approach to preservation which is appropriate to the perceived (estimated) value of the digital object.”

The difficulty of determining the long-term value of material, a crucial issue for legal deposit libraries, is acknowledged (CEDARS 2001, p. 68) and the link between collection management policies and analyses of benefits is identified.

Cost elements identified by CEDARS are:

- Selection for preservation
- Negotiation for the right to preserve
- Negotiation for the right to preserve access
- Applying appropriate preservation strategies (assessing Significant Properties and selecting strategies)
- Quality control and validation of archived information (completeness and integrity checks, accompanying documentation)
- Metadata production (created or provided from various sources, not all populated at the same time)
- Archival storage (procurement and maintenance of hardware and software, periodic backups, integrity checking and media migration)
- Administering the archive (technology watch, addressing technology and legal changes)
- General costs (e.g. staffing, insurance, building overheads, certification or standards activities)

CEDARS further categorised costs into:

- Upfront, long-term – implementing infrastructure; selection; intellectual property issues; validation of integrity; determining technical preservation strategies; preservation metadata.
- On-going - file storage; archive administration; policy evaluation/revision
- Varying over time –
  - Decreasing costs - migration tools become more widely available; cooperative development of metadata standards, automation of metadata generation, standardisation of means of transferring metadata. Collective



rights agreements and standard licences, amendments to rights and legal deposit legislation

While technical strategies may change over time with resultant costs, there may also be periodic costs resulting from updating metadata, for example recording preservation activities over time.

Most of these cost elements are likely to apply to legal deposit libraries, although the new UK legal deposit legislation should pave the way for removing the need to negotiate the right to preserve and to provide limited access. The 2002 legal deposit cost assessment exercise did include some of the cost elements identified by CEDARS, although the cost of implementing technical preservation strategies was omitted. Some of the tasks of administering digital preservation operations, such as watching out for changes in technology could be shared among the legal deposit libraries or even with other organisations involved in the long-term preservation of digital publications. This will be one of the tasks of the newly created Digital Curation Centre in the UK (see chapter 4). According to CEDARS, the extent to which digital preservation activities are integrated into the management of the whole collection as well as automation will influence cost elements (CEDARS 2002a, p. 19).

CEDARS developed a framework for examining the costs of digital preservation (CEDARS 2001, p. 69) involving three variables: diversity in data types to be preserved, the need to identify and negotiate with rights holders and lack of ability to dictate the types of digital objects acquired are likely to increase costs.

The Electronic Resource Preservation and Access Network “cost orientation tool” (ERPANET 2003a) is aimed at any organisation preserving digital material. The tool adds little to existing knowledge of cost elements and it does not provide a means of calculating costs.

### **3.8. Summary**

The issue of definitions - particularly what constitutes a publication and what published in the UK means in the digital environment – makes the implementation of legal deposit for digital publications more complicated. The development and use of technology-neutral definitions will allow legislators to build flexibility into legal deposit laws. Digital publishing poses a number of problems for existing legal deposit systems, including the lack of standard mechanisms for the unique identification of digital publications; an explosion in digital publishing and the development of new types of communication and the difficulty of preserving digital publications. The issue of selection has to be examined in order to decide which types of digital communication should be subject to legal deposit. The sheer volume of material has implications for the comprehensiveness of legal deposit libraries' collecting activities. While information and communication technology could potentially allow legal deposit libraries to improve and widen access to digital legal deposit material, this potential is a source of concern for publishers. There are rights issues to be considered in the provision of access to digital material and in the collection and preservation of digital publications because these activities involve copying. The potential cost of extending legal deposit in the UK has been considered on several occasions, although none of the studies really captures the likely full cost of digital deposit. Other studies of the cost elements involved in digital preservation and the development of cost models may be helpful to legal deposit libraries. However, the real costs of digital preservation also remain unknown.

Some of the issues and questions arising from the legal deposit of digital material concern the development of policy and procedures, for example selection, acquisition and processing of material. Some issues will require cooperation and agreement between legal deposit libraries and publishers, such as how material is deposited and what access is permitted. This chapter has raised many questions; Chapter 4 outlines some of the developments that may provide some of the answers to these questions.

## **4. MANAGING DIGITAL LEGAL DEPOSIT COLLECTIONS**

The focus of this chapter is implementing and managing the legal deposit process. This chapter highlights some developments and emerging standards of relevance to legal deposit libraries. Some tools for developing policies are identified as are supra-organisational infrastructure to support preserving organisations and reduce duplication of effort. Although legal deposit libraries have the central role in the preservation of a nation's intellectual output, they will have to work with others to carry out that role and they are also just one stakeholder in digital preservation. The various stakeholders and roles are discussed.

### **4.1. Digital Preservation Policies**

Legal deposit libraries are likely to have well-developed preservation policies for traditional materials. ERPANET (2003b) has produced a guide to developing digital preservation policies. The basic components are similar to traditional preservation policies: institutional objectives, the aim of the policy, the roles and responsibilities of people involved in preservation and how digital information will be made secure and preserved. One criticism is that the author of the document does not seem to be a native English speaker and meanings are opaque at times for non-experts. For example, it says “preserving [sic] damage and deterioration of the physical media caused by ensuring an environmental control” ERPANET (2003b).

#### **4.1.1. *Digital Information Life Cycles***

The concept of the life cycle of digital resources has been used as the basis of developing a strategic policy framework for the creation, management and preservation of digital resources (Beagrie & Greenstein, 1998). Hendley (1998) added to this original model. The life cycle breaks down into several stages, or modules, as shown below.

	MODULE NAME	SUB-MODULE NAME	SUB-SUB-MODULE NAME
1	Resource creation		
2	Resource selection and evaluation		
3	Resource management		
		Structure	
		Documentation	
		Storage	
		Validation	
			Assessment
			Copying
			Refreshment
4	Resource disclosure		
5	Resource use		
6	Resource preservation		
7	Rights management		

**Table 4-1 Digital resource lifecycle**

This framework summarises the issues identified in chapter 3. While legal deposit was not the focus of the framework, it can be used in this context and Beagrie and Greenstein include legal deposit as a case study illustrating the framework.

#### **4.1.2. *Categorisation for Deposit and Preservation***

The CEDARS project (2001, pp. 109-110) developed a typology for digital resources. While this is mainly aimed at helping academic and research libraries to take preservation decisions, it also is largely relevant for legal deposit libraries. The CEDARS typology contains the following “dimensions” of digital resources:

- Publication source
- Publication format
- Publication type
- Publication medium
- Legal status
- Nationality of publication
- Parallel publications
- Technical environment
- Content level and intellectual value
- Subject area

From a legal deposit point of view, subject may be less relevant. Content level and intellectual value would be relevant if it was necessary to be selective in collecting for legal deposit, for example for Web material. Publication source, nationality of publication and publication type are particularly relevant as initial selection criteria, and should be included in legislation. Publication source relates to the producer of information, nationality is self-explanatory and publication type refers to the type of content, for example monographs, journals, databases, computer games and email. These could then be supplemented by publication format, publication medium and parallel publications. The new UK legal deposit legislation allows legal deposit libraries to select their preferred formats and media if alternatives are available (United Kingdom 2003d). The legal status aspect relates to acquisition, preservation and access. The technical environment required to access the material is important for preservation purposes. This includes hardware requirements, operating systems and software for access and retrieval.

Of course, use of the typology would be dependent upon the development of definitions of terms. While it may be possible to develop technology neutral definitions of “publication” and “publisher” and “place of publication”, applying this typology will require some more technical definitions of data types, publication types, and publication media. Bennett (1997, pp. 12-20) developed a framework of data types and formats. This framework involved four dimensions: type of material; type of file format; type of media and type of platform. These map onto the publication format, publication type, publication medium

and technical environment aspects of the CEDARS typology. Hendley (1998, pp. 27-44) also defined digital resources according to:

- Data type, including alphanumeric data, raster and vector graphic data, audio data and moving image data
- Category of resource, including data sets, structured text, design data, sound recordings and interactive multimedia publications
- Application programmes, including data processing, word processing, SGML editors, Web page design and presentation graphics
- Structuring data types, or data formats, such as ASCII, TIFF, JPEG, WAV and MPEG

Hendley (1998, p. 42) also identified management or distribution systems, “how the digital resources were managed and distributed/used prior to deposit or at the point of deposit”.

While Bennett and Hendley carried out some useful groundwork in this area, the gathering of this type of information will be an ongoing task as technology evolves over time.

#### **4.1.3. Risk Management of Digital Material**

The ERPANET project has created a “risk communication tool” for digital material (Electronic Resource Preservation and Access Network 2003c). This document takes readers through the risk identification and analysis process in a simple way. The processes do not differ from the risk assessment involved in traditional preservation, although the nature of the risks, the practical implementation of risk assessment and how risk is managed may do. Specific risks and how to deal with them is not included in the tool, so it has little to add to existing literature and is of minimal practical use to anyone wishing to carry out such an exercise.

Part of the Cornell University PRISM project involves developing a risk management approach to the preservation of Web-based material (Kenney *et al* 2002). This could be of interest to deposit libraries as a pragmatic alternative to acquiring physical copies of material in large national Internet domains. The Virtual Remote Control project involves monitoring Websites over time, identifying risks and responding to them. However,

capture is the last resort; site owners would be notified of detected risks and possible actions they could take. The objectives of the project are to:

- Develop a risk management preservation model for research libraries
- Adapt the model for use in other settings
- Support a range of responses to detected risks
- Develop an understanding of the nature of Web resources and how they evolve over time
- Identify common risks and how to reduce them (McGovern *et al* 2004)

The PRISM risk management model involves four stages:

- Data gathering and characterisation
- Simple risk declaration and detection
- Contextualised risk declaration and detection
- Automated preservation policy enforcement (Kenney *et al* 2002).

The PRISM team is using both the Mercator Web crawler and manual means of gathering data to identify and classify risks to Web resources. Issues to be considered include the value of material, possible threats, known vulnerabilities, likelihood of loss and potential safeguards. The model involves the development of risk scenarios, probability of these occurring, the likely impact and the costs of recovery. The model also takes into account the willingness of the preserving institution to accept risks and includes the development of “mitigation responses”.

The PRISM team have looked at defining the boundaries of a Web resource that may be useful to legal deposit libraries. These are:

- A stand-alone object, ignoring hyperlinks
- A Web page in a local context considering links into and out of the page
- A Web site as a semantically coherent set of linked Web pages
- A Web site as an entity in a broader technical and organisational context (Kenney *et al* 2002).

Factors indicative of risk, and therefore to be monitored, include the way Web pages are constructed, use of standards, the nature of links and how these factors change over time. Algorithms to work out the level of risk, based on all these factors, will be needed and directed graphs and graph analysis are suggested techniques for developing a set of patterns reflecting good site management. Apparently some of these factors can be monitored remotely, through assessing speed of response, and the use of software to assess security hazards and causes of unresponsiveness (McGovern *et al* 2004). The ability to keep up with technology change is also a risk factor and the Internet Archive (see section 5.5.1) is identified as a good test bed for analysing changes in mark up languages, formats, dynamic behaviour, and metadata over a period of time.

While much of the process will initially be carried out manually, the aim is to automate as much of it as possible by using Web tools to gather baseline data, monitor sites for changes, suggest responses to detected risk and capture material if necessary. The process involves several stages, with both automatic and human actions attached to each.



Stage	Roles
Identification	Human: identify Web sites of interest Toolbox: verify list and expand through searching and crawling
Evaluation	Toolbox: crawl sites, generate characterisations Human: accept/revise characterisations
Appraisal	Human: define/review attributes of value Toolbox: support appraisal, capture results
Strategy	Human: develop/review strategies Toolbox: plot appraisals, compile strategies
Detection	Toolbox: identify/assess risks; propose responses Human: refine risk parameters
Response	Toolbox: propose risk response based on rules; automatic response for some risk categories Human: monitor automated responses; select response based on recommended actions

**Table 4-2 Human/toolbox interaction in VRC organisational stages** (from McGovern *et al* 2004)

The process starts with human identification of some Web sites of interest. For legal deposit libraries, it may be all sites within the national domain or just a selection. For some legal deposit libraries, it may go beyond the “national” output and include sites of national interest located outwith the country. The tool can help in this and in maintaining the list of sites. The tool can also gather information about the scope, structure and status of sites and document the nature and rate of change. Appraisal involves value assessments on relevance and significance. Other issues at this stage include what role the organisation will take in preservation and what role other organisations will have (McGovern *et al* 2004). For legal deposit libraries, there is a question of whether they want to capture all Web sites, and if not, whether there is a need for agreements with other organisations. The tool would assist with the administrative elements of agreements. Preserving organisations will have to decide their strategies, including risk parameters for different classes of materials and organisations, and what to do about risks, including tolerable risk levels and monitoring, risk prevention and response actions. Detection involves ongoing monitoring, assessment of possible loss or damage and identification of possible responses (McGovern *et al* 2004).

The project involves an assessment of the ability of existing tools to meet the requirements of the Virtual Remote Control project. The results of initial identification and evaluation of tools has been captured in an inventory. The inventory contains data on cost, vendor, strengths, capabilities, scalability, adaptability and compatibility amongst other things (McGovern *et al* 2004). The evaluation identified several tool categories: change detectors; HTML validators; link checkers; site managers; site mappers; site monitors; Web crawlers; other tools. Conclusions so far include that Web crawlers will be necessary but insufficient for a risk-based Web preservation programme. There is a correlation between good site management and reduced risk. This has been recognised by the National Library of Australia (NLA) in its advice for Web site managers (National Library of Australia 2002a, 2002c). Legal deposit libraries could use the outcomes of the PRISM research to develop guidelines for Web site managers that will help ensure longevity as well as to identify valuable and at risk Web content that the libraries may want to acquire.

#### **4.1.4.                    *Assessment of Preservation Strategies***

There has already been some discussion of criteria for selecting preservation strategies in chapter 3. Hedstrom (2002, pp. 36-37) argues that experiments should be undertaken, using different preservation strategies on the same material, and making comparisons of the results. Comparisons could be carried out on the basis of effectiveness, user acceptance and costs. These would provide an empirical basis for evaluating different preservation strategies so that decisions on which strategy is most appropriate to use for particular types of resource.

Staff at Cornell University libraries have applied a risk assessment approach to minimising the disadvantages of migration as a preservation strategy. Lawrence *et al* (2000, p. 5) identify three major types of risk to assess when considering migration as a preservation strategy. These are:

- The general collection (including presence or absence of institutional support, resources, IT hardware and software, staff, legal and policy issues)
- Data file format (the internal structural elements that could be modified)

- File format conversion process (severity of conversion errors, conversion may not produce the intended output) (Lawrence *et al* 2000, p. 6)

This is a useful framework that could be adapted to any digital preservation strategy, not just migration. It would need to be used in the light of assessments of Significant Properties to be retained for different types of material. Successful migration is not just about the data, but also the supporting metadata describing the files and how they relate to each other (Lawrence *et al* 2000). During the study, Lawrence *et al* developed a risk assessment workbook for the general collection, reader software to examine files for high-risk format elements and a test file used to test different conversion softwares. Lawrence *et al* (2000) state that their evaluation tools need further development to provide more reliable results.

The study concluded that the risks associated with migration are measurable, but the greatest challenge is to determine when a risk is acceptable. A number of functions for migration software were identified. These are:

- Read the source file format and analyse the differences between it and the target format
- If there is a mis-match, analyse and report on the degree of risk
- Accurately convert source files to target formats
- Provide a record of conversions (Lawrence *et al* 2000, p. 13)

Lawrence *et al* (2000, pp. 13-15) were unable to identify any off the shelf conversion software that could fulfil all these functions, but thought that working with software vendors may result in improved functionality of such software.

Since Lawrence *et al* found that the most difficult part of the study was to require complete and reliable file format specifications, they concluded that there is a need for reliable and sustainable repositories of specifications and related documentation and software (see section 4.1).

ERPANET (2003b) has provided some useful further guidance on evaluation criteria for the selection of technologies. These include the maturity of the technology, existing experience of applying the technology, widespread use and support of the technology, reliability, modularity and flexibility, cost of purchase and maintenance, how well it will deal with the digital objects involved, availability of skills and ease of integration into workflows.

#### **4.1.5.            *Assessment of Costs***

Several studies of the cost elements of digital preservation were discussed in section 3.7.4. While slightly different, they all identified the same sorts of costs. Sanett (2003) proposed a framework to compare costs across cost models. The author's explanation of the framework is opaque at times, but it seems that this author suggests ordering costs according to two sets of categories:

- Whether they are incurred in the acquisition and preservation of “electronic records” or in the use of or, outside access to, these records
- Whether they are capital, direct operating or indirect operating costs.

This is the basic framework within which costs incurred by particular activities can be placed. Sanett (2003) suggests some activities under each of the headings. For example capital costs for acquisition and preservation could include software development, hardware, research and development, facilities and interface design for processing records. Direct operating costs would be incurred in identifying, evaluating, acquiring, processing, producing metadata, preserving, storage, maintenance and deletion. However, the specific activities can be adapted to particular organisational needs. This framework looks very similar to the one used by EPS Ltd (2002) in its survey of costs for UK legal deposit.

#### **4.2.            *Open Archival Information System and its Implementation***

A major initiative relevant to digital legal deposit collections is the development of the Open Archival Information System (OAIS) Reference Model. The Consultative

Committee on Space Data Systems drafted this standard for the International Standards Organisation (Consultative Committee for Space Data Systems 2002; International Organization for Standardization 2003). An OAIS archive preserves information for access and use by a so-called Designated Community. This model is apparently applicable to archives of any type of digital or paper material. The OAIS models the functions involved in the long-term storage of and access to digital information. These functions include acquisition and processing (ingest), archival storage, access, preservation planning, data management and administration of the archive.

The development of the OAIS has influenced other work on the development of digital deposit collections. The CEDARS (CURL Exemplars for Digital Archives) and NEDLIB (Networked European Deposit Library) projects based their work on the OAIS model. The British Library (BL) and the Koninklijke Bibliotheek (KB) in the Netherlands have also used OAIS as the basis of their recent tenders for systems to manage their digital collections. The OAIS Information Model for metadata that should be included along with basic digital content, is also being used as the basis for the development of digital preservation metadata schemas.

#### **4.2.1.            *OAIS Components***

According to Waibel (2003) “the OAIS in and of itself does not specify an implementation – it does not tell you which computers to buy, which software to load, or which storage medium to use.” This is true, but some of Waibel’s other interpretations of the OAIS are questionable. He says:

The standard does tell you, however, how an archive should be organized. In its so-called functional model, it defines the entities (or departments, if you will) in an archive, their responsibilities, and interactions.

The OAIS reference model is conceptual; it is about functions and the interactions between functions (as Waibel states), but it is not how they are implemented within an organisational structure. The OAIS is made up of three main components: roles, information packages and processes. These are shown in the Figure 4-1.

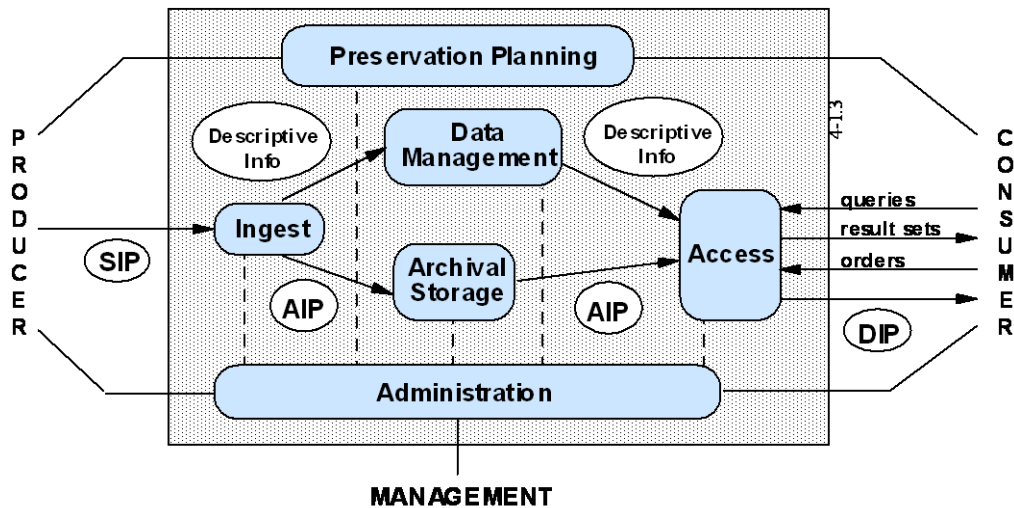


Figure 4-1 The Open Archival Information System Reference Model (Consultative Committee for Space Data Systems 2002)

Three roles are specified:

- The *Producer* role involves entities that provide information to be preserved. In the legal deposit context, this is most likely to be publishers and Web site owners
- The *Management* role sets overall OAIS policy, which is only one component in a broader policy domain. Management is not involved in day-to-day operations; this comes under the administrative functional entity
- *Consumers* use OAIS services to find and acquire preserved information of interest to them. The Designated Community is the set of Consumers who should be able to understand the preserved information. In the case of legal deposit libraries, the Designated Community is potentially huge.

A Producer sends one or more Submission Information Packages (SIP) to an OAIS archive. The OAIS envisages that the form of the SIP will be negotiated between the Producer and the OAIS. SIPs will contain Content Information and some Preservation Description Information (PDI). This will be bound together by Packaging Information.

SIPs are transformed into Archival Information Packages (AIP) for preservation purposes. The AIP should have complete set of PDI for the Content Information. The Packaging Information of the AIP will conform to internal standards. When a Consumer requests a particular item, the OAIS provides all or a part of an AIP in the form of a Dissemination Information Package (DIP); this will be in a form that the Consumer can access. The Packaging Information will be present so that the Consumer can see that they have received the information they requested.

Content Information is the target of preservation. It consists of the Content Data Object (Physical Object or Digital Object, i.e., bits) and associated Representation Information needed to make the Content Data Object understandable to the Designated Community. The PDI specified by the OAIS model reflects the requirements for maintaining authenticity as discussed in sections 3.4.3 and 3.4.4.

- *Provenance Information* describes the source of the Content Information, chain of custody and history of the object, including its processing history.
- *Context information* describes relationships between the Content Information and other information outside the Information Package.
- *Reference information* is one or more unique identifiers. Examples include an ISBN or DOI, or a set of attributes that distinguish one instance of Content Information from another.
- *Fixity information* provides a wrapper or protection against undocumented alteration. This may involve a check sum over the Content Information

Ingest involves the acceptance of SIPs from Producers, or from other internal entities if appropriate (for example if material is migrated, there may be a need to re-ingest material) and prepares the contents for archival storage. Activities involve receiving SIPs, quality assurance, generating an AIP that complies with the archive's data formatting and documentation standards, extracting Descriptive Information for inclusion in the internal database, and coordinating updates to Archival Storage and Data Management. Archival Storage involves the storage, maintenance and retrieval of AIPs. Archival Storage receives AIPs from Ingest, adds them to permanent storage and manages the storage function. This includes managing the storage hierarchy, refreshing storage media, error

checking and disaster recovery capabilities, and providing AIPs to Access. The Access function supports Consumers by providing discovery and retrieval services and functions. Access activities include receiving, coordinating execution of and responding to requests as well as applying access controls where necessary.

Data Management involves populating, maintaining, and accessing both Descriptive Information and administrative data used to manage the archive. Administration manages the overall operation of the archive. Administration functions include negotiating submission agreements with Producers, auditing submissions to ensure that they meet archive standards, and configuration management of system hardware and software. Administration is also responsible for archive standards and policies, customer support, and activating requests. Administration provides for monitoring and improving archive operations and content.

The Preservation Planning process was not included in earlier drafts of the OAIS. This was added in response to a recommendation from the NEDLIB project, which was exploring the application of the OAIS to digital deposit collections. This set of activities is:

- Monitor Designated Community
- Develop Preservation Strategies and Standards
- Develop Packaging Designs and Migration Plans
- Monitor Technology

The Designated Community monitoring activities are aimed at identifying the changing service requirements of the Consumers and trends in the Producer product technologies. These activities help the Develop Preservation Strategies and Standards and the Develop Packaging Designs and Migration Plans functions stay abreast of changing preservation requirements.

The Monitor Technology function includes tracking and assessing emerging standards and technologies, in response to requests from Develop Preservation Strategies and Standards and Develop Packaging Designs and Migration Plans. The relationship between these two functions may be similar to that of conservation and technical staff and preservation managers. However, it seems that neither are involved in the implementation of policies



and plans or of routine maintenance of collections (for example refreshing storage media). They both pass their recommendations to the OAIS Administration function and implementation is carried out by other functions such as Archival Storage. Lavoie (2004b) has provided a clear and straightforward guide to the OAIS.

#### **4.2.2.            *Towards Implementing the OAIS***

The Research Libraries Group is maintaining a list of OAIS implementations. The NEDLIB project work is most relevant to legal deposit libraries. NEDLIB identified the workflow for handling digital publications in the context of legal deposit collections (Werf 2000). The steps are: selection; acquisition; delivery/capture/harvest; registration; verification; description; storage handling; preservation; delivery; access and monitoring. Monitoring is included for quality control of the workflow. The workflow does not include the discovery of publications. The National Library of Lithuania is following the NEDLIB model (Varniene 2002).

In 2000, Harvard University Library began a collaborative project with several publishers to investigate the preservation of scholarly digital journals. As part of this project, the Library commissioned a study on the feasibility of an e-journal archival Document Type Definition (DTD). The aim was to assess whether “a common file structure (DTD or Schema) can be designed and developed into which publishers’ proprietary SGML files can be transformed to meet the requirements of an archiving institution” (Inera Inc 2001, p. 4.). Ten publishers provided their DTDs for analysis and the experiences of organisations already working with DTDs from various publishers were investigated. The value of this study is obvious for UK legal deposit libraries because they will be interested in scholarly journals and will be receiving content from many publishers.

Having a standard structure for preservation purposes could potentially make the process easier for them. The conclusions of this study were that it would be possible to develop a structure to convert content from publishers’ files into a common preservation format. However, the study pointed out the need for quality checks on the converted files because of the inconsistent quality of incoming DTDs (Inera Inc 2001, p. 62). The consultants on

the Harvard study and the National Center for Biotechnology Information (NCBI) of the US National Library of Medicine (NLM) went on to develop an Archiving and Interchange DTD for use by publishers and preserving institutions (National Center for Biotechnology Information 2003).

#### **4.2.3. “Persistent” Archives**

The San Diego Supercomputer Center in the USA has been heavily involved on work on “persistent archives” (Moore *et al* 2000a, 2000b). It involves articulating the essential characteristics of both individual objects and collections in a technology neutral way. This seems similar to some of the preservation approaches described in Chapter 3. This is implemented at the bit stream level by the addition of tags to identify each byte sequence. Each tag is linked to higher-level schema such as data models, data element definitions and DTDs. They are also linked to knowledge level material such as thesauri, topic maps and rules. Tagging and expression of content could be implemented using XML and ASCII, but they could be implemented in some other way. The Persistent Archives architecture is structured to carry out the Ingest, Management and Access functions of the OAIS Reference Model. Items already in persistent format are checked as they come into the archive, other items are converted to persistent format.

The metadata and domain knowledge needed to find, understand and use items are stored in models, dictionaries and knowledge bases. Items are retrieved from Archival Storage and made available in current formats when access is required. Legal deposit libraries could potentially use this infrastructure to implement OAIS-compliant functional preservation systems.

#### **4.3. Preservation Metadata**

The role of metadata in preservation is not new to conservators and archivists. The key role of metadata in digital preservation has been recognised for some time ( for example Chilvers & Feather 1998) and there is much activity in this area. It is required not only to facilitate rendering the digital content over time, but also to manage individual digital objects, and relationships between objects within digital archives. Preservation metadata

overlaps with the more traditional descriptive cataloguing metadata, but it also includes technical and administrative data, including rights data.

#### **4.3.1.            *Preservation Metadata Schema***

Individual institutions and preservation related projects developed metadata schemes for their own purposes. Three prominent examples were the NLA, the CEDARS project and the NEDLIB project. The NLA (1999) developed its specification to facilitate the management of its own digital collections. CEDARS (CEDARS & UKOLN 2000) and NEDLIB (Lupovici & Masanès 2000) based their metadata specifications on the OAIS information model discussed above (Consultative Committee for Space Data Systems 2002). While the NLA scheme was not based on the OAIS information model, it was apparently informed by it. The early work based on the OAIS information model provided some, but not much elaboration of the OAIS AIP.

The Online Computer Library Center (OCLC) and the Research Libraries Group (RLG) have jointly taken a consensus-building role in digital preservation and preservation metadata is one of the areas where these groups have been active. The OCLC/RLG Working Group on Preservation Metadata (2001, p. 1) was tasked with developing “a comprehensive preservation metadata framework applicable to a broad range of digital preservation activity.” All the members of the Working Group were based in libraries. The first task of the Working Group was to produce a review of the state of the art. This review covered the OAIS information model and the CEDARS, NEDLIB, NLA and Harvard University Digital Repository Services approaches. The Harvard University element set was included as an illustration of the use of XML in the representation of metadata (OCLC/RLG Working Group on Preservation Metadata 2001, p.17). The OCLC/RLG state of the art document included a mapping of the NLA, CEDARS and NEDLIB schemas and concluded that achieving consensus on preservation metadata is possible because of the degree of convergence between the schemas.

The Working Group carried out further work on preservation metadata and published interim reports in October 2001 and April 2002 (2001, 2002b) and a final report in June 2002 (2002a). The Working Group’s framework used elements from the different existing

schemes and also added new elements. However, none of the elements were mandatory, reflecting the potential practical difficulties of acquiring the metadata. Another criticism of the metadata framework has been that digital objects are treated more as technical rather than conceptual entities (Hofman 2002, p. 17). While it is necessary to know the technical components in order to preserve conceptual entities, these will change over time and Hofman is not sure whether the changes can be recorded and suggests including an element to capture this data. The conceptual object also needs to be encapsulated in the metadata. The "Significant Properties" element may be used for this purpose, but Hofman seems unconvinced (Hofman 2002, p. 17) and sees a need for greater separation of the technical and conceptual in the framework. The Working Group thought that the framework would be implemented in different ways for different preservation operations.

#### **4.3.2. *Provision and Management of Preservation Metadata***

Preservation metadata schemas are only useful if they can be populated with metadata. While libraries would have to supply some of the metadata, the nature of some elements of the OCLC/RLG schema suggest that cooperation with publishers will be necessary to do this. The publishing industry is involved in the development of metadata schemes, but the focus is on the exploitation of intellectual property rather than its long-term management. The Online Information eXchange (ONIX) is a standard format for the exchange of product information between different players in the information industry.

ONIX defines:

both a list of data fields about a book and how to send that data in an "ONIX message." ONIX specifies over 200 data elements, each of which has a standard definition, so that everyone can be sure they're referring to the same thing. Some of these data elements, such as ISBN, author name, and title, are required; others, such as book reviews and cover image, remain optional. While most data elements consist of text (e.g., contributor biography), many are multimedia files, such as images and audio files. (Editeur, 2002b)

There have been various studies investigating how libraries could use ONIX metadata, including mappings of ONIX to various MARC bibliographic data schemes (Editeur, 2002a). Brindley carried out a mapping between the OCLC/RLG framework and ONIX in 1992 (Brindley, Muir & Probets 2004). This study found that, as it stands, ONIX provides

some, but by no means all the metadata that libraries would require for preservation purposes. There would be some metadata elements that could only be populated by the preserving institutions. ONIX does provide much of the descriptive information. However, it lacks the detailed technical information that preserving institutions would need to implement technical preservation strategies. If publishers adopt ONIX, there is the possibility of exploring the extension of ONIX to provide preservation related information. The British Library has been involved in the development of the SIMONE (Simple ONIX Editing Tool) tool to help publishers produce ONIX metadata. ONIX messages can be sent by email or FTP. There are some other initiatives that could prove helpful to legal deposit libraries in preserving digital content. For example, some standards have been developed for digital content that incorporate metadata elements. The MPEG-7 standard for audiovisual content includes description schemes that include compression methods and access conditions (Day 2003b, p. 3).

OCLC and RLG continued their work by exploring how preservation metadata could be implemented. The objectives of the PREMIS (PREservation Metadata Implementation Strategies) working group were:

- Identifying an implementable set of "core" preservation metadata elements, with broad applicability within the digital preservation community
- Developing a data dictionary to support the core preservation metadata element set
- Examining and evaluating alternative strategies for the encoding, storage, and management of preservation metadata within a digital preservation system, as well as for the exchange of preservation metadata between systems
- Pilot programs for testing the group's recommendations and best practices in a variety of systems settings
- Exploring opportunities for the cooperative creation and sharing of preservation metadata. (OCLC 2004)

The PREMIS work started in June 2003 and is focused on a wide preservation community, including libraries, museums, archives, government and the private sector (Lavoie 2004a). This is necessary to ensure interoperability between preservation institutions and avoid duplication of work. The BL has representatives on the PREMIS working groups. The

bigger problem is that the different stakeholders in the digital publishing industry have different goals and the metadata they produce and use will reflect that. It is probably not feasible to try and standardise the metadata schemes of publishers and libraries, but it would make preservation easier for legal deposit libraries if there was interoperability between them.

In the second half of 2003, the Core Elements sub-group carried out an element-by-element review of the OCLC/RLG framework comparing them with elements in use in the groups members' institutions. This involved deciding which elements were deemed to be "core" and how they could be populated. Some potential new elements were also discussed. The sub-group also produced a report on alternative strategies for modelling the relationships between archived digital objects (Lavoie 2004). The Implementation Strategies sub-group carried out a survey at the end of 2003 to identify "the goals and characteristics of existing and planned digital preservation repositories, with particular emphasis on how metadata is being used to support repository processes, functions and policies." (Lavoie 2004). PREMIS work was due to end in mid-2004, but at the time of writing a report of the outcomes of the work had yet to appear.

One initiative the PREMIS group may have investigated is the National Library of New Zealand Preservation Metadata Scheme. This is "designed to strike a balance between the principles expressed on the OAIS Information Model and the practicalities of implementing a working set of preservation metadata". There is a need "to implement a workable solution within existing resources and organisational structures" (Searle & Thompson 2003). The schema has four entities: object; process; file and metadata modification. The eighteen Object elements describe the logical object, that is, the file or set of files that made up the digital object to be preserved. There are thirteen elements that document the preservation Process, including the objectives, who has given permission, the equipment used and the end results of the process. File metadata provides technical information about each of the files in the logical object. The five Metadata Modification elements document changes made to the preservation metadata. Each set of preservation metadata will relate to a logical object. However, the nature of this logical object may differ. It may be a single file, or several files that have to be present for the logical object to function. A group of objects may also be viewed as a single logical object, for example a set of letters.

The development of the New Zealand metadata scheme was informed by the OCLC/RLG framework work mentioned above, which was seen as useful but not sufficiently applicable to the National Library's working practices. Searle and Thompson (2003) feel that digital preservation can only become incorporated into the Library's routine business if it becomes as "productionised as other processes such as cataloguing". Given limited resources, successful implementation will require limiting the scope of the preservation metadata to data that are key to the preservation function rather than supporting other activities such as discovery. Other requirements are automating the process as far as possible and tracking changes to the metadata.

Another initiative with a possible role in managing preservation metadata is METS. The Digital Library Federation developed the Metadata Encoding & Transmission Standard (METS), "an XML document format for encoding metadata necessary for both management of digital library objects within a repository and exchange of such objects between repositories (or between repositories and their users)" (Library of Congress 2003c). This document includes different types of metadata:

- *Descriptive Metadata* – this may be embedded in the METS document and/or there may be a pointer to external descriptive metadata external to the METS document (e.g., a MARC record)
- *Administrative Metadata* - information on the creation and storage of files, intellectual property rights, provenance of the files (i.e., master/derivative file relationships, and migration/transformation information). This can also be embedded or external
- *File Groups* – a list of all files comprising all electronic versions of the digital object
- *Structural Map* - outlines a hierarchical structure for the object, linking the elements of that structure to content files and metadata for each element.
- *Behavior* [sic] - used to associate executable behaviours with content in the METS object (Library of Congress 2003c).

Waibel (2003) thinks that METS can be used to implement the information packages defined by the OAIS model. However, he also states “METS applies exclusively to the digital realm of images, audio, and video.” He does not give a reason why it would not apply to text as well. It may be because he is thinking about digitised material, rather than born digital material because he refers to “digital surrogates”. Digitised versions of text material may well exist as image files. METS can be used to bind together metadata from different sources. So metadata produced for purposes other than preservation, for example discovery or trade, could be re-used by preserving institutions. For example, a METS document could include or provide pointers to Dublin Core, MARC, ONIX and specialist preservation metadata for a digital publication. It could also provide pointers to sources of information describing the different metadata schemas.

#### **4.3.3.           *The Role of Persistent Identifiers***

The role of identifiers in facilitating discovery of digital publications was discussed in section 3.2.1. In 1998, the Conference of Directors of National Libraries (CDNL) decided to set up a Persistent Identifiers Task Force. A meeting of the Task Force in 1999 concluded that persistent identifiers were important in preserving access. However, there was no satisfactory approach in existence at the time for all players in the information chain. The Task Force thought that diverse requirements of different players meant there would be different systems in place, so interoperability is key (Berthon 2000).

If publishers assign unique and persistent identifiers to their publications, legal deposit libraries could possibly use these identifiers to manage their collections. There has been little discussion in the literature of whether and how legal deposit or other types of library could use publisher systems such as the DOI for managing preserved digital content. From the literature, it looks like legal deposit libraries are assigning their own identifiers to their collections. For example, the National Library of Finland registered its National Bibliography Number (NBN) with the IETF as a URN Namespace Identifier. The CDNL Persistent Identifier subcommittee then accepted this as a potential model for a common URN based identifier for national libraries (Drimmelen 2002). No resolver system has been defined for the NBN. The NLA decided that none of the current schemes was mature enough for its use, at least in the short-term. So the NLA developed its own “managed



URL style naming scheme” (National Library of Australia 2001f). The URL naming scheme will work with a resolver service that will redirect requests for items to the current storage location for the item. The details of the naming system are documented in detailed guidelines (<http://www.nla.gov.au/initiatives/persistence/>). Apparently the way the system is set up means that it could later be used to implement another system such as the DOI system.

Kunze (2003) criticises conventional persistent identifier schemes, including DOIs and PURLs. These schemes work on the premise that persistence can be achieved by using indirect names that resolve to the current URL for a resource. In other words, the identifiers are actionable. However, this only works if the resolver mechanism continues to work. For example not updating resolver databases with new locations will effectively break the link between the identifier and the resources. Kunze (2003, p. 7) does not think existing systems are able to convey the promises that are required in the provision of persistent access to digital material. The promises he identifies are:

- How long the identifier will be valid
- How long access to the object will be available
- To what extent will the content of the object change

According to Kunze (2003, p. 8), the breaking of links is the fault of service providers and that persistence is a matter of service. The authorities involved naming and maintaining links between identifiers and resources in existing schemes change over time and may not have the commitment or resources to maintain persistence. Although if legal deposit libraries were using their own schemes, it is unlikely that they would not maintain them. In his opinion “long-term actionable identifiers” should: deliver users to objects, or to object metadata, or to statements of commitment on the part of preservation service providers and identifiers should be part of the digital object.

The California Digital Library is using an alternative approach, the Archival Resource Key (ARK). This system involves an identifier and the associated services identified above. The identifier is “a special kind of URL” (Kunze 2003, p. 9). The URL includes a domain - the Name Mapping Authority Hostport (NMAH); the ARK label (ark:); the Name

Assigning Authority Number (NAAN), a globally unique registered number for the Name Assigning Authority and the name assigned by the NAA. Depending on how users enter the URL, they can obtain, for example the object (the URL on its own), the metadata (the URL plus '?'), the commitment statement (the URL plus '??'). The ark identifier does not include any meaning. This is a deliberate attempt to avoid language issues; the meaning is contained in the metadata. The NMAH is temporary and replaceable over time, but the rest of the identifier remains over time. So there is still a link to break and persistence involves a new NMAH taking over. If legal deposit library domain names are unlikely to change much over time, so this system might work for them. The fact that the ark can be separated from the domain name also means the identifier can still be used when the Web no longer exists and something else has taken its place.

#### **4.4. Infrastructure**

The RLG/OCLC Task Force on Digital Archiving report (Waters & Garrett 1996) referred to the need to develop a “deep infrastructure” to support the preservation of digital information. The Task Force identified two components of that infrastructure, certified archives and a fail-safe mechanism to rescue digital information in danger of being lost. Clearly, legal deposit libraries have a prominent role to play in the provision of trusted digital archives and capture of at risk material, but they are by no means the only likely players in the preservation of digital information. There is widespread recognition in the literature that even prominent deposit libraries such as The BL and the Library of Congress cannot solve all the issues on their own. There will need to be collaboration between different players, including creators, publishers, technology vendors and other preserving institutions. The supporting infrastructure does not yet exist, but work is being done to create some of the building blocks.

Granger (2002) has set out his view of what this deep infrastructure would involve and discusses some of the issues to be dealt with in order to move towards this infrastructure. Granger envisages an enabling technical infrastructure including the use of standards, open systems and open source software; technical rights protection systems that do not inhibit preservation; and metadata standards and tools. This technical infrastructure will develop as digital preservation requirements continue to be identified. Basically the infrastructure

would provide “tools, techniques or components to make digital preservation easier and cheaper”. There are also organisational, legal and cultural aspects to the deep infrastructure. Granger identifies the intellectual property rights and business model issues discussed in chapter 3 that need to be dealt with. He also identifies the need for culture change so that technologies will not be developed with only the needs of vendors in mind, but those of users and the future preservation of the digital heritage. The preservation community also needs to become more proactive and less resigned to a vendor-driven culture. He agrees that the organisation part of the infrastructure includes the establishment of certified archives, “trusted organizations capable of storing, migrating and providing access to digital collections” and Granger does recognise the need for deposit of digital material. The process of certification is being taken forward (see section 4.4.2).

Granger suggests the need for standards here. The work of the preservation institutions should be supported by research and development centres; this is also being addressed (see Chapter 5). Granger suggests that institutions could collaborate to support centralised centres, carrying out work on, for example developing and maintaining preservation and metadata tools and providing data recovery services. Finally, Granger identifies the need for collaborative structures; this will be needed to achieve the infrastructure Granger envisages and will be difficult to achieve, as he acknowledges. The different actors have different interests and some of these are in conflict. Granger recognises that his model is incomplete, does not take economics fully into account and will be difficult to achieve, but at least it does provide some sort of framework for digital preservation.

The rest of this section reviews work on activities and research relating to the deep infrastructure idea, how they relate to legal deposit libraries and how legal deposit libraries can contribute.

#### **4.4.1.            *Repositories for Metadata and Documentation***

Preservation strategies will require the technical and administration metadata mentioned above, as well as access to documentation on file formats, software and hardware specifications. The potential role of METS for linking content and metadata has been discussed above. Day (2003) has discussed how linking of metadata and its explanatory

documentation could be implemented. The metadata could be linked to the content in an AIP or the metadata and its associated documentation could also be archived as an AIP and linked to the content. He does not think that this approach is efficient because it would multiply the number of AIPs being stored and would result in duplication because metadata being preserved may be held as an AIP, but also be in use in some sort of database.

Like Chilvers (2001, p. 166), Day (2003, p. 4) proposes developing “registries of the better-known metadata standards”. Day elaborates on this by stating the registries could provide authoritative information on individual standards, including definition of terms and how elements should be used. Registries could also provide mappings between different standards (Day 2003b, p. 4). This could help preservation institutions use externally supplied metadata in their internal metadata systems. Day discusses registries in the plural and assumes that there will be a need for interoperability between registries, presumably because the needs of their users will overlap. He gives some examples of existing registries including subject communities, such as environmental or educational metadata. Others cover different types of standard, for example, XML schemas and DTDs or audiovisual content standards and schemas (Day 2003b, p. 5).

There is a question of whether legal deposit libraries could use existing repositories set up for other purposes or whether they should set up their own. The National Library of New Zealand has signalled its intention to develop such a repository (Searle & Thompson 2003). As Day (2003, pp. 5-6) points out, maintaining repositories will be resource intensive, so there may be scope for legal deposit and other types of library to cooperate. The DCMI (Dublin Core Metadata Initiative) Registry Working Group has been working on the development of a generic registry tool (Heery & Wagner 2002) that could perhaps be used by legal deposit libraries, especially since one of the requirements identified is that it should be multilingual, both from a user interface and data point of view.

Lawrence *et al* (2000) found that there are also a number of issues associated with developing file format registries, including the existence and accuracy of this information, and its availability. Researchers at the University of Leeds (2003) carried out a survey of sources of information on file formats and software documentation and came to similar conclusions. The Leeds team also recommended that information should be collected and

preserved in a repository as suggested by Robertson (1996). In the UK, the Science Museum and the Computer Conservation Society are already working towards this and the Leeds team recommend that these organisations should be given assistance as a matter of urgency. The repository should form part of an OAIS “Representation Network” (University of Leeds 2003, p. 44). Open source rendering tools, e.g., logical definitions of file formats and migration tools, should also be deposited as appropriate. The Representation Network could also provide links to material not deposited in the repository and include a technology watch function. The Joint Information Systems Committee funded the research and presumably the recommendations were mainly aimed at the JISC.

In August 2003, JISC published an invitation to tender for a Digital Curation Centre (see section 5.2.4) which would provide the technology watch service and also “develop a reliable, sustained repository of generic tools, software, and documentation, to support curation, preservation and use of digital resources” and “pilot development of services for recording and monitoring file formats and preservation planning tools” (JISC 2003, p. 2).

The UK National Archives (2003) has developed an application for managing information on file formats and rendering software used for electronic records. The National Archives is making the application available to others to wish to use it. So, potentially, the legal deposit libraries could use the application to collect information related to digital publications. The National Archives has built up a database on information for its own use, but supplies the application without this information. There will be a degree of overlap between records and publications, so it would make sense for the National Archives to share their information with at least the UK national libraries and vice versa to avoid duplication of effort and reduce costs. It also makes sense for legal deposit libraries to work with other interested bodies in the UK, such as the JISC and the National Archives to share the burden of gathering information for preservation purposes.

There is a proposal to develop a global registry of digital object format. The proposed registry may involve a repository of format specifications, or links to “persistent external sources” (Abrams, 2002, p. 2). It also involves “persistent, unambiguous binding between unique format *identifiers* and encoded descriptions of these formats” (Abrams, 2002, p. 1). The global repository does not need to be a centralised one, but it does need to be

standardised. Members of the Harvard Library Digital Initiative and the Massachusetts Institute of Technology DSpace project initiated the proposal. An ad-hoc working group was set up to explore issues such as data and service models, governance structure and business issues (Abrams & Seaman 2003). The National Archives and JISC are represented on this working group, but none of the UK legal deposit libraries are taking part at this stage.

The NLA was planning to propose the creation of an international software repository to the CDNL Digital Issues Working Group (Gatenby 2002), but it is not clear from the literature whether this proposal was made and whether it was taken up.

The envisaged Typed Object Model (TOM) architecture looks like a useful contribution to link repositories and preserving institutions. Type brokers maintain information on data types and “services on those data types” (Ockerbloom 2002). The type brokers can give and receive information from clients on data types. They can also mediate between clients requiring services, perhaps conversion, to be carried out on digital objects. The TOM system is envisaged as a distributed one with different brokers exchanging type information.

#### **4.4.2.            “Trusted” Digital Archives**

Legal deposit libraries are likely to be trusted to successfully preserve digital information because they already do this for more traditional information formats. RLG and the Commission on Preservation and Access (CPA) Task Force on Digital Archiving published a report in 1996. One of the recommendations was that there was a need to define the attributes of trusted digital archives and develop a scheme of certification (Waters & Garrett 1996, pp 23-24 & 42 ). An OCLC/RLG working group took these recommendations forward by developing a statement of the responsibilities and characteristics of a digital repository (RLG & OCLC Working Group on Digital Archive Attributes 2002). These included a commitment to digital preservation and operating systems in a well-managed and transparent way. The US National Archives and Records Administration (NARA) and RLG then set up a Task Force on Digital Repository

Certification to identify elements of a digital preservation system that can be certified and develop a plan for the bodies, policies and procedures for certifying digital archives.

This work is probably more applicable to organisations other than legal deposit libraries, but since they hold collections of last resort for perpetuity, they should conform to best practice in the area. In fact, a representative of the BL was a member of the OCLC/RLG Working Group as were representatives of the Bibliothèque Nationale de France and the NLA. The BL is not directly represented on the NARA/RLG Task Force.

#### **4.4.3.            *Stakeholders and Roles***

Legal deposit libraries are just one type of organisation with an interest in digital preservation. Haynes *et al* (1997) examined the attitudes of “originators and rights holders” towards to the issue of their responsibility for digital preservation. The specification for this study was that the consultants should talk to a range of informants representing publishers, libraries, research organisations, data compilers and repositories. They did this, but respondents were based in the UK (with one in Ireland) only. Although the study remit was responsibility for maintaining an archive of digital material in the UK and Ireland, an international perspective may have been useful.

The objectives of the study were to investigate:

- Willingness/perceived ability to preserve/hold data in perpetuity
- Willingness/desire to pass on this responsibility
- Ability to preserve unpublished data where accompanying documentation/bibliographic details are insufficient

The study was to test two models for digital preservation. There were that the publisher retains the database or the database is deposited with a research library/digital archive. The methods used in the study were interviews and focus groups. These methods seem appropriate for identifying and exploring attitudes, although not all of the interviews were face-to-face, but carried out using telephone or email.

The consultants made a number of recommendations in their report. The first was that a body should be established to co-ordinate digital archiving activities. The consultants suggested that this could be an extension to an existing organisation such as the National Preservation Office (NPO). In fact, this role seems to have been taken on by the Digital Preservation Coalition (DPC, see section 5.2.3). Another suggestion was for a distributed approach to digital preservation. Distribution could have a number of bases, including regional, format and ownership. A national register of digital material would facilitate the existence and location of digital material within archives. The suggested National Office of Digital Archiving would coordinate the development of standards and guidelines in cooperation with other agencies.

The consultants conclude that public digital archives should be funded publicly. The consultants suggest that legal deposit should be used as a mechanism for acquiring material, but that publishers should only have to contribute one copy of each publication. This now seems to be the BL's preferred option as long as networked access is permitted (see Chapter 2). There is no recommendation as to what depositories could do with deposited material, for example, whether it should be copied to build redundancy into the system. The consultants suggest that users should not be charged for access, but costs should be shared between research funders, the public (through Government funding), and research communities. There is no mention of any responsibility falling on publishers for maintaining archives.

The DPC commissioned a survey of information technology vendors, carried by Lord (2002) between July and October in 2002. The aims of the study included investigating awareness and perception of digital longevity issues. Other aims related to views on the work of the DPC and cooperation between the DPC and the industry. This study is relevant to legal deposit libraries because they have to deal with published material and are not able to dictate the technologies depositors use for their material. A mixture of large, medium and small companies was invited to complete an email questionnaire. Of the 35 companies approached, 21 completed the questionnaire, and 17 of these also took part in telephone interviews. Findings of particular interest to legal deposit libraries were that there was a perception of a lack of awareness of the problem although most respondents said they did address longevity issues in their own products and services. Companies respond to market requirements, so promoting awareness among the public and customer



was seen as important. Apparently the customers of the vendors surveyed were not asking for digital longevity except when there were specific pressures, for example legal or regulatory obligations. However, respondents did indicate that customer awareness is increasing. Given these findings, it would seem to be in the interest of legal deposit libraries to support awareness-raising activities. Survey respondents showed an interest in the work of the DPC and suggested ways of increasing awareness in the industry, including seminars, workshops and press items. There was some enthusiasm for collaborating with the DPC, but the need to see benefits for the companies was emphasised.

#### **4.4.4.            *Incentives to Preserve***

While raising awareness is useful and necessary it has to result in action. Lavoie (2003) has reported on incentives to preserve digital information. He suggests that current organisational models involve few incentives to preserve and that this needs to be rectified. Like the OAIS Reference Model, Lavoie condenses the main roles in preservation to rights holder, archive, and beneficiary. He also sets out several organisational models ranging from all the roles being held by one entity to separation of all three entities. It is when the rights holder and beneficiary roles are separated that there is least incentive to preserve. However, this model applies to digital preservation in general; publishers will have a legal obligation to deposit. Legal deposit libraries have a desire and obligation to preserve. However, they may not have sufficient resources. At the moment the legal deposit libraries are dependent on public money. Other funding models for legal deposit have been suggested (Ayre & Muir 2004, p. 326), including the provision of endowments by rights holders or preservation services to publishers by legal deposit libraries. Chapman (2003) discussed models for charging for different levels of preservation service. However, in the legal deposit context particularly, such models are dependent on rights holders seeing some benefit for themselves. The benefit may emerge if customers put pressure on publishers to ensure that their material stays accessible.

This may be the underlying motive for the agreements between publishers operating in the Netherlands and the KB. While the terms of these agreements have not been publicised, there may be some payment to the KB for providing services. On the other hand, since

there is no statutory obligation to deposit, the services provided may be in exchange for deposit of content. Asking users to pay for access to preserved material is another possibility, but this could well be unacceptable (Haynes *et al* 1997, Ayre & Muir 2004).

A recent enquiry into scientific publishing in the UK advocated the open access approach to publishing (United Kingdom, House of Commons, Science and Technology Committee 2004). This involves a different business model where access is free and publishers obtain their revenue by other means, including charging authors. Since the aim is to increase access, rather than charge for it, the tension between preservation and access should not exist for open access publishers. Whether the open access ethos extends to actively helping in the preservation of content remains to be seen, but the report strongly supported legal deposit of digital scientific publications and provision of Government funding to support it (United Kingdom, House of Commons, Science and Technology Committee 2004, pp. 90-93).

#### **4.5. Summary**

The importance of life cycle management has been recognised for digital publications and this idea has been developed as an underlying framework for digital preservation policies. The ephemeral nature of digital publications has also been recognised in work on developing models for risk assessment and management. Selection has been addressed through suggestions for preservation selection criteria, categorisation of digital material types and assessment of digital preservation strategies and costs. However, the assessment models developed so far are limited in the types of digital material and the strategies that they cover. The OAIS Reference Model is an influential development in standards for digital depositories, although there is little advice for its practical implementation. Facilitating preservation of digital information through metadata and documentation is another important area of activity. Examples of this include work on “persistent” archives, the development of digital preservation metadata schema, exploration of interoperability of metadata originating in different parts of the information chain and development of persistent identifiers. Developments in the supporting infrastructure include plans for repositories for metadata and documentation, although the management and sustainability of such registries will be problematic. Finally, the relative roles and responsibilities of

different stakeholders, including legal deposit libraries is another issue that is raised in the literature. National and international approaches to legal deposit and preservation of digital material are discussed in Chapter 5..

## **5. RESEARCH, DEVELOPMENT AND INITIATIVES**

This first part of this chapter deals with national and international initiatives in digital preservation. Initiatives in the United States and the United Kingdom are particularly well-documented so are prominent here. Major research and development projects in digital preservation generally and digital deposit are also discussed. Finally, activities in a selection of legal deposit libraries that are leading the way on deposit of digital material are discussed. National libraries have been prominent in activities, although not always leading them. However, the findings of a study trip undertaken by Gatenby in 2002, showed, with some exceptions, a lack of “long-term corporate planning for their new roles in the digital age” by national libraries. Digital-related activities were not integrated into routine activities and the libraries did not have appropriate technical infrastructure in place. National libraries are now addressing some issues through the Digital Issues Working Group of the Conference of Directors of National Libraries (CDNL). The overall picture is very disjointed, but there are signs that this is changing.

### **5.1. United States**

#### **5.1.1. *Task Force on Digital Archiving***

In 1994 the Research Libraries Group (RLG) and the Commission on Preservation and Access (CPA) set up a Task Force “to investigate the means of ensuring 'continued access indefinitely into the future of records stored in digital electronic form' (Waters & Garrett 1996, p. iii). This aim highlights one of the definition problems in this area. As Bearman points out, the Task Force did not spend much time considering records as understood by archivists, but digital archives as understood by librarians.

The report sets out the challenges of preserving digital information, pointing out that the biggest problem is technological obsolescence and advocating migration as a preservation strategy. Bearman (1996) critiqued the Task Force report from an archival perspective. Most of his remarks are not relevant in the legal deposit context, but some are. For

example, the Task Force assumed that migration is “a loss-less” transformation”, but in most cases, there is some loss. As discussed in section 3.4, the question of how much loss is acceptable for different kinds of information requires agreement.

The report contains a number of recommendations for future action. These include setting up projects to advance digital libraries, the development of infrastructure to support digital archives and identification of best practice. A controversial recommendation is the development of fail-safe mechanisms to ensure that digital information in danger of being destroyed can be aggressively “rescued” by digital archives. Bearman is critical of the suggestion that owners have the main responsibility for preserving their information and that accredited archives should have the power to intervene when valuable information is in danger of being destroyed by its owners. He says this does not occur for any other kind of private property. However, in the UK at least, owners of the built heritage are prevented by statute from damaging alteration or destruction of their property. Bearman also points out that the identities and roles of the various stakeholders in digital preservation are not always clear and are often conflicting. Matthews, Poulter and Blagg add that aggressive rescue presupposes that archives have the knowledge to select endangered material and know how to preserve it (Matthews *et al* 1997).

### **5.1.2.            *Research Libraries Group and OCLC Joint Working***

Joint work by RLG and OCLC has been largely covered in previous chapters, so only brief mention is made of it here. Work on preservation metadata and trusted repositories has respectively built on previous work and followed up the Task Force on Digital Archiving recommendations. OCLC and RLG have been working together on preservation metadata since 2001 and have progressed from reviewing the state of the art in this area to synthesising previous efforts at developing preservation metadata schemas to develop their own framework to addressing “the practical aspects of implementing preservation metadata in digital preservation systems” (OCLC 2004).

OCLC set up a new division in 2001 to “support the challenges of creating and sustaining access to and preservation of the global knowledge base’s contents” (Bellinger 2002, p. 39). OCLC is building a digital archive for storing various types of information which may

be used as the basis for providing goods and services. Work on the archive will also identify workflows for capturing and managing digital material. OCLC also wants to implement a metadata set for material in the archive. The archive will be integrated into OCLC's *WorldCat* union catalogue. Users will be able to see items catalogued in *WorldCat* by clicking on a URL. Owners of material in the archive will be able to control access to it. In terms of content, the initial focus was on capturing Web documents. Later phases of building the archive involved historical newspapers, e-journals, and special collections (Bellinger 2002, p. 42). The team developing the metadata set was informed by the joint work in the area with RLG (Bellinger 2002, p. 44) as well as other metadata initiatives. The archive complies with the OAIS Reference Model. Presumably OCLC will benefit from its participation in cooperative metadata activities. Perhaps the organisation will also use its participation in trusted repository work for its own business too. How OCLC would then fit into future preservation infrastructure is not clear. OCLC could conceivably provide a service for national libraries, but as a semi-commercial organisation it may not be considered suitable as a custodian of legal deposit material.

### **5.1.3.            *National Digital Information Infrastructure and Preservation Program***

In 1998, the Library of Congress commissioned a study of how well it was positioned to deal with the challenges of the digital environment. The National Research Council and the National Academy of Science brought together scholarly communities and experts in the archival and information technology fields to answer this question. The resulting report was critical of the Library (Committee on an Information Technology Strategy for the Library of Congress *et al* 2000). However, this enabled the Library, along with the National Archives and Records Administration (NARA), to go to the US Congress with a request for \$100 million to create a national strategy for the collection, archiving and preservation of digital content in cooperation with various partners. Congress agreed to the funding in instalments on the condition that the Library of Congress provides up to \$75 million of matching funding.

The plan to develop the National Digital Information Infrastructure and Preservation Program (NDIIPP) has three phases:

- The development of a master plan
- Development of partnerships with the archival and content creator and provider communities
- Testing and evaluation of partnerships and models in order to return to Congress to discuss the most sustainable options for long-term preservation (Library of Congress 2002)

The planning process started in 2001. The Library of Congress established a National Digital Library Advisory Board including members of the various communities with a stake or role in the development of the national strategy. The Library of Congress is leading the NDIIPP, working with several federal partners including the Secretary of Commerce, the Director of the White House Office of Science and Technology Policy, and the National Archives and Records Administration (NARA). Other partners include the National Library of Medicine, the National Agricultural Library, the Research Libraries Group, OCLC and the Council on Library and Information Resources (CLIR).

The planning process involved different strands of activities: stakeholder meetings; collaborative research; development of a conceptual framework and scenario planning. The Library commissioned a set of surveys on the state of the art in digital preservation of the different types of material the Library will have to deal with (Council on Library and Information Resources & Library of Congress 2002) and national and international initiatives (Beagrie 2003). Staff carried out interviews with various industry groups (Friedlander 2001a) and held workshops (Friedlander 2001b).

The Library of Congress commissioned a draft conceptual framework for the digital preservation infrastructure from PriceWaterhouseCoopers. In this framework, the digital preservation environment is affected by various economic, political, social/cultural, legal and technical factors. The preservation environment involves three levels: organisational, functional and technical. The organisational layer includes players, including preservation institutions, publishers, technology vendors, coordinating bodies, enabling agreements and guidance. The functional level includes preservation and collaboration services and storage and rights clearance services. The technical layer involves standards and supporting technologies (Friedlander 2002). This model sounds like a further articulation of the

concept of “deep infrastructure” discussed by the RLG/OCLC Task Force on Digital Archiving (Waters & Garrett 1996) and Granger (2002) and should facilitate the discussions required to meet the aims of the NDIIPP.

Technical experts also modelled the technical infrastructure in 2001 (Campbell 2002). The architecture involves four layers: repository, gateway, collections and interface. It represents components in a distributed digital preservation operation rather than services provided by individual libraries. The digital bits live in the repository layer, the gateway controls access to the digital content. The collections level includes collection management activities, including digital preservation and the interface provides information to users.

The Library of Congress (2003a) received approval of its master plan for the NDIIPP in February 2003. Elements of the plan include selection and collection development; clarifying intellectual property issues; development of business models for digital preservation and development of standards and best practices. Plans for selection and collection development include working with other national libraries on the collection of both Web-based and licensed materials. The Library of Congress commissioned a report on copyright issues associated with US digital archives (Besek 2003). Lavoie (2003) produced a paper on incentives to preserve (see section 4.4.4). This paper was written for OCLC, rather than under the banner of the NDIIPP. However, since OCLC is cooperating in NDIIPP no doubt the findings will inform work in this area. The Library of Congress was planning to convene a design group to develop the digital preservation architecture (Library of Congress 2002, p. 7).

In addition to the work mentioned above, the Library of Congress is also funding research and development activities through the NDIIPP. Towards the end of 2003, the Library announced a call for proposals for funding for projects to advance the NDIIPP. The projects were to focus on the following goals of the NDIIPP:

- “The selection and collection of at-risk and historically significant digital materials for which no analog equivalent exists and



- the development of a network of committed NDIIPP partners with defined roles and responsibilities to support the long-term collection and preservation of digital content.” (Library of Congress 2003b)

The Library of Congress has also become involved in two research and development projects as part of NDIIPP work. The projects involve working with four university libraries to “explore strategies for the ingest and preservation of digital archives” (Archive Ingest and Handling Test (AIHT)) and testing whether a digital archive can be successfully transferred from the Library of Congress to other libraries (Digital Information Infrastructure and Preservation Program 2004b & 2004c).

In June 2004, the Library of Congress announced that it was launching a research grants programme in partnership with the US National Science Foundation (NSF) (Digital Information Infrastructure and Preservation Program 2004a). The NSF will administer the programme. The Digital Archiving and Long-Term Preservation research programme will initially distribute \$2 million in awards. The research areas to be funded are:

- Digital repository models
- Tools, technologies and processes
- Organisational, economic and policy issues (Digital Information Infrastructure and Preservation Program 2004a).

These areas were identified at a joint workshop held in April 2002 (Hedstrom 2003). At the time of writing, no call for proposals had yet been made. Research funded under this programme is likely to be more basic than the activities the Library of Congress is participating in so far with other libraries and organisations. The director of the NDIIP acknowledges the need for both long and near-term research (Campbell 2002) to help turn ideas into working solutions.

It is likely that this programme will have significant outcomes that will be useful to other national and legal deposit libraries in countries not fortunate enough to attract sufficient funding to take a strategic approach to digital preservation.

## **5.2. The United Kingdom**

In the UK, The British Library (BL) and the Joint Information Systems Committee of the UK Higher Education Funding Councils (JISC) and, to a lesser extent, the National Preservation Office (NPO) have led the library community.

### **5.2.1. *JISC/NPO Studies and the Warwick Meetings***

The British Library commissioned a study on the preservation of digital material, which Tony Hendley carried out in 1995 (Hendley 1996). The same year, JISC and the BL sponsored a workshop on digital preservation that was held in Warwick University and came to be known as 'Warwick 1'. The report of this workshop articulated the reasons for these two bodies' interest in digital preservation. JISC had been funding many projects involving the creation of digital resources through its eLib Programme and presumably wanted to take some responsibility for the preservation of these resources. The BL's reasons for involvement were as a major research library, a (then) funder of library and information research and development and as a major deposit library (Marc Fresko Consultancy 1996, p. 1).

The agenda for the workshop was influenced by the draft version of the RLG/CPA Task Force Report. In particular, there was an interest in exploring the applicability of the report's recommendations in the UK (Marc Fresko Consultancy, p. 1). The objectives of the workshop were to:

- Explore strategic options for developing and managing electronic archives
- Consider possible collection policies for electronic materials
- Consider preservation policies
- Examine practical implications of the various options
- Propose research and action needed, to resolve/shed light on the issues discussed

There is no definition of the use of the term archives. Given the affiliations of the speakers and attendees, it is likely that for the purposes of the workshop "archives" covered both records and library material. Eighteen recommendations came out of the workshop:

- *Facilitate debate* - via existing organisations to encourage moves toward commitment to the long-term preservation of digital information
- *Sensitise IT suppliers* - make leading suppliers aware of preservation issues, the British Library was identified as having a major role
- *Develop guidelines* - preferably suggesting approaches, but not overly prescriptive
- *Influence research funders* - to develop digital archiving policies for data from their projects and make preservation considerations a requirement of proposals
- *Sampling study* - on statistical sampling techniques for collecting electronic ephemera
- *Awareness activities* - encourage awareness of the importance of institutional archiving, including commercial as well as public organisations
- *Promote policies* - encourage universities and research institutes to establish policies for archiving datasets
- *Study archival criteria* - for archiving objects to analyse relevance for digital information
- *Investigate attitudes* - of creators of material to get a view of their attitudes to taking responsibility for preservation
- *Pilot a university/publisher archive* - a university would take responsibility for archiving on behalf of a publisher. The pilot would explore economic, technical and service issues
- *Interpret CPA/RLG report* - analyse the final version of the report to work out how relevant and applicable the recommendations were to the UK
- *Develop data type framework* - of the major data types to allow exploration of various issues associated with each type
- *Digital rescue study* - to assess the costs, practicality and other issues surrounding the rescue of valuable material which is already inaccessible
- *Preservation approaches* - to compare technology preservation, technology emulation and migration strategies
- *Costing models* - to identify the functions and components of costs as a first step towards developing a cost model

- *Media integrity* - commission studies on establishing and monitoring the integrity of storage media
- *Skills* - undertake activities to develop digital preservation skills
- *General awareness* - undertake awareness-raising activities

The recommendations were taken forward in a number of ways. JISC funded an analysis of the CPA/RLG Task Force report (Matthews *et al* 1997). The study consisted of a review of the literature relating to the report, such as comments on the report and an analysis of the report in the light of these comments. The analysis of the report includes a summary of points made as well as the analysis of the report's recommendations discussed in the previous section. The final recommendations that emerged from this study were:

- *Appoint a national Digital Preservation Officer*

The Director of the NPO was tasked with seeking funding for this post. This search was not successful because there is still no dedicated post and the Director has taken over responsibility for digital preservation matters. However, JISC did appoint a digital preservation officer.

- *Establish a national preservation body*

This body, the Digital Archiving Working Group (DAWG) was established as part of the National Preservation Office management structure. The Digital Preservation Coalition was also established some time later.

- *Investigate current and proposed digital archival practice and policy in the UK*

The Chair of the NPO Management Committee was to write to various bodies asking for details of their activities and plans. An upcoming conference was identified as a source of information as the various bodies would be present. There is no indication of what was to be done with the information

- *Identify good practice and gaps in knowledge*

This was to be done via research and/or pilot projects.

- *Devise guidelines on practice and a digital preservation policy*

This recommendation echoes that of the Warwick workshop. The UK Public Record Office and the NPO were identified as already working on this.

- *Raise awareness*

Again, this repeated a Warwick recommendation. The Web site developed for the study was to act as a starting point for this, and there was a recommendation that it should be maintained long-term by the NPO and publicised. An email discussion list was also to be launched. It is not clear whether these specific recommendations were taken up by the NPO, but JISC did establish an electronic mailing list.

- *Promote education and training*

A specific recommendation here was that the NPO should investigate a workshop on digital reformatting that was being planned by the Research Libraries Group with the intention of developing a parallel project on digital archiving.

- *International cooperation*

This was to be mainly through membership of committees.

The JISC and the NPO followed up the recommendation on identifying good practice and gaps in knowledge by commissioning a set of studies on digital preservation: the JISC/NPO Studies on the Preservation of Electronic Materials. The DAWG was set up specifically to oversee the programme. JISC funded the programme and the (then) British Library Research and Development Department administered the studies. There were six studies in all. These covered:

- Data types and formats (Bennett 1997)
- Preservation methods and costing models (Hendley 1998)
- Responsibility for digital preservation (Haynes 1997)
- Digital preservation needs of universities and research funders (Data Archive 1998)
- Development of a strategic policy framework for digital collections (Beagrie & Greenstein 1998)
- Digital archaeology (Ross & Gow 1999)

These studies covered many of the areas for research identified at the Warwick Workshop, but not all. There was no pilot for the preservation of material by a university library on behalf of a publisher, or to devise statistical sampling methods. The final reports of the studies were synthesised by Feeney in a separate publication (Feeney 1999a).

In March 1999, there was a second workshop at Warwick (Beagrie 1999). The focus of this event was preservation strategy. Warwick II produced thirty recommendations compared with the eighteen from the first workshop (Marc Fresko Consultancy 1996). The organisers envisaged using these recommendations to develop a plan of action for the next three years. One of the recommendations was to review the previous set of recommendations to identify any that had not been taken up and determine what action to take. Many of the recommendations related to the adoption of standards, the need for guidelines and to identify and publicise good practice. Another theme running through the recommendations was the need for coordination and cooperation. The workshop endorsed several recommendations from the JISC/NPO studies, including the need for a national forum and for training. The recommendations for research represent a move from the mainly basic exploratory activities of the JISC/NPO studies. For example, there is a recommendation to develop an “Initial Cost Benefit Model”. Presumably, this would build on earlier work on stakeholders and cost identification.

### **5.2.2.            *The CURL Exemplars in Digital Archives Project***

The CEDARS project ran from 1998 until 2001 and was part of the JISC eLib and Digital Preservation Focus programmes. The CEDARS project did not directly address legal deposit of digital publications, but it did focus on academic and research libraries. The aim of the project was to “address strategic, methodological and practical issues and provide guidance in best practice for digital preservation” (CEDARS 1998a). These are issues of interest to legal deposit libraries and several UK deposit libraries were involved in the CEDARS project in some capacity. The project plan (CEDARS 1998b) states that the main deliverables would be:

- Guidelines for developing collection management policies which will ensure the long-term viability of any digital resources included in the collection;
- Demonstrator projects to test and promote the technical and organisational feasibility of a chosen strategy for digital preservation;
- Methodological guidelines developed by the demonstrator projects providing guidance about how to preserve different classes of digital resources;

- Clearly articulated preferences about data formats, content models and compression techniques which are most readily and cost-effectively preserved;
- Publications of benefit to the whole higher education community, available on the WWW.

The project largely fulfilled these objectives, although it perhaps provided less guidance on specific preferences for strategies and formats. In fact the project's own assessment of its achievements included a comment that it "has advocated an overall management approach which will be suitable, regardless of which technical strategy is ultimately adopted" (CEDARS 2002d).

The CEDARS team was a partner in a follow-up project on emulation for preservation funded through the JISC/NSF International Digital Libraries Programme. The other project partner was the University of Michigan. The aim of the Creative Archiving at Michigan and Leeds Emulating the Old on the New (CAMiLEON) project was:

- "To explore the options for long-term retention of the original functionality and 'look and feel' of digital objects.
- To investigate technology emulation as a long-term strategy for long-term preservation and access to digital objects.
- To consider where and how emulation fits into a suite of digital preservation strategies." (CAMiLEON 2004)

The work of this project was discussed in Chapters 3 and 4.

### 5.2.3. *The Digital Preservation Coalition*

The recommendation for a national forum finally materialised in 2001 in the form of the DPC. The DPC's mission encompasses most of the recommendations of the Warwick meetings (Digital Preservation Coalition 2004c). The UK's three national libraries are members of the DPC. Oxford and Cambridge are presumably also represented through the Consortium of Research Libraries (CURL), which is also a member. While no individual publishers are members, publishers are represented through the Publishers Association

(PA) and the Association of Learned and Professional Society Publishers (ALPSP). It is not clear from the current membership how well less traditional publishers such as database producers and open access publishers are currently represented in the DPC. Neither is it clear how well library and information services other than academic and research libraries are represented and whether they should be.

The launch of the DPC was well covered by the UK media for a library and information initiative (Digital Preservation Coalition 2004b). The DPC has taken over the maintenance of the *Preservation Management of Digital Materials Handbook* (Beagrie & Jones 2002) and is working with the National Library of Australia's PADI gateway to produce a current awareness service (Digital Preservation Coalition, 2004d). There have already been a series of seminars and workshops on various topics. An interesting recent development is DPC sponsorship of a new Pilgrim Trust Conservation Award recognising "leadership and achievement in the developing field of digital preservation. It is aimed at highlighting the issues posed by preserving electronic information in the long term." (Digital Preservation Coalition 2004a). So the DPC is working on raising awareness and sharing knowledge and best practice. However, the DPC has a long way to go in this area of its work. Recent surveys on awareness and activities in digital preservation amongst UK libraries and publishers indicated that there is still a lack of awareness and understanding of the issues (Ayre & Muir 2004).

#### **5.2.4.            *Digital Curation Centre***

JISC and the eScience Core Programme are providing £1.3 million pounds annually to support a Digital Curation Centre (DCC). The Centre will be run by the CANDO (Collaborative Associates Network of Data Organisations) consortium, including the Universities of Edinburgh and Glasgow, the UK Office of Library and Information Networking (UKOLN) and the Council for the Central Laboratory of the Research Councils (CCLRC). The Centre will not be a central repository and will not carry out digital preservation. The aim is to support preserving institutions by funding research, promoting knowledge, expertise and good practice. The Engineering and Physical Sciences Research Council (EPSRC) will fund the research programme.



This suggests that the research will be technical in nature and there may be a danger that management and infrastructure issues will not be addressed. However, it will be interesting to see how libraries, including the deposit libraries, will interact with the DCC.

#### **5.2.5. *United Kingdom Web Archiving Consortium (UKWAC)***

Launched in June 2004, UKWAC has resulted from research on the feasibility of and legal issues surrounding Web archiving commissioned by JISC and the Wellcome Trust (Charlesworth 2003, Day 2003a). Both these organisations have an interest in the preservation of Internet material in their areas of interest. However, consortium members also include the National Archives, the BL and the national libraries of Scotland and Wales. The aim is to “share the costs, expertise, and facilities required to archive selected Web sites for future access.” (UK Web Archiving Consortium 2004a). This grouping makes sense in that it should prevent duplication of effort between the national and academic library and archives sectors. The omission of the Bodleian Library and Cambridge University Library implies that these legal deposit libraries will not be actively involved in gathering Internet material for legal deposit purposes. The Consortium will be using the NLA’s PANDAS (PANDORA Archiving System) Web archiving software. UKWAC members have commissioned Magus Research Limited to extend the PANDAS software for UK needs and to provide shared hardware and technical support (UK Web Archiving Consortium 2004b).

.

#### **5.3. *Australia***

The Australians are taking a distributed approach to the preservation of digital information. Although the National Library is taking the lead, it will not be responsible for the long-term preservation of all digital Australian publications (National Library of Australia 2004a). Other institutions taking on roles include state and university libraries, archives, other collecting institutions and publishers.

### **5.3.1.            *National Leadership***

The NLA aims to identify partners and stakeholders to contribute to a coordinated national effort for reliable and sustainable long-term preservation and access to digital information.

The NLA wants to set up effective agreements and also help in the identification, development and promotion of supporting policies, procedures and tools to support such an aim. Partners will include creators, publishers and re-users of digital content who will be encouraged to use practices that will make preservation easier. The NLA also plans to work with policy makers so that legislative and funding frameworks will aid in cost-effective preservation. Cooperative working on digital preservation seems to form part of a wider cooperative network for collection management. While there is no organised programme and funding stream similar to that in the United States, the NLA has been a world leader in digital preservation and there is evidence of implementation of this cooperative policy. This includes the PANDORA (Preserving and Accessing Networked Documentary Resources of Australia) archive of online resources. This was developed and based at the NLA, but now six Australian state libraries, ScreenSound Australia and the Australia War Memorial are also involved (Missingham 2003). PANDORA is discussed in more detail below.

The NLA has also been taking the lead in international cooperative activities. In 1998, the NLA set up an international group with other national libraries and consortia involved in digital preservation “to consult and share ideas” (Berthon & Webb 2000). Consortia members included the perceived world leaders at the time. British legal deposit libraries were notably absent, although the CEDARS project was involved.

### **5.3.2.            *PADI (Preserving Access to Digital Information)***

The NLA has operated the PADI (Preserving Access to Digital Information) Web site since 1997 (Berthon & Webb 2000). This is a subject gateway to resources on digital preservation ([www.nla.gov.au/padi/](http://www.nla.gov.au/padi/)). Material ranges from top-level basic explanations of the topics and issues to links to documents and relevant Web sites. It records “our evolving

understanding of the challenges and solutions of digital preservation” (Berthon, Thomas & Webb 2002). Registered users can add material to the PADI database. However, PADI is now more than just a subject gateway. The NLA and the UK DPC together are taking a proactive approach to making digital preservation information available through a quarterly current awareness service (Digital Preservation Coalition 2004d).

### 5.3.3. *Safekeeping Project*

The NLA set up the Safekeeping project in 2001 as a way of ensuring that important PADI content remains available over time. The Safekeeping operation involves a distributed network of partners. The project involves selection of material that will have long-term value. The categories of material selected include:

- Seminal documents – represent key turning points in thinking
- Documents that define or describe important issues, approaches, projects or studies or raise important issues
- Documents that may have no long-term value in themselves, but which are representative of approaches or opinions from a particular time (Berthon, Thomas & Webb 2002)

Another selection criterion is that material should be closely related to digital preservation; contextual material is not selected. The majority of safe kept items are handbooks, reviews, reports, conference papers and journal articles. However, some policies, strategies and guidelines have also been selected as have projects and case studies and Web sites of interest (Berthon, Thomas & Webb 2002). Selection was initially carried out by NLA staff, but the intention is that this should be distributed world-wide and that “peer’ appraisal of resources will provide an important underpinning to commitments from safekeepers” (Berthon, Thomas & Webb 2002). NLA staff consider that the contribution of this project is highlighting the value of material early in its life so that its preservation can be facilitated. As part of the project discussions were held with resources owners about responsibility for preservation. The aim was to encourage owners to make arrangements for preservation and the discussions often acted as a catalyst for this.

While some would take responsibility themselves, others were looking at other arrangements. The focus is on raising awareness of strategies required for preservation and goodwill rather than formal written agreements (Berthon, Thomas & Webb 2002) because some resource owners said they were unwilling to enter into such agreements. The NLA acknowledges that while the Safekeeping project aims to facilitate preservation, it will not ensure it because of the informal nature of the arrangements. (Berthon, Thomas & Webb 2002). While the NLA has not taken responsibility for preservation of non-Australian material, it has offered to provide back-up storage to support safekeepers' activities.

#### **5.3.4. *International Participation***

Other examples of cooperative activity include participation in CDNL Committee on Digital Issues (National Library of Australia 2003c) and working with the Internet Archive to develop harvesting software for national libraries. The NLA is also monitoring initiatives, such as the US NDIIP and “will attempt to benefit from its outcomes” (Gatenby 2002b).

#### **5.4. *Pan-European Initiatives***

##### **5.4.1. *BIBLINK Project***

The BIBLINK project dealt with the provision of metadata by publishers to national bibliographic agencies. In the UK, the British Library manages the national bibliography and the project objectives were also relevant in the legal deposit context. BIBLINK started in 1996 with support from the European Commission. Although the project came to an end in 1999, work on the initiative continued under the aegis of CDNL. The BL led BIBLINK, which also involved:

- The national libraries of France, Netherlands, Norway and Spain
- The Open University of Catalonia
- Book Industry Communication (BIC) and the CD-ROM Standards and Practices Action Group
- UKOLN

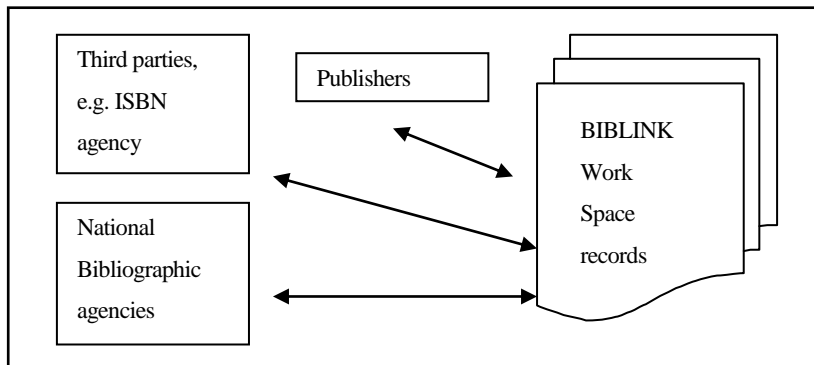
The aim of BIBLINK was to help develop and improve national bibliographic services, focusing on digital publications, especially online publications. The specific objectives were to:

- Create a direct bibliographic link between publishers and national bibliographic agencies
- Develop an agreed bibliographic description for various types of publications that might assist bibliographic control
- Secure the provision of authoritative advance information about new publications
- Exercise bibliographic control over these publications
- Enable publishers to register new electronic publications
- Enable publications to carry enriched bibliographic data as an integral part of the publications
- Help “second level users”, e.g. library utilities meet the increasing demand for information about electronic publications
- Allow unique identification of publications
- Supply “third level users” or libraries with information for their users
- Supply libraries with bibliographic records for integration into their catalogues (Patel & Clayphan 2000)

The objectives were wide ranging and ambitious; the main perceived benefit was that national libraries would not miss the publication of significant publications (Noordermeer 1999, p. 26). The first phase of the project involved information gathering including metadata formats, MARC conversion practices, transmission of data between publishers and libraries, unique identifiers, authentication of publications and metadata and consensus building. All of this fed into the second phase development of a prototype demonstration system. The demonstrator software was called the BIBLINK Workspace (see Figure 5.1)

The BIBLINK demonstrator provided a virtual workspace or “computer mediated work environment” for participating parties (Noordermeer 2000, p. 2). It allowed publishers to create records and participants to access the system to retrieve, update and delete records in the workspace. The system was configurable; preferences were specified in user profiles.

These profiles also governed access rights. The Work Space converted data between the different metadata formats used by the partners. The UNIMARC was used as a switching format between different versions of MARC. The internal metadata format used was “BIBLINK core”, an enlarged version of the Dublin Core metadata set. Authentication of publications and metadata was achieved through calculation of a hash value (Peacock & Powell 1998).



**Figure 5-1 BIBLINK Work Space**

BIBLINK was supposed to be developing an Exploitation Plan to provide a framework for library partners to assess the possibility of incorporating the system into operational procedures (Patel & Clayphan 2000, Noordermeer 2000, p. 2). Since the BL was a partner, the outcome of this assessment could have had implications for the UK legal deposit system. The UK legal deposit libraries currently share responsibility for the provision of MARC records for legal deposit material through the Cooperative Cataloguing Project. There is nothing in the literature discussing the final outcomes of the BIBLINK project. There was some trial use of the software, but it was not considered to be scalable. In any case, the underlying idea of the project, exchange of information between libraries and publishers, was overtaken by commercial developments including ONIX. The main outcome for the partners was the learning experience and it was this experience that was taken forward in internal developments<sup>4</sup>.

<sup>4</sup> Robina Clayphan, British Library, email to Adrienne Muir, 13 June 2001.

#### 5.4.2. *Electronic Deposit (ELDEP) Study*

The European Commission supported the Computerised Bibliographic Records Action (CoBRA) Forum under its Telematics for Libraries programme.

The CoBRA Forum identified digital legal deposit as an important future issue for European national libraries. This concern resulted in the ELDEP study of the issues.

The objectives of the study were to:

- Identify classes of electronic publications and appropriate selection criteria and sampling methods for each
- Assess technical options for storage and access
- Clarify preservation and long-term access issues
- Assess approaches to bibliographic control, including quality issues (Mackenzie Owen & van de Walle 1996, p. xi)

The methods used to gather information were desk research and interviews and the results were disseminated and discussed at a workshop. These discussions were incorporated into the final report. However, there are no details of the desk research carried out. There is no list of interviewees nor are the interviews written up in the report. Therefore it is difficult to assess whether the conclusions made or guidelines developed follow logically from the research carried out.

The report contains a number of general conclusions and recommendations for future activities (Mackenzie Owen & van de Walle 1996, pp. 93-97). The first conclusion is that legal deposit should be extended to cover digital publications because publications are a record of national culture, activities and achievements and should be kept available for future generations. Legal deposit is apparently the only way to guarantee this, but there is no discussion of why this is so. However, the study is in agreement with the Kenny Committee on this point. Two other arguments made for legal deposit are that it facilitates the compilation of national bibliographies and that deposit collections act as a standard that can be used to authenticate publications by comparison with the deposit copy. The latter

point requires that preservation actions either do not result in changes in digital publications or that changes are recorded.

Other conclusions are that:

- Making digital legal deposit work will require a joint effort between the various stakeholders, such as authors, publishers and libraries
- Legal deposit regulations should be broad and flexible to cope with future developments
- Copyright regulations should not cause a barrier to the preservation of digital publications
- There is a lack of theory and practice in selection; selection is done on a pragmatic basis
- Offline publications (e.g. on discs) can be managed in the same way as print publications
- A choice will have to be made between preserving artefacts and information content. The report recommends that the content should be preserved
- Access to digital deposit publications should be restricted. Wider access should only be available through agreement with publishers
- Practical guidelines for dealing with digital publications should be based on legal or voluntary deposit regulations. An effort should be made to extend legal deposit and to run pilot projects to develop practical guidelines

These recommendations are in line with the work of others; only some publishers might disagree with them. The recommendations were aimed at the European Commission, European deposit libraries in cooperation and deposit libraries and publishers in cooperation. The recommendations for the European Commission are to support the development of a European Deposit Library, provide a periodic review of developments in electronic publishing that will have an impact on legal deposit, and identify and coordinate activities in the Telematics for Libraries Programme that are relevant to legal deposit. The first of these recommendations was taken forward through the NEDLIB project.



### 5.4.3. *Networked Electronic Deposit Library (NEDLIB)*

The NEDLIB project started at the beginning of 1998 and finished at the end of 2001. The project leader was the Koninklijke Bibliotheek of the Netherlands. The project partners included eight European national libraries, one national archive, three publishers and two information and communication technology companies. The aim was to “develop a common architectural framework and basic tools for building deposit systems for electronic publications” (Werf-Davelaar 2000). NEDLIB has been an influential project. At least some of the project partners are using lessons and tools from the project to take forward their own digital activities (see sections 5.5.3 and 5.6). The project findings have been well disseminated with interim reports and papers, a series of final reports and guidelines and also some well-designed interactive presentations<sup>5</sup>. The project has also contributed to the development of international standards.

The project aimed to investigate the following areas:

- the functional requirements for a deposit system
- process and data modelling for handling digital publications
- architectural considerations, technical standards and solutions for the implementation of real deposit systems;
- long-term preservation of electronic publications (Koninklijke Bibliotheek 2003b)

The outcomes of the project included a model for a deposit system (Werf, 2000). Initial work included identifying functional requirements common to deposit libraries (or the deposit libraries involved in the project anyway). This work fed into the development of a generic design through defining a common workflow for handling digital material (Werf 2002, pp. 55-56). The aim was that this design could be used as a basis for local implementations. At the end of 1998, the project consortium adopted the OAIS Reference Model as the basis of the design phase. The NEDLIB Deposit System for Electronic Publications (DSEP) was narrower in scope than the OAIS model. This is because some of

---

<sup>5</sup> Available at: <http://www.kb.nl/coop/nedlib/>

the OAIS functions, such as Data Management and Access, were seen as part of the general digital library environment and not specific to the digital depository. The aim was to link the functions of the deposit system and the digital library environment through interfaces and so integrate the depository into general digital library infrastructure (Koninklijke Bibliotheek 2003b).

The NEDLIB work flow involves the following activities:

- selection – individual libraries define their own selection criteria within national deposit regulations
- acquisition – either ordered from the publisher or delivered automatically and deliveries may be registered in the acquisition system
- capture – electronic transfer from publisher to library's deposit system
- registration – incoming material is checked in to the deposit system
- verification – control routine to check physical and logical integrity
- description – title entry created in the library catalogue
- storage – material stored in a digital store which ensures the bit stream remains machine readable
- preservation – preservation strategies are carried out when material is about to become obsolete
- delivery – material retrieved from store and made fit for use on library work stations
- access – the user access environment, material is used according to defined access conditions
- monitoring – the workflow is monitored and quality controlled (NEDLIB 2003)

The project commissioned a study on standards for electronic publishing (Bide 2000). The study involved interviews with 14 European publishers from various sectors. This study found that it was difficult to generalise about trends in the publishing industry because processes were different in different sectors. Another finding was that publishers only adopt standards when there are incentives to do so. These are likely to be commercial and publishers are unlikely to adopt standards solely for deposit and preservation purposes.

However, CD-ROM and other offline publishing was seen to be in decline and online publishing more standardised than offline publishing. For example publishers were using standard Web browsers as the primary interface, PDF was almost universally used for delivering page images, and a significant number of the publishers in the survey had significant SGML and XML text archives. However, customers were having to use additional applications to view material, including PDF viewers, spreadsheet and database software. There was little consistency in the use of multimedia formats, although multimedia elements were little used at the time. Another problem identified was the dependency on middleware for the delivery of material. The middleware might be part of a hosting service and therefore not directly controlled by the publisher. The study found that publishers were becoming increasingly interested in metadata to help them sell their material over the Internet. The conclusion was that although publishers had a lot to learn in this area, this might form the basis of exchange of data between publishers and libraries. This finding is supported by comments from the BIBLINK project and later activities reported in Chapter 4. The NEDLIB metadata schema (Lupovici & Masanès 2000) was developed by the Bibliothèque nationale de France, which is now using the schema internally. As mentioned in Chapter 4, the NEDLIB scheme also informed the framework developed by OCLC and RLG.

A list of standards for the implementation of a deposit system was issued in September 2000 (Feenstra 2000). The report suggest that the selection of standards to use should be part of the specification of requirements and/or design phase for the implementation of a digital deposit system (Feenstra 2000, p. 1).

NEDLIB commissioned Jeff Rothenberg and Rand-Europe to undertake some experimental work on the use of emulation as a preservation strategy (Rothenberg 2000). There is no mention of work on any other type of preservation strategy, although the KB is currently working with IBM on the Universal Virtual Computer approach (see section 3.4.5.3). The project also fed back into the development of the OAIS in that a suggestion of adding another module, Preservation Planning, to the OAIS model was accepted (NEDLIB 2000).

The publishers involved in the project provided some electronic publications for a test bed. A proof-of concept demonstrator was developed and some software tools were also

developed and tested (Koninklijke Bibliotheek 2003b). These were the NEDLIB harvester for harvesting and archiving Web material. The harvester has now been used by the Icelandic and Finnish national libraries. The MMB system for Multimedia Access was developed to support the workflow for digital publications. This is now being used by the Deutsche Bibliothek.

#### **5.4.4.            *The European Library***

The aim of this project to provide access to “the combined resources of the national libraries of Europe” (Woldering 2004). This three-year project ended in January 2004 and was partly funded by the European Commission under the cultural heritage applications strand of Information Societies Technology research programme. The project partners were the Conference of European National Librarians (CENL), the national libraries of Finland, Germany, Italy, Portugal, Slovenia, Switzerland, the British Library and the Italian national central cataloguing institute (ICCU) (Woldering 2004). The project was not merely about searching across catalogues, it was also about European-wide access to the resources described by catalogues so it did have implications for legal deposit in that it could have potentially upset publishers. So one of the strands of the research was the “investigation of the potential agreements such a service might negotiate with publishers concerning the provision of digital resources across Europe” (Woldering 2004).

The work of the project will go on and the European Library service will involve a portal allowing searching for digital and non-digital material in collections. It will also allow delivery of digital material. Some of this material will be free; for other material users will have to pay. The free material will mainly include (presumably) out of copyright material digitised by the national libraries and freely available Web resources gathered through Web harvesting exercises. None of the European Library partners are currently in a position to implement electronic business transactions, so access to other categories of material, including legal deposit material is out of the question (Woldering 2004). However, it remains a possibility for the future.

As part of the research, publishers’ views on this were canvassed. There was a low response, but those publishers who did provide their views varied in their opinions. There

was concern that business interests could be damaged; at the same time there was some recognition that legal deposit, inclusion of publications in catalogues and wider exposure of publications could provide business opportunities. Publishers were willing to become partners as long as they were guaranteed appropriate recompense for access beyond reading rooms (Woldering 2004).

#### **5.4.5.            *The Proposed PATCH Project***

In April 2003, a seventeen partner consortium submitted a proposal to the European Commission under its Sixth Framework Programme for a Permanent Access Toolbox for Digital Cultural Heritage (PATCH) project. The partners included national preservation institutions, preservation research institutes, scientific institutions and technology companies. The proposed project was an attempt to bring together all the work that has been done so far on preservation strategies and implementation of digital archives to develop sustainable integrated working systems. The main aim of the proposed project was to develop a technological framework to facilitate the continuous development of tools to ensure permanent access to digital material (Steenbakkens 2003). The tools envisaged were those discussed in Chapter 3, including migration on request, Jeff Rothenberg's emulation virtual machine idea and the Universal Virtual Computer concept. The project would have addressed a variety of types of digital material, including fixed format files, Web documents, datasets and software and applications. Unfortunately, the European Commission decided not to fund this project. Perhaps the Commission did not want to fund any more digital preservation research projects, but without significant funding and some coordinated effort, it is hard to see how the technological aspects of digital preservation can be taken to this higher level.

#### **5.4.6.            *PRESTO Space***

The PRESTO (Preservation Technology for European Broadcast Archives ) Space project is concerned with the preservation of audiovisual material (<http://presto.joanneum.ac.at/index.asp>). While UK legal deposit will not include film and sound material, the legal deposit libraries are likely to have to deal with multimedia

publications. The British Film Institute and the National Sound Archive at the British Library may also be able to use the findings to preserve their collections.

## **5.5. International Web Archiving Activities**

### **5.5.1. *Internet Archive***

There is another initiative that has developed independently of digital deposit and preservation initiatives, but which is now becoming of increasing interest to legal deposit libraries. The Internet Archive (<http://www.archive.org/>) is currently the most ambitious Web harvesting operation in that it aims to capture the entire Internet. It is a non-profit organisation, providing free, supposedly permanent, access to captured content for researchers. Material is stored in an internal format that the Internet Archive considers to be an archival format, and apparently the Internet Archive is moving towards providing redundancy and media migration for its collection in its entirety (Internet Archive 2004a). It is not clear whether the Internet Archive will be taking steps to combat other aspects of technological obsolescence, but there are plans to collect “software and emulators that will aid future researchers, historians, and scholars in their research” (Internet Archive 2004a).

If the Internet Archive is archiving the entire Internet, there is a question over whether national libraries also need to capture such material. Only publicly accessible Web sites are collected, so material that could be eligible under legal deposit and therefore of interest to deposit libraries is not collected. If the Internet Archive is not carrying out long-term preservation measures, then there is still a need for someone to do this. The Internet Archive is working with partners, including national libraries, and can provide content retrospectively so that they can preserve it. Although it has improved over time, access to material in the Internet Archive is very unsophisticated, and libraries could improve access to material initially captured by the Internet Archive. For example, the Library of Congress is providing access to a snapshot of the Internet in 1997 (Internet Archive 2004b).

The Internet Archive also has total disregard for the copyright status of the material captured, which is a potentially damaging issue. While Brewster Kahle, the founder of the

Archive, is asking for US copyright law to be changed, this is of no use for the capture of non-US material.

### **5.5.2. *International Internet Preservation Consortium***

Created in 2003, the International Internet Preservation Consortium (IIPC) involves the national libraries of Australia, Canada, Denmark, Finland, France, Iceland, Italy, Norway, Sweden; The British Library; the Library of Congress and the Internet Archive.

The aim is to work together on developing Web archiving activities within national legal frameworks (International Internet Preservation Consortium 2004b). The Internet Archive is taking the lead on work on developing “a shared technical basis for web archiving activities” (International Internet Preservation Consortium 2004a). The National Library of Australia is leading on archiving the “deep Web” and the Danish Royal Library is leading on access. These roles are appropriate given the experience and expertise of these libraries. The Bibliothèque nationale de France will no doubt contribute the expertise gleaned from experiments aimed at improving the sophistication of Web harvesting (Masanès 2002). It is not clear whether the BL is taking a specific role. The interesting thing about this initiative is that researchers – the users - will be consulted in order to decide what should be kept in Web archives and how they should be accessed.

### **5.5.3. *Nordic Web Archive***

Scandinavian national libraries have worked together on Web archiving through the Nordic Web Archive (NWA) forum. The national libraries of Denmark, Finland, Iceland, Norway and Sweden have been experimenting with harvesting software, gathering material on both a comprehensive and selective basis and investigating legal and organisational issues (Arvidson 2002, Christensen-Dalgaard 2004, Hakala 2001). However, a particular focus of the NWA has been developing software for access to captured material (Nordic Web Archive 2004). This is open source software and so is available to other legal deposit archives.

## **5.6. Initiatives in Individual National Libraries**

This section includes a small selection of the most prominent or innovative activities. Some libraries are now at the stage of procuring or implementing full-scale systems for storing digital publications. The national libraries of Australia and the Netherlands seem to be the most advanced in this area and a there is lot of information available on their activities. The National Library of Canada was one of the pioneers in exploring the issues associated with digital publications. Other libraries are carrying out research and development work of interest. Scandinavian libraries have been particularly prominent in Web archiving activities as mentioned above, but the national library of Norway is conducting comprehensive research on various aspects of digital legal deposit.

### **5.6.1. *Koninklijke Bibliotheek, Netherlands***

The Koninklijke Bibliotheek (Royal Library) was founded in 1798 and is the national library of the Netherlands. As well as acting as the Dutch deposit library, the KB acts as a research library for the humanities and social sciences. The KB is responsible for the national bibliography and the Netherlands Union Catalogue, coordinates the Dutch inter-library loans system and is a centre of expertise for digitisation, preservation and restoration. Other activities include research and development, supporting cooperation between libraries and international activities (Koninklijke Bibliotheek 2003a).

There is no legal deposit in the Netherlands; voluntary arrangements for print material have been in place since 1974. The KB receives 90% of Dutch publications free of charge and, according to the Library's Web site (<http://www.kb.nl/>), nearly 100% of all digital publications are deposited. If the price of a particular item is above a certain threshold, the publisher is entitled to payment of the price reduced by 25%. The KB staff approach all Dutch publishers, private individuals and businesses that produce their own publications, and ask them to deposit one copy of each newly published work or newly issued periodical (Koninklijke Bibliotheek 2003a).



The KB took the decision to collect digital publications in 1994 (Noordermeer 1997, Steenbakkens 1999, p. 94). Offline publications were collected and stored in the book stacks. From 1995, the KB experimented in handling online publications on a small scale. Three publishers cooperated with the KB by agreeing to deposit some of their electronic publications. Between 1996 and 1998, the KB undertook a project called DNEP-IWI. The IWI is the Steering Group for Innovation of Information Supply and is an initiative of the Dutch universities, the Dutch Organisation for Research and the Royal Dutch Society of Sciences. The project included a market survey of electronic publications. Other areas investigated included selection and acquisition, metadata and identifiers, preservation and copyright issues. The final phase involved a test bed of digital publications. In 1996, the KB reached a provisional agreement with publishers to widen digital deposit. This small-scale deposit system was based on the IBM Digital Library system and became operational in 1998 (Steenbakkens, 1999, pp. 97-98).

- In June 1999, the KB and the Dutch Publishers' Association concluded an agreement for the deposit of digital publications. The agreement took effect on 1 July 1999 and includes first editions and updates. These should be deposited in the form in which they are made available to the public, including the original packaging, retrieval software, manuals or any other accompanying material, whether in print or digital form. If the same content is available in more than one format, for example for different equipment or operating systems, the KB will choose which should be deposited (Koninklijke Bibliotheek 1999). The arrangement covers both offline and online material.

In principle all publications - excluding application programmes and games, but including dynamic databases - are eligible for deposit in what was then called the Deposit of Netherlands Electronic Publications (DNEP, now e-Depot) (Oltmans & van Wijngaarden 2004). However, the KB did not start to collect all publication types immediately because of the practical difficulties.

No price threshold has been set for digital publications because of the relatively low reproduction costs. The agreement is to be reviewed every three years and is still in place at the time of writing.

For print material, KB staff look for items with ISBN numbers in various lists and catalogues. The KB has very good relations with the publishers and has regular meetings with them. There is an organisation in the Netherlands (roughly translated the National Book Corporation) that delivers books to libraries and bookshops. This company maintains databases that the KB staff use to discover the existence of material. A senior member of KB staff interviewed for this research in 2002 was not sure how the existence of all digital publications will be discovered<sup>6</sup>. Many of the traditional publishers that have moved into electronic publishing send digital publications automatically. The KB is actively approaching new, electronic only publishers and there is a systematic programme for this. The interviewee is in favour of the DOI, although many librarians are wary of publisher initiatives and want to use the Dutch National Bibliography Numbers. The interviewee thought that there is no reason why different systems could not be used in tandem. The KB may be able to use publisher descriptive information. There will still have to be some intellectual input, but the KB wants to keep this to a minimum and automate the processing of material as much as possible.

In 2002, the KB was receiving both offline and networked publications and was concentrating on commercial publications. There was a lack of resources, so this is where efforts are being concentrated. The KB was not actively harvesting from the Web, although it had experimented in the past through the NEDLIB project. The KB was setting up agreements with publishers and also has arrangements with universities for theses and their publications.

In 2002, the CDs that were deposited were loaded by hand. At the time there were problems with loading about 30% of the discs. In such cases, the KB contacts the publishers for advice. Journals sent by FTP are loaded automatically, although some are sent on CD for deposit. It was generally smaller publishers that were sending journals by email.

The publishers are happy to allow the KB to carry out whatever copying is necessary for preservation.

---

<sup>6</sup> A. Muir interview with J. Steenbakkens, Directeur Bedrijfsvoering en Informatietechnologie, Koninklijke Bibliotheek, Glasgow, 19 August 2002.

However, since publishers do not always own all the rights in digital material, there may be a need for some legislation confirming that copying can be carried out for preservation purposes without having to spend time and effort tracking down rights holders.

The KB identified the need for a “digital mailroom” where publishers can deposit their material. This would not merely be an address, but would be a controlled area with password, virus checking, mirroring and harvesting capabilities (Werf 2002, p. 58).

The KB was talking to publishers about standards. The KB reports back to publishers on the quality and structure of data and any handling problems. This allows publishers to improve their products. The KB asks publishers to inform them about any changes to their products. Users (especially in professional publishing) are driving what publishers are doing.

In 2002, descriptive and structural metadata were put into existing library systems (Werf 2002, p. 58). The KB recorded details of all deposited digital publications in its online catalogue and the Dutch National Bibliography. Technical metadata was recorded in the data management module of the library’s digital archiving system (Werf 2002, p. 58). The library’s digital archiving system did not support batch loading of metadata received from publishers. The KB wants to be able to process this metadata automatically and convert it to its own XML DTD format (Werf 2002, p. 58).

The NEDLIB work fed into the development of the KB’s digital depository. The KB is now implementing a full-scale system for digital deposit commissioned from IBM Netherlands in 2000. According to the staff member interviewed, the KB philosophy is to outsource as much of the technology development as possible to the market. There are advantages in this in that the KB can benefit from support services and future product developments. A small group of staff from both processing and ICT departments was involved in the development and commissioning of the KB system.

At the testing stage staff became extremely enthusiastic because they had something concrete to work with. The interviewee was leading the development of the system, but handed over responsibility for implementation and developing work flows to the staff with his help. Staff were initially nervous about this. The KB decided to integrate digital

material into existing work flows. Traditional skills are still needed, but staff do need to develop new ones, including ICT skills. Day to day operation of the ‘digital stacks’ will be undertaken as part of the existing processing departments’ duties. The KB identified the need for technical cataloguers who can deal with various file types and IT configuration management tools; technology watch officers to monitor new formats and digital publishing developments; reference platform administrators, digital preservation researchers and quality control managers. The work flows were successfully developed (Oltmans & van Wijngaarden 2004, p. 22). IBM handed over the OAIS compliant Digital Information Archiving System (DIAS) to the KB in October 2002. The core of the system is IBM’s Content Manager for the publications and DB2 Universal Database for the related metadata. IBM also provided the hardware, including online, optical and tape storage (IBM 2002).

Initially the top-level functional components of the system are ingest, archival storage, data management, access and administration. The preservation module is not yet included. There are components that interface with the library’s other systems, as recommended in the NEDLIB model. Delivery and capture and packaging and delivery cover pre- and post-processing of digital objects (Koninklijke Bibliotheek 2002). These modules deal with “locally defined and variable types of things” (Werf 2002, p. 57). These things would include any conversion from publisher file formats to the internal archive standards on the input side (Oltmans & van Wijngaarden 2004, p. 22). On the output side are the users and digital material may be delivered in different ways according to the needs of users. This configuration follows the recommendation of the NEDLIB project.

At the time of the interview, the KB was still putting CD-ROMs on the shelves, but the plan was to load them into the system when it became operational. Publications will be moved from their published environment to a preservation environment as recommended by NEDLIB (Steenbakkens 2002). The DIAS system allows manual loading of material into the system (Werf 2002, p. 57). This is apparently time consuming, as the disc is installed with all its related software on a Reference Work Station and a snapshot of the “entire workstation” is then taken and the resulting image is catalogued manually then loaded into the system. The digital mail room has also been implemented. This is where material received is checked for completeness and where the content and metadata are put together in Publisher Submission Packages (Oltmans & van

Wijngaarden 2004, p. 22). The KB is using publisher metadata, which it is converting to its own internal format (Oltmans & van Wijngaarden 2004, p. 22).

The decision was taken not to include all the required functionality in the IBM system because it would complicate matters and affect the project budget and schedule. There is also the need to future proof the digital archive as much as possible by avoiding unnecessary dependencies. So the system itself is modular to allow updating of different modules without compromising the whole system. Also by having interfaces with the library's other systems, these can also be changed. The KB will only put deposit material into the system and not digitised collections. These will be stored in a separate system. The rationale for this is that they do not require all the functionality needed for the deposit collections.

While the KB now has its basic deposit system in place, it is still carrying out research on preservation. The interviewee was confident that the technical problems are solvable; they just require time and effort. He estimated that it should take a couple of years and that a variety of tools would be needed, not just one strategy. The KB has a contract with IBM to carry out research and was working with the British Library in this area. The interviewee commented that, at the time, the BL's internal reorganisation has taken up a lot of energy and has held the BL back in its developments a little.

The aim of the Long Term Preservation study was to identify and develop methods and techniques in the preservation of digital information. Initial stages involved planning and definition of concepts. The next stage focused on media preservation, including investigation of technological trends, identification of techniques to estimate media quality and management strategies for media migration (Koninklijke Bibliotheek & IBM 2001, p. 3). At the time of the interview work on the Universal Virtual Computer idea had also been carried out (see section 3.4.5.3). A proof of concept had been built and had worked. The next step was to develop the first version of a tool by the end of 2003. At the time of the interview, the research contract with IBM was due to run out, but the KB wanted to renew it. However, there was a need for additional resources. There had just been a change of government in the Netherlands, so the funding situation was not yet clear. The results of the study have been published (Lorie 2002)

In 2002, the KB was looking at how to manage the technical metadata required for digital preservation. The KB is using the “reference platform” concept. This attempts to capture the configuration of a generation of digital publications, including hardware, operating system and viewer applications. The reference platform approach does not aim to capture all possible ways of viewing material; it will be the KB’s preferred view path for a set of publications for a certain period of time. The technical metadata will record the various software and hardware dependencies in order to be able to recreate the technical environment for current and future access (Werf 2002, p. 60-61).

In 2004, it seems that the KB still does not have full operational technical tools and systems in place. However, it does seem to have made some progress. There is now a pilot version of the UVC approach (Oltmans & van Wijngaarden 2004, p. 24). However, the involvement of the KB in the PATCH consortium perhaps indicates that it recognises that it cannot solve the problem of technical preservation on its own.

The KB will be providing the same level of access to digital material as for print material. Access will be on site and users can print out material. There may be restrictions in future. The KB has identified a need for user identification and authentication, authorisation mechanisms and some browsing functionality. This is not included in the functionality of the current DIAS system (Werf 2002, p. 58).

Like the UK deposit libraries, the KB wants immediate, not deferred deposit. The KB also wants to give immediate access, to check that the publications are working if nothing else. The KB works very much in partnership with publishers and sees them as partners and not competitors.

### **5.6.2.            *National Library of Australia***

The NLA has actively, but so far unsuccessfully, pursued extension of legal deposit law. Current national legal deposit law is part of copyright law that dates from 1968 (Australia 1968) and effectively covers print material only, although some of the states have less restrictive laws. Like the UK legal deposit libraries and the British Film Institute, the NLA and ScreenSound Australia (the National Film and Sound Archive) put forward a joint

submission to a Government committee convened to consider simplification of Australia's copyright law (National Film and Sound Archive & National Library of Australia, 1995). The committee supported this submission. Although copyright law was updated (Australia 2000), it took the Government until 2001 to start the process of amending the legal deposit provisions of copyright law (Gatenby 2002). At the time of writing, that process had not yet been completed. The NLA receives digital material through voluntary deposit arrangements and is planning to develop agreements with publishers for the deposit of digital material, similar to those in place at the KB in the Netherlands (Gatenby 2002b).

The NLA set up a Digital Services Project with the aim of developing the infrastructure for the long-term management and preservation of digital material. This infrastructure includes both the technology systems and collection management. The other aim of the project is to develop solutions for digital service delivery, including access to both the Library's collections and the collections of other institutions (National Library of Australia 2001a). The NLA went through a similar process as the KB in its attempts to procure solutions from the market place. The Library was successful in its procurement of a system to provide "Metadata repository and search services", but was not able to procure "an integrated digital collection management system" (National Library of Australia 2001a). The Library published its Request for Tender in August 1999, a few months after the KB, but found that no off the shelf system could meet its needs. It is not clear whether the NLA also received and evaluated a tender from IBM.

The NLA then embarked on three procurement and development projects. These were the PANDORA Collecting System Project, the Digital Object Management System Project and the Digital Object Storage System Project.

The aim of the PANDORA Collecting System Project was to "improve the efficiency of gathering of Web sites for the PANDORA archive" (National Library of Australia 2001a). The NLA set up the PANDORA project in 1996. The aim of the project was to "develop policies and procedures for the selection, capture, archiving and provision of long-term access to Australian electronic publications." (National Library of Australia 2001a). Early work on PANDORA involved modelling work; this was an attempt to map relationships between "processes, the data, database structures, and metadata" The outcomes included a

business process model (National Library of Australia 2003b), logical data model (National Library of Australia 2001d) and national model (see section 5.3.1).

The logical model sets out the information to be gathered and managed in order to run an archive as well as the relationships between data about publications (National Library of Australia 2001e). The processes have apparently changed little since the beginning of PANDORA and so the business process model is still representative of how the NLA is carrying out this work (National Library of Australia 2003b). The business process and logical data models are of interest to other organisations planning to implement similar operations. The national model defines how the PANDORA partners work together. Again, this could provide a template for other legal deposit libraries involved in this activity.

The library then developed a proof-of-concept archive of Australian Internet material, which has been used to develop policies and procedures for the long-term preservation and access to digital publications. Some of the findings of this work are discussed in previous chapters. The archive contains 'born-digital' material rather than digitised paper publications (Relf, 1999). The initial harvesting and management software was developed in-house (Gatenby 2002a). The software supports various functions. These are:

- Managing metadata about both selected and rejected titles
- Initiating gathering of titles
- Managing quality checking and solving problems encountered
- Preparing material for public display
- Managing access restrictions
- Provision of management reports (Phillips 2003, p. 26)

Not all of these functions are fully automated. For example, the quality checking and metadata input has to be carried out manually. This software is now being used by other organisations, such as UKWAC (see section 5.2.5).

PANDORA is still going strong. Collection of Web material, with the permission of rights owners, is carried out in collaboration with most of the Australian State libraries,



ScreenSound Australia, the Australian War Memorial, and the Australian Institute of Aboriginal and Torres Strait Islander Studies according to guidelines. Selection criteria “give emphasis to publications that have research value for studying the history of Australian society” (Gatenby 2002a). This implies that collecting could go beyond material ‘published’ in Australia and therefore any legal deposit remit that may materialise in the future. The NLA’s selection guidelines for online Australian publications confirm this (National Library of Australia 2003a). Public access to this material was always planned; all titles are catalogued and appear in the national bibliography and anyone, anywhere, can access the archive through the Internet. Apparently the majority of titles have no access restrictions imposed; since 98% of the material collected (in 2002) was freely available on the Web, this is not surprising.

The Digital Object Management System Project, which seems to have been renamed Digital Collections Manager, was concerned with metadata. The aim was to find a way to meet the Library’s “needs for recording and processing management and technical information (metadata) about digital collection items and digital surrogates” (National Library of Australia 2001a). A number of existing systems were to be replaced. The NLA developed a database for metadata, as it was unable to find a commercial software product to meet its needs (National Library of Australia 2004b). The NLA developed a metadata schema that was not directly implemented. It was, however, incorporated into the OCLC/RLG work on preservation metadata (see section 4.3.1). The NLA is now involved in the PREMIS metadata work. Other metadata related activities involve supporting abstracting and indexing services that are encountering problems with broken links in citations. The NLA volunteered to archive any Australian online publication on request, with permission of the rights owner, and to provide it with a permanent identifier so that even if the publisher’s site no longer exists, there is a copy that can be linked to (Phillips 2000, National Library of Australia 2004c).

The NLA decided in 2001 that no existing persistent identifier schemes in use or being developed at the time met its needs in managing its digital material, so it decided to implement its own URL naming scheme (National Library of Australia 2001f). Abundant documentation on this work is available via the NLA website (<http://www.nla.gov.au/initiatives/persistence/>).

The aim of the Digital Object Storage System Project was to provide the hardware and software to store digital objects and was to meet requirements between 2001 and 2006 (National Library of Australia 2001a). The NLA took delivery of the new system in 2001 (National Library of Australia 2004b).

As far as preservation is concerned, the National Library of Australia carried out some practical research on the archival properties of recordable compact discs and migrating information stored on floppy discs to CD-R. The NLA claims to have been the first to adopt CD-R as an archival medium and have carried out evaluation and life testing trials on CD-Rs (Woodyard 2002). The research only concerned changing media; technology change issues were not addressed. However, the justification for this exercise was the short life of floppy discs. Some of the discs in the library were already coming to the end of their expected lives and something had to be done (Woodyard 1998). The copying process involved transferring data from the floppy discs to the hard drive of a PC. The research identified some areas for further work. One of these was to establish ongoing copying procedures for floppies, possibly an annual batched procedure. By 2003, the NLA was routinely migrating information on floppy discs (Steenbakkens & Trier 2003). However, the NLA felt that this would be a short-term strategy because the amount of material coming into the Library on floppy disc would decline over time. Other areas of future work were to investigate other forms of migration, the best ways of recording and managing metadata for preservation and to develop guidelines on the selection of material for preservation and preservation procedures.

The NLA has subsequently carried out work on the migration of text and HTML files in the NLA's manuscript and PANDORA collections (National Library of Australia 2003c). This work is aided by software that reads obsolete word processing formats and the NLA is developing formal procedures for such migrations. Work on HTML files involves replacing "dead" HTML tags with current tags so that the HTML files remain readable in newer browsers. The NLA has also carried out emulation trials, focusing on early DOS operating systems (National Library of Australia 2003c)

The NLA has disseminated information on its activities, but it has also used its practical experience to develop a simple brochure for other libraries giving practical guidance on preservation of digital publications. The brochure helps to place digital publications within

the context of wider collections. However, the section on strategies does bring home the fact that there are no proven long-term solutions. This activity supports the NLA's wish to "provide leadership to Australian libraries and other relevant organisations in developing policies and techniques for preserving electronic resources" (National Library of Australia 2001c). The advice offered will only help in the short-term. The NLA has also developed the advice for publishers of Websites (National Library of Australia 2002c) discussed in the previous chapter.

### **5.6.3.            *National Library of Canada***

Legal deposit in Canada was extended to cover all digital publications on physical formats in 1995 (Canada 1995). The National Library of Canada (now Libraries and Archives Canada) was an innovator in experimenting with collecting digital publications. The NLC ran such a project between 1994 and 1995. The purpose of the Electronic Publications Pilot Project (EPPP) was to pilot the acquisition, cataloguing, preservation and provision of access to a few Canadian electronic journals and other publications available via the Internet.

The final report of the project included a set of recommendations on issues including copyright, selection, and long-term preservation (Electronic Publications Pilot Project Team, 1996). The recommendations informed the development of policy and guidelines on networked digital publications (Library and Archives Canada 2001). The NLC made some suggestions for national archiving models that are of interest in the context of legal deposit. These included a distributed model with multiple copies stored at digital locations and a centralised storage with distributed access model.

The National Library of Canada also developed a manual of best practices for digital publications for publishers. This manual includes guidance for publishers that would make publications easier to preserve. It also gives publishers information on how to deposit digital publications (Library and Archives Canada 2004).

#### **5.6.4.            *Library of Congress***

As discussed above, the Library of Congress is now heavily involved in developing a national network of partnerships for digital preservation, advancing the development of technological architectures and preservation tools and working with other libraries on Web archiving. Mandatory deposit law in the United States covers digital publications on physical carriers. However, US legal deposit law is not clear about online publications and they may not be included (Seadle 2001), or if they are, the Library of Congress has the right to claim them, but perhaps no explicit right to download them (Arms *et al* 2001). The UK voluntary deposit scheme used the voluntary agreements on access and use of deposited material developed by the Library of Congress (Working Party on Legal Deposit 1999).

The Library of Congress has been experimenting with Web harvesting. Although the Library of Congress started its experiments later than other libraries, it has taken an interesting approach in that it has concentrated on recording the coverage of particular events on the Internet. The aim of the Minerva (Mapping the INternet Electronic Resources Virtual Archive) prototype was to investigate the practical issues involved and gaining insights into how a full-scale collection and preservation programme could be operated (Arms, *et al* 2001). Snapshots of events collected for Minerva include the 2000 and 2002 elections, the September 11 2001 terrorist attack, and the 2002 Winter Olympics held in Salt Lake City, Utah (<http://www.loc.gov/minerva/>). Material collected is catalogued and made accessible. The MINERVA team worked with the US Copyright Office on clarifying the legal situation. Initially, the procedure used was to download publicly available material without asking for permission (Arms *et al* 2001).

#### **5.6.5.            *Nasjonalbiblioteket, Norway***

Comprehensive legal deposit legislation has been in place in Norway since 1990 (Norway 1989). However, the national library has not been comprehensively collecting all types of digital publication. It has, however, been involved in projects such as BIBLINK and NEDLIB and conducted various pilot projects to investigate the issues and practicalities associated with collecting digital publications. In August 2001, the Nasjonalbiblioteket

initiated a new three-year project “to find the technology, methods and organization for the collection and preservation of electronic documents, and to give the National Library's users access to these documents in compliance with ... (The Legal Deposit Act)” (Nasjonalbiblioteket 2004).

The Preservation of Digital Material (Paradigma) project has eight work packages and covers bibliographic, technical and legal issues (van Nuys 2003). The work packages are: selection criteria; legal framework; harvesting tools; access tools; unique identification and description; promotion of standards for identification and description; test beds and the organisational and economic implications of the project. It is interesting that although there has been comprehensive legal deposit law for many years, the legal issues associated with collection and access to material under legal deposit are only just being examined now. However, the fact that the legislation was so early is probably behind this, since the explosion in Web-based publishing was yet to happen.

One outcome of the project is that the Nasjonalbiblioteket has moved from selective archiving of the Norwegian part of the Internet to comprehensive harvesting (van Nuys 2003). However, this will be done in conjunction with more in-depth event-based harvesting, similar to that carried out by the Library of Congress. The project has also recommended selection criteria and harvesting frequencies based on content and a typology of digital publications to help with selection decisions (van Nuys 2003). These were not publicly available at the time of writing. Interestingly, in an interview with the Project Manager in 2002, it became apparent that preservation is not part of the project even though the term is used as part of the title of the project<sup>7</sup>.

## **5.7. Summary**

Legal deposit libraries started to actively experiment with digital publications in the mid-1990s. While deposit libraries in the UK were putting forward the case for extension of legal deposit, other countries were carrying out experiments on a small scale. These pioneers in the Scandinavian countries, the Netherlands and Australia have made most

---

<sup>7</sup> Adrienne Muir interview with Carol van Nuys, Project Manager, Paradigma project, Nasjonalbiblioteket, Norway, Glasgow, 21 August 2002.

progress in exploring the issues, setting up the technological infrastructure, deposit procedures, and developing policies. The United States and the UK have lagged behind these countries, but have influenced thinking and progress in digital preservation. The Library of Congress is now leading a nationwide cooperative programme. While many early activities involved isolated projects, legal deposit libraries and other institutions are now increasingly cooperating on legal deposit and digital preservation. Such cooperative activities include the US NDIIPP, OCLC and RLG work on metadata and trusted repositories, collaborative Web archiving groups and the Digital Preservation Coalition in the UK.

## **6. METHODS**

### **6.1. Aim and Objectives**

The aim of the research was to explore the potential issues that would arise from the extension of legal deposit in the United Kingdom to digital publications and to identify the implications for existing legal deposit arrangements. The specific objectives were:

1. To review the development of legal and voluntary deposit systems, in the UK and other countries
2. To examine the current voluntary and legal deposit arrangements for printed material in the UK
3. To assess proposals made for the extension of legal deposit to digital publications and progress made on their implementation
4. To identify countries that already had, or were planning to, introduce digital voluntary or legal deposit and investigate their plans and arrangements
5. To identify and explore the potential issues arising from the extension of legal deposit to digital publications
6. To identify all the stakeholders who will be affected by digital legal deposit in the UK and evaluate their interests and concerns
7. To assess preparedness and attitudes of stakeholders who will be affected by digital legal deposit
8. To identify the management implications for legal deposit of digital publications
9. To suggest changes to the current legal deposit arrangements where necessary.

These objectives were addressed through a comprehensive review of the relevant literature and interviews with key players both in the UK and overseas. Soft Systems Methodology (SSM) informed the content and execution of the interviews and data gathered was analysed using a qualitative data analysis software package. Both the literature review and the interviews addressed all of the objectives, although the interviews specifically focused on objectives 2, 5, 6, 7 and 8.

## 6.2. Literature Review

A number of bibliographic tools were used to carry out an initial search of the literature. There is a list of the main databases used in Appendix 10. Both general and specialist sources were used. Specialist databases dealing with library and information science, legal and government information, education, research and news were used. Internet search engines were also used because much of the literature in the field is available online and conventional databases such as *Library and Information Science Abstracts* (LISA) currently have limited coverage of digital resources.

### 6.2.1. Summary of Search Tools Used

The British Library (BL) and National Library of Scotland (NLS) OPACs were used in the hope that grey documents on the subject, such as internal and similar reports, would be identified. The OCLC FirstSearch *WorldCat* search was particularly useful for identifying historical documents and the geographical coverage of the database is good. *LISA* was the main source of articles and conference papers. Unfortunately, the *LISA* search did not identify all relevant articles. In order to improve comprehensiveness, *LibraryLit*, *ERIC*, *GPO* and OCLC's *ArticleFirst* were also searched. There was a great deal of overlap between these databases, with the exception of GPO (US government publications), which retrieved no relevant records. *U K Official Publications* (UKOP) was used to identify legislation, official publications and Government press releases. Since *UKOP* only covers the period from 1980 onwards, only the more recent legislation, Government consultation papers and press releases were retrieved. Her Majesty's Stationery Office (HMSO) catalogues were used to identify older material. The Web version of *Hansard* was used to identify discussions of legal deposit in the UK Parliament. *ABIResearch* was used to identify recent mention of legal deposit in the press. No recent relevant research in the area was identified using *Index to Theses* or *Dissertation Abstracts*. Citation indexes were used in an attempt to identify important works in the area and map the field, although it transpired that journals citing the most important works were not indexed by these services.



## 6.2.2. *Searches*

Various search terms were used in combination. The initial search involved the terms 'legal' and 'deposit' used either as a phrase or using the Boolean AND operator. A number of alternative searches using advanced search features were used to achieve a manageable number of hits from Internet search engines. Other strategies included searching for legal deposit and a particular country name or looking for the legal deposit sections of national library Web sites. This was to compensate for the inability to search the term 'legal deposit' in other languages. The second phase of searching involved following up references from retrieved items and following links in Web pages. The 'legal deposit' search was repeated at intervals to identify new material and other searches were carried out to follow up specific topics. A number of useful Web sites were identified. These sites had lists of pointers to other relevant material. These included:

- *Council for Library and Information Resources (CLIR)* <http://www.clir.org>
- *CURL Exemplars in Digital Archiving (CEDARS)* project  
<http://www.leeds.ac.uk/cedars>
- The Gabriel site had a list of links to European national libraries  
<http://www.bl.uk/gabriel>
- *International Federation of Library Associations and Institutions (IFLA)*  
<http://www.ifla.org>
- *Library Association* <http://www.la-hq.org.uk> (now the Chartered Institute of Library and Information Professionals (CILIP))
- *Preserving Access to Digital Information (PADI)* <http://www.nla.gov.au/padi>
- *Research Libraries Group (RLG)* <http://www.rlg.org>

The *JISC/PADI What's New in Digital Preservation* current awareness service (<http://www.dpconline.org/graphics/whatsnew/index.html>) was also a useful source of information on relevant developments. The Department of Culture, Media and Sport (DCMS) were contacted to ask permission to see responses to the Government's 1997 consultation document. The DCMS would not allow this, but did supply a useful summary of responses and a copy of the voluntary deposit code for microform and digital material (Appendix 1).

### **6.3. Development of the Research**

#### **6.3.1. *Research Problem***

The initial literature search revealed that some work was being carried out in this area. Many of the issues arising from the extension of legal deposit had already been identified. Deposit libraries were carrying out much of the work, often in cooperation with publishers. At the beginning of this research in 1999, the situation in the United Kingdom was not very advanced. While the BL had been leading the deposit libraries' campaign to have legal deposit extended, there was very little other activity except for the development of a voluntary deposit scheme for handheld digital material. While the aim of the voluntary scheme was to work through the practical issues of the deposit of a limited proportion of digital material, there was no evidence in the literature of any active consideration of how a UK digital legal deposit system would work. There seemed to be an assumption that the system would still involve the legal deposit libraries and the Copyright Libraries Agency. It seemed likely that the Government and the players in the legal deposit system would avoid addressing the problem of how to deal with online material, at least in the near future. There did not seem to be any consideration of whether existing arrangements would achieve the main aim of legal deposit: the building, maintaining and provision of access to a national archive of the intellectual output of the nation. Neither was there much discussion on whether legal deposit is still the best means for achieving this aim.

It was obvious from the literature that digital publications pose serious issues for legal deposit systems. It seemed to be an appropriate moment to reappraise the whole concept, rather than just trying to adapt current arrangements to accommodate digital publications. The starting point for this research was a perceived problem situation. There was much discussion of the issues arising from the deposit of digital publications in the literature. These issues fall into several broad categories:

- Technical issues
- Legal issues
- Economic issues

- Management/organisational issues
- Social/Political issues

There are serious technical issues to be resolved, including the selection, acquisition, storage and preservation of digital publications. Work is being carried out in all these areas. However, other issues will also have an impact on whether legal deposit of digital publications can be successfully implemented. Some of the issues are legal, including copyright and privacy issues. There are also substantial political, economic and cultural factors involved. Legal deposit is executed through a complex of relationships between a number of actors.

A major factor in whether these issues can be resolved or not is whether the people involved – government, the publishers, the legal deposit libraries, bibliographic agencies and users – can come to agreements on the way forward. The different stakeholders have different perspectives and interests. For example, while publishers may agree that the overall aim of deposit is a ‘good thing’, they may not agree that legal deposit is the way to achieve this aim, or be unhappy with its practical implementation. For many publishers, the main concern is commercial. Others may just not agree that digital publications can or should be deposited. For resource-limited depositories, legal deposit is the only way that they can hope to acquire the entire UK publishing output. However, acquiring digital material creates additional expense for legal depositories. For users, access is likely to be the main concern. Politicians are also involved: the UK Government supports the continued development of legal deposit. However, the Government has limited resources, more pressing priorities and has to take into account the economic concerns of publishers.

The literature review suggested that a future legal deposit system might well require new roles at an institutional and personnel level and that the relationships between players and processes associated with legal deposit might need to change. These are issues explored in this research. Figure 6.1 shows the model underlying the research reported in this thesis.

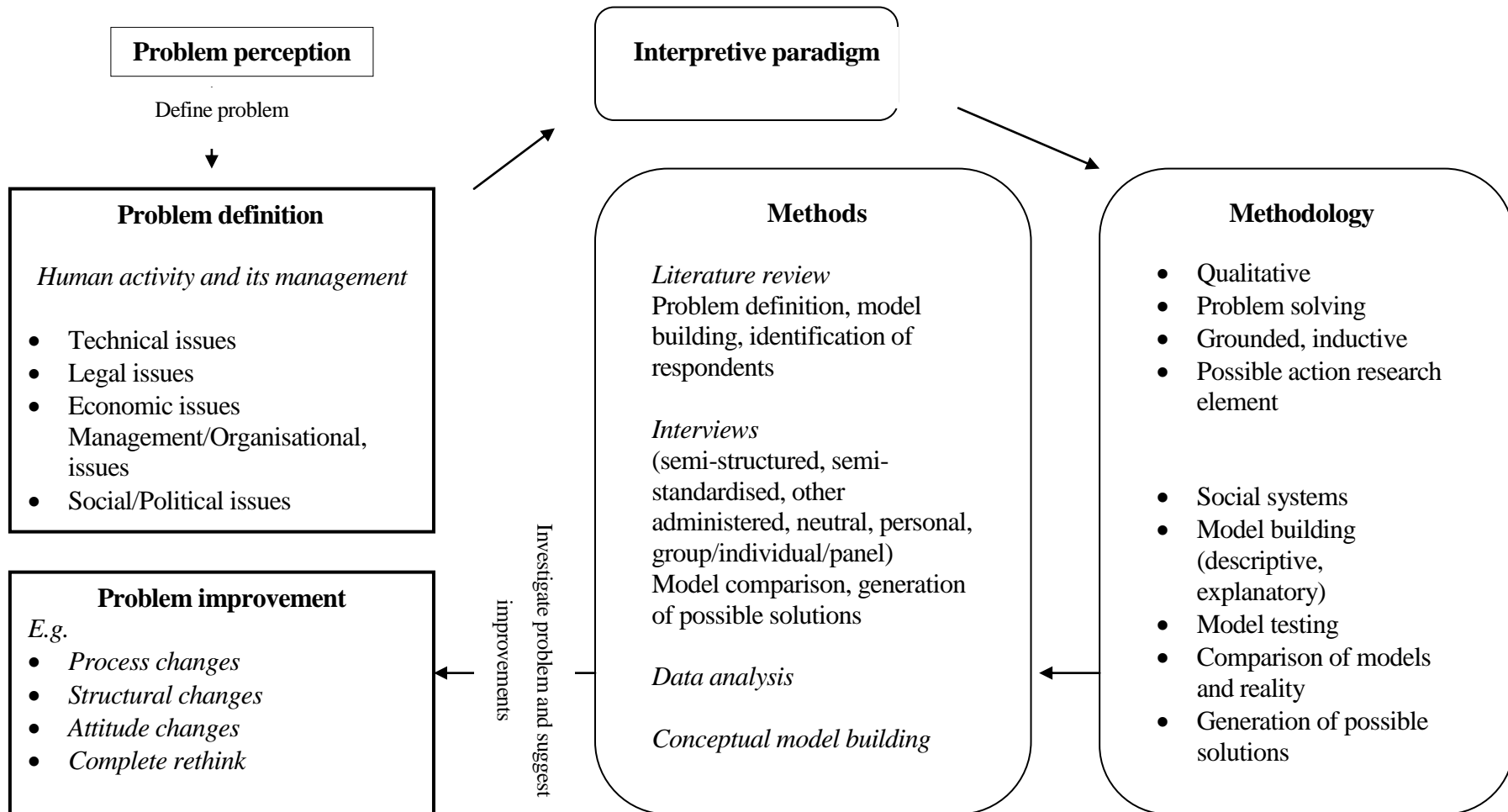


Figure 6-1 Research Model

### 6.3.2. *Methodology*

The focus of this research was examining and attempting to improve a problem situation. The problem situation - legal deposit arrangements - involves a complex of human activities. It is a social system that cannot be 're-engineered' through mathematical modelling. Therefore, the research should be viewed from an interpretative perspective - understanding the human aspects and the different perspectives of the situation - rather than a positivist one. The methodology chosen was therefore qualitative in nature rather than quantitative. Qualitative research involves looking at a particular issue from the perspective of the people involved. The focus is on explanation rather than just counting instances. Human beings have an impact on the world and other people and they in turn are influenced and take action based on their interpretation of the world and the actions of others. There research involved exploring a problem and identifying ways to solve it. While the literature review highlighted a problem situation, the real nature and extent of the problem could only be ascertained through contact with the people involved. These people were also well placed to identify potential solutions that they could implement.

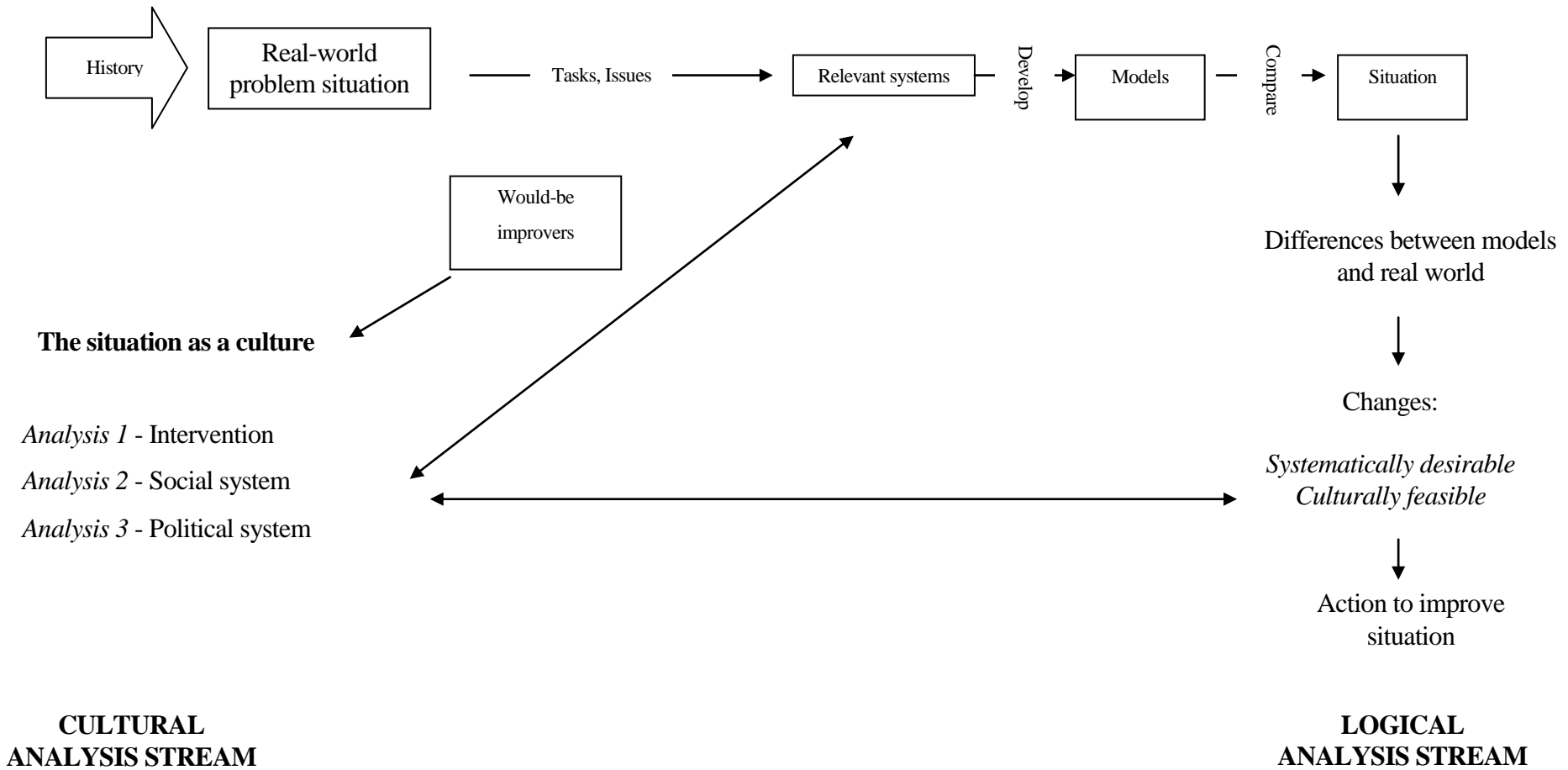
In order to identify ways to improve the situation, the situation had to be articulated in some way. It was possible to model the real world, or the institutionalised arrangements for legal deposit, and then identify problem areas. However, this approach would have restricted thinking and there was the possibility that this would have led only to minor amendments to the existing institutionalised system when more radical action might be needed.

The considerations outlined above suggested that in order to facilitate a more fundamental exploration of the problem, a conceptual model should be developed as well as an empirical model. The model had to be based on the underlying aim of legal deposit rather than legal deposit itself. All activities articulated in the theoretical model should be logical activities required to achieve the aim. The theoretical model could then be compared with the empirical model to explore the issues. A comparison of the two types of models would highlight differences between the 'ideal' and the real situation. These problem areas could then be explored in order to generate suggestions for improvement: both what to do and how to do it.

A literature search for suitable methodological bases for the research identified Soft Systems Methodology (Checkland & Scholes, 1990) as a candidate. It has been used mainly in institutional settings, but it has also been used in settings involving activities involving many different types of stakeholder (Checkland & Scholes 1990, Lea, Uttley & Vasconcelos 1998). SSM has been widely used in information-related work, including the development of information systems. SSM has recently been used for investigating a topic closely related to the topic of this research - the preservation of digital objects through metadata (Chilvers 2000).

As Figure 6-2 shows, SSM starts from a perceived problem situation and uses the formulation of models. The human or 'cultural' aspects are taken into consideration. The comparison of models facilitates structured debate on possible change. SSM also encompasses reaching an accommodation (rather than a consensus) on possible actions that allows change to be carried out. SSM has two streams of enquiry: logical and cultural. Each stream feeds into the other and helps in the process of developing models and identifying culturally acceptable suggestions for improving the problem. SSM recognises task-based systems and issue-based systems. These terms roughly map to the empirical and theoretical models mentioned earlier.

SSM models are called Human Activity Systems. The basis of the models is a Root Definition. This defines a transformation process of inputs to outputs. The root definition takes the form of "A system to do X by Y in order to achieve Z", with Z as the overall aim of the system. The various activities necessary to fulfil the definition are identified. This exercise results in a hierarchy of activities with some activities dependent on others. All the activities can be seen as sub-systems built up into a model of the overall system. The cultural side of the model takes into account the various motivations, value judgements, roles and relationships between the people involved in the real-world situation. The cultural situation can be expressed through a rich picture. This could literally involve a pictorial representation, or a written description or both. The models are compared with the real world situation. The real world situation could also be modelled. The aim of this comparison is to encourage structured debate of the real-world problem and the identification of possible solutions. The cultural stream of analysis feeds into the choice of suggestions for change, by identifying the solutions that are likely to be culturally acceptable as well as feasible.



**Figure 6-2 Soft Systems Methodology**  
 (adapted from Checkland & Scholes 1990)

#### **6.4. Application of Soft Systems Methodology**

This research was guided by SSM rather than being a rigorous application of the methodology. SSM is an action-based activity. The people involved in the situation under investigation actively participate in the process and may carry it out themselves with the assistance of a facilitator. In doctoral research, the researcher has to have a greater involvement. In the later stages of SSM, an accommodation is reached between the different actors. This stage is not always carried out in practice and although methods of achieving this were considered, it was not feasible to attempt this for a doctoral project. The methods explored were a Delphi exercise or a group discussion. An initial scan of the library and information science literature indicated that the Delphi method has not been used very successfully in this field. The main problem is sustaining on-going participation. Another problem identified was that SSM is a learning system, and is an on-going process.

In this research, the researcher had to withdraw once the situation had been investigated, suggestions agreed and put forward. It is up to the participants to decide what to do with the results of the research. Even with these caveats, the main ideas of SSM correspond so closely to the aim and objectives of this research that this methodology was adopted.

There are no fixed methods attached to SSM. The most commonly used method of enquiry is formal questioning of participants. Other methods used include informal discussion, scenario writing (operating a model) and modelling the real world into the same structure as the models. Since this is a fast moving area and the views of all the stakeholders are not represented in the literature this research included a cultural analysis to check and expand on the findings of the literature review. Both the cultural and logical analyses were carried out using face-to-face interviews in order to gather the in-depth information required to understand the human aspect of the problem situation.



#### **6.4.1.                    *Limitations of the Research***

The use of SSM for this research limited the sample size because of the time and cost implications of the methodology. This limitation was not important as far as the legal deposit libraries were concerned; interviews with representatives of each institution were carried out. Rather than providing a statistically representative survey of the views of publishers, the research aimed to explore the range of concerns and views of this group. The response rate to an unsolicited questionnaire survey of publishers was likely to be low and interviews allowed a deeper exploration of publisher perceptions. One limitation of the sample of publishers in the research is that the views of new types of publisher and information providers, such as digital only start-ups, new types of aggregators may not be represented. Although some players were identified and contacted, none consented to be involved in the research. This fact in itself may say something about the attitudes of such players, in that they may not see the legal deposit of digital information as relevant to them or problematic.

The research did not directly address the views of users. Initial discussion with the libraries indicated that they have not given much consideration to what users want in terms of access. In any case, it would not be wise to rely solely on the deposit institutions' perceptions of the issues from the users' point of view.

Given the difficulties of identifying and contacting individual users and the time constraints, the needs of users were only indirectly addressed in this research through representative bodies. This is an area that should be addressed by the libraries in the future.

A major issue for this research is that it is a fast-moving area and that it was carried out part-time over a five-year period. The field work was carried out during 2001 and the first half of 2002. The result is that the literature review provides more up-to-date information than the primary data. A particular issue was that another study of UK legal deposit libraries and publishers was carried out in the second half of 2002 (Electronic Publishing Services Ltd 2002). The focus of this research was the cost of legal deposit rather than exploring issues in depth. However, it did produce some results that could be compared with those of this research and also provided some complementary data.

#### **6.4.2. *Interviews and Selection of Subjects***

Two rounds of interviews were carried out: twenty-four in the first round and eighteen in the second round. The first round of interviews were carried out during 2001 and the second round during the first half of 2002. Fifty-four individuals from twenty-four organisations participated (see Appendix 11). Potential interviewees were identified by various means, including the literature, personal knowledge and recommendations by other interviewees. The main stakeholder groups that were initially identified included publishers, deposit institutions, bibliographic agencies, the Copyright Libraries Agency, Government and the research community, including research libraries. Interviews were held with the trade bodies involved in the Joint Committee on Voluntary Deposit (JCVD) and representatives of a selection of different types of publishers. Representatives of two bibliographic agencies participated. Other research libraries are stakeholders because they depend on the legal deposit libraries to supplement their own provision and because they could potentially become part of the legal deposit arrangements in future. This group of stakeholders was represented by two organisations, the Consortium of Research Libraries (CURL) and the Joint Information Systems Committee (JISC). While not directly involved in legal deposit, digital archives have experience in the long-term management of digital resources and this group was included in the second round of interviews. Relevant research projects and digital preservation experts also fall into this category and were included in the research.

Thirteen organisations participated in both rounds of interviews. There were various reasons for some organisations not participating in the second round, including the participants feeling that they had contributed as much as they could, participants being so unaware of the issues that the researcher did not think it was worth including them in the second round and problems with the UK public transport infrastructure that prevented interviews taking place. Five organisations participated in only the second round of interviews. The researcher selected these on the basis of recommendations by participants or from discussions in the first round of interviews.

The interviews lasted between forty-five minutes and several hours. Some of the interviews were with individuals and others were with groups, this was the choice of the interviewees. The representatives of organisations taking part in both rounds changed in some cases. This was mainly dependent on the main contact; in some cases, this person decided on the most appropriate persons to participate in the second round.

### **6.4.3.            *Cultural Analysis***

The cultural analysis was carried out through the first round of interviews. The data gathered formed the ‘rich picture’ reported in Chapter 7 and informed the development of the conceptual model and the second round of interviews. The aim of the second round of interviews was to compare the conceptual model with the perceived real world situation. A model of the current arrangements for legal deposit in the UK was also developed. The models and the results of the second round of interviews are reported in Chapter 8.

The first round interview format was loosely structured to allow the interviewees to express their views as freely as possible on a number of broad topics. The topics were perceptions of legal deposit, the stakeholders and their roles and the processes involved in legal deposit (see Appendix 12). However, the precise nature of each interview reflected the interviewees' particular area of expertise and responses.

As the interviews progressed, particular issues raised in earlier discussions were incorporated into later interviews. The first interviewee found it difficult to answer the first part of the question on processes because of its abstract nature. Therefore this issue was approached from a more practical angle, anticipating the second round of interviews by asking how the system currently works and whether current processes would be appropriate in the digital environment.

There was a problem in distinguishing between personal opinions and description of institutional policy. In the event, some interviewees gave personal views, especially when the tape was switched off. Even where it seemed that individuals were being circumspect, the responses of all the interviewees combined did give a rounded view of the situation.

The data from the first round of interviews was transcribed and analysed (see section 6.4.4) and used to develop a root definition and a human activity system (conceptual) model. The divergence of opinion between different stakeholder groups was slight enough to allow the development of a single root definition that encompassed the majority of views. Minority views, such as ‘there is no need for legal deposit in the digital environment’ and ‘the costs of legal deposit of digital material outweigh the benefits’ were, however, taken into account in the conceptual model.

#### **6.4.4.            *Logical Analysis***

The findings of the first round of interviews were summarised and sent out to all the first round interviewees (see Appendix 13). A copy of the conceptual model accompanied this report. The report and model was also sent out to new second round interviewees to allow them to familiarise themselves with the research and the conceptual model in advance of the second round of interviews.

The second round of interviews was more structured than the first. Using the conceptual model as the framework for discussion, the interviewees were asked the following questions about each activity in the model (see Appendix 14).

- Does this activity exist in the real world?
- How is it done?
- How is it judged?
- How well does it work?

While the first question resulted in a Yes/No answer, the other three questions elicited less structured responses. The aim of the third question was to find out what the monitoring and control mechanisms are for a particular activity, including criteria and mechanisms for making judgements.

In addition to identifying and delimiting the problem areas, the aim of the second interview was to consider the following questions.

- What changes can be made?
- How can these changes be made?

Again, the interviews were recorded. Detailed notes were made from the recordings rather than verbatim transcripts because the data gathered was more structured than in the first interview. Only obviously irrelevant or repetitious material was omitted and the words of interviewees were still used.

The researcher planned to carry out interviews with staff from legal deposit libraries in other countries that were leading the way in digital deposit. The Australian, Dutch and Norwegian national libraries were identified, but it was only possible to arrange face-to-face interviews with staff from the Netherlands and Norway. However, staff in the National Library of Australia have been diligent in disseminating their activities and a respondent provided some information on legal deposit in Australia via email. The results of the two overseas interviews have been incorporated into Chapter 5.

#### **6.4.5.            *Analysis of the Intervention***

The initial contact letter informed the recipients that findings from the first round of interviews would be written up into a report. This report was mentioned again at the interviews. None of the interviewees objected to this report being circulated. However, some of the interviewees did indicate that they did not want some of their comments to be quoted, either because they were reporting hearsay or because they felt the comments were sensitive. This was dealt with in two ways. Some interviewees requested that the tape recorder be switched off for parts of the conversation. While all the taped conversations were fully transcribed, parts of the interviews where interviewees indicated that they did not want to be quoted were disregarded during analysis of the transcripts. However, one interviewee asked that comments not be included in the analysis and another made exactly the same comments but made no such stipulation. In the latter case, the comments of this individual were included in the analysis.

An abridged version of the analysis was sent to the interviewees (see Appendix 13) along with an executive summary. Interviewees were asked for any comments. Only one

interviewee replied directly with some minor factual corrections. However, a representative from one of the deposit libraries made contact to say that some of the other interviewees from deposit libraries had some concerns about the content of the report and how widely it would circulate. The contact declined to be specific about the exact nature of the concerns. However, the concerns about content seemed to centre on at least one of the deposit librarians being under the impression that they were speaking in confidence and also feeling that they had been misquoted. The interviewee in question had not indicated that the particular comment was to be treated in confidence and another interviewee also made the comment in question. As for the dissemination of the report, the concern was if it was sent to any government departments or to the publishing community, this might jeopardise discussions with publishers and consequently with the Government's commitment to the introduction of legislation.

There was, therefore, an implicit threat of non-cooperation in the second round research. In order to avoid this the deposit libraries were reassured that the report was only being circulated to other interviewees. Only the executive summary was sent to the one interviewee from a government department. A decision was also taken at this point to have this thesis embargoed for three years. None of the publications resulting from this research cover the UK-specific primary research. A request from a UK publisher's journal based on the results was turned down. Ironically, the report of a similar study carried out just after this research (Electronic Publishing Services Ltd 2002) contained comments critical of the legal deposit libraries' ability to cope with digital legal deposit material that could have a bearing on discussions on regulations to implement the new legislation. This report *was* made publicly available.

This problem may have arisen from the fact that stakeholders were interviewed separately, with interviewees feeling more relaxed and expansive than they might have done in group discussions with other stakeholders. Seeing their comments later on paper may have caused alarm. Alternatively, there may have been a genuine misunderstanding. Whatever the explanation, this just reinforces the need to understand the cultural side of problems because of the power of politics to influence problem situations.

## 6.5. Data Analysis

The interview transcripts and notes were analysed using Atlas/ti qualitative data analysis software. The data collected was very rich and the use of software to manage the data and to assist analysis was useful. This particular software was selected for several reasons. A comparison of Atlas/ti with another leading qualitative analysis software, Nudist, suggested that Atlas/ti was more appropriate to this particular piece of research (Barry, 1998). These are the two software packages that are available in the Department of Information Science.

Both packages are useful for data management, for example linking transcription files together, selection of quotations and assignment of codes to quotations. However, Atlas/ti is more flexible. Codes can be assigned as a researcher reads the transcripts and do not have to be developed in advance. Codes can be renamed, deleted and merged easily as coding progresses. This flexibility was important for this research in that the problem was ill-defined and the development of pre-defined codes would not have been appropriate. The researcher's understanding of the data also evolved over the course of the two rounds of research as emerging themes were explored further through arranging interviews with identified experts.

The other important aspect is that while Atlas/ti is capable of producing reports showing hierarchical relationships between codes, the software also allows more complex relationships to be mapped in diagrammatical form. The ability to create "network" views – a complex web of relationships between codes – was helpful in an area where the issues are closely intertwined. Atlas/ti also allows the researcher to write notes on the meaning of codes, relationships between codes and emerging thoughts. The codes, quotations and any associated notes can be viewed together and in context with the full transcriptions. This all helps to ensure that data segments are analysed in their immediate and wider contexts.

Most of the interviews were recorded and transcribed. Three of the interviews were not recorded because the recorder failed, there was too much background noise and in one case the 'interview' involved moving around departments to talk to different people so recording was not practical. Where interviews were not recorded, the notes were written

up as fully as possible immediately after the interview. The transcriptions and notes of interviews were loaded into Atlas/ti and codes were assigned.

The coding process was iterative in that the initial codes were changed over time. Some of the initial codes were very general and were split into more specific codes. Others represented such similar concepts that they were merged. Some codes became categories, or ‘code families’ in Atlas. Some of the code families assigned were purely administrative and followed the structure of the investigation. So the code families included perceptions of legal deposit, stakeholders, processes and problems. Some of the codes were descriptive in order to identify areas in the interviews where specific topics were discussed. Examples of this included the names of stakeholders, or the names of different processes. However, other codes were more abstract. These included themes emerging from the data such as increased complexity, uncertainty, un-coordination and communication. While the number of codes allocated to each quotation were reduced during the coding process, it was impossible to achieve a single code for every quotation because the many issues and problems were closely entwined.

The relationships between major concepts were mapped to give some structure to the research problem area, to develop the conceptual model and identify areas for discussion in the second phase. The transcripts and interview notes from the second round interviews were analysed in the same way as the first round.

## **6.6. Summary**

The literature review revealed a “messy” situation with a number of different stakeholders with different concerns and perspectives. Soft Systems Methodology was selected as an appropriate framework for this research. Data was gathered through two rounds of face-to-face interviews and analysed using a qualitative data analysis software package well suited to SSM techniques.



## 7. CULTURAL ANALYSIS

### 7.1. Definition and Purpose of Legal Deposit

Interviewees pointed out that the 1911 UK legal deposit legislation did not state the purpose of legal deposit; it only dealt with the process of deposit. Interviewees agreed that the purpose of legal deposit is to build an archive of all publications issued in the UK. Most agreed that the main purpose of this is preservation of the recorded heritage. For the legal deposit libraries, the purpose of legal deposit is:

... twofold. One is to develop the national archive of information in whatever format the information is in, currently print, and secondly to give access to that archive. And I wouldn't want to separate the two; the archiving and the access. (Legal deposit library)

The deposit libraries see access as a secondary, but closely related aim. The deposit library perception matches the UNESCO definition of the purpose of legal deposit (Larivière 2000, p.4) quite closely. The differences of opinion between the libraries and publishers really centre on access.

Current access to the legal deposit collection is limited to use within the reading rooms of the individual deposit libraries. Users have to demonstrate a need to use legal deposit collections; they are considered collections of last resort. However, there is a perception that even this limited amount of access may be an economic threat to publishers of high-value material, especially where the value is time-dependent. One view is that if commercial users can easily gain access to this material, other libraries in the area may not purchase it. However, there is evidence from the interviews that the deposit libraries will work with depositors, either by agreeing embargoes for access to time dependent information or by taking fewer copies of very high value, low volume material.

There is some evidence of a lack of awareness of the specifics of how the legal deposit system works rather than misunderstanding of legal deposit and its purpose on the part of depositors. There is some vagueness about the libraries and numbers of copies involved. One misperception was that deposit is a requirement for material to be considered

"published". Another aim identified was to provide a record of the national published output in order to confirm something was published.

The national archive is distributed and there are duplicate copies of publications in the system. This provides redundancy, which aids preservation.

The point of it being distributed I feel is that it survives You can't trust the publishers to keep a copy of everything because they don't. They go out of business .... But if there are copies around the place in libraries, then the chances are you will be able to find one somewhere. If you only keep one copy, if the deposit system said a copy in the British Library, that would be disastrous because if the British Library gets bombed or something. (Publisher)

It is also a mechanism for making print material accessible across the country. There does not seem to be much hard evidence to support the statement that publishers do not keep material, but other research has found that publishers may not have the expertise or will to preserve access to their own digital material (Ayre & Muir 2004). Legal deposit libraries also identified legal deposit as a way of overcoming the vagaries of the publishing industry. There is a perception that this is more important in the digital environment because the commercial life of digital material may be short and it is more difficult to control continuity.

... a lot of these products are going to have a very short retail life if they ever do appear in the market place. There is no guaranteed continuity of access for these products in the way that there is for the print. (Legal deposit library)

There is another perception that the opposite is the case because digital information is much more widely available than printed material.

... is there a need for ongoing legal deposit? Legal deposit originated at a time when, obviously, the output of media in terms of the printing wasn't perhaps so widespread as the huge access now to information in every type of medium meaning that the chances of it surviving somewhere are much, much greater than they would have been perhaps a hundred, two hundred years ago. (Publisher)

However, preserving digital information will require ongoing, proactive intervention. Therefore, there is an argument that while digital information may be more widely distributed, unless it is actively managed it will not survive.

Legal deposit is also seen as a backup to other libraries. Deposit libraries have a long-term role, whereas other libraries may not keep material forever.

...the legal deposit libraries have a role in time way beyond that of any other library because we have material which came to us under legal deposit which everybody else will have thrown away. (Deposit library)

Legal deposit libraries see the fact that older legal deposit material continues to be used today as evidence of the worthwhile nature of legal deposit. Interviewees used terms such as “national good”. Scholars use legal deposit collections to aid their own work. The view of the deposit libraries is that use of the legal deposit collections aids the creation of new publications. Legal deposit material is listed in the British National Bibliography. This is seen as free advertising, which can stimulate sales. No publishers mentioned this as a potential benefit for them. While publishers must contribute their material without payment, the cost to the libraries of the long-term care of this material is considerable. As mentioned above, many publishers cannot undertake the long-term archiving of their material themselves. There is a view that preservation is outwith the remit of commercial publishers.

Publishers in general do not object to legal deposit in principle. One deposit librarian commented that there seems to be an attitude change among publishers. They are becoming more cooperative, and keen for libraries to preserve material since they may not want to do it themselves. Some publishers do object to aspects of the practical implementation of legal deposit.

Well publishers complain about it in that they have certain criticisms of the system. They don't see why it's necessary to have six depositories. But I don't think they would question the idea of the copies being retained somewhere. (Publisher)

These are likely to be small publishers and publishers of high-value, low-volume material. There is a view amongst some publishers that their legal deposit obligation is a burden both financially and in staff time.

## 7.2. Alternative Mechanisms for Building a National Archive

One interviewee questioned the ongoing need for legal deposit because someone will find it commercially desirable to maintain and provide access to digital information for as long as it is wanted.

It might be that whole idea of legal deposit, in trying to widen legal deposit, is something that needs to be questioned as to whether that is a vast undertaking that will cost the libraries a lot of money and is not really achieving anything because that information would have existed anyway in other ways. Their resources could be better directed at other types of activity ... you've got to weigh up the cost of and the practicality of being able to capture all that information against the probability of it being of interest to somebody in the future. We simply don't know and it could well be that most things are of interest to somebody at some point in time. But it's just to know whether the effort of doing it is disproportionate to the benefit that will be available to future generations (Publisher)

This argument does not address the fact that interest in publications could increase again a long time after it has initially diminished. By this point, commercial interest in the information has waned and important information may already be lost.

Apart from this, there was general acceptance that legal deposit is the only way to build a comprehensive national archive. This mirrors the conclusion of the Kenny Committee.

I think you need some very formal structure. Because otherwise, people don't do it, or if they do it is a mess. I think that if it was a legal requirement, then once you've got that legislation in place, people are less likely to argue about doing it. If it's a voluntary thing, people say "can't be bothered" or "I forgot that one, it doesn't matter, it's too expensive for us to do", whatever. Whereas, if it's legal there is a formal structure to it. I think it has to be structured. ... It does have to be formally structured in a way that people can see what that structure is. (Publisher)

Publishers would prefer not to have to provide publications for free, but the structure of the UK publishing industry is seen as a major reason why it would be difficult to get publishers to cooperate with a voluntary system.

I'm aware that legal deposit in some other countries is voluntary. But the UK is quite unique in its publishing industry, I think. It has a very large and diverse publishing industry ... Those that I am most aware of actually have got relatively small publishing industries. Or where they are big, like in the Netherlands, we are talking about two big publishers, who are world international publishers. Effectively they are only taking in material in from two publishers.

A voluntary arrangement on that sort of basis is probably much easier to do than if you are dealing with hundreds where the publishers are small to medium size rather than really large. (Specialist agency)

The Dutch interviewee did not fully agree with this assessment and said that the Koninklijke Bibliotheek (KB) made an effort to cultivate relationships with publishers<sup>8</sup>. While there may be too many publishers for UK legal deposit libraries to deal with to make a voluntary system work, there is recognition that in the digital environment, a legal deposit system would have to be complemented by negotiations with publishers.

I think without it [legislation] it will be difficult to bring individual publishers to the table to work together. But I'm certain that the only realistic way forward is genuine collaboration with the publishers, with the rights owners. That a deposit system that was based on obligation on the publishers without paying proper attention to their concerns and to their rights would not work and wouldn't come into being anyway. So therefore one has to find a way in which the proper needs, concerns, at the national level all this stuff to be captured, and for it to be guaranteed to be available for posterity has got to be addressed. (Deposit library)

It is clear from the literature that legal deposit for print material was imposed on publishers against their wishes (Feather 1994). While it took some time and legal action to make legal deposit function in the UK, publishers now generally comply.

Another point raised is that having a formal scheme will make it easier to get funding from government to support the national archive.

...any statutory obligations that we have are probably more likely to be funded by Government than voluntary arrangements. So I think we'd like to see our obligations there enshrined in statute, which would strengthen our hand in seeking new sources to allow us to cope with it. (Deposit library)

In theory, legal deposit should result in a comprehensive national archive. While the rationale for keeping copies of some types of material may not be immediately obvious, it is difficult, if not impossible to know what will have value in the long-term.

And no other way of collecting would be as effective in the sense that material comes in; it isn't selected, it isn't subject to contemporary views of what's important and what's not. If you only see your role as collecting the scholarly then you only capture a very small part of the published output.

---

<sup>8</sup> Adrienne Muir interview with Johan Steenbakkens, Directeur Bedrijfsvoering en Informatietechnologie, Koninklijke Bibliotheek, Glasgow 19 August 2002.

And what turns out to be significant to the researcher of the future, certainly if you look at Victorian publications, is very often what would have been seen at the time as ephemeral or not important, but actually now gives the insight into the life, into the culture of the time. So it's a very important mechanism at a national level for capturing in the present and recording, preserving into the future. (Deposit library)

However, legal deposit is not a panacea. Legal deposit brings in a high proportion of the published output, but coverage is not complete. Problematic areas identified by interviewees include local publications (apart from in Wales) and organisations that don't see themselves as publishers.

A final point raised here was that of international harmonisation of legal deposit. These comments were in relation to digital deposit. Large multinational companies may want to deposit in the same way in different countries. They may also want to avoid depositing the same material in different countries. On the other hand, an interviewee thought that the publishers are still anxious to limit access to deposited material, so they are likely to be concerned about international collaboration between deposit libraries. This concern also arose in the context of The European Library project (see section 5.4.4). So there is a paradox here. While international cooperation on legal deposit is seen as desirable, some interviewees thought it would be too complicated and therefore unrealistic to implement.

### **7.3. Stakeholders**

There are several groups who have a role or interest in legal deposit. The central group includes the deposit libraries, publishers and users. Another important stakeholder is government. Other libraries and authors were also identified as having an interest.

#### **7.3.1. *Legal Deposit Libraries***

The legal deposit librarians see themselves as a heterogeneous group and described their similarities and differences in various ways.

- Trinity College Dublin is isolated in terms of funding because it cannot benefit from UK funding sources or initiatives
- An “outer” group comprising the three “Celtic” libraries (National Library of Scotland (NLS), National Library of Wales (NLW) and Trinity College Dublin (TCD)) and the inner group, the “Golden Triangle” or the English libraries (The British Library (BL), the Bodleian and Cambridge (CUL))
- The BL is unique, the NLS and NLW are similar, the Bodleian and CUL are similar and TCD, while a university library like Oxford and Cambridge, is in a foreign country and somehow different in some way

Comments from interviews indicated there is tension between the different roles of the legal deposit libraries. There is a view that the BL has a dual role: it is the national library of England - a role it plays down - as well as being the British national library. The BL, NLS and NLW are under pressure from Government, which is their direct funder. This is seen to be particularly strong in the case of the BL.

Funding is an important issue for the Bodleian and CUL. The two libraries do receive extra funding for their legal deposit role, but it is not adequate. There is a view from outside these libraries that the libraries cannot cope with the volume of legal deposit material they have to handle and that there is a tension between their roles as university libraries and this ‘national’ role. The result is that

... they see the imperative of cutting costs as very, very important and they have constant struggles, I think, politically in their institutions for funds and backing. For something, which, as I say, is a national role, it’s not necessarily seen as one which is directly relevant to the mission of those universities. (Legal deposit library)

On the other hand, the Bodleian and CUL are glad to have legal deposit collections. Purchase budgets can be used for foreign material and the richness of the collections means that the institutions attract donations, which enriches them even more. In return, the libraries stress that they care for the collections and make them accessible.

The NLW is unusual amongst the legal deposit libraries in that it is also the national archive of Wales. The NLW uses its Royal Charter as well as its legal deposit privilege as a basis of its collection policy. It tries to collect Welsh material as comprehensively as

possible and it collects the recorded heritage in whatever form it takes, published or not. Wales is a small country; this makes it is easier to collect locally produced material and the NLW seems to have a close relationship with Welsh publishers. Relations are either direct, through the Welsh Books Council, or through the public library network in Wales.

One interviewee summarised the difference between the role of the deposit libraries and publishers thus

But it's a matter for us of getting as much as we can for our clientele and the community at large. For them it's a matter of securing their profits or protecting their profits or whatever. (Deposit library)

### **7.3.2. *Publishers and Authors***

The term publisher was used in a very wide sense by interviewees, especially when talking about the digital environment. In the digital environment, anyone can be a publisher and therefore it will be difficult to identify more than a fraction of publishers who could be obliged to deposit material. This point is discussed further below.

Like the libraries, publishers do not have a unified view on legal deposit. While many accept the principle of legal deposit, some object to aspects of its implementation. The major objections are the number of depositories and consequently the number of free copies of a publication that have to be deposited. A reply to this is

And if they're publishing in significant runs, then the six copies that go to the legal deposit libraries are just neither here nor there. Generally they send them out along with the review copies. (Deposit library)

This view was also expressed in the literature. Apparently there are occasional objections to the inclusion of TCD, a library in a foreign country. A reply to this is that UK publishers deposit one copy in an Irish Library, whereas Irish publishers have to deposit five copies in the UK. Another objection mentioned is the unfairness of the inclusion of the Oxford and Cambridge University libraries in the system to other libraries. This view also emerged during discussion of the draft new UK legislation. However, it does not recognise, or separate, the dual role of these libraries; while Oxbridge scholars have access to legal



deposit material, the libraries also have to give access to anyone else who wants to use the material.

Two groups that were identified as being particularly concerned about digital legal deposit are CD-ROM and database publishers. Their concerns are primarily economic. As mentioned above, these concerns stem from the high value, low volume nature of their offerings. There seemed to be uncertainty about how these concerns can be dealt with. Mechanisms such as limited claiming of material and embargoes on deposit and access are being used for print deposit.

Academic and professional publishers were seen as having the strongest interest in the current legal deposit system and it was thought that this will continue to be the case in the digital environment.

... in terms of which publishers are most interested, in terms of attitudes to, it would be academic and professional publishers who have the highest interest in this whole process. They produce by far the largest number of titles ... say there were a hundred thousand titles of books produced in the country, about half of those will be academic professional. It is unlikely that the situation will change in electronic publications because of the width of publication that is necessary to cover the needs of that particular market. (Publisher)

Authors were also identified as stakeholders who have multiple roles. The concept of intellectual infrastructure comes up again here; the authors want their work to be part of the corpus of knowledge, so have an interest in the preservation of their own work and this greater corpus. Related to this was the view that authors are also users of the legal deposit collection. This point was made in relation to higher education institutions; not all users are authors and vice versa. The third point made about authors is that they have rights in their intellectual property: possibly copyright if they have not assigned this, as well as moral rights. Interviewees thought that moral rights would also be relevant to digital preservation strategies may alter material in ways that may infringe moral rights. This view is reflected in the literature.

### 7.3.3. *Users and Other Libraries*

The range of potential users is seen to be wide. While some interviewees mentioned specific groups such as students, academics and researchers, the public in general is a potential user. Some of the deposit libraries mentioned the need to widen access to collections and the BL has made this explicit in its strategic documents (BL 2001). The libraries emphasised that while they are not public libraries, users merely have to demonstrate a need to use the legal deposit collections. The legal deposit libraries take a long-term view of their user community; they are catering for current and also future users. The needs of future users are unknown. This is given as the reason why material must be kept and why material that may seem of little research value now is included in legal deposit collections

Legal deposit libraries also consider publishers to be potential users. For example,

... one can quote all sorts of examples of publishers who simply do not do this [archive] with current publications: publishers who have to come back to the libraries and say could we have a copy of something we published twenty years ago because we no longer have a file copy. (Deposit library)

Other points made about users were that they want legal deposit libraries to keep all material regardless of format. Academic users have more of an idea of what they want in terms of access to digital material because they already have electronic access to information through their institutions. There was a comment from one of the academic legal deposit libraries that there is academic pressure to give digital access to purchased material and this may also be the case for legal deposit material. So user expectations will have to be managed in some way. This issue is discussed further in the following chapter. While not stated explicitly, it seems that the main focus at the moment is setting up the technological infrastructure for acquiring, processing and storing digital material and the high-level arrangements between the libraries and the depositors, rather than on user needs.

As well as being users, the research community were seen as partners. The national libraries are part of university and research library consortia in the UK. Legal deposit collections are only a part of a national resource and there seems to be a lot of activity focused on finding ways of collaboratively managing this resource.

I think that legal deposit probably does need to be seen within the context of discussions in terms of distributed national collection or the distributed national electronic resource, and so on. (Specialist agency)

#### **7.3.4. Government**

Interviewees identified the Government or the state as a major stakeholder in legal deposit. Government provides the majority of funding to deposit libraries, through central and devolved government departments and funding councils. The legal deposit libraries, especially the national libraries, have to work within the agenda of their funding bodies. This causes some problems because the libraries have wider remits than their sponsoring departments, or because government policy objectives clash with the objectives and implementation of legal deposit. Interviewees pointed out that the introduction of legislation to extend legal deposit ultimately depended on the Government.

The government bodies identified as stakeholders were:

- The Department for Culture, Media and Sport (DCMS)
- The Department of Trade and Industry (DTI)
- The Patent Office (part of the DTI)
- The Department for Education and Skills (DFES), through the Higher Education Funding Council for England (HEFCE) – this is the funding body for Oxford and Cambridge. It is only mentioned in passing and so is not discussed further
- The Scottish Executive and the Scottish Executive Education Department
- The National Assembly for Wales and the Culture, Sport and Welsh Language Division
- The government of the Republic of Ireland (not discussed further here)

The DCMS was the lead department for legal deposit legislation. A DCMS spokesperson said that the DCMS was working closely with the DTI and the devolved governments in Scotland and Wales in this area. Relations with the Irish government are through the Northern Ireland Office and are apparently good. Interviewees saw the DCMS as representing user interests and libraries. However, there was also a perceived conflict of

interests because the DCMS is also the department dealing with sections of the media that would come under extended legislation and also some export activities. There is no clear evidence in the literature that the exclusion of sound and films from the new legal deposit legislation was related to any such conflict of interest.

The DCMS was committed to new legal deposit legislation before the General Election of 2001. The reason given by interviewees for the delay in the introduction of legislation was the need to find a slot in the legislative timetable. According to interviewees, a slot was allocated; legal deposit legislation was to be part of a *Culture and Recreation Bill*. This did not happen and interviewees gave reasons for this. They thought that the DCMS was still committed to legislation, and the DCMS representative confirmed that the Department was still looking for a slot on the legislative timetable at the time. Interviewees perceived the DCMS as not being a big or powerful department and legal deposit as not being a high-profile issue and this may be why there were delays in finding slots.

Some interviewees expressed the view that that the Government did not want to introduce controversial legislation and wanted libraries and publishers to sort out their differences – on the access issue mainly – before going ahead. One interviewee said that the DCMS did not want to act as an arbitrator between publishers and libraries, although the view was expressed that this *should* be part of the role of government.

Chris Smith [the then Secretary of State for Culture] has made it absolutely clear that he is not going to put forward any legislation or any bill which is contentious. So there are certain bits where there are difference of emphasis, differences of view between the libraries and the publishers, particularly on access, and I think we will be expected to have ironed those out before we get to the point where a bill goes to parliament because it's been made clear that DCMS doesn't want to act as arbitrator between publishers and the libraries. Which I would have thought is its role. (*Deposit library*)

As discussed in section 2.5, the fast-track legislative route that was eventually taken for legal deposit was best suited to non-controversial legislation, although publishers had concerns about the draft legislation.

As well as being the champion for legal deposit, the DCMS is the sponsoring department for the BL. Consequently, this emphasises the cultural role of the BL. However, the BL does have connections with other government departments, such as the DFES. The DCMS

has no direct relations with any of the other deposit libraries, although they were dependent on it for the extension of legal deposit. The BL acted on behalf of the other deposit libraries on this matter.

The interest of the DTI in legal deposit is the interests of the publishing industry, including the intellectual property rights of the publishers and authors. Various interviewees mentioned that a previous attempt to introduce legislation by the then Department of National Heritage (DNH) was blocked by the DTI because

... the publishers were not happy with, they felt it was being bounced without proper consideration of the impact on their business. And there was quite determined opposition and lobbying by publishers to DTI ... (Deposit library)

The DTI was seen at that time as a strong lobby for the publishers, and the unknown economic impact on publishing businesses seen as the reason for blocking moves to legislation. However, from the deposit libraries' point of view, the DTI was more zealous on this than the publishers themselves.

There's evidence within the DTI that they have supported the publishers' case, or what they have perceived to be the publishers' case to be. (Deposit library)

While interviewees expressed this view, some stressed that they had heard this was the case, rather than knowing for certain. In addition, no interviewees stated what evidence there actually was for this view.

The Patent Office's interest is that it has responsibility for copyright law and it is concerned that legal deposit provisions do not compromise copyright law. Interviewees thought that both the DTI and the Patent Office (with regard to the rights issues specifically) were happy about the extension of legal deposit in principle, but were concerned about how far legislation would be extended. Interviewees thought at the time that there would only initially be a limited extension of legal deposit, it was probably safe to assume that the DTI would be comfortable with the proposed legislation. While there were renewed claims of hasty drafting and pushing legislation through Parliament in 2002 and 2003, the DTI does not seem to have strongly objected (see 2.5.5). It would be interesting to know how much influence the DTI had in the matter.

No doubt the Patent Office will be involved in the drafting of Regulations to implement the new UK legal deposit legislation.

The Scottish Executive and the National Assembly for Wales fund the NLS and NLW respectively. While the NLS is sponsored by the Scottish Executive Education Department, the Environment Minister has responsibility for the NLS. This situation was seen as “confusing”. In the case of the NLW, the sponsoring department is a new Culture, Sport and Welsh Language Division. There was no indication of any objections to the BL taking the lead, with the DCMS, on extension of legal deposit. There were mixed feelings about the NLW coming under the Welsh Culture Department. The NLW has done well financially recently, but there was concern that the NLW is conceived only as a cultural institution when it also has an information role. There was regret about the severing of the connection with the Welsh Education Department, the NLW’s former sponsoring department. The NLW has a remit to serve the higher education community through its Royal Charter.

#### **7.4. Relationships Between Stakeholders**

##### **7.4.1. *Deposit Libraries and their Sponsoring Departments***

As mentioned above, the Government was seen as setting aspirational agendas. The three national libraries are under pressure from their sponsoring departments to fit in with Government policy agenda.

... what struck me very much is the political angle which has to drive the BL. Because they’re much closer to, well the National Library of Scotland and Wales are close to the Scottish Office, the Scottish Parliament and the Welsh Assembly now, so there is a political agenda and you have to jump when the strings are pulled because you don’t have a choice really. Which can lead to problems arising. You can’t guarantee that your policies can stay the same because the political forces change then you often have to change your management. (Deposit library)

An example given of this political pressure is the potential danger of a lack of support for activities the Government department does not see as directly contributing to its own objectives.

We see totally eye-to-eye on the question of greater accessibility and keying into lifelong learning and the rest of it. We've no divergence of ideology at all there. What's more difficult is to get the ... [devolved government]<sup>9</sup> to see the relevance in the future and now of things which don't directly key into their strategy. When we put the corporate plan forward the first thing we have to do is to say how our vision and strategy keys in to [that] so we have to go to great pains to construct these great tables which shows x initiative keys into your strategy. Anybody can play that game. What is difficult is when you have got something that has got nothing to do really or is only periphery with the aims of ... [devolved government]. (Deposit library)

The National Libraries of Scotland and Wales have to deal with agenda in their home countries that may be geographically narrower than their traditional role.

... particularly in Scotland I think, mainly that with devolution, national, that is country, agendas tend to brook that much larger now than perhaps they did in the past and the national libraries are seen as national resources by the national governments and therefore there are other agendas, more powerful agendas going on in Wales and Scotland which might override any ambitions to collaborate. I don't know how strong that is. (Deposit library)

... there may be aims within [country] that don't have anything in particular to do with the [devolved government], but worse than that is where we are a national or an international body; there is no responsibility for international affairs. Where we're seen as an extra -[country] body that's when it becomes difficult or could become difficult for example with legal deposit in the future with getting [the devolved government], there's a kind of parochialisation at work with devolution. There's lots of good things about devolution but it's also a kind of narrowing, or danger of a narrowing, of view. I can't say it's happened to us in the sense that the [devolved government] has refused to fund something that we have been doing or that we want to do. But you can see that that might well happen in the future. (Deposit library)

This "parochialisation" may manifest itself in a reluctance to support the NLS and NLW's current collection policies in the digital environment. These libraries currently do not just take Welsh and Scottish legal deposit material and leave the rest to the BL. If a secure network is built between the libraries then this situation may change, and these libraries may only provide access rather than store publications not published in their country.

Another point related to mismatches between government policy and legal deposit of digital publications is on widening access. While the libraries may be able to take advantage of digital and networking technologies to provide wider access to some of their collections, access to legal deposit material will still be restricted, at least for a period.

---

<sup>9</sup> Name of the devolved Government department removed to maintain anonymity.

There is a question as to whether the devolved governments perceive the difference between the current legal deposit and older collections and whether they would be willing to provide support for activities that do not “add value” or increase access.

#### **7.4.2.           *Between the Legal Deposit Libraries***

The feeling amongst the deposit libraries was that they have, on the whole, good working relationships. The libraries felt there has been a greater degree of cooperative working over the last few years. Lack of resources has been a catalyst for this; legislation does not stipulate any cooperation between them. There is also some collaboration between TCD and Queen’s University Belfast in an effort to compensate for the lack of a deposit library in Northern Ireland.

The legal deposit libraries admitted that there are differences between them that mean that they have different perspectives that are reflected in their attitudes to joint working. While they stressed that it is not a serious problem, the BL was seen as having a very different approach to the others; much of this due to external pressure. There were comments that the BL has been frustrated in the past with the scope of collaborative working. It was not clear from interviewee comments what the BL’s then current position on sharing responsibilities was. On one hand, there was a perception that the BL had recently retreated to an extent from the position of wanting a greater degree of shared responsibility. There were hints that this might be, at least partially, due to bad publicity over disposals of foreign newspaper material. Current BL involvement in initiatives such as the Digital Preservation Coalition (DPC) and Web harvesting groups would suggest that the BL has some commitment to a cooperative approach to preserving digital information.

There was a view, expressed by outsiders, that the legal deposit libraries only talked about cooperation and that there was not much successful cooperation actually going on at the time. Shared cataloguing was considered a success by the libraries. However, they admitted that efforts at shared acquisition have been less successful. There was a suggestion from outsiders that there are psychological barriers to more effective cooperative activity between the libraries.



... basically you've got to give something up for the greater good ... How do you get people to say "well we don't do that, we won't spend our money on this because we can see that if we pool our resources to do that, everybody will get a better benefit?" ... How are we going to get over that psychological barrier? (Specialist agency)

Another outsider expressed the view that the aims and desired outcomes of cooperative activity were not explicit and there was a lack of mechanisms for evaluating success.

Unless you, it seems to me, have a pretty good idea of what it is you want, how you're going to measure the efficacy of whatever you put in place to achieve that, you don't know whether there's any benefit in co-operating. (Specialist agency)

Apparently there has been an element of shared acquisition between the Agency libraries for some time, in that not all five libraries will take a copy of every publication. However, a publisher thought that the Copyright Library Agency is making too many block requests for material; a view also expressed by Whitaker (see section 2.5.4). All the legal deposit libraries were involved in a project for leisure journals. The rationale seems to have been to reduce the burden on publishers and cost savings for the deposit libraries. The aim was for the BL to take 50% of the journal output whilst the other libraries were to take 10% each. Allocation was to be on a subject basis. Apparently there was no reduction of costs. The libraries did not seem to be entirely clear about why this was the case, but the number of journals designated as 'leisure' was quite small. Another reason seems to be that not all of the libraries felt that they could adhere to their quotas because they would be unable to meet the needs of their users. It is more difficult for the national libraries to only take a proportion of material because users expect them to have everything, or at least everything in their geographical area. There is also the need for redundancy in print collections to ensure preservation. Another comment was:

Cost savings is quite an interesting one. It depends on how much they tried to do again. Quite often for cost savings, it has to be quite radical, doesn't it? (Specialist agency)

While shared acquisition may be difficult for print material, this may not be the case for digital material because access can be networked.

The legal deposit libraries said that they had a number of committees or working groups with legal deposit in their remits.

The three main groups were:

- The Copyright Librarians or Principals meeting
- The Standing Committee on Legal Deposit (SCOLD)
- The Copyright Libraries Agency Management Committee

A number of other groups were also mentioned. These included a sub-committee of SCOLD dealing with preservation issues (SCOLD Pres) and two sub-committees of the Joint Committee on Voluntary Deposit (JCVD, see section 7.4.4), one technical and the other a librarian's group. The NLS, NLW, Universities of Oxford and Cambridge and TCD jointly fund the Copyright Libraries Agency to claim material on their behalf.

#### 7.4.2.1. Principals' Meeting

Interviews said that that this group met infrequently in the past.

... two and a half years ago it was very much a club, gentlemen's club, really. It only met twice a year, well one proper meeting a year, and it was all wining and dining and so on. (Deposit library)

None of the interviewees who mentioned this meeting could remember its correct title. While this might be because this title had changed recently, it might also indicate how seriously this was taken as a working group in the past. It seems that the, then, new Chief Executive of the BL had been a catalyst for change here. At the time of the interviews, there were two meetings a year, and although there is still "a nice dinner" involved,

The frequency of meetings has improved and the quality of debate and range of issues that that body deals with has extended as well. So I think that's going to be a more important body. (Deposit library)

The purpose if this meeting is now very high-level and strategic.

That's where we look at the broad issues and then try and decide how we go forward and what SCOLD's programme ought to be and what the Agency should be doing and all the rest of it. (Deposit library)

There's a committee of the principals of the legal deposit libraries that meets twice a year. So that's the chief librarian in each case. They delegate the specific activity relating to legal deposit to a committee called SCOLD, which is the Standing Committee on Legal Deposit. (Deposit library)

#### 7.4.2.2. Standing Committee On Legal Deposit

The role of SCOLD was not clear from the interviews. On the one hand

SCOLD is the kind of strategic body, and the Agency management committee deals very much with the practicalities of managing legal deposit on behalf of the Agency libraries. (Deposit library)

However, SCOLD was also described as the committee

That tends to look at issues of practice and procedure, not nitty gritty things but it's a practitioners' forum. (Deposit library)

...that focuses more on the operational aspects of the system and the principals' focus is obviously more on the policy and strategic aspects. (Deposit library)

While the language used to describe the role of SCOLD is ambiguous, it does seem to occupy a role in between the high-level Principals group and the Agency Management Committee, at least from the point of view of the Agency libraries. SCOLD was also seen as being important and carrying out good work. However, SCOLD had no power or resources to do anything itself; it could only allocate responsibilities. The membership of this committee included heads of legal deposit or heads of collections and there were representatives of all six libraries.

#### 7.4.2.3. The CLA and the Agency Management Committee

The Copyright Libraries Agency (CLA) operates on behalf of the five so-called "Agency libraries". The CLA claims material on behalf of the Agency libraries, receives this material, and then distributes it to the libraries. The CLA shares bibliographic data with the BL Legal Deposit Office (LDO) daily. This cooperation is to help both operations identify all new publications.

A new Agent had recently joined the CLA; the previous incumbent had been in post for a long time. In fact many employees, including the previous deputy Agent had worked at the CLA for a number of years. The new Agent had spent a lot of her time on modernising procedures. The introduction of new claim forms increased the volume of material coming to the CLA. With its current set-up the CLA cannot cope with this. The accommodation is cramped and relocation was being considered.

The rationale for the CLA is to reduce the burden of legal deposit for the Agency libraries and publishers.

And one of the key things for all of us, all five libraries, is that as well as being extremely efficient on behalf of what it does for all of us, it's the most cost effective way of doing it. I think if each of the five libraries tried to run its own operation separately, it would be far less effective, it would cost us far more money and it would be far more inconvenient for the publishers. It's much easier to deliver five copies to one location in London or another centre, than sending five copies to libraries throughout these islands. (Deposit library)

However, the CLA was seen as a drain on the resources of the libraries.

There's been a fair amount of under-funding of the CLA, under investment in the CLA; on the other hand we need to make sure it's as slim-line an organisation as it can be. (Deposit library)

Another point made was:

You see it all relates, Adrienne, to politics because ... if they're going to transfer resources, from their own institutions to the Agency, obviously to me it's much more cost effective. But if they do that, then their own internal organisation has got to undergo a tremendous change. That I can't see happening. (Specialist agency)

The BL and the other libraries have two separate operations for acquiring legal deposit material. There is some cooperation between LDO and CLA staff already: the sharing of data mentioned earlier and the two operations shared software at the time of the interviews. There was some puzzlement as to why there has to be two organisations carrying out the same kind of work.

To me it's extraordinary that you've got two organisations doing the same kind of work and yet there has not been, on the operational side, I'm not talking about the managerial side ..., about legal deposit policy. I'm talking on the operational side, thinking strategically about workflows or productivity levels and communicating about how we train staff on the software. (Specialist agency)

There is the question of course of whether the CLA should or should not be joined in some way, more formal way, with the BL's Copyright Office. It has always stuck me as very strange even granted the fact that there are legal differences between, in terms of the legal framework, between the BL and the rest of us. Nevertheless it has always seemed to me rather odd to have two bodies ... (Deposit library)

Relative newcomers to the legal deposit system made these comments. Another interviewee explained that there is no objection in theory to closer cooperation between the CLA and the LDO. This issue was reviewed a number of years ago and the conclusion was that it was not an option at that time. The differences between the legal entitlements of the BL and the other libraries mean that the LDO and the CLA have different procedures. The projected payback time from changing the procedures was too long to be acceptable. Another interviewee disagreed that there were any real differences between the procedures of the CLA and the BL LDO, saying that the LDO does have to claim for some material. There is a difference in perception of differences between the two operations here.

The role that the CLA will play in digital deposit is not clear. There was a view that the case for a central agent for distribution is stronger. The role of the CLA could be expanded to include some processing of material, for example cataloguing or the provision of central storage. On the other hand, there was a view that the CLA has no role; the BL will take on a central role, or material will be go directly to the libraries. The issue of the lack of willingness to invest in the CLA came up again in this context. There did seem to be agreement that the CLA will continue to have a role in the print environment.

Like the high-level group, the Agency Management Committee also seemed to be changing its character. Until recently, this committee had been very "hands-on"

The Agency management committee looks after the nuts and bolts of how the Agency is run, whether it's in terms of how it's functioning generally or in terms of the day-to-day decisions that may need to be made. (Deposit library)

However, this was changing

... because of the change in the identity of the agent. I think I'm right in saying there's a movement towards the management committee being more hands-off in the sense of leaving the management questions to the manager and really concentrating on the policy side of how the CLA works (Deposit library)

We're going through a phase of change. We've appointed a new agent last year and she quite rightly feels that she doesn't really want micromanagement as has been provided historically. So we are looking at how that relationship works out. (Deposit library)

In fact the roles of the various committees were being reviewed because

... we have become conscious that there is a plethora of committees and sub-committees around the legal deposit thing and I have suggested to the librarians ways in which they can be rationalised. (Deposit library)

While not stated explicitly, comments from interviewees give the impression that despite all the committees, communication on digital deposit between the six libraries was poor. It was not clear where the communication break-downs were. One example was that there were only two legal deposit library representatives on the JCVD (described below), one for the BL and one for the Agency libraries. Interviewees from legal deposit libraries were not always clear about what other legal deposit libraries were doing. Another explanation may be that representatives on committees were not disseminating information well within their organisation. The size of the BL was also seen as a problem: there were lots of different BL staff involved in different, but related things, who do not always talk to each other.

The proposed reorganisation of committees discussed by interviewees looked as if it might overcome some of the communication problems. SCOLD was to have (or retain) a strategic role. The Agency Management Committee was to have more of a strategic role. SCOLD and the Agency Committee were to have cross-membership. Ad hoc committees were to be set up for specific purposes. SCOLD and the JCVD were working semi-independently and the plan was that the JCVD sub-committees would report to SCOLD as well.

### 7.4.3. *Relationships Between Legal Deposit Libraries and Publishers*

It seems that there is not much formal contact between legal deposit libraries and publishers for the deposit of print material.

I'm not sure that there's a forum in which all the deposit libraries talk to all the publishers or the publisher representatives. (Deposit library)

I'm not sure whether on the legal deposit print side there is a mechanism whereby the Agent and the head of our Legal Deposit Office have actually got an opportunity to talk to publishers as a group or contact with individuals as it were. (Deposit library)

The BL had an annual meeting with the PA, but the remit of this meeting was much wider than legal deposit. The NLS claims print material directly from many Scottish publishers. The NLW takes a proactive approach to claiming Welsh material and works through the Welsh Books Council and the Welsh public library system. Apart from this, the libraries did not have direct relationships with publishers for print deposit. Interviewees expressed the view that the role of publishers is likely to become more important in the digital environment because of the anticipated need for a greater degree of collaboration and negotiation between publishers and deposit libraries. This view is confirmed by the literature.

Because there was no existing mechanism for publishers and deposit libraries to discuss matters of mutual interest, such a mechanism had to be built to discuss the deposit of non-print material. There was a strong impetus for both sides to collaborate closely. The Government was perceived to not want to introduce controversial legislation, the deposit libraries wanted legislation to be introduced and the publishers seemed to accept that this was inevitable and wanted to ensure that it was sensible from their perspective.

... part of the premise for doing the voluntary code is that if we get out there, well we, whoever, with a sensible code then if you get something working and you're happy with it, then maybe things will progress along the lines that you want instead of leaving it to someone else. (Publisher)

The libraries felt that the less commercial publishers are less threatened by the deposit of digital material and more willing to work in partnership.

There are some that would take the view that the state has no right to interfere in their right to make commercial profit. There are others, and it tends to be on the learned society, non-commercial front, whose main concern is to ensure the intellectual record, recognise that their own resources to ensure that are limited and would very much welcome a partnership approach with the deposit libraries, and others if necessary, to achieve it. So I think there's a strong interest and willingness to ensure the long-term archive among all those publishers who see their role as being part of the scholarly communication process. I think it's more difficult where people see their role primarily as commercial, making money. (Deposit library)

The libraries realised that they have to build up a trust relationship with the publishers in the digital environment. An interesting comment was that:

If we got our act together better, and collaborated more, then potentially we are likely to get further with the publishers, plus it is going to be easier for them to comply. (Deposit library)

However, reactions to the draft new legislation suggest that the trust is not yet there.

#### **7.4.4. *Joint Committee on Voluntary Deposit***

The JCVD met quarterly to monitor the voluntary deposit scheme and work on the issues. None of the legal deposit libraries commented on why there were only two library representatives on the JCVD. Initially, the Copyright Libraries Agent was not involved, but later joined. Three publisher associations were members: the PA, the Association of Learned Professional and Learned Society Publishers (ALPSP) and the Directory Publishers Association (DPA). The general feeling is that this group worked well

...I think that's quite an effective forum; it's open and it's honest and there's good discussion on issues and problems. (Deposit library)

It's a very civil discussion. Librarians and information publishers co-exist. We're different sides of the same coin. We publish information for librarians to use. If they can't use it, they're professional people, they understand content and the value of it. (Publisher)

However, there was a comment from a publisher trade body that although what needed to be addressed was understood, practical progress was difficult.



The publishers associations represented traditional publishers who have moved in to digital publishing. By their own admission, they did not represent newer types of publisher or electronic only publishers.

... the reason for that that is not to get overburdened with having to deal with the whole wide range of possible publishing associations. These are quite a cohesive group; they work together a lot, these three associations. They're traditional print publishers developing into online. They've obviously been the main associations concerned with the legal deposit system as it currently stands in terms of discussions over that. And of course there are all sorts of publications beyond the membership of these associations or the sort of publishing interests they represent. (Publisher)

The voluntary scheme is nominally limited to handheld material. Originally, online was to be included but publishers could not agree to this, so this was taken out of the main part of the scheme, but left in the appendices to leave open the option of experimenting with online material (see Appendix 1). Offline was seen as a declining electronic publishing area, a view confirmed by other research (Electronic Publishing Services Ltd 2002). However, it is felt that there are some issues that will apply to all digital publications and this was seen as a practical way forward. The researcher's view is that it might have been helpful to have a representative of the 'new' types of electronic publishers.

There were various views on the purpose of the voluntary scheme. It was generally viewed as a useful period before legislation to develop policies and procedures and to "get things right".

But we have to lay down a series of, well I shouldn't say we lay down the guidelines, agree a series of guidelines with the publishers as to how it will be done, what kind of records will be provided and so on because there was no point in simply saying great send us a whole pile of CDs. (Deposit library)

The voluntary scheme was also seen as a confidence building exercise.

One of the virtues of having a period of voluntary deposit is that it will help build confidence with people who are cautious about how all of this is working; the dangers there might be for the commercial viability of products. We hope we will be able to build up confidence and demonstrate that what the legal deposit libraries do is essentially for the long term benefit of everyone who is working with information, including the publishers of various products. (Deposit library)

The BL is central to the voluntary scheme; it set up the scheme on behalf of all the libraries and it seems that most material coming in is going to the BL. The general opinion of the success of the voluntary scheme at the time of the interviews was that it was going slowly in terms of the amount of material being deposited. The figure most quoted for the first year of operation (2000) was around 700 items. The deposit libraries themselves were not entirely clear because at the time of the interviews they were not getting a clear picture from the BL. Reasons given for the lack of clarity over the success of the scheme were:

... all we have done is really at a certain point is to say is from well January the fourth two thousand all electronic publications received by the Agency and passed onto the libraries or received directly have come under the terms of voluntary deposit. In practice, we don't know. I suspect that many of them have simply come by default as it were because the publishers have deposited their total output (Deposit library)

... we don't know how many CD-ROMs are published in the UK each year. The publishers don't know because they've got no central, there isn't a Whitakers equivalent for electronic publications. (Deposit library)

However, a Whitakers representative said that they do include digital material, including handheld material in their database.

... if we want to be an aggregator for bibliographic data, covering the whole range of media, then it's vital that we carry ebook data, print on demand data and so on, as well as printed books. \*Interviewer And you already do? \*Interviewee Yes. (Specialist agency)

However, the database may not have included all digital publications.

One interviewee thought that the scheme was going really well and that targets had been surpassed. This was one of the later interviews carried out, so the situation may have changed over time. Later research on the proportion of material received is reported in section 3.7.3.

Several reasons were given for the slow progress of the voluntary scheme. One was that there might be some active hostility to the scheme and that publishers were actually ignoring it. The publisher associations on the JCVD publicised the scheme to their members. However, there was a view that this had not reached the right people within publishing organisations.

There was some evidence of a lack of awareness of the scheme in the interviews with publishers who were members of the publishing associations represented on the JCVD. Publishers who are not members of any of the organisations were probably not aware of the scheme at all. As well as a lack of awareness, lack of process may have been a problem as far as publishers were concerned. While the deposit of print material may be well integrated into the publishing process, deposit of non-print material is not. There were suggestions that in future there would be some more direct contact with identified publishers to find out why they were not depositing material.

Another point made here was that the mechanisms for deposit were cumbersome. Publishers were asked to fill in two forms. One of these covered details of the publications and whether the publishers were willing to deposit. The other covered access arrangements granted by the publisher. The libraries were not getting back the technical details they required. Most publishers who were depositing were allowing access in all six libraries.

Another problem identified was the lack of progress on setting up the technological infrastructure in the libraries, in particular the BL. The secure network and the BL's Digital Library System were not yet in place. This was holding up experimentation with online material. There was a feeling that the volume of deposit would improve when these problems were solved.

Other cooperative activities with publishers mentioned by interviewees were discussions with publishers about online digital material and about an appropriate forum for discussing non-commercial Web material.

Now that there is new legislation whose detailed provisions will have to be implemented by subsidiary legislation, there will be a continuing need for a forum for the deposit libraries and publishers to work together. As discussed in section 2.5.5, a new Joint Committee on Legal Deposit has now been set up.

#### **7.4.5. *Legal Deposit Libraries, the Research Community and Publishers***

There seemed to be an impetus towards closer collaboration between these different sectors and a belief that this can overcome some of the potential problems of long-term access to digital collections that are discussed in the next chapter. The Publishers Association (PA) was having meetings with the BL and the Joint Information Systems Committee (JISC). The higher education sector and the BL had a joint Task Force at the time. JISC was funding the CEDARS project. Sir Brian Follett was leading work on the provision of information to researchers and the DPC was newly established. Legal deposit libraries and publishers expressed a desire for more coordination of activities.

However, there was a view that collaborative activity between the legal deposit libraries and other players in the research community was at a very early stage.

Collaboration and things like distributed national collections are actually ideas at the moment. All the work to actually put them together is just beginning to happen. And much of what has happened so far has been in a research environment. Implementing and building that into services is going to be quite a painful process I think. I think libraries don't really know how to deal with it ... collaboration has been sort of around the edges, an add-on and meetings, an exchange of views. (Specialist agency)<sup>10</sup>

#### **7.5. *Extending Legal Deposit Legislation***

Legal deposit provisions were, at the time, part of the Copyright Act of 1911. The general view from the first round of interviews was that these provisions would be repealed or superseded and that new legislation would be drawn up. At the time, it seemed likely that future legal deposit provisions would be part of a larger Act of Parliament rather than a standalone piece of legislation. The DCMS was looking at other draft legislation to see if legal deposit could be incorporated. This did not happen, and new legislation was a dedicated legal deposit Act.

The DCMS was not looking at legislation of other countries; it was relying on the BL to advise on this. The DCMS considered the UK to be a leader in this area, although several

---

<sup>10</sup> The Research Libraries Network should take this activity forward.

other countries had already introduced extended legislation. Interviewees expected that experience gained under the voluntary scheme would influence legislation. The DCMS had consulted the BL on legislation. In April 2002, during the second round of interviews, the DCMS approached the BL because the Government was now actively interested in taking the extension of legal deposit legislation forward through a "hand out" bill. This type of draft legislation is explained in Chapter 2.

Legislation apparently cannot be drafted unless all the relevant government departments agree. There also has to be a Regulatory Impact Assessment of the likely effects of extending legislation on commercial information suppliers. This does not just cover the cost of material supplied, but also other costs of compliance, such as staff time. The publishers were not happy with the outcome of a previous exercise and are concerned that it was carried out "properly" this time. This looked like a potential future area of contention. It has now been carried out and so far there have been no major complaints about the methodology or data collected. Assuming legislation was successfully introduced, 2004 was given as a realistic estimate of when digital material would be collected under legal deposit. At the time of writing (September 2004) there was no sign of this.

Interviewees thought that new provisions were likely to take the form of some brief enabling legislation that would be general enough to accommodate any future developments in the publishing industry. The detailed regulations would then be laid down through a series of secondary legislative instruments. The purpose of this approach is to allow for new formats to be added without having to periodically redraft legislation. It was expected that legislation would set out the purpose of legal deposit – to extend the concept of a national published archive. The enabling legislation does not really do this.

At the time of the first round of interviews, it was not clear to interviewees what the initial scope of legislation would be. It was expected that it would cover offline digital publications. There was apparently pressure for it to cover online and this is why the JCVD wanted to carry out experiments with online material. During the first round of interviews, the DCMS thought that provisions were likely to include offline material initially and that electronic journals, electronic books and reference material might follow this.

Interviewees thought that the current legal deposit libraries would retain their deposit status and it was likely that the British Film Institute would be the only new institution added. The new law actually excludes the National Sound Archive and the British Film Institute. The deposit of sound and film material will continue to be on a voluntary basis. There was some speculation in the interviews about the respective privileges of the current legal deposit libraries. One interviewee thought that the new legislation would not have the historical basis of the current legislation for print material and that there would be an attempt to reduce the number of differences between the deposit libraries. This would involve a single category of deposit library implying that there would no longer be the distinction between the automatic right to receive and the right to claim material. Another interviewee did not think that the differences would be evened out.

Other possibilities that were discussed in the interviews were the possibility of Trinity College Dublin retaining its privileges for physical format material and online material being deposited in Northern Ireland.

Interviewees were not sure whether the proposed secure network between the deposit libraries would have a node in Northern Ireland. There was a view that there should be a deposit library in the North of England because there is a perceived gap in geographical coverage. This option came up during discussions for the 1988 copyright legislation, but it was rejected. Interviewees thought that the number of deposit libraries would remain the same, except for the addition of the British Film Institute for moving images, for continuity or political reasons. However, there was a view that if any deposit libraries were vulnerable, it would be the two UK university libraries.

... the 1997 Legal Deposit of Publications [government consultation document] did raise the possibility, which is of course has been raised before, of cutting back on the number of repositories and I think it's clear that the two vulnerable ones would be Oxford and Cambridge. For political reasons the repositories in Ireland and Scotland and Wales would be the ones likely to remain. But of course, an alternative scenario is that you reduce the number of copies and share them among the six. That's something that would be maybe more acceptable, if you have to reduce the number. (Deposit library)

Since the new Bill was not publicly available at the time of the interviews, it was not clear what the relative deposit library privileges would be. However, the new legislation retains the pre-existing provisions for print.

## **7.6. New Players**

Apart from the above, there were no particular views on the possibility of new players in the legal deposit system, apart from new types of publishers. There was a view that a body could be set up with specific expertise to deal with digital publications. The only other point made was that there is a lot of expertise in the higher education sector. However, there was a view that it is questionable whether higher education libraries would have the resources to take on legal deposit responsibilities. The BL has had discussions with the Internet Archive, the concern there is the long-term sustainability of what they are doing. The BL has been comparing what it has with what the Internet Archive collects. Representatives thought that setting up agreements with other organisations at the moment might interfere with legislation and there are not as yet any clear economic or service models for such arrangements, although the KB seems to be going ahead anyway.

## **7.7. Summary**

Most of the interviewees agreed that there is a need for a national archive of digital publications and that legal deposit is the best mechanism to achieve this, although there is disagreement on how it should be implemented. While interviewees acknowledged expertise in digital archiving and preservation in the higher education sector, there were no serious suggestions of alternatives to the existing legal deposit libraries. Stakeholders in legal deposit include legal deposit libraries and publishers. When interviewees talked about “publishers”, they did so in a broad sense including commercial, non-commercial, traditional and new types of publishing organisations and individuals. Authors and users were also identified as having a stake in legal deposit, although discussions focused on the libraries and publishers. Government is an important stakeholder. The libraries were dependent on Government for progressing legislation and for funding. Interviewee comments made it clear that these stakeholder groups are not homogeneous. The legal

deposit libraries have different approaches to legal deposit and are responding to different external pressures. Different types of publishers have different views on legal deposit, with some of them perceiving it as a burden and potentially detrimental to their business and others welcoming the deposit of their publications for posterity. There was some evidence of a lack of communication within and between the different stakeholder groups. Along with a lack of resources, this was holding back the development of a system for digital deposit in the UK. However, there was a successful forum for cooperation between libraries and publishers that was working on addressing the issues that will arise from the extension of legal deposit to digital publications. The issues identified by interviewees are discussed in Chapter 8.



## 8. LOGICAL ANALYSIS

### 8.1. Legal Deposit – the Real World

Figure 8.1 models the current legal deposit system. This model has been developed from descriptions of the current system by interviewees, the author's personal experience and the literature. It includes the main stakeholder groups, the flows of funding, publications and metadata (at the moment, mainly descriptive metadata) between the different stakeholders, and the provision of access through each of the legal deposit libraries. The model does not show the internal processes in each of the stakeholders because there are differences between individual publishers and deposit libraries. The details of how the current system works and the issues and problems of introducing legal deposit for digital publications are discussed according to the framework of the conceptual model (Figure 8.2).

### 8.2. Developing the Conceptual Model

As discussed in Chapter 6, Soft Systems Methodology involves developing a root definition of the system under investigation, then developing a human activity system (conceptual model) based on this definition. The root definition takes the form of a transformation of inputs to outputs - "A system to do X by Y in order to achieve Z". Based on the findings of the research, the following definition was developed:

*A system to build and maintain a national archive of the nation's intellectual output in digital form by collecting, preserving and making available digital publications to support research and scholarship*

The archive could be centralised or distributed and it could be physical or partially or wholly virtual. The conceptual model in Figure 8.2 breaks this definition down into its constituent activities. The choice of activities was particularly informed by the literature (see Chapter 3), personal knowledge and Part 3 of the first round interviews (see Appendix 12). With hindsight, the conceptual model also could have included use of material in the legal deposit collections. While issues of use were discussed by participants, it was more

from the point of view of the rights issues associated with use rather than user needs and wants.

This chapter is divided into sections reflecting the activities within the conceptual model. Building the national archive involves the following activities: discover publishers and publications; select publications; provide publications (publishers) and assemble publications (deposit libraries). Maintaining the archive includes: store publications and preserve publications. Making digital publications available involves describe publications and provide access to publications. Create publications is another activity that has to be taken into account in the digital environment, since the way they are created affects the way they are provided, stored and preserved. The provision of metadata is covered by the create, provide and describe publications activities.

The conceptual model also includes the monitoring and evaluation activities advocated by SSM. Processes 10-12 relate to how well the system is working, whereas activities 13-16 are a more fundamental evaluation of the value of having a national archive at all. This model takes into account all the views expressed in the literature, even those sceptical of legal deposit as a mechanism to build a national archive or the need to collect and preserve publications at all.

The conceptual model was used as a framework for discussion rather than a perfect representation of the activities involved in legal deposit. It tried to take into account the different ways in which legal deposit could work including claiming, rather than receiving publications (the linkages between activities 1-4). It could, however, have also had a link between activity 3 Identify Publications and activity 6 Describe Publications because the data used to discover the existence of publications for selection purposes can also be used to catalogue publications.

Each section of this chapter includes a description of how an activity is currently carried out, where appropriate. This includes known problem areas for print material. This is followed by comments on the issues raised by introducing digital publications into the existing legal deposit arrangements. Each section ends with suggestions for overcoming problems if these have been identified.

The issues and suggested solutions are summarised in Appendix 15. There are also comments, where appropriate, on the degree of preparedness of stakeholders to deal with the legal deposit of digital publications.

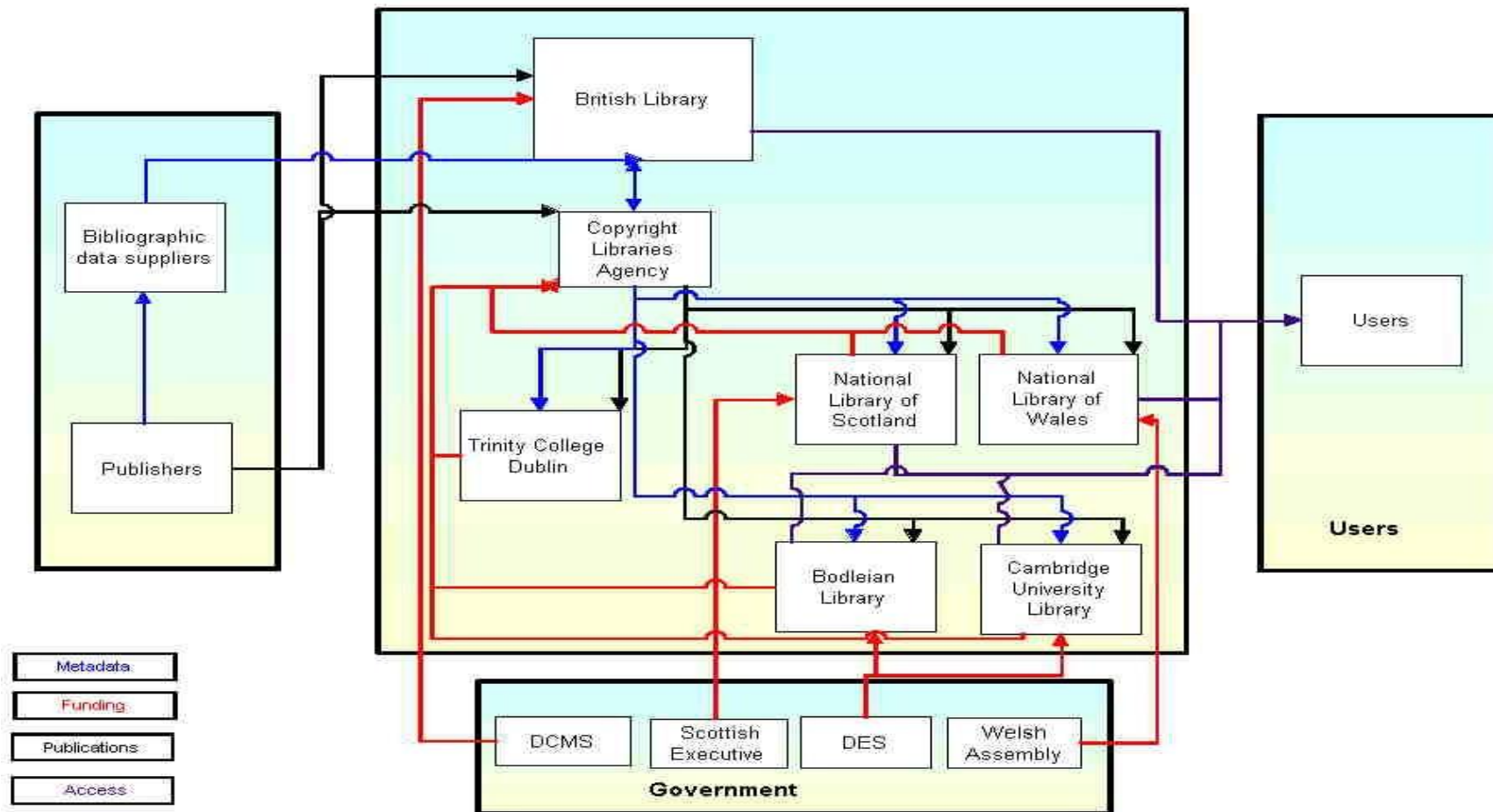
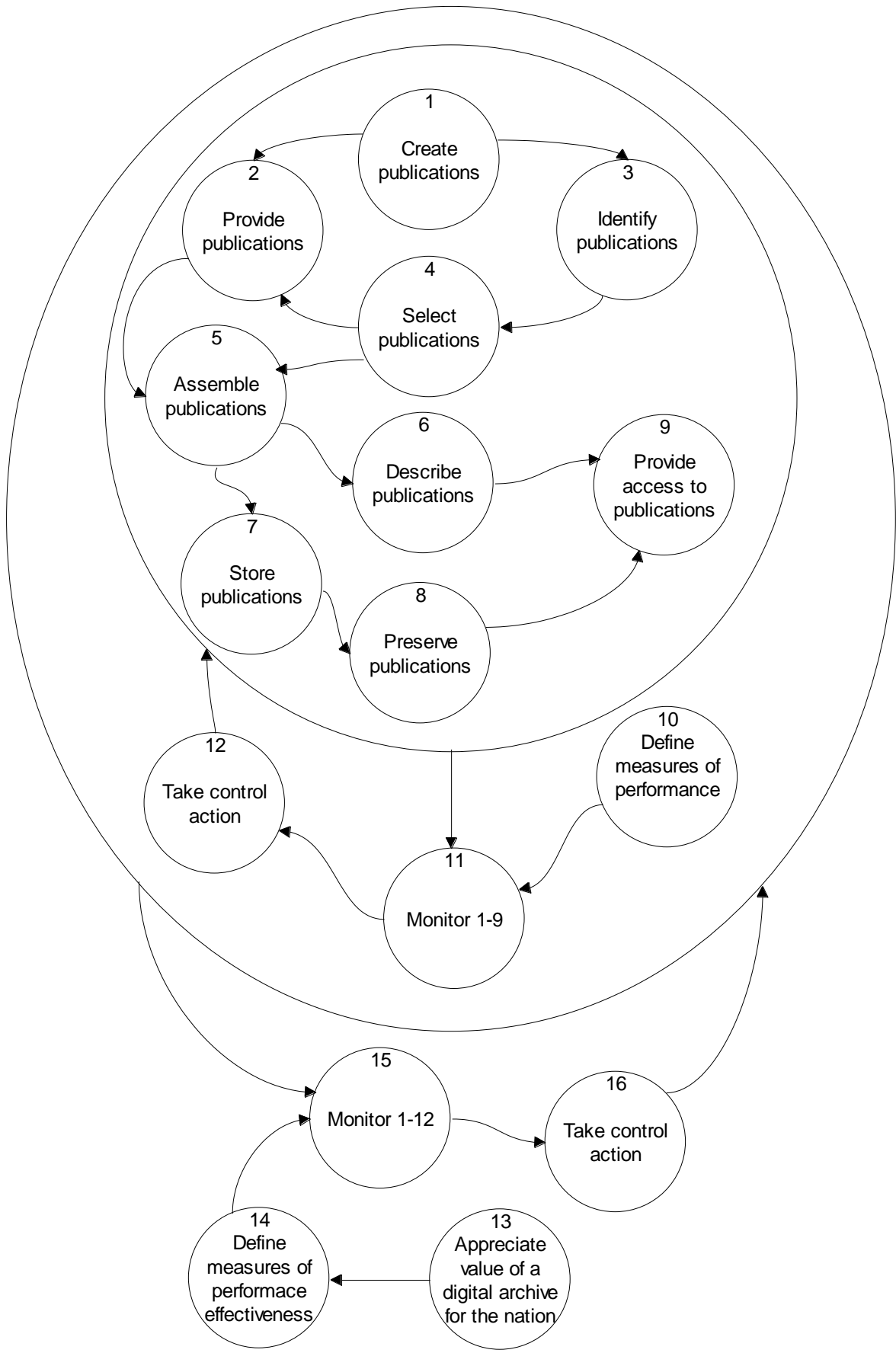


Figure 8-1 Model of UK Legal Deposit System



**Figure 8-2 Human Activity Model of a National Archive of Digital Publications**

### **8.3. Create Publications**

#### **8.3.1. *Print Publishing***

Interviewees from legal deposit libraries pointed out that the volume of material made available in print form does not seem to be diminishing. Digital legal deposit will be an additional burden on deposit libraries.

...we are still receiving an enormous amount of material in conventional form and trying to deal with problems of long-term access to that. And whatever we do on the electronic side the access to traditional material is going to remain as a problem and something that has to be dealt with. (Deposit library)

During the first round of interviews, a scholarly publisher commented that the volume of digital publishing was low because the market could not support digital only material at the time. At the end of the second round of interviews, a different scholarly publisher commented that for them, things were changing fast. At the time, all their journals were available in both print and digital form. Their customers did not want to continue taking paper in many cases. They were also being more selective about what they wanted to take on paper and what they wanted to take online only. There will be a point when it is more economic to supply journals in digital form and print will become a high cost optional purchase. Whether this scenario will apply to other publishing sectors is not clear. Other interviewees mentioned volume of material as being a potential problem for both libraries and publishers in the digital environment. The way digital publications and publishers are defined will affect the volume of material eligible for legal deposit.

#### **8.3.2. *Definitions***

Interviewees questioned what ‘published’ means in the digital environment. This perceived lack of definitions amongst stakeholders in legal deposits mirrors concerns raised in the legal deposit and digital preservation literature. Interviewees commented on whether Web material was published. If ‘publishing’ means making copies available to the

public (which is the definition in the UK Copyright Act), this means that Web sites are publications. Given the Web archiving activities being carried out across the world, it seems that legal deposit libraries consider Web-based material to be publications. Another issue with Web based material is the blurring of boundaries between records and publications. An example is material from public organisations. Material that would traditionally been part of a records management to archives continuum is now been made available through this channel. It seems from the literature (see Chapter 5) that legal deposit libraries are addressing this issue through working with national archives and other organisations in their Web harvesting activities.

So-called ‘publishing on demand’ was another example of a new way of making material available that raises questions.

... the underlying data is now our product and the output of that database is user defined. Now I don't really think that is what anyone sees as a publication under the meaning of the legal deposit part of the copyright legislation. It is a publication in that once it's output, it has a whole and is protectable under copyright. (Publisher)

Interviewees also raised questions about whether material that is made available through interaction with a user is also published. Examples include database driven Web sites where pages are created in real time and interactive television where people can open up several screens at once in different combinations. A final point here is the extent of publications. This particularly applies to hyperlinked material. The question raised was where the publication begins and ends.

Interviewees identified various problems for later activities, such as discovery and selection, caused by the lack of definitions at the creation and publishing stages. Even though stakeholders were only too aware of this issue, there does not seem to have been much progress towards solutions. The NLA (2001b) has acknowledged that it does not know how to deal with database Web material, but has taken decisions on where Web publications begin and end.

### 8.3.3. *Lack of Standards and Technological Dynamism*

Interviewees confirmed the lack of standards in the digital publishing industry that is identified in the literature. One component of this is quality control, which is part of the traditional publishing process, but not necessarily part of online publishing.

In the printed world there is a reasonably easy definition to the published work and you've got all the filters, the publishing process that goes on as part of the publishing industry, which helps as well. In the online world that's just not so and particularly when you get to online rather than just the digital world. What is a publisher and what is a published work? (Deposit library)

Interviewees saw the lack of, or rapidly changing, standards throughout the publishing process as a problem. Interviewees commented on CD publishing.

... just producing sort of one-off CD-ROMs or two or three CD-ROMs and every time they're doing something different. It really won't bed down until CD-ROM, or let's hope something else, becomes the standard form, means of disseminating information. (Deposit library)

Since this technology is in decline, this will be less of a problem in future.

There was also a perception that the rate of change is rapid, although some interviewees saw this as more of a problem than others.

Well all I can say is with ten years involvement in IT, I see no sign of it changing and slowing down. I see signs of it getting more and more diverse. (Deposit library)

One interviewee commented that in the case of broadcasting, the software developers are dictating technological change with the content creators following. This may also be the case in sectors of the digital information industry because the formats mentioned are also used in digital publishing.



For TV and radio there has always been movement to new formats, but the digital area is just a mess at the moment. This process has been relatively slow for TV and radio and the BBC has been at the forefront in making industry decisions on standards. This is not the case for digital. There are other players, such as Real and QuickTime. (Publisher)

One interviewee commented of publishers

I don't think they are aware of the sort of difficulties that they are creating for us. They're not doing it deliberately, but they're maybe not aware of the difficulties they are creating for themselves. It's a new world for them, for many of them, just as it is for us. I think that at a purely technical level, no one's really got it. (Deposit library)

Another thought that the big publishers would set standards because they have the resources to do so.

While legal deposit libraries could try and influence publishers on standards, interviewees thought that long-term preservation is not the prime motivation for most publishers. While publishers may be following standards set by other players, the paying customers of publishers may have more influence. The legal deposit libraries may have to work with other types of library and information service in order to stimulate demand from the market for digital information that is created in such a way that it can be preserved.

#### **8.3.4. *Formats and Distribution Channels***

The complexity of digital publications was discussed in Chapter 3. While publishers may hold content in an internal format, such as SGML or XML, they may distribute that content in different forms. For example:

... with ebooks, the user can choose what format they want and it can be as proprietary as they want, and it can actually have a lot of copyright content protection mechanisms wrapped around it. (Specialist agency)

As well as incorporating anti-copying mechanisms, digital publications may be dependent on retrieval software and particular technological platforms.

Other areas of complexity include the content itself. Digital publications may contain multimedia, which are in a variety of formats. While the formats for text, for example XML or HTML may be robust, it is not clear whether this would be the case for image formats.

*With MP3, sound is well ahead of vision. In fact, there is a feeling that sound has already developed standards and protocols that cannot be bettered. Video is more complicated because it incorporates images and sound. Graphics is yet another completely different thing and is not stable yet. (Publisher)*

Interviewees commented that digital publications that are based on print formats, for example scholarly journals, are also becoming more complex and more dynamic. Digital journals contain reports of primary research in the form of articles. These reports are a record of research achievements and are frozen in time. So the core content of electronic journals is really being added to over time rather than being replaced. This static content is also now accompanied by more ephemeral material, such as “news and views and magaziney parts and classified ads and display ads” that now can be updated and changed. So an ‘issue’ of a journal no longer has the same meaning. Information that originally accompanied the articles may disappear over time. Electronic books can be static, but they can also be changing. An example is reference works, which are updated over time. While interviewees did not specifically discuss these, there are new means of disseminating scholarly information such as preprint and e-print archives and scholarly work that can be changed over time as authors and commentators add to a particular piece, and it is not yet clear how these initiatives will fit into the legal deposit framework.

Publications can be added to, changed or deleted. Financial databases change all the time and very rapidly; unless the data is captured it may be lost. For some types of digital material, the publisher may not want changes to be archived, for example previous versions of a company’s Web site. There may be cases when digital information is not meant to have long-term value.

*\*Interviewee A:* The difficulty comes in not recognising the fundamental differences between print publications and the world of electronic publication. In electronic publication on the Internet, it's all about constantly updated, ever changing information. It's not the same and you have different considerations and different market requirements in putting together a product for that market than a book which is going to be there and reprinted for posterity for ever after.

*\*Interviewee B:* A lot of it is fairly short-term stuff, it doesn't necessarily have long-term application. (Publisher)

This is the view of a publisher driven by profit. Publishers are not necessarily the best judges of the future value of information. Newspapers are not meant to last, but the topical and other information they contain are of interest to future researchers.

Legal deposit currently applies to new material. In the digital environment, a lot of material is reused and repackaged into new products and services. News information, including images, can be made available in various forms by different organisations. The same information may be created by a press organisation and then used in printed newspapers, Web sites, broadcasts and commercial databases. While the information may be the same, value may have been added by the way the information is indexed, organised or presented. So there is a question of what is to be deposited and by whom. A related point made was whether digitised versions of non-digital originals should be counted as new publications and therefore subject to deposit.

Interviewees commented that the electronic book industry is developing slowly. One interviewee commented (in 2001) on the commercial aspects of ebook publishing as follows.

... many of the commercial arrangements for ebooks in the widest sense are still being formed. So on the one hand you have publishers like Harper Collins or Helicon, making plans to commercially publish appealing trade titles for handheld devices, whether they're Rocketbooks, Palm Pilots or whatever. And in this country, we're at least six to twelve months behind America in this respect. But now Harper Collins, for example, have plans to publish a hundred trade titles by the end of the year. Helicon much the same.

Now in the academic sector, big publishers like Taylor and Francis, Wiley, McGraw-Hill also at CUP, also have very ambitious plans, but they're less directed at handheld devices than at PC delivery. And consequently they're making commercial arrangements with, how can I put it, library- focused aggregators like Ebrary, Exlibris and Questia, and so on. And the commercial models underpinning those arrangements are selling textbooks by the chapter, charging students only when they print. (Specialist agency)

While online publishers may disseminate material directly to users, many do this through intermediaries. For example, a small independent publisher has

... been contacted by the big companies that are going into ebooks of one sort or another, NetLibrary, Ebrary and so on and we've taken those quite seriously.

Intermediaries, such as hosts, may add functionality to products. The same product may be offered via different hosts, so the functionality a user experiences may depend on the supplier. Online delivery of material may mean that users never physically own it, although

Other customers prefer to have material loaded locally. Local load customers are fed material from the central servers. They then have to take responsibility for the user interface and context of use. (Publisher)

#### **8.3.5.            *Create Metadata***

Interviewees from legal deposit libraries reported difficulties in obtaining both conventional bibliographic and technical metadata from publishers under the voluntary scheme. Their view is that the libraries need to get as much metadata from the publishers as possible and that they also need to automate the process of metadata generation in order to bring down costs because manual production of metadata is expensive. The need for publisher-created metadata is reflected in the literature (see Chapters 3 and 4) and will require good relations with publishers.

So much of the metadata, the provision of the materials, the format in which they are provided I think comes essentially down to a good relationship with publishers.

(Specialist agency)

There are a number of problems with this ideal.

Publishers' ability to supply consistent and standard metadata, their ability varies hugely. (Specialist agency)

There has been a historical lack of data standards in the publishing industry. Agencies supplying bibliographic services to the book trade and libraries have had to carry out a lot of work to produce the sort of records needed by their customers. One such agency commented that it has been impossible to accept data in digital form from publishers. It is actually simpler to accept hard copy data and re-key it from scratch than to try and modify electronic publisher data. Another agency confirmed this view

Several interviewees commented that not all publishers are aware of the benefits of metadata or are in a position to produce what libraries may require. Smaller publishers, or organisations whose publishing activities are not their main focus of activity are less likely to be able to comply with library metadata requirements.

The bigger publishers, who are likely to be more organised and perhaps developing their own digital stores, will need the metadata they could be supplying to the libraries for their own purposes. It might be more difficult with the small to medium publishers.

(Specialist agency)

Publishers' processes were also mentioned as a problem.

*\*Interviewee 1 ... what we want is to do is to try and get as much metadata as possible with the items themselves. And the trouble is that the metadata is created at a different stage from the stage at which the stuff is shipped out either to the customers or the legal deposit libraries. \*Interviewee 2 At a few stages. It's created by the publishers and by the developers and by people who are taking all the content and putting it into something handheld and then it goes off to the shippers and the shippers know nothing.*

(Deposit library)

Interviewees discussed the potential of ONIX International. An interviewee from a bibliographic data agency commented that while the ONIX scheme includes a lot of purely bibliographic data, it does not cover all the metadata needed by libraries, such as classification information. The bibliographic information will not be authority controlled to library standards. ONIX also includes additional information that is not relevant to deposit libraries. Other research has shown that ONIX does not contain all the data needed for preservation purposes (see section 4.3.2). However, some of the interviewees thought that if the deposit libraries work with the ONIX developers on the addition of preservation-related metadata, agreements on vocabulary could be made. The BL thought that it may be practical to receive ONIX data from publishers and carry out any necessary conversions. At the time of the interviews, the British Library (BL) was in a dialogue with the ONIX developers.

Several interviewees mentioned that there is currently little incentive for publishers to create the metadata that deposit libraries would need to manage digital publications. The ability to provide metadata will depend on the amount of detail needed. Several interviewees mentioned the possibility of developing tools for small publishers for metadata capture. Apparently there is a precedent for this; the Public Record Office (now National Archives) has a tool set available for government departments. At the time of the interviews, the BL was looking into developing a tool for publishers. The BL has now gone ahead with this by supporting the development of the Simple ONIX Editing Tool (SIMONE). One interviewee pointed out that the deposit libraries would have to listen to what the publishers say they need; it has to be partly driven by the needs of the publishers. The publisher processes, including the use of third party providers, may introduce complexities. Another interviewee, a metadata expert, thought that most of the data needed by the deposit libraries is likely to be technical. A lot of the publishers generate their metadata from their digital workflow, so will probably have a lot of that information. Publishers already have to pull data from different parts of the operation to put into a central system from which they could output their ONIX data. There are apparently very few integrated publishing systems at the moment.

#### **8.4. Discover Publishers and Publications**

While the BL is entitled to print publications sent to it automatically, in practice some publishers do not comply. The other deposit libraries have to claim material. Therefore, there is a need to identify material published in the UK to allow the libraries and the CLA to claim material.

Figure 8.1 includes a simplified representation of the flow of metadata between the different stakeholders. Publishers provide bibliographic data to bibliographic data agencies. The BL LDO and the CLA use the products of these agencies to discover new publications. Bibliographic Data Services Limited (BDS) provides Cataloguing in Publication (CIP) Data to the BL and the Copyright Libraries Agency uses Whitaker's data. The LDO and the CLA share this data.

The aim of the CIP programme is to provide as much pre-publication information for new books as possible. At the time of the interviews, data was released six to eight weeks prior to publication, in accordance with the CIP contract with the BL. The data was delivered on a weekly basis. BDS provided at least 50 000 records per annum as specified by the BL. The data is used in the BL's bibliographic products and is used as the basis of records for the BL online catalogue as well as for legal deposit purposes. BDS also scan information on changes in the publishing industry including personnel movements, ownership changes and new players and make contacts where appropriate. BDS also analysed the coverage of its sister company Book Data and identified any gaps in its own coverage. However, the aim is not really to identify every new UK publication because the main target audiences for BDS products are public and academic libraries rather than legal deposit libraries.

Whitaker, now part of Nielsen, also received a great volume of information from publishers on paper. Larger publishers are increasingly sending digital files. Whitaker also had various measures for working out what they are not receiving, by using their BookTrack system, which analyses sales data. Whitaker contacts the publishers of any titles included in the sales data that is not already covered in the bibliographic database.

As well as claiming, the CLA receives many unsolicited publications from publishers as they often send out six copies automatically. The CLA passes bibliographic data to the

legal deposit libraries for use in cataloguing. The libraries also sometimes identify material through publisher catalogues and other sources and send requests to the CLA.

There was a perception that there will be a problem identifying publishers and publications in the digital environment.

At this stage we probably haven't identified more than a small proportion of the publishers of electronic publications. (Deposit library)

... it's a question of knowing what is published and that again is one of the difficulties that we face. (Deposit library)

For the small publishers, individuals, the difficulty there is going to be finding out who has published what. (Specialist agency)

... anyone can publish now with electronic publication, many people are doing just that and I think they have no conception of legal deposit, nor do they have any conception of any other aspects of publishing such as copyright in some respects. So there are all sorts of difficulties there. (Deposit library)

The CIP data supplied to the BL at the time of the interviews only covered print publications, but the supplier was willing to consider supplying records if the CIP contract were to be extended. The CIP data was supplied as a spin-off from BDS Ltd's main business and there was not yet any commercial incentive to supply data on digital publications to customers. While Whitaker claimed to cover digital publications, not all digital publishers were part of the system; only a small proportion of electronic publishers had been identified. The output of new players is also not tracked and is difficult for deposit libraries to discover. Some publishers email the CLA notification of new web sites and cessation of publication of print material. On the other hand, the BL sometimes was only finding out by accident about material that was published in hard copy and is now only available digitally.

So there are gaps in the existing system used to discover publications. The problem of discovery was examined through the voluntary scheme, but no solution identified. There was a comment from a deposit library that there is a need to talk to publishers about how



to get digital publications into the British National Bibliography, even if they are not all collected by the deposit libraries. One view was that legislation will mean that formal identification channels will be set up. There was one suggestion that may improve discovery of publications. Electronic item release messages to publicise availability may be incorporated into ONIX. These messages could be sent to legal deposit libraries, either directly or indirectly.

It was pointed out that for non-commercial material, it is likely that the libraries will have to carry out some proactive work to discover publications. This may be easier for some of the libraries than others. For example, National Library of Wales (NLW) staff are very close to the small Welsh publishing industry and already proactively seek out publications.

When it comes to harvesting Web publications, there will be difficulties for the libraries with geographic responsibilities in the UK because there is only one .uk domain and no separate designations for Scotland and Wales. One potential solution would be to manually search the Web rather than harvest material, but this would be resource intensive. The NLW had developed guidelines for identifying web sites through its OCLC CORC project work. Since there is no Welsh domain, the Library used directories of organisations. The NLW was also looking at the example of Scandinavian countries: the Scandinavian deposit libraries automatically interrogate telephone numbers included in Web pages.

#### **8.4.1.            *Identifiers***

There are well-established systems for identifying print material, including the International Standard Book Number. The ISBN is also used for handheld digital material. Given that the same underlying digital content can be issued in different versions and that works can contain references to other material, there has to be a way of uniquely and precisely identifying different manifestations of that content.

... if we envisage that there's going to be a supply chain for ebooks, in the widest sense of ebooks, then there has to be a way of uniquely identifying each and every manifestation of an ebook in the supply chain. (Specialist agency)

Interviewees were not clear about how this is going to be achieved in the digital environment, but the trade was aware that it is an issue that has to be solved.

Now there is considerable debate and quite a lot of friction about whether the ISBN is the right identifier, and if it is the right identifier, how you apply it or whether, in fact, you need an actionable identifier like the DOI. I think views are becoming polarised, not least because of the Andersen Consulting report last year that the AAP sponsored. But what is so striking in this respect is the supply chain's needs demand an immediate resolution of this, not some incredibly leisurely response. If you're Amazon.com or BarnesandNoble.com in the US, you're already selling considerable volumes of ebooks. And their systems cannot cope with anything other than ISBNs. So there's an undeniable tension in this area at the moment. (Specialist agency)

So while the deposit libraries may have problems getting the metadata they need from the publishing industry, the industry does have an incentive to decide what sort of identifier should be used. However, getting all digital publishers to use an industry standard is another matter. For example, it may be difficult for those that have never been involved in the traditional book world or who are not publishing on a commercial basis. These publishers may be unaware of the need for unique identifiers and may not have any incentive to apply them.

As discussed in section 3.2.1, the Digital Object Identifier is an actionable identifier that overcomes the lack of persistence in URLs and the need to uniquely identify different versions of material. At the moment, DOI use is almost entirely confined to journals because this was its first application area. An expert in identifiers and metadata felt that DOIs could help in the identification of different versions of material for discovery purposes. Initially, there was some confusion amongst publishers about how to use identifiers in these cases. So the idea of identifying the underlying work and manifestations as expressions of this work of this came about. This approach has apparently worked well. According to the expert, if the identifier can be also be linked to structured metadata, the discovery of mainstream publications at least is likely to become easier for legal deposit libraries. There would have to be a generic structure to cover all types of material and provide interoperability between the different players in the

information industry. All metadata elements need to be precisely identified and relationships well described. The DOI coupled with metadata such as ONIX may be a potential solution here. ONIX is an implementation of the INDECS data dictionary. The data dictionary at the moment is set up for selling information, but other communities can add to it, including libraries wishing to preserve information. This would allow different communities to keep their own metadata schemes.

As long as there was a mapping from these schemes to the DOI metadata kernel, the system could be interoperable. An interviewee mentioned that there was a proposal to map ONIX to the metadata requirements developed by the BL.

The well-established infrastructure for the identification and discovery of print publications does not yet exist in the digital environment. There are some potential building blocks emerging from the book trade, but there is no coherent system. It does seem that legal deposit libraries could benefit, but it would take some collaborative work with the publishing industry. Extension of legal deposit may provide the incentive for publishers for collaboration. None of the interviewees expressed a view on the need for international consensus on this topic.

## **8.5. Select Publications**

Given that there is general agreement that the purpose of legal deposit is to maintain an archive of the UK published output, it is reasonable to assume that there is no need to select material other than by country of publication. While publishers may complain that the CLA may make blanket claims for material, some of the deposit libraries themselves say that they do not take everything they are entitled to claim.

There seemed to be an acceptance that at least some of the deposit libraries will have to be more selective than in the past. One interviewee commented that:

You are moving towards a situation, which is much more akin to what archivists are dealing with than the world of libraries and traditional ideas about collecting and comprehensiveness of collecting within professional librarianship. To archivists, the

idea of selection is utterly crucial to the whole profession. It's something, which comes naturally; it's bred in archivists that you don't collect everything. . . . And I think we need to be moving toward that kind of archivist's view of the world in the electronic era. In other words, the element of judicious selection is going to be much, much more important than that of comprehensiveness. It's no longer relevant. But that's within the context where there's a vast spectrum, isn't there, of electronic material publicly available from near print materials to Joe Bloggs' personal web site. (Deposit library)

The root of this perception of the need to be selective is a lack of resources and perceived volume of material to be dealt with.

I think it's such a big task that the real work that needs to be done is to identify the information that is really, looks really, even if you might have got it wrong, it's better to try and decide what information would be of value in the future. And then the question that then has to be asked is this information going to be preserved in any other way? If it is, then you have to say fingers crossed and then with a fair wind that will exist two hundred years hence. If it's not, or if it looks uncertain, then I suppose you can say that then the state has a responsibility to try in some way to preserve that information.  
(Publisher)

The selection issue is also influenced by political developments in the UK, namely devolved government.

I think when we come onto electronic materials there's either a danger or an opportunity, whichever way you look at it, of the national library becoming more selective geographically in its collecting policies. In other words, there's a question about the extent to which we involve ourselves in being an archive for the whole of the UK. I don't think the debates' even started on this yet. (Deposit library)

### 8.5.1. *Selection Criteria*

Two of the deposit libraries said that they were likely to use existing selection guidelines for digital material. At the moment, there is some selection of print material. This does not just apply to the Agency deposit libraries. An example that was mentioned was newspapers: even the British Library does not collection every edition of every newspaper published in the country.

The voluntary scheme for non-print material includes a trigger point for the deposit of CDs based on the number of sales. Legal deposit of digital material may initially start with deposit of the same formats as the voluntary scheme. There is also material that is now available only in digital form that was collected when it was available in hard copy.

The problem with the combination of a long-term responsibility to preserve and the need to select what is preserved is that the librarians of today do not necessarily know what the researchers of the future will find valuable. One potential criterion mentioned was quality.

There are issues as to whether that's got to go through a quality control, peer review process to warrant its preservation. (Specialist agency)

However, future researchers will not just be looking for peer reviewed scholarly material. Material, such as the Web sites of individuals, may not have undergone any quality process, but still be of value.

Harvesting of important Web sites is an area that is often ignored, but it may have a greater long-term social and cultural value to do this than for keeping commercial products. (Deposit library)

Due to the potential volume of material involved, assessment of quality by the deposit libraries may have to be a condition of selection for non-commercial material. However, another approach could be making a judgement of how likely material is to survive

without intervention by a deposit library. The non-commercial material may be least likely to survive. If it has no commercial value, there may be no motivation for the publisher to maintain material that is no longer current. There was a suggestion that at the moment, local and central government, clubs and societies are less likely to archive Web sites and data. Another point made was that aspects of publications that may seem ephemeral, for example adverts, should not be disregarded because they may be of value in the future.

Other selection criteria could be based on use and a hybrid approach could be taken. An example would be comprehensiveness for scientific material; future researchers need access to previous work. On the other hand, material that may be useful for social history may be collected on a more selective basis. The aim would be to have a representative sample.

Yet another suggestion was that material could be collected comprehensively but preservation of material collected would be prioritised. Some material may be higher priority than others, but the material that is not prioritised is not completely lost. In the author's opinion, material that is just collected and left on some storage media may be as good as lost. One deposit library commented that it is likely that the material for each type of publisher or each publisher will have to be looked in terms of whether it is worth archiving and the difficulty and cost of archiving.

If the deposit libraries are going to be more selective, there is the question of what the role of the publishers will be in preservation in the digital environment. One deposit library mentioned that it catalogues Web sites it considers important and is considering whether to preserve catalogued material or leave it to publishers. Even if the publishers were willing to preserve their material, it is not clear how they would decide what to keep. An example given was newspapers. A UK newspaper publisher of many local titles is archiving only unique content, not individual titles. A lot of content is duplicated between the different titles.

A final point on selection is that more selection of material is likely to mean more human intervention. Human intervention introduces costs that have to be considered against the costs of dealing with large volumes. Another point was that over time, the cost of gathering and processing digital material may decrease if more of the activities become

automated. The points were acknowledged, if not fully addressed, in the costing modules discussed in 3.7. If costs go down, selection may not be necessary.

## **8.6. Assemble Publications**

There are established procedures for getting printed objects to legal deposit libraries, accessioning, cataloguing and shelving material. There are new challenges for both the deposit libraries and publishers in the digital environment. As well as the problem of how to identify who is publishing what, there are the questions of who should deposit material, and what, when and how should they deposit. Libraries and publishers have to take decisions on these issues and then develop policies and procedures.

### **8.6.1. *Number of Copies Deposited and Shared Acquisition***

At the moment, publishers deposit one copy of a new printed publication with the BL LDO, and up to five copies with the CLA. Some publishers do not follow this pattern, for example, they may deposit directly with the other legal deposit libraries or deposit six copies with the Copyright Libraries Agency. Comments from interviewees suggested that this led to a lack of coordination in the system, with the Agent sometimes claiming for material that had already been sent directly to libraries.

Given that the publishers would prefer to deposit fewer copies of material, for digital publications the deposit of fewer than six, or even one copy with networked access seems to be a possible way forward. At the moment, each library owns the physical items in its collection. The idea of shared ownership of digital material was raised: the libraries would share ownership of a distributed legal deposit collection.

It was not clear from the interviews exactly how shared acquisition of material would be handled. The deposit libraries mentioned one or two copy deposit of digital material. At the same time, two of the libraries mentioned that they thought digital material on physical carriers would be deposited in several copies, the same as print material. This might not be necessary if the content of physical media can be copied and networked.

The possibilities include the BL becoming a central store for a digital material, with networked access for all the other libraries. The potential problem with this is that the other national libraries, at least, may find this unacceptable because of their national roles.

I think we all feel that if there is a central warehouse or series of warehouses nobody is really going to be too bothered about where the bits and bytes are kept. There is an obligation of a national kind for the national libraries of Scotland and Wales that is not felt by Oxford and Cambridge about our obligation to secure and ensure the preservation of particularly Scottish or Welsh material. So I don't think we would necessarily agree to a single mega-warehouse in Milton Keynes, but I think we would be quite happy to see quite a lot of stuff in Milton Keynes as long as were able to retain ... the material we felt was part of our patrimony. (Deposit library)

This comment from the first round of interviews implies that that library would want to have physical ownership of material from their country. However, in the second round, the same interviewees commented that for them having primary responsibility does not necessarily mean that material has to be housed in their library. For online material, it may be more a case of knowing about the material and making it part of the collection in a virtual way. Presumably, this would mean including the material in the library catalogue. In fact, there was a suggestion that the rights of these libraries may change in future in that they may be given privileges similar to the BL with respect to Scottish and Welsh material and retain the right to claim other material. It is not clear what the role of the BL would then be under such an arrangement.

While some interviewees thought the BL would act as a central store for all materials with all activities and expertise being centralised, others said that the BL wants a more distributed system. Interviewees from different parts of the BL expressed both these opinions, although more senior members of staff talked of a distributed system with the BL taking a large proportion of the load. The question then is how responsibility for the storage and preservation of digital material should be allocated.

By their own admission, the UK legal deposit libraries do not have a great track record on shared acquisition of print material. It might be that sharing responsibility for digital material is easier than for print because networking would take care of access problems.



In some ways electronic deposit might actually ease the situation in terms of the inconsistencies or the lack of ability to rationalise which often comes with printed materials in that, we're talking in effect of depositing one copy and then making it available electronically across the system. So electronic deposit will, I think, from a political point of view be easier to handle. (Deposit library)

If legal deposit libraries were to take responsibility for parts of the national output, there would be problems with identifying what constitutes Scottish or Welsh material. Under the current system, material has to be published in the UK. Material by Scottish or Welsh authors or Scottish or Welsh material that is published outwith the UK has to be purchased. There is also a problem with defining what 'Scottish and Welsh material published in the UK' means. It could mean material published by Scottish or Welsh publishers. However, Scottish and Welsh authors may have English or Northern Irish publishers and publishers in these countries may publish material of Scottish or Welsh interest. If the BL were only to take responsibility for English material, the same problems of definition may arise.

Other questions raised about the distribution of responsibilities included the role of the three academic libraries. Interviewees at TCD thought that they would still act as the interface between the Irish and UK systems and as a conduit for Irish publications to British legal deposit libraries. Other suggestions included Oxford and Cambridge taking responsibility for academic material and the other libraries taking the rest. Alternatively, responsibility might be by physical proximity, with libraries dividing up the national output by geographical location of the publisher. The issue of distribution of responsibilities between the legal deposit libraries seems complex and will require much work to resolve. The libraries seemed to be discussing this, but were at an early stage in their discussions.

... sooner or later it's going to have to be faced. Certainly we have not faced up to it yet. It's very early days. It's impossible to see what the landscape's going to be like, whether there is going to be legislation. So I think we can be forgiven for not having grasped it so far. During the next five years it's going to have to be looked at because it's going to affect the whole outlook of the library. (Deposit library)

Without a central receiving point, publishers would have to send their material to several different locations, rather than the current two. Publishers who commented on this point preferred one point of entry into the legal deposit system. While some of the deposit libraries also preferred a single point of entry to a distributed system, they pointed out that some publishers might prefer to deposit with a particular library. It is interesting that at the moment, for email electronic journals each of the libraries has a mailbox and the libraries are bypassing the CLA. It seems that the CLA will only be used for print material, but this is not certain. If this is the case, then the complexity in the system will be compounded with different collection operations for different types of information.

It's hard to say, to be quite honest what the role of the Agency will be because in the end it might be that a lot of the stuff will simply come electronically to whichever library is looking after it. So it might completely bypass the Agency. (Deposit library)

#### **8.6.2.            *Who Should Deposit What***

The thing is in the digital environment lots of people are creating digital resources and it's less clear who the publisher is and on whom the responsibility is to actually deposit that material. So, whereas we know in a print on paper environment there is a recognised list of people who are actually publishing books and periodicals. In an electronic environment literally everyone is a publisher. So there are issues in terms of even if there were legal deposit how you actually enforce that. In other words, are you simply enforcing that in terms of commercial products, which can be fairly easily identified, or across the board? (Specialist agency)

There was a suggestion that the deposit libraries have not yet come to terms with new types of publishing.

There are a number of issues related to responsibility for deposit that will have to be resolved. Under the current arrangements only new material should be deposited. This is difficult in the digital environment because so much digital information is reused. So there is a question of what is to be deposited and by whom. If there are parallel paper and online versions, there is the question of who decides how substantial differences are and whether only one or both versions are to be deposited. One suggestion was it should be up to

publishers to deposit the fullest version. Aggregated information services contain information that originated elsewhere. A related point is whether digitised versions of non-digital originals should be counted as new publications and therefore subject to deposit. There are already problems with depositing all versions of newspapers and the libraries do not take all editions of a title. Depositing the publishers' final digital archive version rather than the print copy may get over this problem.

There were suggestions that information distributed on physical carriers would be easier to deal with because it would fit more easily into established procedures. Working out the extent of publications is seen as a problem in the digital environment, because of extensive linking between pieces of information. In addition, one interviewee questioned whether Web-based material should be deposited at all because

... the web is very much a real-time experience rather than an archived historical sort of experience. (Publisher)

Interviewees asked this question for other categories of material, including financial or other numeric data, ephemeral material, and information created 'on the fly'. The dynamic nature of much digital information was also seen as problematic. For example interviewees questioned how frequent changes could be dealt with. It is not just a question of frequency of deposit. If databases are updated rather than changed, then there is the question of whether the updates rather than whole database should be deposited each time.

Another problem that causes repercussions at the preservation stage is what aspects of a publication are to be deposited. Digital publications can be much more complex than print. The questions here are what aspects of digital publications can and should be deposited: the basic information or the package as it is made available to the public, that is the look and feel and functionality or other added value added by the publisher or distributor. It may not be desirable or even possible to deposit the packaging. Software may be licensed from third parties or bespoke. For some products, it may be the packaging that is the selling point, the basic intellectual content is elsewhere, probably in the legal deposit libraries. These issues will all have to be dealt with in the forthcoming legal deposit Regulations.

Another question was whether and how other preserving institutions should be working with legal deposit libraries in the digital environment. It may no longer be a case of cooperation between deposit libraries, but also between a wider range of institutions to ensure the digital information heritage is safeguarded.

### **8.6.3.            *Lack of Established Processes and Procedures***

While there is some lack of coordination in the current legal deposit system, there are established processes and procedures in place. Publishers send or are asked for material according to the timescales set down in legislation. There are established conventions for the provision of bibliographic material accompanying material and cataloguing. Staff have the requisite skills to deal with material. This is not yet the case in the digital environment.

Publishers were not ready for digital deposit.

It's very easy because the economies of scale of, just send one more book out of a thousand you're sending out. Everything's geared up; it has been geared up over years to economies of scale. Whereas everything's still wild west style isn't it really still with electronic stuff? (Publisher)

... the analogue system is well established and there are protocols and processes for dealing with it. Whereas, the electronic is different. Within most companies it's not even, I think we found that in the analogue it could almost be the despatch manager who would be sending something out. Or the promotions manager would be alerting the despatch and so on. There doesn't seem to be that process yet within publishers. (Publisher)

This also seems to be the case for the deposit libraries and the voluntary scheme was seen as a way of working out procedures and dealing with issues.

#### 8.6.4 *Provision of Publications*

Assuming that material will still have to be physically deposited, it seems that the most problematic material to deal with will be online material, especially dynamic online material.

... there are practical difficulties, particularly as you move from a single discrete object to a changing service as to what you deposit. So the online dimension as opposed to discrete CD-ROMs and so on is tricky. (Publisher)

This view contrasts with the policies of the National Library of Australia, which seems to prefer the online version of digital publications.

While electronic journals may be the first category of online material to be included in legal deposit provisions, even here there are some issues that have to be dealt with. There is an issue about whether and how the more ephemeral content of electronic journals, for example conference announcements, news items and editorial material, should be deposited. This type of material is likely to be regularly replaced, and decisions have to be taken as to whether this material is of value.

Although that core scholarship issue is clear, those papers are clear. All the other stuff - what do you want to capture for deposit, legal deposit, and how do you capture it? (Publisher)

In terms of the history of science, you don't know what somebody's going to want to look at in a hundred years time. It might help to show the way science develops by the way conferences went. And therefore those sorts of tat end bits now could become somebody's valuable data in fifty year's time. So I guess for completeness, it would be nice if they could be kept as well, but I think the priority is the data and intellectual content. (Journal editor)

The real problems come with how and when to capture frequently updated products. Real-time products were identified as not being suitable for deposit by their very nature.

However, the question was raised whether periodic capture would be useful to provide time series data, for example for financial data. Although not mentioned by interviewees, some companies, such as Reuters, already offer this to their customers. The question is then why such companies could not simply deposit what they create for customers. There are also questions as to the frequency of deposit.

... are you trying to actually get legal deposit on a continuous basis or a snapshot basis? (Specialist agency)

One interviewee suggested this decision might have to be a pragmatic one. Factors underlying decisions may be the time and resource implications and the volume of material to be dealt with. These issues are discussed below. Commercial factors may also impinge on the timing of deposit of static as well as dynamic material, especially in the case of high-value material. The voluntary code has provisions for the deposit of material after a certain number of sales. An alternative is to deposit at the time of publication but embargo access for a period of time. These alternatives are discussed below, but one point here is that delaying deposit may mean that material never gets deposited at all and ends up being lost.

You'd have the problem from the point of view of the public good, the danger of that sort of situation is that rather than requiring the publisher to deliver it at the point of most likely compliance, which is when they're sending out in the print environment all the free copies, all the early copies that are going to people right left and centre, and it is part of that machinery so it is likely to happen, you would be asking them to pass it across when it's at its last gasp and the publisher might be at their last gasp, the product might be at their last gasp, it's less likely that you'll actually get it into the system. So there's that danger. It's also not just a last gasp issue; it's also a change of ownership issue. So that you, ten years later it might have gone through three different hands, how do you keep a grip on that? (Publisher)

Another difficulty is what to do about Web pages that are created on the fly. This problem is similar to that of on demand publishing mentioned earlier. Is the underlying database to be deposited? Another issue is whether publishers send the material to deposit libraries in

some way or whether the libraries pull material from publisher servers. At the moment the libraries are being sent material on discs and also email attachments. Some of the libraries were planning to experiment with acquiring Web-based material, but they seemed to be taking a selective approach and asking publishers' permission, rather than attempting Web harvesting. However, the BL at least would like the latitude to be able to harvest material deemed appropriate for the BL's collections from the UK produced part of the Web. This is not likely to be all UK-based Web material and it is unlikely that Web producers will have to "deposit" material.

#### **8.6.5.            *Provision of Metadata***

Print publications contain bibliographic information as part of their packaging. The libraries have found that material received under the voluntary scheme often lacks conventional publishing information, such as imprint. There is also a lack of overall technical information and metadata. It has also been difficult to obtain this information from the publishers. There have been some bibliographic identity issues. An example is that the same CD ROMs may come with different packing or accompanying material.

While it is not unknown for the information on the inside and outside of books to differ, it is a simpler matter to find the information and rules for which information to use in cataloguing. Doing this for a CD ROM requires installation.

It is not clear whether the deposit libraries would want metadata to be embedded in digital publications or sent separately. Generally, ONIX metadata is sent separately from the publication. There is no reason why some cannot be embedded in the publications, but the publishers are unlikely to see any great benefit in doing this. Using an actionable identifier such as a DOI should get over the problem of metadata separate from the publication.

As mentioned above, publishers have not historically been good at supplying good quality bibliographic metadata. If the publishing industry adopts the ONIX standard, this situation may improve. However, ONIX might only be used for commercial publications. Non-commercial publishers may not be interested in supplying ONIX records. The deposit libraries would like to get as much metadata as possible from the publishers and ideally

they would like to be able extract it automatically. Even if ONIX does take off, it may only be of limited use to libraries in its present form. ONIX is aimed at the publishing world and does not take into account library requirements, both in terms of metadata for discovery and long-term management purposes. The British Library has worked out a scheme for its metadata requirements.

The big barrier to publishers supplying the metadata that libraries want is lack of incentive, especially for commercial publishers. Several interviewees commented that there is a need for publishers to see some sort of benefit to supply anything more than they need for their own purposes. One large international publisher commented that it might work with the British Library on the provision of metadata, but would prefer that libraries came with a unified approach. Publishers and libraries should both have a single standard for metadata, which is better than different institutions inventing their own solutions. A worldwide approach would be the best. The rationale for this comment is that while this may be difficult and slow to achieve it would avoid duplication of effort and tying up of resources. While large publishers may be aware of and willing and able to create and supply metadata, this is less likely to be the case for smaller publishers.

In terms of skills, authors and editors know most about what the publication is and how it is used, but they do not have the skills to make it retrievable. Since not all publishers will be able to create metadata, there was a suggestion that there is perhaps an opportunity for a service to sell to libraries and/or publishers here. Possible service providers could be the likes of OCLC, abstracting and indexing services or bibliographic data agencies. Metadata is discussed further in section 8.7.5.

#### **8.6.6.            *Processing Publications***

The legal deposit libraries would like to automate as much of the accession and processing of digital publications as possible. It became clear in the interviews with the libraries, that they have not yet decided how they will deal with digital publications and to what extent the processing operations for different types of material will be integrated or separate. At the time of the interviews, they had not yet had much practice, but they had already identified various problems with processing digital publications. The libraries thought that



there will be more steps in processing digital material, so it may well take longer and human intervention is expensive. The metadata aspects were specifically mentioned as being expensive.

We've done some costings and it seems quite clear that the big cost is at that point, is at the metadata level. The actual technical aspects of storing the material are relatively cheap. Or not as expensive as the handling of the metadata. What we're trying to do is to find ways now of driving down of what CEDARS calls ingest. So that as much as possible is actually embedded in the item itself so that when it's taken into one of the libraries the material is already there, the information is already there and doesn't have to be either created by a person which means it's expensive, or even extracted by a person, it just flows into the system. But when we shall reach that point is a question.  
(Legal deposit library)

If publications are not going through the CLA, then the agency libraries will not receive CLA records. There is an immediate problem in that digital materials may have in-built copyright protection mechanisms that prevent access for library staff who want to catalogue or otherwise process material. One library had attempted to automate the receipt of one type of publication. This had gone well, but was not without problems, caused by lack of semantic structure, lack of metadata (including data of publication) and inconsistent use of HTML. The library had a back file of material and the indexing was taking a long time even though it was automated because some human intervention was needed.

It is not clear whether centralised or distributed processing would be most cost effective. An argument for centralisation was that

Ideally we would want the central location not only to coordinate the collections of material, storage and all the rest of it, but the creation of the metadata and the distribution of that to the stakeholders for inclusion in the local lists. Resources are not deployed in that way at the moment ... I would have thought it would make sense to concentrate all that activity at one central place and just distribute the records outwards, not just for electronic material but for printed material as well ... There's much more harmonisation being developed now, so I think in practical terms there's less of an

obstacle for using centrally produced records in the local situation than there perhaps has been in the last twenty or thirty years. (Deposit library)

It may be simpler if only one site was dealing with processing and technical installation of the digital publications. At the time of the interviews, cataloguing of monographs was distributed with each of the libraries cataloguing a section of the legal deposit intake. In the past, the BL carried out all cataloguing and distributed records to the other libraries. While the BL may be the most sensible candidate for any future centralisation, the CLA or a merged CLA/LDO operation were other possibilities suggested.

The libraries were investigating cataloguing of digital publications, setting up working groups and recruiting staff with digital cataloguing skills who could train others. The libraries also mentioned taking part in wider initiatives in this area. One library mentioned its participation in the OCLC CORC cataloguing project.

... because we've been working on cataloguing web sites we've acquired experience of cataloguing online material as well as CD-ROMs. So we have cataloguers who have been learning the techniques of metadata and so on and so forth. So we'll try to develop that. (Legal deposit library)

There was also some activity on developing cataloguing standards with other libraries. MARC is being used as the basis for discovery information at least, but it may need some modification.

The volume of digital material and the time needed to catalogue material was an issue. It is not likely that the BL would list everything available on the Web, but would list everything acquired. An interviewee from another legal deposit library made the following comment.

... there's an interesting parallel I think between the requirements of electronic publications and practices in the archive world because archivists have been used to the concept of cataloguing levels and how detailed do you, in what detail do you describe collections? How far do you go? And that is a question for electronic publications. (Legal deposit library)

However, at least in one library there may be some concern about changes in practices.

For some staff there is a concern that doing things in new ways, for example cataloguing, means sacrificing quality and this is a perception that has to be overcome. (Legal deposit library)

## **8.7. Store Publications**

There was a view that digital output is growing fast and a concern that the libraries will not be able to cope with it. The inability to cope was confirmed by other research (Electronic Publishing Services Ltd 2002). Interviewees mentioned big databases in business and science especially, and taking this sort of material would have storage implications. All the libraries mentioned plans for digital storage. The three national libraries, at least, had plans to acquire greater storage capacity in the next year or two. Of these, the BL seemed to be the most advanced in its plans, although these have not yet materialised. Interviewees said that financial resources were a factor for the NLS and NLW.

Some of the libraries mentioned that they had or were planning hierarchical storage systems, i.e. that is using a mixture of online and offline storage for material, depending on demand. Software control mechanisms will have to be developed and also management information systems to give data on use. This will help the libraries decide how to archive material in hierarchical storage

The libraries were working together on storage. The technical model for the secure network being put in place at the time of the interviews was based on a distributed model rather than a centralised one. It gives the choice of deciding where to put things even if the choice is to keep them in one place. The respective roles of the libraries were not clear from the interviews. There was a view that the BL would act as a central store for legal deposit material. The possibility of the Welsh and Scottish national libraries wishing to store their own material and of publishers wishing their material to be stored in a

particular geographic location was also raised. At the time of the interviews, no decisions seemed to have been taken on how material will be stored within the system.

The storage problem is not one of space as it is with print material. It is more about the costs.

I think the cost of storage is not to be taken lightly either. There seems to be a sense with librarians that digital storage solves the problems of space of physical storage. I'm always amazed just how much it costs me, how much my production people tell me it costs to store this damn stuff. (Publisher)

From the publisher point of view, the potential volume to be submitted could cause problems.

It is never as simple as you think to deliver material in digital form and volume is always a problem. And we would have quite high volumes. We would find ourselves with clogged up ISDN lines, or having to load a great chunk of it onto hard discs and send it in. And all that sort of complication takes time in my experience. And expensive. (Publisher)

This comment was contradicted by subsequent research carried out for the Regulatory Impact Assessment that accompanied the new UK legal deposit legislation (see section 3.7.3). Publishers surveyed for that study thought the cost of deposit would be marginal.

## **8.8. Preserve publications**

Interviewees summed up the different reasons for preserving information on a national basis.

... if you preserve you preserve for a purpose. There's no point in preserving if people can't get access in the future. And access will depend on the right to access and it's a very different world than the world that we're used to. (Legal deposit library)

While access to print material may result in damage and therefore preservation risks, in the digital environment access does not necessarily involve this risk, especially if access to material is online.

It would be the ideal situation to have a copy that you didn't use, that you simply archived forever, as it were, without using it on the grounds that the best way to preserve it is simply not to use it. But I think in a way that almost defeats the object of deposit, which is access and preservation. And I think if you get to that sort of principle with electronic material, it seems to me in a sense it would be self-defeating because the only way you are going to preserve the material is by accessing it. In other words, you can't trust that it is going to be there. You've got to keep checking to see that things as you expect them to be. And electronic preservation is going to be a rather more complex and interactive process than simply saying this is a book that we are going to shut in a vault and not look at for a hundred years. That's not going to work with electronic materials. (Legal deposit library)

The difference between archiving and preservation from the publisher perspective was pointed out.

There is a difference between preservation and archiving and these get muddled up. From the publisher perspective, the short-term issue is archiving. This is ensuring continued access to material through substitution copies. This may not link up to preservation, which is more of a deep cave issue. (Publisher)

Preservation is

... maintaining long term access to intellectual assets. It is more than perhaps just cultural; it could also be, I guess, in certain circumstances, existing traditional material is reutilised both by scholars and the commercial world. It's maintaining that and carrying forward infrastructure. Intellectual infrastructure. National intellectual infrastructure, let's call it that or something similar. (Specialist agency)

Several interviewees mentioned the issue of preserving information so it can be reused, including publishers who discussed how they preserve their own material and commented

that they may want to reuse their own material in the future. Deposit libraries also have a role.

The problem arises when the stuff ceases to be financially timely, somebody might come back to those articles and be able to do things. This is happening all the time... people are going back to the articles of the twenties, scientific articles and find the basic work that was done then is now, they're now able to take it further because technology and so on have moved on. Now that's all right because places like this store the paper. In the future we need to do it for electronic materials. And if the likes of Elsevier are not selling access to those articles, because nobody wants to read them five years hence, somebody at some stage might want to read them twenty-five years hence, so we have to store them. Or somebody has to store them on behalf of the community. (Legal deposit library)

How legal deposit fits into the wider question of who takes responsibility for the preservation of research collections both nationally and internationally was raised. One reason given for the problems here is that the technology is still immature. The pace of change has meant that there is a lack of coordination in the digital environment. Some of the interviewees were not optimistic about the rate of change slowing down.

... one of the difficulties we have with this is that electronic communication has been of enormous benefit in many ways, but it's meant that things developed at such a quick rate that people aren't able to communicate any better with each other and I think that there is this danger that all sorts of things are happening without people understanding the big picture and what is happening in other areas. It seems to me that complex technological world in which we live has made that more rather than less likely. But I do think that the uncoordination of the electronic world is such that I think it is to be very difficult to get things back on track. (Publisher)

Well all I can say is with ten years involvement in IT, I see no sign of it changing and slowing down. I see signs of it getting more and more diverse. (Deposit library)

Other interviewees were more optimistic.

At some point, things are going to have to settle down. And when they settle down, we are going to have to pick up the pieces. Meanwhile we're trying to keep track of what's going on. It can't go on going crazy forever; it'll go on crazy for a while but not forever. No technology is crazy forever. The preservation of digital products is very problematic for libraries and it seems that to overcome these problems libraries and publishers will have to work together. (Deposit library)

### **8.8.1.            *Redundancy***

Redundancy is an important part of preservation. Having more than one copy of traditional material means that it is more likely to survive.

The principles that SCOLD has established over a period of time is that we really need two copies. We need a primary copy and a secondary copy. That would be an accepted principle in preservation circles for conventional material. (Deposit library)

As one interviewee commented.

It's far easier to burn the books when they're electronic, just wipe the disc, when there's one copy to which you're providing access. (Specialist agency)

This would suggest that redundancy is just as important in the digital environment. Several options came up in discussions; each had perceived advantages and disadvantages in terms of cost effectiveness. One interviewee summed these up.

But I think when you are dealing with electronic materials, then maybe you are really dealing with something rather different. If you had two copies, could you really afford to do the same process twice? If you were talking about migration of data et cetera. If it can be done absolutely securely one hundred percent accurately in one place, you needn't do it twice. But I'm not sure that having, do you really want two physical copies as an insurance? (Deposit library)

The options seem to be preserving a single copy and distributing this copy periodically to one or more remote locations. The single copy preservation option may be carried out

centrally or be a distributed operation. Copies could be generated automatically. The Lots of Copies Keeps Things Safe (LOCKSS) project was mentioned in this context. The BL was participating in this project, but it is an archiving rather than a preservation initiative. Current access can be maintained by the use of cached copies in the case of corruption or failure. The planned BL system was to include the facility to periodically check the accessibility of material stored. Another option is to have multiple versions created through different strategies and held in multiple locations to provide more security against information loss. However, preserving multiple copies of digital material brings its own problems. With print material, it is just a case of providing good storage and handling conditions. With digital materials it is much more complicated because preservation requires more intervention. While there was concern that carrying out preservation in different institutions would be expensive and probably result in different objects, there was also a view that this would be crucial for the survival of the information.

### **8.8.2.            *What to Preserve***

The digital preservation literature stresses the need to decide on the important aspects of digital material to be preserved as a first step in deciding preservation strategies.

My gut feeling is that we probably will never be able to retain all of the look and feel of every electronic product. There will have to be some loss. (Legal deposit library)

From the deposit libraries' perspective, it is the content that is important. For some digital publications, the functionality only adds value to content that is already in other forms. Functionality may depend on software and it may not be possible to load the content onto another platform without recoding it. The NLW was exploring how to retain the look and feel of Web sites. From comments made by both library staff and publishers, it seems that the BL will be, at least initially, taking a "shoe box" approach. Content will be saved, but discovery, retrieval and delivery of material to users will have to be dealt with later. It is not just a case of acquiring material, there has to be an infrastructure to make the collection usable. Interviewees speaking from publisher and author perspectives also commented that the content should be the main focus of legal deposit. One interviewee commented that there does not seem to be a lot of value for the UK if the aim is to preserve the state of the



art, snapshots of the digital publishing process and the way in which it has evolved. A scholarly publisher, who is also an academic author, commented

I think in the long-term what we're about is disseminating content, disseminating people's ideas, results. So as long as both the ideas and results can survive, I'm not too wedded to the precise format.

...it doesn't matter to me the precise format it's in, the question is how people see what I've said, so that the ideas that I've put down or the results I've put down, and know when I said it.

There is a question of whether all types of publishers and authors would feel the same way.

Focusing on the content seems to be the pragmatic way forward. Publishers who commented on this issue were not willing to take responsibility for future proofing their material. Some mentioned the steps they were taking to keep their material usable for their own purposes, including the use of formatting standards and making security backups. The cost of adhering to standards that facilitate preservation is likely to affect adoption by publishers. If it costs more there may be no incentive to do so. The deposit libraries were trying to influence publishers on standards.

And one of the things we're trying to do through this joint committee is to set up a dialogue with the publishers so that if, if they can choose between let's say two options, each of which is apparently equal and costs the same, and one of them is far more beneficial to us than the other, then we would try and encourage them to use that one rather than the other. Clearly we're not going to get them to change if there's a cost implication, but if there isn't a cost implication, then they are open to being persuaded that one is more advantageous than the other. (Deposit library)

A point that was made was that while material was commercially valuable, publishers wanted any access to be to the material they maintain and not to an archived version. However, when the publisher no longer had a commercial interest in the material, then there is less interest in whether functionality is maintained.

I think if it's got to the point where it is beyond commercial value to the publishers, then the publisher has got to acknowledge that they can have no call over that. And then it is simply a matter of costs versus benefits. Losing the look and feel would be a loss to future scholarship naturally, but that may be a necessary sacrifice. (Publisher)

### **8.8.3.           *How to Preserve Publications***

None of the libraries said they were actively preserving digital material. At most, they were copying it to different media and storing it. The BL was seen as taking the lead on developing preservation capabilities. It had a close working relationship with KB in this area. The two other national libraries seemed to be watching and waiting, although the NLW was appointing someone who will develop policies in this area. The BL Digital Library System was due to be ready in July 2002. However, at both interviews with BL staff, there were hints that there were problems with the procurement of the system. These hints turned out to be justified. At the time of writing, the BL still did not have a system in place.

The others libraries seemed to be behind the BL in terms of procuring digital storage systems. Procurement of storage systems will not solve preservation problems. None of the libraries had a clear idea of what would be involved in digital preservation. Some interviewees suggested that the BBC is leading the way on this issue, but a discussion with that organisation revealed that was not actively preserving material either.

The legal deposit libraries were focusing on capturing and storing material only. The BL was copying material onto CDs, online was being backed up and there was some printing out of material.

While interviewees discussed the problematic types of material and had no clear view of the way forward with these, the need to just press on and do something was expressed by more than one interviewee.

I think somebody's just got to start doing it at the end of the day. And we're going to lose stuff and we're going to get it wrong and we're going to have to park stuff and we're not going to know what we need to do with some stuff. (Specialist agency)

Some ways forward that were mentioned were following the CEDARS idea of saving the bits and creating extensive metadata to facilitate future migrations on demand. This approach could be more cost effective than periodic migration. However, the user might have to wait some time until the material is made accessible, depending on how quickly the conversion can be carried out. Once the material is accessible, it may be possible to supply it remotely. This would depend on the restrictions agreed with the publishers.

If look and feel was deemed important in some circumstances, one interviewee suggested using robots to create virtual tours. The data would be preserved and search engine software could be written perhaps using some sort of specification to try and reconstruct the product. The virtual tour could give a visual representation of the system being used using screen dumps. A subset of the actions carried out by the robots could be recreated to reconstruct look and feel.

One interviewee commented:

... at the end of the day it'll come down to how you manage the structure and how you manage the, it'll all be about people at the end of it. What sort of structures you need in your organisation, what sort of management structures, what sort of resources, what sort of skills, whether this is going to work or not, isn't it? (Specialist agency)

#### **8.8.4.            *Rights Issues***

While discussion of rights issues was mainly related to access, interviewees also identified some preservation rights issues. The main issue was whether legal deposit libraries have the necessary rights to preserve information. The voluntary scheme allows some reformatting of material, that is copying it onto another medium, but it was not clear to interviewees whether library privileges in UK copyright law allow the sort of copying that would be needed for preservation of legal deposit material. There were different views on this ranging from current legal provisions being sufficient to there being no legal basis for digital preservation copying.

One issue that interviewees did agree on is that the copyright aspects of legal deposit would not be dealt with under new legal deposit legislation. The imminent implementation of the European Union copyright directive was a source of uncertainty, because it was not clear to interviewees at the time what library privileges the UK Government would implement. In practice, there was no change except for restrictions on copying for commercial purposes under sections of the law outwith the preservation copying exceptions.

Another complicating factor was that different government departments are responsible for legal deposit and copyright legislation. The impression interviewees gave was that the DCMS, which is responsible for legal deposit, hoped that the provisions of copyright legislation would be sufficient. If this was not the case, an alternative would be agreements with publishers rather than trying to amend copyright law.

Interviewees had views on preservation copying agreements with publishers. One point made is that publishers may not be in a position to grant rights.

Publishers are not necessarily in a position to grant the rights needed for long-term archiving and access. In theory, one would hope that they would be, but they quite often commission third parties to provide things like images and to provide software. They may be able to provide the rights to archive the text of the material they have commissioned or bought from the author, but it doesn't mean they will have as extensive rights in other material that is published. ... to get the long-term archiving permissions will be actually quite difficult. I know people who have tried that. (Specialist agency)

... we carry a lot of third party content, so it would probably require us to revisit all of our contracts with third party (Publisher)

While there was a comment by publisher representatives that they trusted the British Library on copyright matters, the libraries expressed concern that a system of agreements would be onerous. If the libraries could not secure agreements with all publishers, then some material may be lost.

Interviewees who commented on the legality of harvesting Web sites were either asking for advice or thought that this would not be covered by existing preservation privileges. One publisher commented that they would want libraries to ask permission before copying the publisher's Web sites. The deposit libraries that were experimenting with collecting Web-based material were asking for permission.

Another publisher representative thought that deposit libraries would need to be given special rights to do this. During a later interview when the hand bill process had started, interviewees reported that the DCMS was thinking over some issues, the most substantive issue of which was dealing with the non-commercial part of the Web and how best to handle this. The BL wanted the latitude to be able to harvest material deemed appropriate for the BL's collections. This is not likely to be all UK-based Web material and it is unlikely that Web producers will have to "deposit" material. Therefore, the DCMS was thinking about whether legal deposit is the best way to deal with this issue or whether copyright legislation should be adjusted, perhaps by extending fair dealing provisions to allow the BL to capture, preserve and make available the material it wants.

Subsequent research found that preservation copying provisions in UK copyright law probably would not allow preservation copying (Ayre & Muir 2004). The new UK legal deposit legislation makes specific provisions for changes to the relevant laws to allow digital legal deposit.

#### **8.8.5. *Identifiers and Preservation Metadata***

The role of identifiers in the discovery of digital publications was discussed above. Another suggested role was the long-term management of digital material. For example, redundancy could be implemented by identifiers. They could be used to keep track of multiple versions of material created through different preservation strategies and held in multiple locations. Use of actionable identifiers and preservation metadata could facilitate digital preservation by linking different versions of migrated content and related migration and access tools. Another application suggested was the delivery of appropriate versions of material to users.

Elements of the Digital Object Architecture approach could also be used in a preservation context. This includes the Handle system, which is a means of pointing to something. It also includes repositories, which is an abstraction in the digital information environment. The content and the access mechanisms are separated. The Handle system needs to be configured to different applications or services and the systems need to be defined. The legal deposit system could be defined as an application of the Handle system. Digital Object Identifiers and OpenURLs could be used as part of this application.

There would be a pointer to the contents of an object and another pointer to the application tool. If using migration, both of these pointers could be updated to the migrated object and the new access tool. The relationship between the two versions could be maintained and either could be accessed. There could be multiple resolution where a DOI points to multiple objects for different purposes. Every DOI could have details of all the associated copies in existence. The DOI could point to multiple URLs or a library archive copy. The pointer to the archive could be a default service at some point. Controlling this centrally apparently would be unmanageable.

The local institution should be able to dictate what the DOI resolves to - the appropriate copy. DOIs could be used for global resolution, coupled with local resolutions such as OpenURL. In principle changes in what is the appropriate copy over time can be dealt with. The resolver system is just a means to point to current state data. The metadata can be updated. Or the Handle system could be used to add a new URL and add a condition on the previous URL to indicate its status. The DOI could be published and stay the same, whereas the URL would change all the time. DOIs could be initially assigned by publishers, although it was suggested that libraries could assign DOIs as a last resort.

The creation and provision of metadata by publishers was discussed above. The focus in this section is on library requirements for preservation metadata. There was a feeling that MARC will have to evolve to cope with digital publications. At the time of the interviews, SCOLD was basing its administrative metadata on the CEDARS model. It is not clear whether this is still the case.

As mentioned above, the legal deposit libraries were exploring the use of ONIX. There was interest in getting more preservation elements into ONIX. This means looking at what

libraries need and whether it is practical for publishers to supply this data. Some of the technical elements are already incorporated into ONIX2 because they are needed for making purchasing decisions. A point related to trade metadata is that the rights aspects are not yet well-defined even for trading of material, and that this sort of metadata certainly does not cover preservation-related rights. If legal deposit libraries were required to agree with publishers what they can do with digital material for preservation purposes, this could be a useful way of reducing the burden of administering these agreements. In principle, preservation rights metadata could be included in ONIX metadata. The Rights Data Dictionary underpinning ONIX relates to any transaction. Transactions could be loans or deposits. The dictionary is a general model for events, A (usually a user) doing something to B (a resource). The events need to be defined, but they are not restricted to e-commerce. Libraries and identifier agencies were involved in the development of the INDECS (Interoperability of Data in E-Commerce Systems) dictionary.

One of the bibliographic data suppliers interviewed was adding records for digital material to its databases and was apparently willing to enhance records received from publishers for the benefit of library clients.

#### **8.8.6. *International Responsibility for Digital Preservation***

The sharing of responsibility for legal deposit material between the deposit libraries is discussed above. This section deals with wider issues, namely the international sharing of responsibility for legal deposit material and responsibilities for the preservation of the national research collection. The issue of international publishers was raised and the question of how far should there be international duplication of international output. There seems to be a potential trade-off between having international redundancy and the costs of duplication.

... in the digital environment it's going to be interesting to see what attitudes people take to international collaboration, particularly on selection for long-term archiving. We're all dealing with a group of international publishers. Do we all try and archive nationally the international output? ... To what degree do we need to have redundancy within this framework? Is it enough to have it all in one legal deposit

library? What are the potential risks over decades of doing that and not having the degree of redundancy that we've taken for granted in the paper environment in international libraries? ... There's also a cost attached to duplicated materials internationally. There's quite high risks. (Specialist agency)

As mentioned earlier, there was some scepticism as to how workable any international legal deposit collection would be.

While interviewees thought that the purpose of legal deposit is to build a national archive, interviewees pointed out that a national archive is more than just its legal deposit collection. As one interviewee put it, it is "legal deposit plus"; researchers also require access to material published in other countries. There was a view that responsibility for the preservation of digital research materials should be shared.

I think the responsibility has to be a shared one. I think that the national libraries and the copyright libraries who are also research libraries will continue to have a very close interest in how this develops. There may well be other members within the research library community who have an interest in preservation and access of a particular collection or something which is of particular relevance to their local environment or local community. I think it is absolutely the case that it is not something that only one institution can take responsibility for. (Specialist agency)

There is a trend in the UK to look at the "distributed national collection as a whole", with collaboration between deposit and other research libraries. The impression given by interviewees is that we are currently in a transitional period; responsibilities have not yet been mapped out. One point that was made here is that consideration of preservation of both digital and traditional collections should be done together, even if responsibilities for these do not map directly on to each other.

Several interviewees commented that there were several strands of activity, including the deposit libraries, higher education and research and development, in this area that need to be drawn together. The (then) proposed Digital Preservation Coalition was seen as a way of doing this.



## **8.9. Provide Access**

Legal deposit libraries currently provide access to most of the items in their legal deposit collections as soon as they are processed. There are some embargoes on access for commercially sensitive material, such as market research reports. Access is mainly reference only in the reading rooms of the libraries. Readers are able to identify material in the library catalogues and it is delivered to them, subject to any embargoes.

At the time the interviews were carried out, only Cambridge University Library (CUL) loaned material by acting as a backup library to the British Library for inter library loans. However, the intention was only to loan material that is no longer in print. There seemed to be an assumption among interviewees that if legal deposit legislation was extended, the British Film Institute will be brought into the legal deposit system. This did not happen; neither will there be a node in Northern Ireland or the north of England to make up for geographical gaps in the current print system.

### **8.9.1. *Networked Access***

The legal deposit libraries wanted to retain access to legal deposit material in each library. The use of networking between deposit libraries has the potential to get over the expense of multiple deposits and provide transparent access for users. Networked access was a concern for at least some publishers, with trust as the major factor here.

So it's [the voluntary scheme] been going very slowly for publishers, partly because of the mechanisms are not as we would wish them to be at the moment in terms of mechanisms, the secure network is not there at the moment. Publishers are asking what's this secure network going to be because I want to be sure it is actually going to be a secure network. (Publisher)

And I think publishers have expressed their concerns about the way the information is then available because for high priced networked items if legal deposit moves to being national networking, you've killed your market in an extreme case. So we need solutions to combat that. And there's still discussions about standalone networked

stations and this sort of thing for deposit libraries. That's been what's been discussed in the past. (Publisher)

For publishers

The crucial issue is the use of the publications actually at the various libraries. (Publisher)

Publishers had commercial concerns.

It would depend entirely on what limits they put on use. If it was only for use in their library and it's not copyable in any shape or form, it wouldn't compete at all. But if it was able to be used by anybody who could get onto their library Web site from anywhere in the world, then it would compete. (Publisher)

Progress on setting up the secure network was slow during the period of the research and a proposal to Government for funding to build the network was turned down. Therefore the libraries were unable to give publishers reassurance that they could control access to material. However, by the summer of 2002 there was a working demonstrator. The access model that publishers seemed to be comfortable with is one simultaneous user anywhere in the system only.

### **8.9.2.            *Current Access to Material***

While the differences of view between libraries and publishers will make access arrangements more complicated, the nature of the digital material itself also adds complexity. A major point of contention was current access to deposited material. The libraries think that current access is necessary and seem willing to compromise with publishers on limitations to current access. Under the voluntary scheme, the publishers have various options on the access they will permit. The models put forward by publishers involve embargoes, deferred deposit, deposit after certain economic thresholds are crossed, and controlling the number of simultaneous users.

One concern for publishers includes any sort of current access to high-value material, especially to commercial users through London reading rooms because of the threat of reduced sales. This concern seems to be limited to a small group. The deposit libraries may press for immediate deposit with embargoed access.

The only publishers who are really worried about it are those who have high value material and there are different types of publishers there. For instance the directory publishers, they are involved in our discussions and they produce material at great cost and research which is only available at a high price for a short period of time.

They are concerned about people going and getting access to that material while it's at its most valuable ... so they want a time limit. (Publisher)

Access triggers could be built into the digital system in a similar way to print material. However, for material whose value is not diminished over time and whose potential market is small, even embargoes may not be acceptable to their publishers. An element of the voluntary code is deposit only after a certain number of sales. Some publishers suggested deposit without access up to an agreed number of sales. Deposit without any guarantee of future access is unlikely to be acceptable to deposit libraries because of the investment they will have to make in preserving something that no one will be able to use.

A point made by publishers is that current access to online material should be to the publisher's web site, even if the material is deposited. This brings up the question of whether current access to such material should be charged for or whether publishers would be willing to provide some level of free access to their material. For some publishers this may be seen as an alternative to deposit; for others this would be difficult and not commercially viable.

### **8.9.3.            *Access Restrictions***

Publishers did not want legal deposit libraries to act as competitors. Access that goes beyond the equivalent of current arrangements was a concern. The provision of networked access to anyone anywhere or even within library buildings was a concern. The libraries seemed to accept the need to restrict access within their institutions.

... we recognise, the obligation by making those collections available to outside readers and clearly it has to be done with restrictions. We can't give carte blanche, but on the other hand, the view that I have of the way that I think we do it is to make material available to those who can demonstrate the need to do, to see it, to view material.  
(Deposit library)

Currently the suggestion is that there would be a secure network linking the UK legal deposit libraries. There would be perhaps a single workstation where people could access securely the material which had been deposited. So that imposes really very severe limitations on the wider distribution of material. (Deposit library)

It would have to be clear who could and could not use the system and restrictions on physical locations.

There is a particular problem for the three academic libraries within the legal deposit system because of their dual role of serving their own communities as well as legal deposit users. For example, it is not likely that networked access to legal deposit materials within Oxford University will be allowed as it is for other digital resources. It may be necessary to develop a special access category for legal deposit material, but access is still likely to be within one building. Oxford users are used to using the network to get access to purchased digital material from various locations. So it will be difficult to explain different access arrangements for legal deposit material and it adds another layer of complexity.

However, access may be gradually widened over time.

In the short to medium term, I would hope that even where rights holders retained their rights, after a certain period they might be willing to allow wider access to what in commercial terms would be historical products rather than any commercial viability.  
(Deposit library)

Some material remains commercially valuable for a long time, whether or not it is current. The length of time material remains commercially valuable differs by subject. Some publishers will exploit back files or older material such as images from news material for some time. So there will be a problem with deciding cut off points for access. There was apparently no reliable way of triggering this at the time of the interviews. In the digital

environment when a regularly published item no longer arrives, this could be investigated. For one-off items, it may be complaints from users that would be the trigger. This would not exactly be a desirable solution. There was a suggestion that the Copyright Licensing Agency and similar organisations may be able to provide information here.

The libraries were confident they could meet publisher concerns. At the beginning of the research, one interviewee commented that:

We don't really think that there is a technical problem in giving the publishers what they need. It's very easy to do things which will ensure that the publishers will not lose commercial value. But making this clear to the publishers is taking an awful lot of time.  
(Deposit library)

In the second round of interviews, the view from this institution had changed in that they now felt that they had convinced the publisher organisations.

There was a view that a set of rules or arrangements would have to be developed regarding access. Experimentation would help develop these rules and it could also help overcome distrust amongst publishers about the intentions of networking. It seems that the arrangements will probably end up being a hybrid of all of the options that will have to be pragmatically applied. The legal deposit libraries will have to monitor use for the purpose of internal management of digital storage. It was suggested that aggregated user data could be provided to publishers on how their products are being used and reassure them that legal deposit use is not undermining their sales. Publishers would not just want to know about how many users were accessing their material, but how they were using the material.

Solutions to access will need negotiation between libraries and publishers. Discussions on access arrangements between the legal deposit libraries and publishers were planned for the summer of 2002. It may be that that current access to material is very restricted and any wider current access is licensed and paid for.

We're not going to gain any of the advantages from the technology; it'll have to be used exactly as a book. From the publishers' point of view, I can understand why. But if we

want to network something, we will have to buy a separate subscription to it. That's the way forward. (Deposit library)

In addition,

Publishers are not necessarily in a position to grant the rights needed for long-term archiving and access. In theory, one would hope that they would be, but they quite often commission third parties to provide things like images and to provide software (Specialist agency)

As discussed in section 2.5.5, this point has been acknowledged in the new UK legal deposit legislation.

The main point on access is that arrangements are likely to be more complex than the current system.

Maybe as a model, you create a legal deposit net of some sort, a central agency with perhaps half a dozen possible methods of depositing depending on the nature of the product [inaudible], it's very time sensitive, it's very valuable or not very valuable. And with the restrictions on its use accordingly. (Publisher)

The discussions on access focused on the publisher and library perspectives; the needs of users were not really considered. While the nature of the access and use will be subject to negotiation between publishers and libraries, the main point made about users is that they will want access to information regardless of format.

## **8.10. Looking Forward**

### **8.10.1. Skills**

The BL was working on developing a preservation layer for its system in a project with the Koninklijke Bibliotheek at the time of the interviews, although this system did not come to fruition.

... the people who are running with preservation are the BL. There's no doubt about that. Although I suppose you could say that the CEDARS project has given Oxford and Cambridge a lead start on understanding the issues. But I'm not aware of anything happening really in the other three legal depositories. I'm sure there's discussion, but they would say that themselves that they are not so far ahead. (Specialist agency)

Some of the deposit libraries felt they were lacking in skills to deal with digital publications. The NLW staff in particular felt that they had a low skills base and lack of experience, although one area of strength was the cataloguing of Web sites. Recruiting new or temporary staff with the requisite skills was mentioned as one way to rectify lack of knowledge and skills amongst existing staff. These staff would lead the development of strategy and policy and/or train staff.

The study carried out by Electronic Publishing Services Ltd (2002), however, showed that recruiting new staff did not automatically solve the problems. The Bodleian Library and Cambridge University Library had expertise in secure networking and preservation issues. One interviewee commented on the difficulties of retaining any staff recruited because of the low levels of reward in the sector.

#### **8.10.2.            *Preparing for Digital Legal Deposit***

While the libraries were looking at recruitment as a way forward, they were also tackling this through cross-institution working groups and organisational restructuring. So they were beginning to address the issue of management and organisation mentioned earlier.

One of the problems, you're absolutely right about in the library is because we've had a functional arrangement for the last few years. When you're dealing with something that has a lifecycle that covers many functions, it's actually very different managerially to assign responsibility and make sure it all happens. (Deposit library)

The British Library had a Digital Materials Working Group.

So what I did when I was first involved in the deposit stuff, starting to think about voluntary deposit was to set up an internal working group that covered the full range of

activities, so acquisitions, processing, cataloguing, preservation, storage, reading room access. So that committee had the key representative from each of those functional activities. (Deposit library)

The NLS had set up various working groups to look at issues.

And I think one of the most interesting things is that because it is something that is affecting departments right across the library is that you're getting mixes of staff who would perhaps not normally be brought together into discussion groups so there's quite a lot of horizontal rather than vertical thinking going on and I think people are finding that both challenging, but enjoyable as well. (Deposit library)

The NLW had a working group addressing digital strategy that covered legal deposit and it was also restructuring itself. The NLW was changing its curatorial divisions based structure towards one where all collections are dealt with through one department.

### **8.10.3.            *Costs and Financial Resources***

It is clear that increased complexity and the need for closer cooperation will be major features of any future legal deposit system. It is also clear that costs and a lack of financial resources will limit what can be done. The issue of costs pervaded discussions with interviewees. The general feeling was that it will cost more to deal with digital material, in fact a great deal more, at least in the short term.

... we're talking about at least a factor of ten over and above what it costs to preserve printed material and that's a very significant investment. (Specialist agency)

If standards and procedures are developed, the technology matures and processes are more automated, then costs may come down, but:

In the longer term the processing is supposed to be automated so the costs should drop towards, but probably not actually reach, the costs of processing print material. (Deposit library)



At least for the foreseeable future, the cost of dealing with digital material will be an additional cost for the deposit libraries. The amount of non-digital material being made available is not diminishing at the moment.

Several interviewees raised the question of whether the likely costs will outweigh the benefits of digital deposit. If the answer is yes, then there is the question of who provides the funding. Various interviewees thought the responsibility for providing additional funding lies with government. Various views on the government's attitudes were expressed.

I suspect they are concerned about the resource implications. I suspect that their concerns probably don't match what would, in fact, be the resource implications. They think it will be expensive but I think it would be even more expensive ... nobody, including government, I think, would be willing to pay what it costs. (Specialist agency)

I think one of the issues about extending deposit is an underlying assumption if you go back to the DCMS consultation paper or the DNH consultation paper, that extending deposit almost should be cost free to the tax payer so that it should be done within existing resources ... It is their official stated line and one to which they hold. We certainly don't consider it to be realistic once you move into online publications. (Deposit library)

In my view, if the DCMS is the lead department on this then it should be aiming to make provision to support the system because there are a number of implications of this obviously. (Publisher)

I think it [funding] should come from national government because I think you're talking about preserving the country's heritage. Now if governments can create partnerships along the way to recover some of that money, that's fine. But I think the fundamental responsibility has to be with the government. If industry won't cough up then the government has to pay. (Publisher)

The DCMS official interviewed said that the British Library would have to present a case for additional funding. The uncertainty of the willingness of devolved governments to pay for initiatives that go beyond Scottish or Welsh material has already been mentioned.

Some alternatives were mentioned, including contributions from publishers through taking care of their own material, or libraries providing services to publishers, but this was seen as problematic. Value Added Tax on digital publications could be used to finance digital deposit. If funding is more than individual countries could take on, then perhaps a supranational body such as the EU could be used as a way of sharing costs between countries.

### **8.11. Summary**

Appendix 15 summarises the issues and possible solutions identified by interviewees. While there are various technical problems to be solved, including putting technological infrastructure in place and the difficulties of automating procedures, the main issues are more to do with cooperation between stakeholders, developing appropriate management structures, development of skills and lack of resources. The way digital publications are created will cause problems for legal deposit libraries. Commercial publishers have little incentive to make things easier for the libraries. The libraries were not fully prepared for digital legal deposit in terms of skills or infrastructure. However, libraries and publishers will have to work more closely together to successfully implement digital deposit. The main underlying theme of discussions with all interviewees was one of a new and uncertain situation for everyone involved. While the Government has supported new legal deposit legislation, it is not backing this up with additional funding. These findings are discussed further in Chapter 9.

## **9. DISCUSSION**

### **9.1. The Extension of Legal Deposit to Digital Publications**

The UK has a new legal deposit law. This is something for the UK legal deposit libraries and their users to celebrate. However, this legislation does not mean that there will be legal deposit for all kinds of information. Sound and film material is specifically excluded. The implementing Regulations are not yet in place. The Government has not even yet set up its Advisory Panel, many months after the law came into force. Implementing legal deposit for digital publications will require further negotiations between libraries, publishers and Government. It has now been some years since the UK legal deposit libraries first expressed the need to extend legislation to avoid the loss of the nation's digital intellectual heritage. While it took some time to get the legislation in place, the UK Government always expressed its commitment to this. So while it was clear it would happen, libraries and publishers were still not able to work out all the various issues that would allow the new law to be implemented immediately.

The UK is still behind several other countries in terms of the scope of its legal deposit law. Several countries already have laws that cover all non-print material and several more than that have laws that cover digital publications distributed on physical carriers at least. As a Government minister hinted, Regulations to implement legal deposit for this class of material could be drawn up quite quickly because of the experience gained from the current voluntary scheme in the UK. This has not been done. However, the regulatory impact exercise concluded that this dissemination method would decline over time anyway. The more important issue for the future is deposit of online publications.

Arguably, UK legislation should have been drafted so that its provisions were similar to those of Norway or South Africa. In these countries, any type of publication would theoretically be within scope of legal deposit without the need of further Regulations. However, even the countries with comprehensive legislation are not necessarily collecting all types of material. Where they are, there have been problems of incompatibilities between legal deposit and copyright law that have had to be rectified. The UK has learned

from the experiences of others in this respect and has made provision for legal issues in its legislation.

It seems that the need to avoid controversy was the driving force behind the mode of introducing new legislation. This research found little disagreement over the need for a national archive and that legal deposit is the best mechanism for building and maintaining such an archive. However, there was some evidence that there is disagreement over how legal deposit should be implemented. It is not as simple as saying that the attitudes of publishers and libraries are different, that they are looking at legal deposit from commercial versus heritage perspectives. Publishers do not hold a unified view, and the views that appear in the literature were more anti legal deposit than those gathered through the interviews. The interview discussions and subsequent comments made in Parliament suggest that there is general support among publishers for legal deposit as a concept. Until the passage of the *Legal Deposit Libraries Bill* through Parliament, it seemed that the position of publishers was changing. It seemed that opposition to legal deposit and its extension to newer forms of publication was lessening and the cooperation and understanding between publishers and libraries was increasing. However, it is clear that tensions remain.

There were still lingering concerns among certain categories of publishers, including those producing low volume, high value material. One of the main themes that emerged from the research is the tension between publishers and libraries around the issue of access. This tension is not restricted to the legal deposit context, but is a feature of digital library developments in general. Publishers generally accepted the need for and value of preservation, but had concerns about the need for access. The counter-argument is that there is no point in preserving unless there is some access. For cases where rights owners are concerned, the most sensible option seems to be negotiating acceptable access terms. Other issues were raised, such as the perception that having to provide material without payment was a burden on publishers. There is a burden on libraries as well. Legal deposit is not just a source of free collections; they have to manage these collections for the long-term.

There is no mention in the literature of similar disagreements between deposit libraries and publishers in other countries. Of course, this does not mean that there have been no

disagreements. However, this does seem to be a particular feature of the UK context and it is not clear why this is so. The Netherlands has a successful *non-statutory* deposit system. A representative of the Dutch national library did not ascribe this to the nature of the Dutch publishing industry as another interviewee had suggested, but to good communication between the Library and publishers. It seems that communication is the crucial point. Relations between publishers and libraries in general seem to have improved in recent years. Groups such as Publisher and Library/Learning Solutions (PALS) have had a major role to play here.

Interviewees thought that the Joint Committee on Voluntary Deposit was an effective group doing good work, but when draft legislation was introduced there were complaints from publishers. The concerns expressed by publishers resulted in amendments that did improve the draft legislation. However, the subsequent requests for assurances on the implementation of the legislation strongly suggested that the trust built up between legal deposit libraries and publishers is very fragile. The fact that the Government had to repeat assurances suggests that the publishers do not trust the Government to keep its promises. That is not surprising, given general mistrust of governments. However, the fragility of the trust between publishers and legal deposit libraries is a cause of concern if digital legal deposit is to be successfully implemented in the UK.

It was clear the legal deposit libraries' view that there should be some current access to most digital publications is worrying to publishers. It is also clear that delaying deposit to avoid this is not satisfactory from a preservation perspective. The fact that the libraries were slow in demonstrating their ability to ensure secure access to and control use of deposited material was also cited as a cause for concern. Discussions with deposit libraries staff showed them to be dedicated to protecting the interests of publishers even though they are equally committed to making material accessible. From the comments made and questions asked by publisher interviewees, it seemed that their lack of confidence in legal deposit libraries has various origins. These include concerns about the way the print legal deposit system works, for example the need to deposit multiple copies. Some of the concerns were based on ignorance of the true situation, for example a perception that legal deposit material is routinely used in interlibrary lending systems. Interviewees saw the British Library Document Supply operation as a source of competition. There was concern about ready access to high value material in London reading rooms. Perhaps if legal

deposit as a whole was reviewed and the legal deposit libraries addressed any gaps between perceptions and reality, progress on digital deposit would be easier. The legal deposit libraries need to persevere in informing and reassuring publishers of their intentions and activities and the benefits of legal deposit

Interviewees mentioned negotiations between libraries and publishers on issues such as embargoes on access. A more flexible attitude to access on the part of libraries and publishers may be a useful compromise to make sure the material is archived and not lost altogether. Maintaining so-called ‘dark archives’ of material that is preserved, but not immediately accessible to users, may be more acceptable to publishers. However, it could be difficult to justify this long-term strategy to funders with short-term perspectives and users that cannot, or do not want to, access material elsewhere. Legal deposit libraries already agree embargoes on access with publishers of certainly types of material. Libraries and publishers have to find access arrangements that are acceptable to both sides for certain categories of material. Discussion of the draft legislation in the House of Lords also mentioned that there would be ongoing consultation and discussion between government, publishers and libraries and that the publishers of aggregated content would not be caught by legal deposit. So another way of speeding up implementation of the new law would be to exclude, at least in the short to mid-term, the difficult categories of material that are the main source of complaints from publishers. This could mean the loss of some valuable material, but it would be better than protracted negotiations that would result in an even greater loss of the digital heritage.

## **9.2. Readiness for Digital Legal Deposit**

While the UK legal deposit libraries were committed to the extension of legal deposit to digital material, at the time this research was carried out they were not prepared to deal with this. The regulatory impact assessment indicated that there was a backlog in processing the material coming in under the voluntary scheme. At the time of the interviews, none of the libraries had a suitable system in place for the long-term storage and management of digital publications. The BL seemed closest to this at the time, but at the time of writing, the Digital Object Management (DOM) system, was not due to be in place until 2005.

The interviews revealed that material was not being fully catalogued or made accessible. This remained the case when the research for the legislative regulatory impact exercise was carried out in late 2002. While the libraries were investigating digital preservation, the only preservation measures in place that were discussed by any of the libraries was printing out email journals and saving materials to CDs that were kept in locked cupboards.

In fairness, there is little evidence that any of the world's legal deposit libraries are really preserving digital material, although some are closer to doing it than others. The BL was working closely with the KB on system and preservation matters. This seemed a sensible thing to do at the time, given that the KB was ahead of the BL in procuring a system from the same supplier and investigating preservation options. The interviews indicated that the BL, at least, was aware of developments in other countries, such as Australia and Scandinavia. However, it is a matter of concern that the BL, especially, is only now starting to catch up with other, often smaller, legal deposit libraries in its readiness for digital deposit. While Australia, Sweden and Finland, for example, were experimenting with different methods of Web archiving in the mid to late 1990s, the BL first experiment did not take place until 2001.

Where the UK has had some influence is in its voluntary scheme, which has been used as a model for other countries as an interim measure. Reasons suggested for the lack of preparedness on the part of the BL included its concentration on other priorities for the last few years. The new building at St Pancras was beset with problems and when the new Chief Executive arrived at the Library in 2000, the focus of attention became internal restructuring. The Library of Congress seems to be in a similar position and has only recently become very active in the preservation of the digital heritage. The national libraries of Scotland and Wales were similarly at an early stage of preparation for digital deposit. However, they seemed to be less aware of developments than the other legal deposit libraries. By contrast, the Bodleian Library and Cambridge University library seemed to be more prepared, perhaps because as academic libraries they already had extensive experience of managing digital collections. They were also closely involved with the CEDARS digital preservation project. Trinity College Dublin seemed to be less prepared than the Bodleian or CUL. However, staff seemed acutely aware of the need to

prepare because of TCD's role as a sole repository of British publications in the Republic of Ireland.

### **9.3. Organising the Digital Legal Deposit System**

Currently, up to six copies of print material are stored separately in the legal deposit libraries. In the digital environment, the physical location of collections is less important as long as access can be networked between the libraries. In the current system, the deposit of multiple copies enables access to material all over the country and builds redundancy into the system for preservation purposes. It is not clear how many copies of digital material are needed for redundancy purposes.

Options include:

- one copy preserved and backed up by one institution
- one copy preserved by one institution with backup copies made and stored remotely
- two or more copies preserved and stored at separate locations

The legal deposit libraries have to decide the degree of centralisation or distribution in the digital deposit system. The current system involves much duplication of effort in order to improve access and build in redundancy. For digital publications, a centralised storage and preservation facility with networked access for the other libraries could replace this duplication of effort for digital publications. Alternatively, a future digital legal deposit system could use the existing distributed system to share the load of storage and preservation. Under the voluntary scheme, the BL was receiving most material, but the other libraries were storing the publications they received. Given the concerns of publishers and the lack of resources available to legal deposit libraries, the most sensible option seems to be one copy deposit, somewhere in the system, with access across a fully functioning secure network.

Interviewees mentioned the technical considerations that would influence decisions on how the digital legal deposit collection could be managed, including controlling network



loads so individual libraries are not overwhelmed by network traffic. The impression given by interviewees was that the BL was likely to take on the bulk of the storage and preservation responsibilities for digital material. An interviewee heavily involved with developing the secure network between the libraries said that the network was designed to be used either in a centralised or distributed way. The roles of the different libraries are also likely to be a factor. The BL is the prominent library, but the other national libraries have geographical roles. There is a question of whether Trinity College Dublin would have to physically host copies of British publications to fulfil the reciprocal arrangements between the UK and Ireland or whether networked access would be enough. The role of the two British academic libraries is not clear, especially since they have two user communities to serve. The possibility of networked access only to digital material should be seriously considered.

So, the roles and responsibilities of the legal deposit libraries really need to be determined before a digital legal deposit system can be successfully implemented. Then decisions on numbers of copies to be deposited and where they should be deposited can be taken. At their own admission, the legal deposit libraries have not been very successful in rationalising the way they collect print publications through their leisure journal experiment. The requirement to deposit multiple copies of print material has already been mentioned as an issue for publishers. However, the libraries have difficulties in that they are the intermediaries between the publishers and the users. One of the reasons mentioned by the libraries as a reason for failure of collaborative acquisition of material is that their users expect them to have material. If material is networked, one barrier to collaborative acquisition disappears. The libraries would then be moving towards the concept of a jointly owned collection. The issue of responsibility for proactive collecting – collecting material from the Internet – also has to be resolved. The UK Web Archiving Consortium is now working on this and it looks like the three ‘national’ libraries of the UK, rather than all the legal deposit libraries, will be involved in this activity in partnership with other organisations.

However the system is set up, the aim should be to make the acquisition of legal deposit material as simple as possible, especially for the publishers. Publishers should currently send print material to two locations, the BL Legal Deposit Office and the Copyright Libraries Agent. The new UK legal deposit legislation mentions the possibility of there

being different arrangements in place for different categories of material. The nature of these arrangements will have to be worked out between publishers, libraries and the government. The prospect of publishers having to send different categories of material to different physical and virtual locations in different volumes at different times is not an encouraging one for the future success of the UK legal deposit system. This could cause publishers even more irritation.

Now that there is new legislation in the UK, the publishers and libraries need to work towards agreements about what publishers and publications will be eligible for legal deposit. The issues of definitions of publishers and what constitute UK publications will have to be dealt with. There have been investigative studies that have explored the issues and made some suggestions. There are a few examples of policy making by legal deposit libraries in other countries, the most notable being Australia. As mentioned above, the more difficult cases could be set aside for the moment to give the publishers concerned some reassurance. This issue has been discussed at length for years. The time has come to take decisions. The obvious priority is material made available in the UK by organisations based in the UK. This arrangement would most closely correspond to legal deposit for print material. It seems clear that material about or by Britons based outwith the UK would be beyond the traditional scope of legal deposit. Collecting this material may well involve negotiations with and payments to rights holders. This material could be lower priority, but the legal deposit libraries should investigate if legal deposit institutions in other countries are collecting this material.

#### **9.4. Managing Digital Legal Deposit Collections**

The interviewees found it easy to identify the various problems involved in building and maintaining digital legal deposit collections, they found it more difficult to suggest solutions. The literature reveals various building blocks that could be brought together to manage digital legal deposit collections, but as yet there does not seem to be one full-scale, fully functioning system in any legal deposit library in the world. This section brings together the findings of the research and the literature review.

#### **9.4.1.            *Identification of Digital Publications***

A number of issues relating to the nature of digital publications and lack of bibliographic control in electronic publishing were identified. Whilst the nature of most print publishing is well understood, it is not clear whether new ways of making information available constitute “publishing”. The identification of what digital publications exist was identified as a problem with no obvious solutions. Traditional bibliographic tools currently do not cover all digital publications. This situation makes enforcement and the compilation of performance indicators difficult for digital publications. A UK interviewee expressed the hope that a statutory system will catalyse developments. Given the size of the UK publishing industry, it would be resource intensive to search out publications. It was suggested that more widespread use of unique identifiers will help, but for some material libraries may have to search for material proactively.

#### **9.4.2.            *Selection***

The legal deposit libraries may have to be more selective than in the past because of resource limitations. Even in the print environment, not all libraries take all potential legal deposit material. As discussed earlier, networking could mean that they do not have to. It could also mean that some material that is subject to legal deposit may not be acquired by any of the legal deposit libraries. It is to be hoped that if legal deposit libraries do become more selective, they will make the right decisions about future value. Interviewees thought that there was a need for priority setting and selection guidelines.

Another issue is that human intervention in selection will increase costs, so the libraries will have to look carefully at whether increased selection will actually save costs overall. It looks likely that traditional selection criteria will still apply, but additional criteria could be number of sales, likelihood of survival without intervention and the need for comprehensive coverage in all areas. The libraries may need to introduce some ways of minimising human effort, perhaps by selecting categories of material or particular types of publishers rather than on a case-by-case basis. Work on automation of selection of Internet material may be helpful here.

At the moment, offline material is being sent to libraries and processed in a similar way to print material. The voluntary system did not cover online material, although some material was being deposited using online means. None of the interviewees had established procedures for online material, although the situation in the libraries may have changed since then. If not, procedures and systems have to be set up for deposit of online material by publishers or capture by libraries. Problems with online material include: capturing ephemeral and frequently changing material and material created on demand; the size of databases and exactly what to capture. Since it is likely to take some time to determine what online material should be deposited, they have time to think about these procedures.

It seems likely that the sequence of events will be agreement of Regulations, with legal deposit libraries taking the lead in developing policy and setting up procedures, then publishers setting up procedures to comply with Regulations and library requirements. Of course, the libraries will have to consult with publishers whilst setting up procedures so that proposed arrangements will be workable. Both parties do need to develop procedures if, as predicted, most digital material is online in the future.

Policy decisions need to be made on what to capture, when and how. Trigger events and mechanisms need to be developed for collection or deposit of online material. Other legal deposit libraries have experience that UK libraries and publishers could draw on when advising on the drafting of Regulations and developing procedures. Perhaps the biggest and most important change that is needed is that there will have to be much closer cooperation between publishers and libraries than there has been for print material.

The output of the print publishing industry is apparently not diminishing and the issues listed above all have resource implications for legal deposit libraries.

#### **9.4.3.            *Preservation***

If legal deposit libraries are to receive digital publications as they are published, they will have to deal with the range of formats used to encode data and software required to access digital publications. It would be easier if relevant standards were developed and applied in

the digital publishing industry. However, it is not clear if legal deposit libraries could influence publishers.

The UK legal deposit libraries were not clear about what is the best approach to digital preservation, including what aspects of publication to preserve and how to maintain the “intellectual infrastructure”. Content is of prime importance to them and they felt that it may not be possible or even necessary to preserve functionality or look and feel. The libraries did not seem to have thought about the current and future needs of users of digital material in order to answer these questions.

Research on preservation strategies continues and it seems that the various strands of activity are beginning to converge with hybrid migration and emulation solutions. The direction of work on the development of so-called logical views of documents looks promising, but it looks as if there is still a lot of work to be done on decisions of what to preserve and how to preserve it. At the time of the interviews, it seemed the UK deposit libraries were not directly engaging with digital preservation and were looking at what was going on elsewhere. The British Library was working closely with the KB, but it is not yet clear whether the Universal Virtual Computer approach really will work.

The UK legal deposit libraries could learn from work in other legal deposit libraries relating to providing access in time and space to Web based material and the way material is managed in storage systems.

#### **9.4.4.            *Metadata***

Interviewees identified cataloguing as the one area in which the legal deposit libraries have successfully collaborated, through the Shared Cataloguing Project. Descriptive metadata is provided as part of the packaging in print information. It might not be entirely satisfactory, but it is at least easily identified. The lack of bibliographic sources for the discovery of the existence of digital publications was identified as a problem. Interviewees reported patchy provision of metadata required for describing and managing digital information for the voluntary deposit scheme. The legal deposit libraries have been working in this area, including some cataloguing of digital publications and investigative work in metadata

standards for bibliographic material in Scotland. The BL developed a schema for the metadata it requires that was apparently based on the CEDARS framework. The legal deposit libraries catalogue print material using MARC, so there is the question of whether MARC can accommodate digital publications and if and how preservation metadata can be integrated with descriptive metadata. Given that the central role of the legal deposit libraries is the long-term preservation of information, they should be leading the way in this area, including shaping and adopting standards. There is some evidence that they are. For example, there is a British Library representative on the PREMIS working group.

However, the libraries cannot deal with metadata issues on their own. They need to work with publishers to get the metadata they need. Unlike the library sector, there is no tradition of metadata standards in the publishing industry. However, ONIX is an emerging standard for the industry that has potential as the basis of exchange of information between publishers and libraries. However, only some of the ONIX metadata is useful to libraries and it does not include much of the metadata they would need for preservation purposes. So this is another area where the legal deposit libraries have to work with publishers. Interviewees did point out, however, that some publishers are not aware of the importance of metadata and that there may be no incentive for publishers to provide metadata over and above their own needs. This issue needs to be addressed when Regulations are being drawn up; it may need the influence of the Government to persuade publishers to provide preservation related metadata to legal deposit libraries.

#### **9.4.5.           Access**

In the main, the legal deposit libraries provide reference only access in library reading rooms. Most current material can be accessed immediately, but some embargoes are negotiated. Libraries want to retain access in each institution, but this could be networked access to material held in one location. For some categories of material there will have to be negotiated restrictions on current access if it would compromise a publisher's business. The libraries need to convince the publishers that they really can control access to and use of material, then the publishers could accept networked access. However, as mentioned earlier, the libraries and publishers need to work out all the access issues and options, including embargoes and trigger events for widening access.

Interviewees mentioned that the BL was already in discussions with some publishers on deposit of their material. The agreements that the KB has with large publishers could be used as a model in the UK, but the UK legal deposit libraries would have to develop the capacity to manage large amounts of digital material. The legal deposit libraries could promote such arrangements as a service to publishers. This approach could bring publishers and legal deposit libraries closer together. The legal deposit libraries would receive the material they need to preserve under legal deposit. They would also take on the burden of preservation of the publisher's output, and possibly provide interim services if publishers suffered a disaster situation. However, the legal deposit libraries would have to be careful that publishers do not perceive such arrangements as a potential source of competition. There are also commercial issues that would have to be dealt with if legal deposit libraries were to provide interim services. There is the question of whether it would be necessary and acceptable to publishers for them to hand over subscriber information. The other question is whether the libraries would be able to maintain security so that only authorised users have access to the interim service. In addition, there is the question of whether the UK legal deposit libraries have the resources to provide services beyond their legal deposit remit.

#### **9.4.6.            *Costs***

The cost of the publications is only one element of the total costs of legal deposit. An issue that emerges strongly from both the literature and interviews is that the costs of digital preservation are likely to be high, at least in the short term. It is too early to know how much digital legal deposit will cost, but some of the other cost elements include infrastructure and providing access to material. Increased standardisation in digital publishing and establishment and automation of procedures may bring down costs over time and these will all involve cooperation between legal deposit libraries and publishers.

Digital publications will be an additional cost for legal deposit libraries for the foreseeable future. Whatever the costs turn out to be, limited resources will restrict what can be done, although it is more feasible for the libraries to collaborate in the digital environment. Government seems to be the most likely source of additional funding, but Ministers may not recognise the need for additional funding or have sufficient budgets to provide what is

required. The Department for Culture, Media and Sport, which funds the BL, is not a big department and there are higher profile issues on the Government agenda.

Interviewees raised the question of whether the costs outweigh benefits in digital legal deposit. The benefits of preserving information are difficult to quantify, although the legal deposit libraries have some models from the library and management literature to draw on when making the case for more funding. The UK legal deposit libraries could point to other countries, such as the Netherlands and the USA, where governments have recognised the importance of preserving the intellectual output of the country and have provided extra funding for the deposit library. If this funding is not forthcoming in the UK, then the legal deposit libraries will have to deal with the situation through automation of processing material, selection and cooperation.

#### **9.4.7.           *Legal Deposit in the UK: the Wider Context***

UK legal deposit libraries do not work in a vacuum; they form an important part of the distributed research collections of the UK and the rest of the world. The legal deposit libraries need not only to work out responsibilities between themselves and cooperate with publishers. They also need to be involved in the allocation of responsibilities for preservation between other research libraries and other heritage institutions.

The UK legal deposit libraries are members of groupings of research libraries such as the Consortium of Research Libraries and the Scottish Consortium of Research Libraries. Some are also involved in the UK Digital Preservation Coalition. While it is reasonable to expect the three national libraries to lead the way on digital preservation, resource constraints provide an impetus to work with other libraries. Legal deposit libraries are joining efforts to influence publishers on standardisation in electronic publishing.

The interviews raised the issue of international cooperation on legal deposit. The BL is involved in The European Library project that aims to share the resources of European deposit libraries. International publishers also raised this issue, but there are a number of difficulties that may make this unfeasible.



While the size of the task and lack of resources may provide an incentive to set up cooperative arrangements for managing legal deposit and the wider research collection, it is important that arrangements made are not too complex, are well-managed and sustainable. Given the global nature of digital communication, it seems clear that international cooperation on legal deposit has to be investigated. It may not be appropriate to provide immediate access to legal deposit collections beyond national borders or even legal deposit institutions within a country. However, ensuring that material of value is collected by someone, somewhere, is something that could be done cooperatively by legal deposit libraries. This will require international understanding between legal deposit libraries, even if requirements differ in some ways. There is also the need to make it as easy as possible for rights holders to participate in legal deposit.

## **9.5. Overview**

The United Kingdom now has legislation to enable the extension of legal deposit to non-print publications. However, drawing up Regulations to implement legislation will require negotiations between the Government, legal deposit libraries and publishers. These groups generally support legal deposit and library and publisher groups have worked well together in developing and overseeing a voluntary deposit scheme. However, a prominent issue that emerged from this research is that tension remains between the two primary stakeholder groups. This tension mainly arises from different perspectives on the related issues of preservation and access. Misconceptions and resentment of other British Library activities on the part of publishers also contribute to this tension. Communication between legal deposit libraries and publishers to mitigate concerns and negotiate access agreements are options to address this. Focusing initial Regulations on non-controversial material could also facilitate implementation of legal deposit for digital publications.

While the UK voluntary deposit scheme has been influential, the research revealed that the legal deposit libraries were quite unprepared in some areas to deal with the deposit material that they receive. How the legal deposit system will operate for digital publications needs to be determined. The use of networking technology would, in principle, reduce the need for the deposit of multiple copies. However, there are political issues that have to be taken into account when considering which legal deposit library

takes responsibility for particular publications. Other issues to be addressed are the development of a mechanism for identifying publications eligible for deposit, selection policies and procedures for submission, processing and preserving material. This will require closer cooperation between legal deposit libraries and publishers and identifying user needs. Awareness raising and lobbying to dispel misconceptions and encourage publisher cooperation will need to be done on a national, if not an international, basis in cooperation with other bodies and institutions. There are also serious resource implications associated with the extension of legal deposit to digital publications. However, despite all the problems that need to be resolved, the over-riding factor is the need to preserve the digital heritage..

## 10. CONCLUSIONS AND RECOMMENDATIONS

### 10.1. Aims and Impact of the Research

The following research objectives have all been addressed in the research.

1. To review the development of legal and voluntary deposit systems, in the UK and also in other countries
2. To examine the current voluntary and legal deposit arrangements for printed material in the UK
3. To assess proposals made for the extension of legal deposit to digital publications and progress made on their implementation
4. To identify countries that already have, or are planning to introduce, digital voluntary or legal deposit and investigate their plans and arrangements
5. To identify and explore the potential issues arising from the extension of legal deposit to digital publications
6. To identify all the stakeholders who will be affected by digital legal deposit and evaluate their interests and concerns
7. To assess preparedness and attitudes of stakeholders who will be affected by digital legal deposit
8. To identify the management implications for legal deposit of digital publications
9. To suggest changes to the current legal deposit arrangements where necessary

The review of the literature provided much of information relating to the first five objectives and publications based on this synthesis of the issues (Muir & Davies 2000, Muir 2001a, Muir 2001b – see Appendix 16) were well-received by the information community. For example, the Library of Congress is using Muir 2001a to provide an overview of the issues for legal deposit libraries for its IT suppliers<sup>11</sup>.

---

<sup>11</sup> Amy Friedlander email to Adrienne Muir, 14 May 2003.

Other objectives were addressed through primary data gathering methods. The research used parts of the Soft Systems Methodology (SSM) approach. These were the cultural analysis, development of a conceptual model and use of this model to discuss aspects of the real situation. The research methodology selected was particularly useful for identifying stakeholders and their perspectives. The research highlighted various political and cultural issues not openly discussed in the literature and that are unlikely to have been identified through other methods. Some of the issues were not directly related to digital legal deposit, but did seem to affect attitudes in a way that had an impact on this issue. While there are very significant technical problems that need to be addressed, dealing with these other issues will be crucial if legal deposit is to be successfully extended to digital publications in the UK. While SSM has been applied to information management situations, it has been applied mainly within organisations, whereas this study used SSM to look at national issues encompassing many different organisations.

## **10.2. Major Issues and Recommendations**

A number of central themes emerged from the research. Although the UK situation has moved forward over the course of the research, it still lags behind other countries in terms of legislation. Notwithstanding the new *Legal Deposit Libraries Act 2003*, much detail remains to be determined. A number of constraints on the successful implementation of digital legal deposit were identified:

- Lack of agreement on which publications should be covered
- Lack of awareness and trust, on the part of publishers particularly
- Lack of communication and cooperation between stakeholder groups
- Lack of preparedness in legal deposit libraries
- Lack of resources to implement digital legal deposit

The three main interested parties are the legal deposit libraries, the publishers and the Government. Overcoming these restraints will require better communication and cooperative working amongst the legal deposit libraries, and between the legal deposit libraries and publishers. The only realistic approach to funding the new commitments that digital legal deposit will bring is increased Government funding for legal deposit.

Advocacy and lobbying by legal deposit libraries and other stakeholder organisations will be required to raise awareness of the importance and costs of digital legal deposit in Government and perhaps also among the general public. The following recommendations address the constraints mentioned above

### ***10.2.1. Scope of Legal Deposit***

There has been much discussion in the literature on what constitutes a UK digital publication for the purposes of legal deposit. The time has come to take decisions. The digital environment is not the same as the print environment. It is likely to be difficult to develop a working system for digital deposit if thinking is conditioned by the print paradigm. Some of the principles underlying legal deposit will have to be abandoned, or at least modified. The interested parties have to decide what material is to be covered. The goal of comprehensiveness is even more difficult in the digital environment. The interviews revealed that the legal deposit librarians were already adjusting their thinking and recognised this. Decisions on the scope of legal deposit have to be taken before the libraries and publishers can move on to deciding on deposit procedures.

Dealing with ‘easier’ categories of material first may be the best way to proceed. From the literature and the interviews, it seems that material distributed on discs and ‘static’ online material, however distributed might be easiest to collect and manage. The fact that the use of offline media for distributing digital publications is in decline might make it more important to collect this material and ensure it does not disappear or become stranded in obsolete technological environments. As for online material, the priority should be ‘born digital’ rather than digitised material that has already been collected in paper formats, for example journal articles. It might be in the interests of legal deposit libraries to collect the digitised versions in preference to hard copy to save space, but this might complicate matters. The UK legal deposit libraries cannot, however, ignore more ‘difficult’ digital material, and there is evidence from the literature that they are involved in activities exploring how to deal with this material, for example, involvement in Web archiving initiatives and working with publishers (see section below).

### **10.2.2.                    *Deposit Libraries and Publishers: Cooperation and Communication***

The hurried passage of the new Act through Parliament was a cause of concern to publishers. The new Joint Committee on Legal Deposit has taken over from the Joint Committee on Voluntary Deposit. This Committee includes representatives of rights owner organisations not represented on the JCVD, including the Newspaper Licensing Society and the Digital Content Forum. So, joint working between the legal deposit libraries and the publishers is becoming more inclusive. Including publishers of the more “difficult” categories of publication is likely to result in a lot of difficult discussion and negotiation. However, this is necessary in order that the, hopefully, soon to be formed Advisory Panel can advise on the drafting of Regulations. The interested parties need to really listen to each other in order to understand their different perspectives and concerns and be willing to make compromises. It is to be hoped that the JCLD, or any other Committee that emerges in this area, will also be suitable fora for libraries and publishers to solve any implementation problems as they arise.

Some publishers may have little motivation to engage with this issue other than an impending statutory duty. The legal deposit libraries will have to change publisher perceptions by convincing them that they can be flexible and that they will be responsible and effective custodians of digital legal deposit material. There is also a role for Government here. It is important that the Government remains committed to legal deposit and keeps its promises to publishers. Government also has to make it clear to publishers if they are being unreasonable in their attitudes or behaviour. The legal deposit libraries will have to be flexible in the comprehensiveness of collecting and on access. They also need to put the necessary technical infrastructure in place to preserve and provide secure access to their digital legal deposit collections.

The need for closer communication and cooperation is perhaps the greatest change that will be required to implement digital legal deposit.

### 10.2.3. *Communication and Preparedness in Deposit Libraries*

The legal deposit libraries will have to fully develop their policies and procedures for digital deposit. They will have to take the lead so that the publishers can develop their own procedures for compliance. However, it is likely that there will still have to be some discussion so that the procedures developed are realistic and as easy as possible for the publishers to implement. The legal deposit libraries have to develop the supporting technical and organisational infrastructure.

The research revealed a lack of communication between the legal deposit libraries. Some action was being taken to address this at the time of the research and it is important that effective communication between the libraries is developed in order to implement digital legal deposit in an efficient and effective way. Even if the libraries put the infrastructure in place to collect and safely store digital publications over time, they still have to ensure that the publications can be read. It seems that not even the most advanced deposit libraries have solved this problem, but it is crucial that they do at some point or legal deposit will be a worthless exercise for digital publications.

Ideally, publishers would cooperate with the legal deposit libraries system to make the processing and preservation of digital publications easier for the legal deposit libraries. The ultimate goal for the legal deposit libraries would be publications that are, if not future-proofed, preservation-friendly. The Digital Preservation Coalition (DPC) has studied this area and is committed to lobbying the technology sector. However, it may need an even bigger and concerted effort from customers and users to convince the technology vendors and publishers that there is a need that they have to meet.

The DPC has brought together different types of memory organisation and government departments in the UK. This effort should be coordinated on an international basis. The DPC could be the link between the UK and other countries on this, with active involvement of the UK legal deposit libraries. Other examples of cooperation are provision and organisation of metadata for discovery purposes and provision of metadata for preservation. As some of the interviewees pointed out, there is currently no compelling business case for commercial publishers to do this.

The BL has tried to make it easier for publishers by developing a metadata schema and metadata creation software. Again, legal deposit and other interested libraries should be working on an international basis to make it easier for publishers to provide metadata.

#### **10.2.4.            *Increased Funding***

There have been discussions of sources of funding to support digital preservation in the literature and in the original research carried out here. It seems clear that legal deposit libraries will require additional funding to take on digital legal deposit. Various suggestions of funding sources have been made, including contributions by publishers to preservation, charging by deposit libraries for ‘services’ and re-direction of taxation income. These are all worth exploring, but it seems likely that additional Government funding is the most realistic option.

The legal deposit libraries and their supporters have to press their case for additional funding. Legal deposit libraries in other countries, for example the USA and the Netherlands, seem to be getting more support from their governments. Again, there is a need for national and perhaps international advocacy efforts to raise legal deposit on government agendas. The world’s national libraries should work toward this, perhaps through groups such as the Conference of European National Libraries, Conference of Directors of National Libraries, International Federation of Library Associations and the United Nations Educational, Scientific and Cultural Organization. The library sector can also work with sympathetic publishers.

### **10.3.            Recommendations for Future Research**

This research has identified various issues arising from the extension of legal deposit to digital publications and preservation of digital material generally. The management implications have also been explored and the preceding chapter has identified the building blocks for solutions that already exist. There is also a need for further research in the following areas.



### **10.3.1.            *Review of Legal Deposit of Print Publications***

It may be useful to review the operation of the legal deposit system as a whole. This research highlighted that some publisher hostility to the extension of legal deposit to non-print material is due to dissatisfaction with the operation of the print system. If such concerns could be identified and explored in more depth, action could be taken by the legal deposit libraries to address them. This could make the implementation of digital legal deposit easier to accomplish.

### **10.3.2.            *Review of Digital Legal Deposit***

The implementation of digital legal deposit in the UK should be reviewed at some point in the future in order to assess how well it is functioning in terms of issues such as relationships between publishers and legal deposit libraries, comprehensiveness of collecting against set targets and accessibility issues. The needs of users of legal deposit collections have not been directly addressed in this research. These should inform collection and preservation decisions. Therefore, the legal deposit libraries should survey users, and other interested parties, to find out their needs and expectations. Although it is difficult for legal deposit libraries to anticipate future needs, they do have to address the authenticity issues discussed in this report. Legal deposit libraries have to endeavour to preserve digital publications in forms that will be, if not identical, at least acceptable to future users. The legal deposit libraries could carry out at least some of this work under the aegis of the new UK Research Libraries Network<sup>12</sup>.

### **10.3.3.            *Sustainable and Integrated Research on Preservation***

An increasing amount of research and development work is being carried out on addressing preservation challenges.

---

<sup>12</sup> At the time of writing, the formation of the RLN had recently been announced. Although early activities will focus on increasing knowledge of and access to resources, preservation of digital archives is part of longer-term plans. British Library press release, 29 July 2004.

The PATCH consortium recognised that work is being carried out through individual short-term projects and that a more integrated and sustainable approach has to be adopted.

The PATCH consortium failed in its bid for matching funding from the European Commission. However, this does not mean that there is not a need for such an approach. The European Union is still funding initiatives such as DELOS and ERPANET<sup>13</sup>, and European national libraries are carrying on with The European Library initiative. DELOS, in particular, has a specific remit for integrating and coordinating research activities and has a preservation strand, so this activity could be funded in the longer term. The United States Congress and the National Science Foundation are funding research in the United States and the Joint Information Systems Committee and the Electronic and Physical Sciences Research Council will also be funding digital preservation research in the UK. There are some examples of large scale funding of preservation research and evidence that research funders are now taking notice of the digital preservation problem. These initiatives could be built upon, and put on a more sustainable basis, by being incorporated into funding programmes and mainstream library activities.

#### **10.3.4. Sustainability, Cost and Funding Models**

Research costs money, as does implementing the findings of research in operational systems. Sustainability issues, cost and potential funding models need to be explored. There is a vicious circle here: to know what levels of funding are necessary, cost and funding models have to be developed. To develop cost and funding models, more has to be known about how material will be preserved. However, until there are some fully operational deposit systems in place, it will be difficult to explore the cost and funding models. This research would have to be done on an international basis, bringing together all that is already known in this area and drawing on the experience of the most advanced institutions. Ideally, this would involve funding bodies working together. The problem is that the funders will have many other priorities for scarce resources.

---

<sup>13</sup> The DELOS Network of Excellence on Digital Libraries is funded under the European Union's 6<sup>th</sup> Framework Programme and ERPANET is funded under the Information Society Technologies programme.

### **10.3.5.            *Creation and Provision of Preservation Metadata***

Legal deposit libraries and publishers will continue discussions on how to implement digital legal deposit in the UK. This research also revealed a need to explore the possibility of the provision of metadata by publishers to assist legal deposit libraries. The researcher has since made a contribution to the metadata question by investigating the potential of the ONIX metadata standard for the exchange of metadata between publishers and legal deposit libraries (Brindley, Muir & Proberts 2004). There is a lot more work to be done in this area. Individual legal deposit libraries and groups such as OCLC and the Research Libraries Group are developing metadata requirements, but publishers have to be persuaded to cooperate. The automation of metadata creation and provision would benefit from further research.

### **10.3.6.            *Legal Issues***

The literature and interviewees discussed the legal issues arising from the acquisition and preservation of and provision of access to digital publications by legal deposit libraries. Since the primary research was carried out, researchers have addressed some of the issues and the new UK legal deposit legislation also recognises the copying and liability issues. The researcher secured an Arts and Humanities Research Board grant to study the copyright and licensing issues for digital preservation. The study ran from September 2002 until March 2004. The aim of this study was to investigate whether copyright law and agreements governing access to licensed digital content will affect the abilities of libraries to carry out their traditional preservation role. This study found that the type of copying that is required for digital preservation is unlikely to be permitted under copyright law. The study also found that even where publishers have included guarantees to libraries that they will continue to make content accessible over time, they may not be able to honour these guarantees. The new UK legal deposit legislation includes provision to change the law to allow legal deposit activities, including copying Web sites, preservation copying and providing access, to be carried out legally. Potential solutions to overcome limited access to legal deposit collections included the use of trusted repositories (perhaps including deposit libraries) and the development of hybrid licensing regimes. Such licences could address the concerns of both libraries and publishers in that there would be standard

clauses for preservation to save libraries from having to seek permission to preserve from individual publishers and negotiable access clauses (Ayre & Muir 2004).

The researcher is also submitting a funding bid for a project that would build on work on authenticity requirements for preserved digital material and how these requirements can be incorporated into digital preservation strategies.

#### **10.4. Concluding Remarks**

The primary research reported here was carried out in 2001 and 2002 and some of the findings have been overtaken by subsequent events and activities. The value of the research has been that it has provided an overview and synthesis of the issues, not only of digital legal deposit, but of digital preservation more generally. It has gone beyond investigating what is happening in the UK and has attempted to explore the underlying reasons for the way this issue has evolved in the UK. The research has highlighted the issues and management implications and has also gone some way to suggesting ways forward. The extension of legal deposit in the UK is now within sight, but there is still a lot of work to be done before it can be achieved.

## REFERENCES

- Abrams, S.L., 2002. *Proposal for a format registry for digital library preservation*.  
Version 1.2, July 26 2002.  
<[hul.harvard.edu/formatregistry/meetings/20030124/proposal.doc](http://hul.harvard.edu/formatregistry/meetings/20030124/proposal.doc)>, [accessed 30.6.04].
- Abrams, S.L. & D. Seaman, 2003. *Towards a global digital format registry*.  
<[http://www.ifla.org/IV/ifla69/papers/128e-Abrams\\_Seaman.pdf](http://www.ifla.org/IV/ifla69/papers/128e-Abrams_Seaman.pdf)>, [accessed 1.7.04].
- Adams, M., 2001. Letter to the editor. *The Bookseller*, **1 June**, 22.
- Adkins, L. & R. Adkins, 2001. Letter to the editor. *The Bookseller*, **1 June**, 22.
- Allan, R. 2003. *Legal deposit libraries bill. Standing Committee F, Wednesday 4 June 2003*.
- Altick, R.D., 1957. *The English common reader: a social history of the mass reading public 1800-1900*. Chicago: University of Chicago Press.
- Arber, E., ed., 1950. *A Transcript of the registers of the Company of Stationers of London, 1554-1640 A.D.* [New York: P. Smith].
- Arms, W. Y., *et al.*, 2001. Collecting and preserving the Web: the Minerva prototype. *RLG DigiNews* [online], **5(2)**. <<http://www.ohio.rlg.org/preserv/diginews/diginews5-2.html> >, [accessed 29.6.04].
- Arvidson, A., 2002. The Collection of Swedish web pages at the Royal Library - the Web heritage of Sweden. *68th IFLA Council and General Conference*. Glasgow, Scotland: IFLA.
- Arvidson, A. & F. Lettenstrom, 1998. The Kulturarw3 Project - the Swedish Royal Web Archive. *The Electronic Library*, **16(2 April)**, 105-108.

Aschenbrenner, A., 2004. The bits and bites of data formats: stainless steel design for digital endurance. *RLG DigiNews* [online], **15**(8).

<[http://www.rlg.org/preserv/diginews/v8\\_n1\\_feature3.html](http://www.rlg.org/preserv/diginews/v8_n1_feature3.html)>, [accessed 8.7.04].

Australia, 1968. *Copyright act 1968*. S. 201.

<[http://www.austlii.edu.au/au/legis/cth/consol\\_act/ca1968133/](http://www.austlii.edu.au/au/legis/cth/consol_act/ca1968133/)>, [accessed 29.6.04].

Australia, 2000. *Copyright amendment (digital agenda) act 2000*. No. 11, 2000.

<<http://scaletext.law.gov.au/html/comact/10/6223/top.htm>>, [accessed 29.6.04].

Ayre, C. & A. Muir, 2004. *The right to preserve: copyright and licensing for digital preservation: a report for the Arts and Humanities Research Board*. <

[http://www.lboro.ac.uk/departments/ls/disresearch/CLDP/Project\\_reports.htm](http://www.lboro.ac.uk/departments/ls/disresearch/CLDP/Project_reports.htm)>, [accessed 1.7.04].

Barry, C.A., 1998. Choosing qualitative data analysis software: Atlas/ti and Nudist compared. *Sociological Research Online*, **3** (3).

<<http://www.socresonline.org.uk/socresonline/3/3/4.html> >, [accessed 1.7.04].

Beagrie, N., 1999. *The Warwick2 Workshop on Digital Preservation Strategy 2-4 March 1999*. <<http://www.ukoln.ac.uk/services/papers/bl/rdr6238>>, [accessed in 2000, no longer available online].

Beagrie, N., 2003. *National digital preservation initiatives: an overview of developments in Australia, France, the Netherlands, and the United Kingdom and of related international activity: commissioned for and sponsored by the National Digital Information Infrastructure and Preservation Program, Library of Congress*. Washington, DC: Council on Library and Information Resources; Library of Congress.

<<http://www.clir.org/pubs/reports/pub116/contents.html>>, [accessed 29.6.04].

Beagrie, N. & D. Greenstein, 1998. *A strategic policy framework for creating and preserving digital collections*. British Library Research and Innovation report. London: British Library Research and Innovation Centre.

<<http://www.ukoln.ac.uk/services/papers/bl/framework/framework.html>>, [accessed 1.7.04].

Beagrie, N. & M. Jones, 2002. *Preservation management of digital materials: a handbook*. <<http://www.dpconline.org/graphics/handbook/>>, [accessed 3.8.04].

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 1.7.04].

Bell, R., 1977a. Legal deposit in Britain (Part 1). *Law Librarian*, **8**(1 April), 5-8.

Bell, R., 1977b. Legal deposit in Britain (Part 2). *Law Librarian*, **8**(2 August), 22-26.

Bennett, J.C., 1997. *A framework of data types and formats, and issues affecting the long term preservation of digital material*. British Library Research and Innovation Report 47. London: British Library. <<http://www.ukoln.ac.uk/services/papers/bl/jisc-npo50/bennet.html>>, [accessed 1.7.04].

Berthon, H. & C. Webb, 2000. *The moving frontier: archiving, preservation and tomorrow's digital heritage*. <<http://www.nla.gov.au/nla/staffpaper/hberthon2.html>>, [accessed 1.7.04].

Berthon, H., 2000. *The moving frontier: archiving, preservation and tomorrow's digital heritage*. <<http://www.nla.gov.au/nla/staffpaper/hberthon2.html>>, [accessed 13.7.03].

Berthon, H., S. Thomas & C. Webb, 2002. Safekeeping: a cooperative approach to building a digital preservation resource. *D-Lib Magazine* [online], **8**(1). <<http://www.dlib.org/dlib/january02/berthon/01berthon.html>>, [accessed 1.7.04].

Besek, J.M., 2003. *Copyright issues relevant to the creation of a digital archive: a preliminary assessment*. Washington, D.C.: Council on Library and Information Resources; Library of Congress.

<<http://www.clir.org/pubs/reports/pub112/contents.html>>, [accessed 1.7.04].

Bide, M., 2000. *Seminar report: digital preservation: standards issues surrounding the deposit of non-print publications: a Book Industry Communication Seminar, sponsored by the British National Bibliography Research Fund: the Policy Studies Institute, London, 5th July 2000*. <<http://www.bic.org.uk/digpres%20seminar.doc>>, [accessed 1.7.04].

Bide, M., 2000. *Standards for electronic publishing: an overview*. The Hague: Koninklijke Bibliotheek.

Bide, M. *et al.*, 1999. *Digital preservation: an introduction to the standards issues surrounding the archiving of non-print material*. Library and Information Commission Research Report 23; British National Bibliography Research Fund Report 97. London: Book Industry Communication.

Blum, R., 1991. *Kallimachos: the Alexandrian Library and the origins of bibliography*. Madison; London: University of Wisconsin Press.

Board of Trade, 1952. *Report of the Copyright Committee*. Cmd 8662. London: Her Majesty's Stationery Office.

Board of Trade, 1977. *Copyright and designs law: report of the Committee to consider the law on copyright and designs*. Cmnd 6732. London: Her Majesty's Stationery Office.

Borbinha, J. *et al.*, 2000. *NEDLIB glossary*. <<http://www.kb.nl/coop/nedlib/glossary.pdf>>, [accessed 1.7.04].

G. Brindley, A. Muir & S. Proberts, 2004. Provision of preservation metadata: a role for ONIX? *Program*, **38**(4), 240-250.



Brindley, L., 2000. Open book. Letter to the editor. *The Guardian*, **November 27**, 21.

British Library, [n.d.]. *Legal deposit of publications: a consultation paper: response from The British Library*. <<http://www.bl.uk/information/legal-deposit.html>>, [accessed 1.12.99, no longer available].

British Library, 1996. *Proposal for the legal deposit of non-print publications to the Department of National Heritage from the British Library*. London: British Library.

British Library, 1999. *Digital Library System: briefing document*. [London]: The British Library Digital Library Programme.

British Library, 2001. *New strategic directions*. <<http://www.bl.uk/about/strategic/planfor.html>>, [accessed 15.5.03].

Brown, A.D., 1992. Grounding soft systems research. *European Journal of Information Systems*, **1** (6), 387-395.

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. <<http://www.clir.org/pubs/reports/pub121/contents.html> >, [accessed 1.7.04].

CAMiLEON, 2004. *camileon*. <<http://www.si.umich.edu/CAMiLEON/about/aboutcam.html>>, [accessed 3.8.04].

Campbell, L., 2002. The National Digital Information Infrastructure and Preservation Program (NDIIPP) and its implications for a research agenda for digital preservation. *Library & Information Research News*, **26/84**, Winter 2002, 32-40.

Canada, 1995. *National Library book deposit regulations, 1995* (SOR/95-199). <<http://laws.justice.gc.ca/en/N-12/SOR-95-199/index.html>>, [accessed 30.6.04].

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://www.parl.gc.ca/PDF/35/2/parlbus/chambus/house/bills/government/C-32\\_4.pdf](http://www.parl.gc.ca/PDF/35/2/parlbus/chambus/house/bills/government/C-32_4.pdf)>,  
[accessed 13.11.02].

Canfora, L., 1989. *The vanished library*. London; Sydney; Auckland; Johannesburg:  
Hutchinson Radius.

CEDARS, 1998a. *Cedars project summary*.  
<<http://www.leeds.ac.uk/cedars/documents/MGA04.htm> >, [accessed 1.7.04].

CEDARS, 1998b. *Project plan - baseline, July 1998*.  
<<http://www.leeds.ac.uk/cedars/documents/ABA02.htm#DESCRIPTION> >, [accessed  
1.7.04].

CEDARS, 1999. *Working definitions of commonly used terms (for the purposes of the Cedars Project)*. <<http://www.leeds.ac.uk/cedars/documents/PSW01.htm>>, [accessed  
1.7.04].

CEDARS & UKOLN, 2000. *Metadata for digital preservation: the CEDARS Project outline specification*. March 2000. <<http://www.leeds.ac.uk/cedars/OutlineSpec.htm>>,  
[accessed 1.7.04].

CEDARS, 2001. *The Cedars Project report: April 1998-March 2001*. June 2001.  
<<http://www.leeds.ac.uk/cedars/admin/CedarsProjectReportToMar01.pdf>>, [accessed  
1.7.04].

CEDARS, 2002a. *Cedars guide to digital collection management*.  
<<http://www.leeds.ac.uk/cedars/guideto/collmanagement/guidetocolman.pdf/>>, [accessed  
1.7.04].

CEDARS, 2002b. *Cedars guide to intellectual property rights*.  
<<http://www.leeds.ac.uk/cedars/guideto/ipr/guidetoipr.pdf>>, [accessed 1.7.04].

- CEDARS, 2002c. *Cedars guide to preservation metadata*.  
<<http://www.leeds.ac.uk/cedars/guideto/metadata/guidetometadata.pdf>>, [accessed 1.7.04].
- CEDARS, 2002d. *Project overview*.  
<<http://www.leeds.ac.uk/cedars/overview/overview.html>>, [accessed 1.7.04].
- Charlesworth, A., 2003. *Legal issues relating to the archiving of Internet resources in the UK, EU, USA and Australia: a study undertaken for the JISC and Wellcome Trust*.  
<[http://www.jisc.ac.uk/uploaded\\_documents/archiving\\_legal.pdf](http://www.jisc.ac.uk/uploaded_documents/archiving_legal.pdf)>, [accessed 1.7.04].
- Chapman, S., 2003. Counting the costs of digital preservation: is repository storage affordable? *Journal of Digital Information* [online], **4**(2).  
<<http://jodi.ecs.soton.ac.uk/Articles/v04/i02/Chapman/chapman-final.pdf>>, [accessed 30.6.04].
- Checkland, P. & J. Scholes, 1990. *Soft Systems Methodology in action*. Chichester: John Wiley & Sons.
- Chilvers, A., 2000. Critical issues in the use of Soft Systems Methodology: a case study in the long-term management of digital data objects. *Journal of Librarianship and Information Science*, **32**(4), 167-177.
- Chilvers, 2001. The super-metadata framework for managing long-term access to digital data objects: a possible way forward with specific reference to the UK. *Journal of Documentation*, **58**(2), 146-174.
- Chilvers, A. & J. Feather, 1998. The management of digital data: a metadata approach. *The Electronic Library*, **16**(6), 365-372.
- Clarke, A., 1986. *The British Library's compact disc experiment*. London: British Library.

Clavel-Merrin, G., 2000. *NEDLIB list of terms*. NEDLIB Report Series 7. The Hague: Koninklijke Bibliotheek. <<http://www.kb.nl/coop/nedlib/results/NEDLIBterms.pdf>>, [accessed 1.7.04].

Clayphan, R., 2001. *The British Library digital publications metadata profile*. V0.1 draft for comment. Unpublished document.

Commission 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*. Brussels: European Commission.

Committee on an Information Technology Strategy for the Library of Congress *et al*, 2000. *LC21: a digital strategy of the Library of Congress*. Washington, DC: National Academy Press. <<http://books.nap.edu/books/0309071445/html/R1.html#pagetop>>, [accessed 29.6.04].

Conference of Directors of National Libraries, 1996. *The Legal deposit of electronic publications*. <<http://www.unesco.org/webworld/memory/legaldep.htm>>, [accessed 1.7.04].

Conference of European National Libraries & Federation of European Publishers, 2001. *International declaration on the deposit of electronic publications*. <<http://www.ddb.de/news/epubstat.htm>>, [accessed 1.7.04].

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://ftp.ccsds.org/documents/pdf/CCSDS-650.0-R-2.pdf>>, [accessed 10.7.02].

Corporation for National Research Initiatives, 2002. *Handle System: introduction*. <<http://www.handle.net/introduction.html>>, [accessed 1.7.04].

Council of Europe & EBLIDA, 1999. *Guidelines on library legislation and policy in Europe*. <[http://www.coe.int/T/E/Cultural\\_Co-operation/Culture/Resources/Texts/DECS\\_CULT\\_POL\\_BOOK\(99\)14\\_EN.pdf?L=EN](http://www.coe.int/T/E/Cultural_Co-operation/Culture/Resources/Texts/DECS_CULT_POL_BOOK(99)14_EN.pdf?L=EN)>, [accessed 14.7.04].

Council on Library and Information Resources, 2002. *The state of digital preservation: an international perspective: conference proceedings*. Washington, D.C. April 24-25, 2002. Washington, D.C.: Council on Library and Information Resources.

<<http://www.clir.org/pubs/reports/pub107/contents.html>>, [accessed 1.7.04].

Council on Library and Information Resources & Library of Congress, 2002. *Building a national strategy for digital preservation: issues in digital media archiving*. Washington, D.C.: Council for Library and Information Resources; Library of Congress.

<<http://www.clir.org/pubs/reports/pub106/contents.html>>, [accessed 1.7.04].

Court of Star Chamber, 1637. *A Decree of Starre-Chamber, concerning printing, made the eleventh day of July last past*. London: R. Barker.

Cuba, 1999. *Legal deposit law*. Decree 265.

<[http://www.bnjm.cu/bnjm/english/about/library/decree\\_law.htm](http://www.bnjm.cu/bnjm/english/about/library/decree_law.htm)>, [accessed 1.7.04].

Data Archive, 1998. *An investigation into the digital preservation needs of universities and research funders: the future of unpublished research materials: report to the Digital Archiving Preservation Group*. British Library Research and Innovation report 109.

[London]: British Library Research and Innovation Centre.

<<http://www.ukoln.ac.uk/services/papers/bl/blri109/datrep.html>>, [accessed 1.7.04].

Davies, J., 1998. Safe deposit: a UK publishing view. *Alexandria*, **10**(2), 159-166.

Day, M., 1999. Metadata for digital preservation: an update. *Ariadne* [online], 22.

<<http://www.ariadne.ac.uk/issue22/metadata/intro.html>>, [accessed 1.7.04].

Day, M., 2003a. *Collecting and preserving the World Wide Web: a feasibility study undertaken for the JISC and Wellcome Trust*.

<[http://www.jisc.ac.uk/uploaded\\_documents/archiving\\_feasibility.pdf](http://www.jisc.ac.uk/uploaded_documents/archiving_feasibility.pdf)>, [accessed 29.6.04].

Day, M., 2003b. *Integrating metadata schema registries with digital preservation systems to support interoperability: a proposal*.

<[http://www.siderean.com/dc2003/101\\_paper38.pdf](http://www.siderean.com/dc2003/101_paper38.pdf)>, [8.7.04].

Dellar, G., 2001. Legal deposit accounted for. Letter to the editor. *Library Association Record*, **103**(9), 542.

Department for Culture, Media and Sport, 1998. *Legal deposit of publications: summary of responses to the consultation paper issued by the Department of National Heritage, the Scottish Office, the Welsh Office and the Department of Education Northern Ireland on 11 February 1997*. London: Department for Culture, Media and Sport.

Department for Culture, Media and Sport, 2003. *Regulatory impact assessment: legal deposit libraries bill*.

<<http://www.culture.gov.uk/NR/rdonlyres/eceqogxdlxwiljppmsf4qvb5fyq7grn6pbkjit27tv aovwc64hbdkbyslrbwuiujqhd6tsnda4m65uzwzyiflkkxb/LegalDepositLibrariesBill.pdf>>, [accessed 16.6.04].

Department of National Heritage *et al*, 1997. *Legal deposit of publications: a consultation paper*. DNHJ0198NJ February 1997. London: Department of National Heritage.

Department of Trade, 1981. *Reform of the law relating to copyright, designs and performers' protection: a consultative document*. Cmnd 8302. London: HMSO.

Digital Preservation Coalition, 2004a. *Digital Preservation Coalition and Pilgrim Trust announce new digital preservation award*. July 2003.

<<http://www.dpconline.org/graphics/awards/>>, [accessed 1.7.04].

Digital Preservation Coalition, 2004b. *DPC advocacy campaign: media and PR*.

<<http://www.dpconline.org/graphics/advocacy/>>, [accessed 1.7.04].

Digital Preservation Coalition, 2004c. *Mission and goals*.

<<http://www.dpconline.org/graphics/about/mission.html>>, [accessed 1.7.04].

Digital Preservation Coalition, 2004d. *What's new in digital preservation*.  
<<http://www.dpconline.org/graphics/whatsnew/index.html>>, [accessed 29.7.04].

Dodd, M., 2003. Publishers protest libraries bill. *PA Mediapoint* [online], **7 April**.  
<[http://pamediapoint.press.net/services/media law/updates/2003/library.html](http://pamediapoint.press.net/services/media%20law/updates/2003/library.html)>, [accessed 11.4.03].

Drimmelen, W. van, 2002. *The CDNL Committee on Digital Issues: report*.  
<[http://www.nla.gov.au/initiatives/meetings/cdnl/committee\\_on\\_digital\\_issues.html](http://www.nla.gov.au/initiatives/meetings/cdnl/committee_on_digital_issues.html)>, [accessed 28.7.04].

Eden, P. *et al*, 1999. Legal deposit: local issues in a national context, *Library Review*, **48(6)**, 271-277.

Editeur, 2002a. *ONIX mappings to MARC*. <<http://www.editeur.org/onixmarc.html>>, [accessed 1.7.04].

Editeur, 2002b. *What is ONIX International?*  
<<http://www.editeur.org/ONIX%20International%20FAQ.html>>, [accessed 1.7.04].

El-Abbadi, P., 1990. *The Life and fate of the ancient Library of Alexandria*. Paris: UNESCO.

Electronic Publishing Services Ltd, 1996. *The legal deposit of online databases*. British Library R&D Report 6244. London: British Library Research and Development Department.

Electronic Publishing Services Ltd, 2002. *The impact of the extension of legal deposit to non-print publications: assessment of cost and other quantifiable impacts: study report prepared for the Joint Committee on Voluntary Deposit*. London: EPS Ltd.

Electronic Resource Preservation and Access Network, 2003a. *ERPAtools: cost orientation tool*.

<<http://www.erpanet.org/www/products/tools/ERPANETCostingTool.pdf>>, [accessed 7.7.04].

Electronic Resource Preservation and Access Network, 2003b. *ERPAtools: digital preservation policy tool*.

<<http://www.erpanet.org/www/products/tools/ERPANETPolicyTool.pdf>>, [accessed 7.7.04].

Electronic Resource Preservation and Access Network, 2003c. *ERPAtools: risk communication tool*.

<<http://www.erpanet.org/www/products/tools/ERPANETRiskTool.pdf>>, [accessed 13.7.04].

Electronic Resource Preservation and Access Network, 2003d. *ERPAtools: selecting technologies tool*.

<[http://www.erpanet.org/www/products/tools/ERPANETSelect\\_Techno.pdf](http://www.erpanet.org/www/products/tools/ERPANETSelect_Techno.pdf)>, [accessed 7.7.04].

*Entire French web to be archived*, 2001.

<<http://www.europemedia.net/shownews.asp?ArticleID=4075>>, [accessed 18.11.02].

Evans, Lord., 2003. *Lords Hansard*, 12 Sept. 2003, Cols. 576-580.

Ezard, J., 2000. Library halts dumping of newspapers. *The Guardian*, **24 November**, 13.

Feather, J., 1994. *Publishing, piracy and politics: an historical study of copyright in Britain*. London: Mansell.

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.



- Feenstra, B., 2000. *Standards for the implementation of a deposit system for electronic publications*. The Hauge: Koninklijke Bibliotheek.  
<<http://www.kb.nl/coop/nedlib/results/NEDLIBstandards.pdf>>, [accessed 1.7.04].
- Fell, C., 2004. *Legal deposit libraries bill – update*. <<http://www.alpsp.org/legdep.htm>>, [accessed 1.7.04].
- Frankel *et al.*, 2000. *Defining and certifying electronic publications in science: a proposal to the International Association of STM Publishers*.  
<<http://www.aaas.org/spp/sfirl/projects/epub/define.shtml>>, [accessed 13.7.04].
- Fresko, M. and K. Tombs, 1998. *Digital preservation guidelines: the state of the art in libraries, museums and archives*. Luxembourg: European Commission.
- Friedlander, A., 2001a. *Background summary of results from interviews and essays*.  
<[http://www.digitalpreservation.gov/repor/interviews\\_summary.pdf](http://www.digitalpreservation.gov/repor/interviews_summary.pdf)>, [accessed 29.6.04].
- Friedlander, A., 2001b. *National Digital Information Infrastructure and Preservation Program (NDIIPP) convening sessions, November 5-6, 7-8, 15-16 November 2001 summary report*. <[http://www.digitalpreservation.gov/repor/cs\\_summary\\_012802.pdf](http://www.digitalpreservation.gov/repor/cs_summary_012802.pdf)>, [accessed 29.6.04].
- Friedlander, A., 2002. The National Digital Information Infrastructure Preservation Program: expectations, realities, choices and progress to date. *D-Lib Magazine* [online], 8(4), April 2002.  
<<http://www.dlib.org/dlib/april02/friedlander/04friedlander.html>>, [accessed 29.6.04]
- Gatenby, P., 2002a. Legal deposit, electronic publications and digital archiving – the National Library of Australia’s experience. <<http://www.ifla.org/IV/ifla68/papers/071-124e.pdf>>, [accessed 1.7.04].
- Gatenby, P., 2002b. *Report on Senior Executive Fellowship to research digital archiving in national libraries*. <<http://www.nla.gov.au/nla/staffpaper/2002/elect.html>>, [accessed 1.7.04].

Geustyn, M. & J. De Beer, 2001. National libraries around the world 1999-2000: a review of the literature. *Alexandria*, **13**(2), 71-102.

Giannattasio, I., 2002. *From off-line to on-line: legal deposit in France*. <[http://www.europeanlibrary.org/doc/wp1\\_workshop\\_den Haag\\_141202\\_giannattasio.doc](http://www.europeanlibrary.org/doc/wp1_workshop_den Haag_141202_giannattasio.doc)>, [accessed 13.7.04].

Gilheany, S., 1998. *Preserving information forever and a call for emulators*. <<http://www.archivebuilders.com/aba010.html>>, [1.7.04].

Granger, S., 2001. *Digital preservation & emulation: from theory to practice*. <<http://www.leeds.ac.uk/cedars/pubconf/papers/ichim01SG.html>>, [accessed 1.7.04].

Granger, S., 2002. Digital preservation and deep in frastructure. *D-Lib Magazine* [online], **8**(2) February 2002. <<http://www.dlib.org/dlib/february02/granger/02granger.html>>, [accessed 1.7.04].

Great Britain, 1662. *Press licensing act 1662*. 13 & 14, Car. II, Ch. 33, S. XVII.

Great Britain, 1664. *Press licensing act 1664*. 16 & 17, Car. II, Ch. 7.

Great Britain, 1665. *Press licensing act 1665*. 17, Car. II, Ch. 4.

Harris, P.R., 1991. *The library of the British Museum: retrospective essays on the Department of Printed Books*. London: British Library.

Haynes, D. et al, 1997. *Responsibility for digital archiving and long term access to digital data: a JISC/NPO study within the Electronic Libraries (eLib) Programme on the preservation of Electronic publications*. London: Library Information Technology Centre. <<http://www.ukoln.ac.uk/services/papers/bl/jisc-npo67/digital-preservation.html>>, [accessed 1.7.04].

Hedstrom, M., 2003. *It's about time: research challenges in digital archiving and long-term preservation: final report, Workshop on Research Challenges in Digital Archiving*

and Long-Term Preservation, April 12-13, 2002.

<[http://www.digitalpreservation.gov/repor/NSF\\_LC\\_Final\\_Report.pdf](http://www.digitalpreservation.gov/repor/NSF_LC_Final_Report.pdf)>, [13.7.04].

Heery, R. & H. Wagner, 2002. A metadata registry for the Semantic Web. *D-Lib Magazine* [online], **8**(5). <<http://www.dlib.org/dlib/may02/wagner/05wagner.html>>, [8.7.04].

Helsinki University Library, 1999. *Legal deposit services*.

<<http://hul.helsinki.fi/hyk/hul/kte/legaldep.html>>, [accessed 2.11.99, no longer available online].

Hendley, T., 1996. *The Preservation of digital material*. British Library R&D report 6242. London: British Library Research and Innovation Centre.

Hendley, T., 1998. *Comparison and methods & costs of digital preservation*. British Library Research and Innovation report 106. London: British Library Research and Innovation Centre.

<<http://www.ukoln.ac.uk/services/elib/papers/tavistock/hendley/hendley.html>>, [accessed 1.6.04].

Hensher, P., 2000. There's a lot to be said for reading yesterday's papers: the British Library has no business guessing what its users will need in future. *The Independent*, **August 25**, 15.

Hoare, P., 1996. *Legal deposit of non-print material: an international overview, September-October 1995*. British Library R&D Report 6245. London: British Library Research and Development Department.

Hoare, P., 1997. Legal deposit of electronic publications and other non-print material: an international overview. *Alexandria*, **9**(1), 59-79.

Hofman, H., 2002. Review: some comments on preservation metadata and the OAIS model. *DigiCULT.Info* [online], **2**, 15-20.

<[http://www.digicult.info/downloads/digicult\\_info2.pdf](http://www.digicult.info/downloads/digicult_info2.pdf)>, [accessed 13.7.04].

Holdsworth, D., 2000. *Emulation: C-ing ahead*.

<<http://cedarsgw.leeds.ac.uk/CAMiLEON/dh/oldcing.html>>, [accessed 1.7.04].

Holdsworth, D. & Wheatley, P., 2000. *Emulation, preservation and abstraction*.

<<http://cedarsgw.leeds.ac.uk/CAMiLEON/dh/ep5.html>>, [accessed 1.7.04].

House of Commons, Standing Committee F, 2003. *Legal deposit libraries bill*.

<<http://www.publications.parliament.uk/pa/cm200203/cmstand/f/st030604/am/30604s02.htm>>, [accessed 1.7.04].

Howells, K., 2003. *Legal deposit libraries bill. Standing Committee F, Wednesday 4 June 2003*.

IBM, 2002. *Royal Dutch Library preserves culture with Content Manager and DB2*.

<<http://www-5.ibm.com/nl/dias/resource/rdl.pdf>>, [accessed 1.7.04].

Inera Incorporated, 2001. *E-journal archive DTD feasibility study*. Newton, MA: Harvard University Library Office for Information Systems.

<<http://www.diglib.org/preserve/hadtdfs.pdf>>, [accessed 8.7.04].

Institute for Information Law, 1998. *Copyright aspects of the preservation of electronic publications*. Amsterdam: Institute for Information Law, University of Amsterdam.

<[www.ivir.nl/publications/koelman/kbeng2.doc](http://www.ivir.nl/publications/koelman/kbeng2.doc)>, [accessed 1.7.04].

Interim Action Committee on the Film Industry, 1978. *Proposals for setting up of a British Film Authority: report of the Interim Action Committee on the Film Industry*. Cmnd 7071. London: HMSO.

International DOI Foundation, 2002. DOI overview.

<[http://www.doi.org/overview/sys\\_overview\\_021601.html](http://www.doi.org/overview/sys_overview_021601.html)>, [accessed ].

International Internet Preservation Consortium, 2004a. *Framework Working Group*.

<<http://netpreserve.org/about/framework.php>>, [accessed 29.6.04].

International Internet Preservation Consortium, 2004b. *Netpreserve.org*.

<<http://netpreserve.org/about/index.php>>, [accessed 29.6.04].

International Organization for Standardization, 2003. *Space data and information transfer systems: Open archival information system: reference model*. ISO 14721: 2003. ISO.

Internet Archive, 2004a. *About us*.

<<http://www.archive.org/about/about.php?PHPSESSID=d80e99d2d6585eb3753f07afc2cf97e6#storage>>, [accessed 29.6.04].

Internet Archive, 2004b. *Web collaborations*.

<[http://www.archive.org/web/collaborations.php#LoC\\_sculpture](http://www.archive.org/web/collaborations.php#LoC_sculpture)>, [accessed 29.6.04].

Ireland, 2000. *Copyright and related rights act 2000*, Ch. 22, S. 198. Government of Ireland. Oireachtas. <<http://www.irishstatutebook.ie/ZZA28Y2000S198.html>>, [accessed 13.7.04].

Jasion, J.T., 1991. *The international guide to legal deposit*. Aldershot: Ashgate.

Joint Information Systems Committee & Publishers Association, 1998. *Guidelines for fair dealing in an electronic environment*. Bristol: Joint Information Systems Committee; London: Publishers Association.

<<http://www.ukoln.ac.uk/services/elib/papers/pa/fair/intro.html>>, [accessed 1.7.04].

Joint Information systems Committee, 2003. *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*. JISC Circular 6/03 (Revised) July 2003.

<[http://www.jisc.ac.uk/index.cfm?name=funding\\_digcentre](http://www.jisc.ac.uk/index.cfm?name=funding_digcentre)>, [accessed 1.7.04].

Kahle, B., 1996. *Archiving the Internet*. <[http://www.archive.org/sciam\\_article.html](http://www.archive.org/sciam_article.html)>, [accessed 1.7.04].

Kenney, Anne R. *et al*, 2002. Preservation risk management for web resources. *D-Lib Magazine* [online], **8**(1). <<http://www.dlib.org/dlib/january02/kenney/01kenney.html>>, [accessed 1.7.04].

Kent, A. *et al* (eds), 1975. *Encyclopedia of information and library science*. Vol. 14. New York: Marcel Dekker, pp. 140-181.

Koninklijke Bibliotheek, 1999. *Arrangement for depositing electronic publications at the Deposit of Netherland Publications in the Koninklijke Bibliotheek*. <<http://www.kb.nl/kb/dnp/overeenkomst-nuv-kb-en.pdf>>, [accessed 1.7.04].

Koninklijke Bibliotheek, 2000. *Koninklijke Bibliotheek and IBM Nederland work on digital archiving*. <[http://www.kb.nl/kb/resources/frameset\\_kb.html?/kb/pr/pers/pers2000/ibm-en.html](http://www.kb.nl/kb/resources/frameset_kb.html?/kb/pr/pers/pers2000/ibm-en.html)>, [accessed 1.7.04].

Koninklijke Bibliotheek, 2002. *IBM/KB joint study on long term preservation: general DNEP architecture*. <<http://www.kb.nl/kb/ict/dea/ltj/jointstudy/page3.html>>, [accessed 11.3.03].

Koninklijke Bibliotheek, 2003a. *Deposit collection*. <[http://www.kb.nl/kb/resources/frameset\\_collecties-en.html](http://www.kb.nl/kb/resources/frameset_collecties-en.html)>, [accessed 1.7.04].

Koninklijke Bibliotheek, 2003b. *NEDLIB*. <<http://www.kb.nl/coop/nedlib/>>, [accessed: 1.7.04].

Koninklijke Bibliotheek & IBM, 2001. *The IBM/KB Long-term preservation study*. <<http://www.kb.nl/kb/ict/dea/ltj/ltjstudy-overview.pdf>>, [accessed 11.3.03].

Kreher, H., 1993. Critique of two contributions to soft systems methodology. *European Journal of Information Systems*, **2**(4), 304-308.

Kunze, J.A., 2003. Towards electronic persistence using ARK identifiers. In: Masanès, J., A. Rauber & G. Cobena, eds., 2003. *3<sup>rd</sup> Workshop on Web Archives*. Trondheim, Norway, August 21<sup>st</sup>, 2003: proceedings. <>, [accessed].

Lakeman Fraser, C., 2001. Closing the deposit gap. *The Bookseller*, **6 July**, 27-29.

Larivière, J., 2000. *Guidelines for legal deposit legislation*. Rev., enl. and updated ed. Paris: Unesco.

Lavoie, B., 2003. *The incentives to preserve digital materials: roles, scenarios and economic decision-making*. <<http://www.oclc.org/research/projects/digipres/incentives-dp.pdf>>, [accessed 7.7.04].

Lavoie, B., 2004a. Implementing metadata in digital preservation systems: the PREMIS activity. *D-Lib Magazine* [online], **10**(4). <<http://www.dlib.org/dlib/april04/04lavoie.html>>, [accessed 2.7.04].

Lavoie, B., 2004b. *The Open Archival Information System Reference Model: introductory guide*. Technology Watch Report 04-01. Online Computer Library Center & Digital Preservation Coalition. <[http://www.dpconline.org/docs/lavoie\\_OAIS.pdf](http://www.dpconline.org/docs/lavoie_OAIS.pdf)>, [accessed 8.7.04]

Lea, W., P. Uttley & A.C. Vasconcelos, 1998. Mistakes, misjudgements and mischances: using SSM to understand the Hillsborough Disaster. *International Journal of Information Management*, **18**(5), 345-357.

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.

LeFurgy, W.G., 2003. PDF/A: developing a file format for long-term preservation. *RLG DigiNews* [online], **15**(6). <[http://www.rlg.org/preserv/diginews/v7\\_n6\\_feature1.html](http://www.rlg.org/preserv/diginews/v7_n6_feature1.html)>, [accessed 7.7.04].

*Legal deposit*, 2000. <<http://www.katak.gl/vo/eng/groenlandica/pligtaflevering.htm>>, [accessed 2000, no longer available].

Lehmann, K.-D., 1996. Making the transitory permanent: the intellectual heritage in a digitized world of knowledge. *In*: Graubard, S.R. & P. LeClerc, eds. *Books, bricks & bytes*. Cambridge, Mass.: American Academy of Arts and Sciences, pp. 307-329.

Library and Archives Canada, 2001. *Networked electronic publications policy and guidelines*. <<http://www.collectionscanada.ca/9/8/index-e.html>>, [30.6.04].

Library and Archives Canada, 2004. *Electronic publishing: guide to best practices for Canadian publishers*. <<http://www.collectionscanada.ca/9/13/index-e.html>>, [accessed 30.6.04].

Library of Congress, 2002. *Preserving our digital heritage: plan for the National Digital Information Infrastructure and Preservation Program: a collaborative initiative of the Library of Congress*. <[http://www.digitalpreservation.gov/rep/ndiipp\\_plan.pdf](http://www.digitalpreservation.gov/rep/ndiipp_plan.pdf)>, [accessed 29.6.04].

Library of Congress, 2003a. *Library announces approval of plan to preserve America's digital heritage*. Press release, 14 February 2003. <<http://www.loc.gov/today/pr/2003/03-022.html>>, [accessed 29.6.04].

Library of Congress, 2003b. *Library of Congress seeks to capture at-risk digital materials and build a network of partners*. Press release, 12 August 2003. <<http://www.loc.gov/today/pr/2003/03-141.html>>, [accessed 29.6.04].

Library of Congress, 2003c. *METS: an overview and tutorial*. <<http://www.loc.gov/standards/mets/METSOverview.html>>, [accessed 13.7.04].

Lor, P., 1995. Legal deposit: some issues in the international scene. *Mousaion*, **12**(1/2), 94-111.

Lord, M., 2003. Commons *Hansard*, 4 July 2003.



Lord, P., 2002. *Public draft – extracts from a survey of information technology vendors*. <<http://www.dpconline.org/graphics/reports/isreportp3.html>>, [accessed 13.7.04].

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 13.7.04].

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. <[http://www.kb.nl/kb/hrd/dd/dd\\_onderzoek/reports/4-uvc.pdf](http://www.kb.nl/kb/hrd/dd/dd_onderzoek/reports/4-uvc.pdf)>, [accessed 13.7.04].

Lunn, J., 1981. *Guidelines for legal deposit legislation*. Paris: Unesco.

Lupovici, C. & J. Masanès, 2000. *Metadata for the long-term preservation of electronic publications*. NEDLIB Report Series 2. The Hague: Koninklijke Bibliotheek. <<http://www.kb.nl/coop/nedlib/results/NEDLIBmetadata.pdf>>, [accessed 13.7.04].

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 7.7.04].

Lynch, C., 2000. Authenticity and integrity in the digital environment: an exploratory analysis of the central role of trust. In: *Authenticity in a digital environment*. Washington, D.C.: Council on Library and Information Resources, pp. 32-50. <<http://www.clir.org/pubs/reports/pub92/pub92.pdf> >, [accessed 13.7.04].

Mackenzie Owen, J.S. & J. van de Walle, 1996. *Deposit collections of electronic publications*. EUR 16910 EN. Luxembourg: Office for Official Publications of the European Communities.

Mandel, C., 1996. Enduring access to digital information: understanding the challenge. *LIBER Quarterly*, 6, 453-464.

Marc Fresko Consultancy, 1996. *Long term preservation of electronic publications: a JISC/British Library Workshop as part of the Electronic Libraries Programme (eLib) organised by UKOLN, 27th and 28th November 1995 at the University of Warwick*. British Library R&D Report. London: British Library; Bristol: Joint Information Systems Committee of the Higher Education Funding Bodies. <<http://www.ukoln.ac.uk/services/papers/bl/rdr6238/paper.html>>, [accessed 13.7.04].

Marcum, D., 2000. *State of legal deposit legislation*. <<http://www.icsti.org/icsti/2000workshop/marcum.html>>, [accessed 13.7.04].

Martin, D., 1996. *Definitions of publications and associated terms in electronic publications*. British Library R&D Report 6243. London: British Library Research & Development Department.

Martin, E., 2001. *Management of networked electronic publications: a timetable of status in various countries*. <<http://www.collectionscanada.ca/obj/r7/f2/r7-100-e.pdf>>, [accessed 13.7.04].

Masanès, J., 2002. Towards continuous Web archiving: first results and an agenda for the future. *D-Lib Magazine*, **8**(12). <<http://www.dlib.org/dlib/december02/masanes/12masanes.html>>, [accessed 13.7.04].

Matthews, G. *et al.*, 1997. *Preservation of digital materials policy and strategy issues for the UK: report of a meeting on the CPA/RLG report, December 1996*. [London]: British Library Research and Innovation Centre.

Mauritzen, I. and S.A. Solbakk, 2000. *A study on copyright and legal deposit of online documents*. Helsinki, NORDINFO.

McCormick, P. & Williamson, M., 1990. Legal deposit and electronic publishing: results of a survey. *Alexandria*, **2**(3), 51-56.

McCue, J., 2000. Library to get rid of its historic newspapers. *The Times*, **7 August**, 17.

McGovern, N.Y., 2004. Virtual remote control: building a preservation risk management toolbox for Web resources. *D-Lib Magazine* [online], **10**(4).

<<http://www.dlib.org/dlib/april04/mcgovern/04mcgovern.html>>, [accessed 7.7.04].

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 13.7.04].

Mills Wade, A., 2003. *Quoted in: British Library, 2003. Historic change in legal deposit law saves electronic publications for future generations – Bill to extend legal deposit to UK non-print materials receives Royal Assent*. Press release 31 October 2003.

<<http://www.bl.uk/cgi-bin/press.cgi?story=1382>>, [accessed 13.7.04].

Missingham, R., 2003. *Creating digital futures together: collaboration between the national library and Australian libraries*.

<<http://conferences.alia.org.au/shllc2003/papers/048.pdf>>, [accessed 13.7.04].

Mole, C., 2003a. *Commons Hansard*, 14 March 2003.

Mole, C. 2003b. *Commons Hansard*, 4 July 2003.

Mole, C. 2003c. *Legal deposit libraries bill. Standing Committee F, Wednesday 4 June 2003*.

Moore, R. *et al*, 2000a. Collection-based persistent digital archives: Part 1. *D-Lib Magazine* [online], **6**(3). <<http://www.dlib.org/dlib/march00/moore/03moore-pt1.html>>, [accessed 7.7.04].

Moore, R *et al*, 2000b. Collection-based persistent digital archives: Part 2. *D-Lib Magazine* [online], **6**(4) April 2000. <<http://www.dlib.org/dlib/march00/moore/04moore-pt2.html>>, [accessed 7.7.04].

Morris, E., 2003. *Commons Hansard*, 4 July 2003.

Moss, M., 2003a. *Commons Hansard*, 14 March 2003.

Moss, M., 2003b. *Commons Hansard*, 4 July 2003.

Moss M., 2003c. *Legal deposit libraries bill. Standing Committee F, Wednesday 4 June 2003.*

Muir, A., 2001a. Legal deposit of digital publications: a review of research and development activity. *Journal of Documentation*, **57**(5), 65-682.

Muir, A., 2001b. Legal deposit of digital publications: a review of research and development activity. *In: Proceedings of first ACM/IEEE-CS Joint Conference of Digital Libraries, June 24-28, 2001, Roanoke Virginia, USA.* New York: ACM, 2001, pp. 165-173.

Muir, A. & J. E. Davies (2000). "Legal deposit of digital material in the UK: recent developments and the international context." *Alexandria*, **12**(3), 151-165.

Muir, A., 2003. Copyright and licensing for digital preservation. *Library & Information Update*, **2**(6), 34-36.

Nasjonalbiblioteket 2004. *About the project.*

<[http://www.nb.no/paradigma/eng\\_om\\_prosjektet.html](http://www.nb.no/paradigma/eng_om_prosjektet.html)>, [accessed 30.6.04].

National Archives, 2003. *PRONOM.* <<http://www.nationalarchives.gov.uk/pronom/>>, [accessed 13.7.04].

National Center for Biotechnology Information, 2003. *XML DTD describes standard content model for electronic archiving and publishing of journal articles.* Press release, 2 June 2003. <[http://www.nlm.nih.gov/news/electronic\\_archiving.html](http://www.nlm.nih.gov/news/electronic_archiving.html)>, [accessed 13.7.04].

National Diet Library, 2001. The amended legal deposit legislation. *National Diet Library Newsletter*, **117**. <[http://www.ndl.go.jp/en/publication/ndl\\_newsletter/117/171.html](http://www.ndl.go.jp/en/publication/ndl_newsletter/117/171.html)>, [accessed 13.7.04].

National Digital Information Infrastructure and Preservation Program, 2004a. *Digital preservation program launches research grants initiative*. Press release, 16 June 2004. <[http://www.digitalpreservation.gov/about/pr\\_061604.html](http://www.digitalpreservation.gov/about/pr_061604.html)>, [accessed 29.6.04].

National Digital Information Infrastructure and Preservation Program, 2004b. *Digital preservation project testing transfer of digital archive*. Press release, 26 February 2004. <[http://www.digitalpreservation.gov/about/pr\\_022604.html](http://www.digitalpreservation.gov/about/pr_022604.html)>, [accessed 29.6.04].

National Digital Information Infrastructure and Preservation Program, 2004c. *Library of Congress announces joint digital preservation project with four universities*. Press release, 8 June 2004. <[http://www.digitalpreservation.gov/about/pr\\_060904.html](http://www.digitalpreservation.gov/about/pr_060904.html)>, [accessed 29.6.04].

National Film and Sound Archive & National Library of Australia, 1995. *National Film and Sound Archive and National Library of Australia submission to the Copyright Law Review Committee on Legal Deposit*. <<http://www.nla.gov.au/policy/clrclld.html>>, [accessed 29.6.04].

National Library of Australia, 1999. *Preservation metadata for digital collections: exposure draft*. <<http://www.nla.gov.au/preserve/pmeta.html>>, [accessed 10.7.02].

National Library of Australia, 2001a. *Digital Services Project*. <<http://www.nla.gov.au/dsp/>>, [accessed 13.7.04].

National Library of Australia, 2001b. *Guidelines for the selection of online Australian publications intended for preservation by the National Library of Australia*. <<http://www.nla.gov.au/scoop/guidelines.html>>, [accessed 14.7.04].

National Library of Australia, 2001c. *National Library of Australia: electronic information resources strategies and action plan 2001-2002*. <<http://www.nla.gov.au/policy/electronic/resourcesplan2001report.html>>, [accessed 13.7.04]

National Library of Australia, 2001d. *The PANDORA logical data model*. Version 2, 1997. <<http://www.nla.gov.au/ldmv2.html>>, [accessed 29.6.04].

National Library of Australia, 2001e. *The PANDORA Project: a summary of progress*. <<http://pandora.nla.gov.au/documents.html>>, [accessed 13.6.01, no longer available]

National Library of Australia, 2001f. *Persistent identifier scheme adopted by the National Library of Australia*. <<http://www.nla.gov.au/initiatives/nlapi.html>>, [accessed 8.7.04]

National Library of Australia, 2002a. *Managing Web resources for persistent access*. <<http://www.nla.gov.au/guidelines/persistence.html>>, [accessed 8.7.04].

National Library of Australia, 2002b. *PADI selection criteria*. <<http://www.nla.gov.au/padi/selection.html>>, [accessed 28.7.04].

National Library of Australia, 2002c. *Safeguarding Australia's web resources: guidelines for creators and publishers*. <<http://www.nla.gov.au/guidelines/webresources.html>>, [accessed 8.7.04].

National Library of Australia, 2003a. *Online Australian publications: selection guidelines for archiving and preservation by the National Library of Australia*. <<http://pandora.nla.gov.au/selectionguidelines.html>>, [accessed 29.6.04].

National Library of Australia, 2003b. *The PANDORA business process model*. <<http://pandora.nla.gov.au/bpm.html>>, [accessed 29.6.04].

National Library of Australia, 2004a. *Digital preservation policy*. <<http://www.nla.gov.au/policy/digpres.html>>, [accessed 13.7.04].

National Library of Australia, 2004b. *Digital Services Project*. <<http://www.nla.gov.au/dsp/>>, [accessed 29.6.04].

National Library of Australia, 2004c. *Service to indexing and abstracting agencies*. <<http://pandora.nla.gov.au/abstracting.html>>, [accessed 29.6.04].

National Library of Canada, 2000. *Consultation on online publications: summary report*. <<http://www.collectionscanada.ca/8/7/index-e.html>>, [accessed 13.7.04].

NEDLIB, 2000. *NEDLIB contribution to the review of OAIS*. <<http://www.kb.nl/coop/nedlib/results/OAISreviewbyNEDLIB.html>>, [accessed 13.7.04].

NEDLIB, 2003. *Preserving our digital heritage*. <<http://www.kb.nl/coop/nedlib/>>, [13.7.04].

New Decree, 2002. *New decree for Kulturarw3, 2002:287*. <[http://www.kb.se/Info/Pressmed/Arkiv/2002/020605\\_eng.htm](http://www.kb.se/Info/Pressmed/Arkiv/2002/020605_eng.htm)>, [accessed 13.7.04].

Noordermeer, T.C., 1997. Deposit for Dutch electronic publications: research and practice in the Netherlands. In: C. Peters & C. Thanos, eds. *Research and advanced technology for digital libraries: first European Conference, ECDL'97, Pisa, Italy, September 1-3, 1997: proceedings*. Berlin: Springer, 1997, pp. 361-373.

Noordermeer, T.C., 1999. A bibliographic link between publishers and national bibliographic agencies concerning electronic publications: project BIBLINK. *International Cataloguing and Bibliographic Control*, **28**(1), 26-29.

Noordermeer, T.C., 2000. A bibliographic link between publishers of electronic resources and national bibliographic agencies: Project BIBLINK. *Exploit Interactive* [online], **4**. <<http://www.exploit-lib.org/issue4/biblink>>, [accessed 13.7.04].

Nordic Web Archive, 2004. *NWA*. <<http://nwa.nb.no/>>, [accessed 29.6.04].

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](http://www.pliktavlevering.no/html/legal_deposit.html)>, [13.7.04].

Ockerbloom, J.M., 2002. *What is TOM? (and what's it good for): coping with diverse data formats*. <<http://tom.library.upenn.edu/intro.html>>, [accessed 13.7.04].

OCLC, 2002. *PURLS*. <<http://www.purl.org/>>, [accessed 13.7.04].

OCLC, 2004. *PREMIS (PREservation Metadata: Implementation Strategies)*.  
<<http://www.oclc.org/research/projects/pmwg/default.htm>>, [accessed 13.7.04].

OCLC/RLG Working Group on Preservation Metadata, 2001. *Preservation metadata for digital objects: a review of the state of the art: a white paper*.  
<[http://www.oclc.org/research/projects/pmwg/presmeta\\_wp.pdf](http://www.oclc.org/research/projects/pmwg/presmeta_wp.pdf)>, [accessed 13.7.04].

OCLC/RLG Working Group on Preservation Metadata, 2001. *A recommendation for Content Information*. <<http://www.oclc.org/research/pmwg/contentinformation.pdf>>, [accessed 1.11.01, no longer available].

OCLC/RLG Working Group on Preservation Metadata, 2002a. *Preservation metadata and OAIS information model: a metadata framework to support the preservation of digital objects*. Dublin, Ohio: OCLC. <[http://www.oclc.org/research/pmwg/pm\\_framework.pdf](http://www.oclc.org/research/pmwg/pm_framework.pdf)>, [accessed 13.7.04].

OCLC/RLG Working Group on Preservation Metadata, 2002b. *A recommendation for Preservation Description Information*.  
<[http://www.oclc.org/research/pmwg/pres\\_desc\\_info.pdf](http://www.oclc.org/research/pmwg/pres_desc_info.pdf)>, [accessed 30.9.01, no longer available].

Oltmans, E. & H. van Wijngaarden, 2004. Digital preservation in practice: the e-Depot at the Koninklijke Bibliotheek. *Vine*, **34**(1), 21-26.

Oskamp, L., 2003. *Report on current practices among publishers regarding the deposit of digital publications*. Issue 1.0 16 July 2003. The Hague: Koninklijke Bibliotheek.  
<[http://www.europeanlibrary.org/pdf/tel\\_d1.2v1.0.pdf](http://www.europeanlibrary.org/pdf/tel_d1.2v1.0.pdf)>, [accessed 13.7.04].

Parsons, E.A., 1952. *The Alexandrian Library: glory of the Hellenic world: its rise, antiquities, and destructions*. London: Cleaver-Hume Press.



Partridge, R.C.B., 1938. *The history of the legal deposit of books throughout the British Empire: a thesis approved for the Honours Diploma of the Library Association*. London: Library Association.

Patel, M. & R. Clayphan, 2000. *Project BIBLINK: linking publishers and national bibliographic agencies*. <<http://www.ukoln.ac.uk/metadata/publications/biblink/proj-biblink.html>>, [accessed 13.7.04].

Peacock, I. & A. Powell, 1998. Metadata: BIBLINK checksum. *Ariadne* [online], **17**. <<http://www.ariadne.ac.uk/issue17/biblink/>>, [accessed 13.7.03].

Phillips, M.E., 2000. *Managing chaos in the cyberworld*. <<http://www.nla.gov.au/nla/staffpaper/mphillips5.html>>, [accessed 13.7.04].

Phillips, M.E., 2003. PANDORA, Australia's Web archive, and the digital archiving system that supports it. *DigiCULT.Info* [online], **6**. <[http://www.digicult.info/downloads/dc\\_info\\_issue6\\_december\\_20031.pdf](http://www.digicult.info/downloads/dc_info_issue6_december_20031.pdf)>, [accessed 13.7.04].

Price, J.W., 1985. Optical disk and demand printing research at the Library of Congress. *Information Services & Use* **5**, 3-20.

Prime Minister's Working Party, 1976. *Future of the British film industry: report of the Prime Minister's Working Party*. Cmnd 6372. London: HMSO.

Prys-Davies, Lord, 2003. Legal Deposit Libraries Bill. *Lords Hansard*, **12 Sept**.

Publishers Association, 1997. *Legal deposit of publications: representation to the Department of National Heritage, the Scottish Office, the Welsh Office, the Department of Education for Northern Ireland*. London: Publishers Association.

Publishers Association, 2003. *Legal deposit libraries bill*. <<http://www.publishers.org.uk/paweb/paweb.nsf/pubframe>>, [accessed 15.6.04].

RLG/OCLC Working Group on Digital Archive Attributes, 2002. *Trusted digital repositories: attributes and responsibilities: an RLG-OCLC report*. Mountain View, CA. <<http://www.rlg.org/longterm/repositories.pdf>>, [accessed 13.7.04].

Robertson, S., 1996. *Digital Rosetta Stone: a conceptual model for maintaining long-term access to digital documents*. MSc dissertation, Air Force Institute of Technology, Air University. <[http://www.au.af.mil/au/database/projects/ay1996/afit\\_la/rober\\_sb.pdf](http://www.au.af.mil/au/database/projects/ay1996/afit_la/rober_sb.pdf)>, [accessed 2000, no longer available].

Ross, S. & A. Gow, 1999. *Digital archaeology: the recovery of digital materials at risk*. British Library Research and Innovation report 108. [London]: British Library Research and Innovation Centre.

Rothenberg, J., 1995. Ensuring the longevity of digital documents. *Scientific American*, **272**(1), 42-47.

Rothenberg, J., 1999a. *Ensuring the longevity of digital information*. <<http://www.clir.org/pubs/archives/ensuring.pdf>>, [accessed 13.7.04].

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 13.7.04].

Rothenberg, J., 2000a. *An experiment in using emulation to preserve digital publications*. Den Haag: Koninklijke Bibliotheek.

Rothenberg, J., 2000b. Preserving authentic digital information. In: *Authenticity in a digital environment*. Washington, D.C.: Council on Library and Information Resources, pp. 51-68. <<http://www.clir.org/pubs/reports/pub92/pub92.pdf>>, [accessed 13.7.04].

Rugaas, B., 1990. Legal deposit and bibliographic control of new media in Europe. *LIBER Bulletin*, **35**, 156-170.

Russell, K., 1999. Digital resources into the future: digital preservation and the CEDARS project. *Serials*, **12**(3), 259-267.

Sanett, S., 2003. The cost to preserve authentic electronic records in perpetuity: comparing costs across cost models and cost frameworks. *RLG DigiNews*, **7**(4).

<<http://www.rlg.ac.uk/preserv/diginews/diginews7-4.html#feature2>>, [accessed 8.7.04].

Seadle, M., 2001. Copyright in the networked world: digital legal deposit. *Library Hi Tech*, **19**(3), 299-303.

Searle, S. & D. Thompson, 2003. Preservation metadata: pragmatic first steps at the National Library of New Zealand. *D-Lib Magazine* [online], **9**(4).

<<http://www.dlib.org/dlib/april03/thompson/04thompson.html>>, [accessed 2.7.04].

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 14.7.04].

Shuler Shaw, T., 1975. Legal depository libraries. In: Kent, A., ed. *Encyclopedia of Information and Library Science*. New York: Marcel Dekker, pp. 140-181.

Sjøvoll, R., 1996. *Presentation of the National Library of Norway, Rana branch (NB Rana) and legal deposit of electronic documents at the NNC Workshop in Lund, Sweden October, 10th, 1996*. <<http://www.nbr.no/aktiv/foredrag.html>>, [accessed 27.10.99, no longer available].

Smith, C., 1998. *Publications*. House of Commons Hansard Written Answers, 17 December 1998, Part 43, Column 682.

Smith, G., 2001. Review of: Larivière, Jules. Guidelines for legal deposit legislation. Paris:UNESCO, 2000. (CIL.2000/WS/7). *Alexandria*, **13**(1), 51-53.

Smith, R., 1994. Legal deposit in France: the new law of 20 June 1992. *The Law Librarian*, **25**(3), 143-145.

Sollins, K.R & L. Masinter, 1995. *Requirements for Uniform Resource Names*. Network Working Group RFC 1737.

South Africa, 1997. *Legal deposit act 54 of 1997*. <[www.dac.gov.za/legislation\\_policies/acts/0rb~LEGAL%20DEPOSIT%20ACT%2054%20OF%201997.doc](http://www.dac.gov.za/legislation_policies/acts/0rb~LEGAL%20DEPOSIT%20ACT%2054%20OF%201997.doc)>, [accessed 27.7.04].

Spivey, C., 2002. Online archiving with the British Library: the Emerald experience. *Serials*, **15**(3), 211-217.

Steenbakkens, J., 1999. Developing the Depository of Netherlands Electronic Publications. *Alexandria*, **11**(2), 93-105.

Steenbakkers, J., 2000. *Setting up a deposit system for electronic publications: the NEDLIB guidelines*. NEDLIB Report Series 5. The Hague: Koninklijke Bibliotheek. <[http://www.kb.nl/kb/resources/frameset\\_kb.html?/kb/hrd/dd/dd\\_projecten/nedlib\\_publicaties-en.html](http://www.kb.nl/kb/resources/frameset_kb.html?/kb/hrd/dd/dd_projecten/nedlib_publicaties-en.html)>, [accessed 13.7.04].

Steenbakkers, J., 2002. *Electronic publications: working on long term access*. <[http://www.kb.nl/kb/resources/frameset\\_kb.html?/kb/pr/pers/pers2000/ibm-en.html](http://www.kb.nl/kb/resources/frameset_kb.html?/kb/pr/pers/pers2000/ibm-en.html)>, [accessed 11.3.03, no longer available].

Steenbakkers, J. and Trier, G. van, 2003. *Report on the CDNL Committee on Digital Issues*. <<http://www.nla.gov.au/initiatives/meetings/cdnl/2003/07cdi.html>>, [13.7.04].

Steenbakkers, J., 2003. *Permanent archiving of electronic publications: research & practice*. International Summer School on the Digital Library 2003. Course 3: Libraries, Electronic Resources, and Electronic Publishing. <[http://www.kb.nl/kb/hrd/dd/dd\\_links\\_en\\_publicaties/publicaties/summerschoolticer2003.pdf](http://www.kb.nl/kb/hrd/dd/dd_links_en_publicaties/publicaties/summerschoolticer2003.pdf)>, [accessed 8.7.04].

Thibodeau, K., 2001. Building the archives of the future: advances in preserving electronic records at the national archives and records administration. *D-Lib Magazine* [online], 7(2). <<http://www.dlib.org/dlib/february01/thibodeau/02thibodeau.html>>, [accessed 7.7.04].

Thomas, S., 2003. *Legal deposit libraries bill*. Standing Committee F, Wednesday 4 June 2003.

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 13.7.04].

Todd, M., 2003. *Legal deposit libraries bill*. Standing Committee F, Wednesday 4 June 2003.

United Kingdom, 1710. *Copyright act, 1709*. 8, Anne, Ch. 19.

United Kingdom, 1775. *(University) copyright act 1775*. 15, George 3, Ch. 53.

United Kingdom, 1801. *Copyright act 1801*. 41, George 3, Ch. 107.

United Kingdom, 1814. *Copyright act 1814*. 54, George 3, Ch. 156.

United Kingdom, 1836. *Copyright act 1836*. 6 & 7 Wm 4, Ch. 110. London: George Eyre; Andrew Strahan.

United Kingdom, 1842. *Copyright act 1842*. 5 & 6 Vict., Ch. 45, Sc. 6-10. London: George Eyre; Andrew Spottiswoode.

United Kingdom, 1911. *Copyright act, 1911*. 1 & 2 Geo. 5, Ch. 46, S. 15. London: HMSO.

United Kingdom, 1915. *Copyright (British Museum) Act, 1915*. 5 & 6 Geo. 5, Ch. 38. London: HMSO.

United Kingdom, 1925. *National Library of Scotland Act 1925* (15 & 16 Geo. 5, Ch. 73). London: HMSO.

United Kingdom, 1932. *British Museum Act 1932*. 22 & 23 Geo. 5, Ch. 34. London: HMSO.

United Kingdom, 1956. *Copyright act 1956*. 5 Eliz. 2, Ch. 74. London: HMSO.

United Kingdom, 1968. *Theatres Act 1968*. 17 Eliz. 2, Ch. 5. London: HMSO.

United Kingdom, 1969. *Film (Statutory deposit) bill*. 18 Eliz. 2, Ch. 78. London: HMSO

United Kingdom, 1972. *British Library act 1972*. 20 & 21 Eliz. 2, Ch. 54. London: HMSO.

United Kingdom, 1987a. *National Library of Wales (Delivery of books) (Amendment) order 1987*. Statutory Instrument 1987/698. London: HMSO

United Kingdom, 1987b. *National Library of Wales (Delivery of books) (Amendment) regulation 1987*. Statutory Instrument 1987/918. London: HMSO

United Kingdom, 1988. *Copyright, designs and patents act 1988*. Eliz. 2, Ch. 48. London: HMSO.

United Kingdom, 1989. *Copyright: the copyright [librarians and archivists] [copying of copyright material] regulations 1989*. 37 & 38 Elizabeth II. London : HMSO.

United Kingdom, 1995. *The duration of copyright and rights in performances regulations 1995*. 43 & 44 Elizabeth II. London : HMSO.

United Kingdom, 1997. *Copyright rights in databases: the copyright and rights in databases regulations 1997*. 45 & 46 Elizabeth II. London : HMSO.

United Kingdom, 2003a. *Legal deposit libraries bill*. Bill 26 53/2., London: The Stationery Office.

United Kingdom, 2003b. *Legal deposit libraries bill* [as amended in Standing Committee F] (Bill 118 53/2). London: The Stationery Office.

United Kingdom, 2003c. *Legal deposit libraries bill* (HL Bill 91 53/2). London: The Stationery Office.

United Kingdom, 2003d. *Legal deposit libraries act 2003*. Eliz. II, Ch. 28. Norwich: The Stationery Office.

United Kingdom, 2004. *Legal deposit libraries act 2003 (Commencement) Order 2004*. Statutory Instrument 2004/130 (C. 5). Norwich: The Stationery Office.

United Kingdom, House of Commons, 2002. *Weekly Information Bulletin*, 14 December 2002 (Session 2002-03).

<<http://www.publications.parliament.uk/pa/cm200203/cmwib/wb021214/ind.htm>>, [accessed 1.7.04].

United Kingdom, House of Commons, Science and Technology Committee, 2004.

*Scientific publications: free for all?* 10<sup>th</sup> Report, HC 399-I.

<<http://www.publications.parliament.uk/pa/cm200304/cmselect/cmsctech/399/399.pdf>>, [accessed 28.7.04].

UK Web Archiving Consortium, 2004a. *Project overview*.

<<http://www.webarchive.org.uk/>>, [accessed 29.6.04].

UK Web Archiving Consortium, 2004b. *Wonders of the Web captured forever ...* Press release, 21 June 2004. <<http://www.webarchive.org.uk/>>, [accessed 29.6.04].

United States Copyright Office, 2004. *Copyright basics*.

<<http://www.copyright.gov/circs/circ1.html>>, [accessed 14.7.04].

United States of America, 1976. *Copyright act of 1996*, Ch. 4. United States Code, Ch. 17. US Government Printing Office.

United States of America, 1998. *Digital millennium copyright act*.

<[www.loc.gov/copyright/legislation/dmca.pdf](http://www.loc.gov/copyright/legislation/dmca.pdf)>, [accessed 13.7.04].

University of Cambridge, 2003. *Preserving our past: a joint digital repository project for University and MIT*. Press release, 10 February 2003.

<<http://www.admin.cam.ac.uk/news/press/dpp/2003021001>>, [accessed 13.7.04]

University of Leeds, [2003]. *Survey and assessment of source of information on file formats and software documentation: final report*.

<[http://www.jisc.ac.uk/uploaded\\_documents/FileFormatsreport.pdf](http://www.jisc.ac.uk/uploaded_documents/FileFormatsreport.pdf)>, [accessed 8.7.04].

*UPF Home*, 2000. <<http://info.wgbh.org/upf/>>, [accessed 13.7.04].



Van Bogart, J., 1995. *Magnetic tape storage and handling*.  
<<http://www.clir.org/pubs/reports/pub54/>>, [accessed 7.7.04].

Van Nuys, C., 2003. The Paradigma Project. *RLG DigiNews* [online], 7(2).  
<<http://www.rlg.ac.uk/preserv/diginews/diginews7-2.html>>, [accessed 30.6.04].

Varniene, R., 2002. Bibliographic control of electronic resources within Lithuanian Integrated Library Information System. In: *The National Bibliography: from print to the digital age, Riga, September 12-13, 2002*, pp. 38-42. Riga: National Library of Latvia.

Vitiello, G., 1993. Legal deposit throughout the European Community: results of an enquiry. *Alexandria*, 5(1), 41-52.

Waibel, G., 2003. *Like russian dolls: nesting standards for digital preservation*. *RLG DigiNews* [online], 7(3).  
<[http://www.rlg.org/legacy/preserv/diginews/v7\\_n3\\_feature2.html](http://www.rlg.org/legacy/preserv/diginews/v7_n3_feature2.html)>, [accessed 27.1.05].

Waters, D. & J. Garrett. *Preserving digital information: report of the Task Force on Archiving of Digital Information commissioned by the Commission on Preservation and Access and the Research Libraries Group*. Washington, D.C.: Commission on Preservation and Access, 1996.

Webb, C., 1997. *Long-term management and preservation of publications on CD-ROMs and floppy disks: technical issues*. <<http://www.nla.gov.au/niac/meetings/tech.html>>, [accessed 13.7.04].

Werf, T. van der, 2000. *The deposit system for electronic publications: a process model*. Nedlib Report Series 6. The Hague: Koninklijke Bibliotheek.  
<<http://www.kb.nl/coop/nedlib/results/DSEPprocessmodel.pdf>>, [accessed 13.7.04].

Werf, T. van der, 2002. Experience of the National Library of the Netherlands. *In: The state of digital preservation: an international perspective: conference proceedings*. Washington, D.C.: Council on Library and Information Resources, pp.54-64.

Werf-Davelaar, T. van. der, 2000. *DNEP 1995-2000: the Dutch Deposit of Electronic Publications*. <<http://www.bic.org.uk/Titia.ppt>>, [accessed 13.7.04].

Wheatley, P. Migration – a CAMiLEON discussion paper. *Ariadne* [online], **29**. <<http://www.ariadne.ac.uk/issue29/camileon/>>, [accessed 13.7.04].

Whitaker, D., 2001a. Nobody expects the Spanish Inquisition. *Library Association Record*, **103**(8), 492-493.

Whitaker, D., 2001b. Time to move on – free legal deposit has had its day. *The Bookseller*, **18 May**, 24-25.

White, B. 2003. *Legal deposit libraries bill*. *Standing Committee F, Wednesday 4 June 2003*.

Wille, N.E., 1994. Legal deposit of electronic publications: questions of scope and criteria for selection. *In: Legal deposit with special reference to the archiving of Electronic publications: proceedings of a seminar organised by NORDINFO and the British Library (Research and Development Department): held at Castle Hotel, Windsor, England 27-29 October 1994*. NORDINFO.

Williamson, R., 1987. *Knowledge Warehouse*. Library and Information Research Report 65. London: British Library.

Williamson, R., 1994. Access and security issues from a publisher's perspective. *In: Legal deposit with special reference to the archiving of electronic publications: proceedings of a seminar organised by NORDINFO and the British Library (Research and Development Department): held at Castle Hotel, Windsor, England 27-29 October 1994*. NORDINFO.

Woldering, B., 2004. The European Library: integrated access to the national libraries of Europe. *Ariadne* [online], **38**. <<http://www.ariadne.ac.uk/issue38/woldering/>>, [accessed 30.6.04].

Woodyard, D., 1998. *Farewell my floppy: a strategy for migration of digital information*. <<http://www.nla.gov.au/nla/staffpaper/valadw.html>>, [accessed 13.7.04].

Woodyard, D., 2002. *Digital preservation: the Australian experience*. <<http://www.nla.gov.au/nla/staffpaper/dw001004.html>>, [accessed 13.7.04].

Working Party on Legal Deposit, 1998. *Report of the Working Party on Legal Deposit*. <<http://www.bl.uk/about/policies/workreplegdep.html>>, [accessed 13.7.04].

Working Party on Legal Deposit, 1999. *Code of practice for the voluntary deposit of non-print publications*. Revised version, September 1999. <<http://www.bl.uk/about/policies/codeprac.html>>, [accessed 13.7.04].

Yakel, E., 2001. Digital preservation. In: Williams, M.E., ed. *Annual Review of Information Science and Technology*. Medford, N.J.: Information Today, 2001, pp. 337-378.

## OTHER WORKS CONSULTED

Armstrong, C. J. & R.E. Lonsdale, 1998. *The publishing of electronic scholarly monographs and textbooks*. <<http://www.ukoln.ac.uk/dlis/models/studies/elec-pub/elec-pub.htm>>, [accessed 19.7.04].

Arvidson, A., K. Persson & J. Mannerheim, 2000. *The Kulturarw3 project – the Royal Swedish Web Archiv3e – an example of “complete” collection of Web pages*. 66<sup>th</sup> IFLA Council and General Conference, Jerusalem, Israel, 13-18 August. <<http://www.ifla.org/IV/ifla66/papers/154-157e.htm>>, [accessed 6.8.04].

Aschenbrenner, A., 2001. *Long-term preservation of digital material: building an archive to preserve digital cultural heritage from the Internet*. Diplomarbeit. Institut für Softwaretechnik und Interaktive Systeme, Technischen Universität Wien.

Association of Learned and Professional Society Publishers, 2004. *Legal Deposit Libraries Bill - update*. <<http://www.alpsp.org/legdep.htm>>, [accessed 8.6.04].

Atkinson, C.J., 2000. The ‘Soft Information Systems and Technologies Methodology’ (SISTeM): an actor network contingency approach to integrated development. *European Journal of Information Systems*, **9**(2), 104-123.

Bazán, C.B., 2003. Legal deposit and the collection of national publications in Argentina. *IFLA Journal*, **29**(3), 227-229.

Beagrie, N., 2001. Preserving UK digital library collections. *Program*, **35**(3), 217-226.

Beagrie, N., 2001. Towards a Digital Preservation Coalition in the UK. *Ariadne* [online], **27**. <<http://www.ariadne.ac.uk/issue27/digital-preservation/>>, [accessed 6.8.04].

Bearman, D., 2000. Intellectual property conservancies. *D-Lib Magazine* [online], **6**(12). <<http://www.dlib.org/dlib/december00/bearman/12bearman.html> >, [accessed 6.8.04].

British Library, 2003. *Support legal deposit extension*. London: British Library. <<http://www.bl.uk/cgi-bin/news.cgi?story=1322>>, [accessed 12.2.03].

Brygfjeld, S.A., 2002. *Access to Web archives: the Nordic Web Archive Access project*. Paper presented at the 68th IFLA Council and General Conference. Glasgow, Scotland. <http://www.ifla.org/IV/ifla68/papers/090-163e.pdf>, [accessed 30.6.04].

Byford, J., 2002. Publishers and legal deposit libraries: cooperation in the United Kingdom since 1610: effective or not? *IFLA Journal*, **28**(5/6), 292-297.

Byford, J., 2003. Making the Bill a reality. *Managing Information*, **10**(2), 12, 14.

CEDARS, 1999. *Collection management issues and the Cedars Project*. <<http://www.leeds.ac.uk/cedars/documents/CIS04.htm>>, [accessed 30.6.04].

CEDARS, 2002. *Cedars guide to: digital preservation strategies*. 2 April 2002. <<http://www.leeds.ac.uk/cedars/guideto/dpstrategies/dpstrategies.htm> >, [accessed 30.6.04].

CEDARS, 2002. *Cedars guide to the Distributed Digital Archiving Prototype*. <<http://www.leeds.ac.uk/cedars/guideto/cdap/guidetocdap.pdf>>, [accessed 30.6.04].

Charlesworth, A., 2002. *The CAMiLEON Project: 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 30.6.04].

Checkland, P. & C. Tsouvalis, 1997. Reflecting on SSM: the link between root definitions and conceptual models. *Systems Research and Behavioural Science*, **14**(3), 153-168.

Chilvers, A., 2000. *Managing long-term access to Digital Data Objects: a metadata approach*. PhD thesis, Loughborough University.

Christensen-Dalsgaard, B., 2004. Web archive activities in Denmark. *RLG DigiNews* [online], **8**(3).

<[http://www.rlg.org/en/page.php?Page\\_ID=176661&Printable=1&Article\\_ID=1241](http://www.rlg.org/en/page.php?Page_ID=176661&Printable=1&Article_ID=1241)>, [accessed 23.6.04].

Connell, N.A.D., 2001. Evaluating soft OR: some reflections on an apparently 'unsuccessful' implementation using a Soft Systems Methodology (SSM) based approach. *Journal of the Operational Research Society*, **52**(2), 150-160.

Council on Library and Information Resources, 2004. *Access in the future tense*.

<<http://www.clir.org/pubs/reports/pub126/pub126.pdf>>, [accessed 13.7.04].

Day, M., 2003. *Collecting and preserving the World Wide Web: a feasibility study undertaken for the JISC and Wellcome Trust*.

<[http://www.jisc.ac.uk/uploaded\\_documents/archiving\\_feasibility.pdf](http://www.jisc.ac.uk/uploaded_documents/archiving_feasibility.pdf)>, [accessed 30.6.04].

Denmark, 1997. *Act on copyright deposit of published works*. Act no. 423, 10 June 1997

(1997.2301-1). <<http://www.kb.dk/kb/dept/nbo/da/pligtafl/Law-text.htm>>, [accessed 16.6.04].

Digitale Bewaring Testbed, 2001. *Digital Preservation Testbed white waper: migration: context and current status*. Den Haag: ICTU.

<<http://www.digitaleduurzaamheid.nl/bibliotheek/docs/Migration.pdf>>, [accessed 30.6.04].

Duval *et al*, 2002. Metadata principles and practicalities. *D-Lib Magazine* [online], **8**(4).

<<http://www.dlib.org/dlib/april02/weibel/04weibel.html>>, [accessed 2.7.04].

Easterby-Smith, M., R. Thorpe & A. Lowe, 1991. *Management research: an introduction*. London: Sage Publications.

Feather, J., 1988. Publishers and politicians: the remaking of the law of copyright in Britain 1775-1842 part I: legal deposit and the battle of the library tax. *Publishing History*, **24**, 49-76.

Fell, C., 2003. *Legal deposit*. ALPSP Advice Note, 20.  
<<http://www.alpsp.org/legalduk.pdf>>, [accessed 16.6.04].

Field, C., 2002. Web archive aims of British Library. Letter to the editor. *The Times*, **11** May 2002. <<http://www.timesonline.co.uk/article/0,,59-293057,00.html>>, [accessed 28.9.04].

Finland, 2003. *Proposal for an act of Parliament: legal deposit act*.  
<[http://www.minedu.fi/julkaisut/tiede/2003/tr14/Legal\\_Deposit\\_Act.pdf](http://www.minedu.fi/julkaisut/tiede/2003/tr14/Legal_Deposit_Act.pdf)>, [accessed 16.6.04].

Flecker, D., 2001. Preserving scholarly e-journals. *D-Lib Magazine*, **7**(9).  
<<http://www.dlib.org/dlib/september01/flecker/09flecker.html>>, [accessed 13.7.04].

Friedlander, A., 2002. Digital preservation looks forward. *Information Outlook*, **September 2002**, 12-18.

Fullerton, J., 1998. *Developing national collections of electronic publications: issues to be considered and recommendation for future collaborative actions*.  
<[http://www.nla.gov.au/nla/staffpaper/int\\_issu.html](http://www.nla.gov.au/nla/staffpaper/int_issu.html)>, [accessed 30.6.04].

Fox, P., 2002. Archiving of electronic publications – some thoughts on cost: guest editorial. *Learned Publishing*, **15**(1), 305.

Glaser, B.G. & A.L. Strauss, 1967. *The discovery of Grounded Theory: strategies for qualitative research*. New York: Aldine de Gruyter.

Granger, S., K. Russell & E. Weinberger, 2000. *Archive models, cost elements & strategies for digital preservation*.

<[http://www.si.umich.edu/CAMILEON/Emulation%20papers%20and%20publications/Cost\\_Elements.doc](http://www.si.umich.edu/CAMILEON/Emulation%20papers%20and%20publications/Cost_Elements.doc)>, [accessed 13.6.01].

Greenstein, D. & D. Marcum, 2000. *Minimum criteria for an archival repository of digital scholarly journals*. Version 1.2, May 15, 2000.

<<http://www.diglib.org/preserve/criteria.htm>>, [accessed 13.7.04].

Gregory, W.J. & G. Midgley, 2000. Planning for disaster: developing a multi-agency counselling service. *Journal of the Operational Research Society*, **51**(3), 278-290.

Groenlandica, 2003. *Legal deposit*. <<http://www.groenlandica.gl/html/pligt2.htm>>, [accessed 6.8.04].

Hakala, J., 2001. Collecting and preserving the Web: developing and testing the NEDLIB harvester. *RLG DigiNews* [online], **5**(2).

<<http://www.ohio.rlg.org/preserv/diginews/diginews5-2.html>>, [accessed 30.6.04].

Hakala, J., 2002. Principles of identification: European perspectives. In: *The National Bibliography: from print to the digital age, Riga, September 12-13, 2002*, pp. 98-107. Riga: National Library of Latvia.

Hallgrímsson, Þ., 2002. The new legal deposit law in Iceland. In: *The National Bibliography: from print to the digital age, Riga, September 12-13, 2002*, pp. 24-27. Riga: National Library of Latvia.

Haynes, D., 1998. *Electronic publications: an agenda for publishers and libraries*.

<<http://www.ddb.de/gabriel/projects/pages/cobra/epubrep.pdf>>, [accessed 19.7.04].

Haynes, D. & D. Streatfield, 1998. A national co-ordinating body for digital archiving? *Ariadne* [online], **15**. <<http://www.ariadne.ac.uk/issue15/digital/>>, [accessed 19.7.04].

Hedstrom, M. and Lampe, C., 2001. Emulation vs. migration: do users care? *RLG DigiNews* [online], **5**(6). < Hedstrom, M. and Lampe, C., 2001. Emulation vs. migration: do users care? *RLG DigiNews*, **5**(6).

<<http://www.rlg.org/preserv/diginews/diginews5-6.html#feature1>>, [accessed 6.8.04].



Hedstrom, M. *et al*, 2003. *Invest to save: report and recommendations of the NSF-DELOS Working Group on Digital Archiving and Preservation*. <<http://delos-noe.iei.pi.cnr.it/activities/internationalforum/Joint-WGs/digitalarchiving/Digitalarchiving.pdf>>, [accessed 8.7.04].

International Federation of Library Associations and Institutions and International Publishers Association, 2004. *Preserving the memory of the world in perpetuity: a joint statement on the archiving and preserving of digital information*. <<http://www.ifla.org/V/press/ifla-ipa02.htm>>, [accessed 13.7.04].

Ireland, 1927. *Industrial and commercial property (protection) act 1927*, S. 178. <<http://www.irishstatutebook.ie/ZZA16Y1927.html>>, [accessed 30.6.04].

Joint Committee on Voluntary Deposit, [n.d.]. *Overview of the British Library's Digital Library System*. < <http://www.alpsp.org/bldig.pdf> >, [accessed 28.9.04].

Jones, M., 2003. Archiving e-journals consultancy – final report: report commissioned by the Joint Information Systems Committee (JISC). October 2003. <[http://www.jisc.ac.uk/uploaded\\_documents/ejournalsfinal.pdf](http://www.jisc.ac.uk/uploaded_documents/ejournalsfinal.pdf)>, [accessed 30.6.04].

Kavčić-Čolić, A., 2003. Archiving the Web: some legal aspects. *Library Review*, **52**(5), 203-208.

Kelle, U., 1997. Theory building in qualitative research and computer programs for the management of textual data. *Sociological Research Online* [online], **2**(2). <<http://www.socresonline.org.uk/2/2/1.html>>, [accessed 3.5.2002].

Kranch, D.A., 1998. Beyond migration: preserving electronic documents with digital tablets. *Information Technology and Libraries*, **17**(3), 138-145.

Kunze, J. A., 2001. A Metadata kernel for electronic permanence. *Journal of digital Information*, **2**(2). <<http://jodi.ecs.soton.ac.uk/Articles/v02/i02/Kunze/>>, [accessed 8.7.04].

Ledington, P. & J. Ledington, 1999. Extending the process of comparison in Soft Systems Methodology. *Journal of the Operational Research Society*, **50**(11), 1149-1157.

LeFurgy, W.G., 2002. Levels of service for digital repositories. *D-Lib Magazine* [online], **8**(5). <<http://www.dlib.org/dlib/may02/lefurgy/05lefurgy.html>>, [accessed 13.7.04].

*Legal deposit in Denmark: published on the occasion of the 63<sup>rd</sup> IFLA Council and General Conference, Copenhagen 1997*. [Copenhagen: Royal Library].

Manzoni, M., ed. 1992. *A synthesis on legal deposit and its practice in the EC Member States*. Luxembourg: Commission of the European Communities.

Mattila, R., 2002. The main points of the future legal deposit act in Finland. In: *The National Bibliography: from print to the digital age, Riga, September 12-13, 2002*, pp. 28-30. Riga: National Library of Latvia

McCue, J., 2002. Can you archive the Net? *The Times*, **29 April**, 4-5.

Mole, Chris, 2003. *MP in bid to capture electronic publications for future generations*. Press release, 6 February 2003. <[http://www.chrismolemp.org.uk/news\\_60203](http://www.chrismolemp.org.uk/news_60203)>, [accessed 6.2.03].

Moore, T.T. & F.H. Gregory, 2000. Cultural problems in applying SSM for IS development. *Journal of Global Information Management*, **7**(4), 14-19.

National Library of Scotland, 1999. *National Library of Scotland Department of General Collections legal deposit*. Edinburgh: National Library of Scotland.

New Zealand, 2003. *National Library of New Zealand act (Te Puna Mātauranga o Aotearoa) act 2003*. <<http://www.natlib.govt.nz/files/Act03-19.pdf>>, [accessed 19.7.04].

Pandit, N.R., 1996. The creation of theory: a recent application of the Grounded Theory method. *The Qualitative Report* [online], **2**(4). <<http://www.nova.edu/QR/QR2-4/pandit.html>>, [accessed 3.5.02].

- Paskin, N., 2003. On making and identifying a "copy". *D-Lib Magazine*, **9**(1). <<http://www.dlib.org/dlib/january03/paskin/paskin/01paskin.html>>, [accessed 8.7.04].
- Pinion, C., 1986. *Legal deposit of non-book materials*. Library and Information Research Report 49. London: British Library Research and Development Department.
- Powell, D. J., 2003. Voluntary deposit of electronic publications: a learning experience. *Learned Publishing*, **16**(2), 149-152.
- Samzugui, A., 2003. An appraisal of legal deposit laws in Tanzania. *University of Dar es Salaam Library Journal*, **5**(1). <<http://www.inasp.info/ajol/journals/udslj/vol5no1.html>>, [accessed 13.7.04].
- Sergeant, D.M., 2001. *Synthesis of findings from the test site phase of Cedars*. <<http://www.leeds.ac.uk/cedars/testsitesSynthesis.html>>, [accessed 30.6.04].
- Seville, C. & E. Weinberger, 2000. Intellectual property rights lessons from the CEDARS project for digital preservation: presentation to the CEDARS Conference, York 8 December 2000. <<http://www.rlg.org/events/pres-2000/weinberger.html>>, [accessed 30.6.04].
- Seville, C. & E. Weinberger, 2000. Intellectual property rights lessons from the CEDARS project for digital preservation. *New Review of Academic Librarianship*, **6**, 217- 223.
- Shenton, H., 2000. *From talking to doing: digital preservation at the British Library. Preservation 2000: An International Conference on Preservation and Long Term Accessibility of Digital Materials. 7/8 December 2000, York, England*. <<http://www.rlg.org/events/pres-2000/shenton.html>>, [accessed 28.9.04].
- Siebert, F. S., 1965. *Freedom of the press in England 1476-1776*. Urbana: University of Illinois Press, p. 51.

- Steenbakkers, J., 2002. Preserving electronic publications. *Information Services & Use*, **22**, 89-96.
- Strauss, A. & J. Corbin, 1998. *Basics of qualitative research: techniques and procedures for developing Grounded Theory*. 2nd ed. Thousand Oaks: Sage Publications.
- Vitiello, G., 2001. A European policy for electronic publishing. *The Journal of Electronic Publishing* [online], 6(3). <<http://www.press.umich.edu/jep/06-03/vitiello.html>>, [accessed 6.8.04].
- Watson, N., 2000. *The preservation and legal deposit of electronic resources: is legal deposit desirable or possible in the case of electronic resources*. MSc dissertation, Department of Information Studies, University of Sheffield.
- Weinberger, E., R. Clayton & R. Anderson, 2002. *A security policy for a digital repository*. <<http://www.cus.cam.ac.uk/~ew206/security-policy.html>>, [accessed 8.7.04].
- Willison, I.R., 1999. Legal deposit: a provisional perspective. *Publishing History*, **45**, 5-34.
- Winter, M.C., D.H. Brown & P.B. Checkland, 1995. A role for soft systems methodology in information systems development. *European Journal of Information Systems*, **4**(3), 130-142.
- Woodyard, D., 2004. Significant property: digital preservation at the British Library. *Vine*, **34**(1), 17-20.

## **APPENDICES**