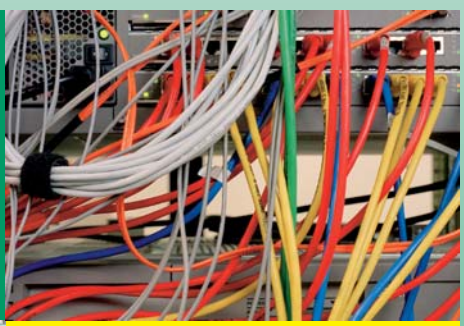
IFLA

Publications 119

Networking for Digital Preservation



Current Practice in 15 National Libraries

Ingeborg Verheul

K · G · Saur

International Federation of Library Associations and Institutions
Fédération Internationale des Associations de Bibliothécaires et des Bibliothèques
Internationaler Verband der bibliothekarischen Vereine und Institutionen
Международная Федерация Библиотечных Ассоциаций и Учреждений
Federación Internacional de Asociaciones de Bibliotecarios y Bibliotecas

IFLA Publications 119

Networking for Digital Preservation: Current Practice in 15 National Libraries

Ingeborg Verheul

K · G · Saur München 2006

IFLA Publications
edited by Sjoerd Koopman

Recommended catalogue entry:
Networking for Digital Preservation: Current Practice in 15 National Libraries. Ingeborg Verheul / [International Federation of Library Associations and Institutions]
– München : Saur, 2006, 268 p. 21 cm
(IFLA Publications ; 119)
ISBN 3-598-21847-8

Bibliographic information published by Die Deutsche Bibliothek
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the Internet at http://dnb.ddb.de.

♾

Printed on permanent paper
The paper used in this publication meets the minimum requirements of American National Standard – Permanence of Paper for Publications and Documents in Libraries and Archives ANSI/NISO Z39.48-1992 (R1997)

Printed / Bound by Strauss GmbH, Mörlenbach

ISBN 13: 978-3-598-21847-7
ISBN 10: 3-598-21847-8
ISSN 0344-6891 (IFLA Publications)

Foreword

Increasingly, libraries are having to deal with digital materials that need to be safeguarded not only for our generation, but also for the generations to come. Digitised images and born-digital objects need to be preserved for future access and use. For national libraries, safeguarding the digital heritage is a major issue because of their legal task to preserve the national heritage of a country in paper or digital form. One particular problem with digital material is the very short lifespan of the carriers. Moreover the hardware and software needed to render digital materials is undergoing constant technological development so that existing systems rapidly become obsolete. Therefore safely storing the digital heritage whilst still ensuring access for future use, requires that libraries not only need to have a trusted digital repository system in place, but also an ongoing R&D programme aimed at developing preservation strategies.

Despite being a relatively new field in the library sector, digital preservation is becoming increasingly important in the everyday routine of the library. Cooperation and knowledge dissemination activities on digital preservation issues are starting to emerge, but can still be intensified. An overview of recent developments in the field of digital preservation could be a valuable aid when planning digital preservation activities: Does the day-to-day practice in storing and accessing digital objects illustrate a mutual need for certain standards? Are there currently any standards for the development and building of digital repositories, and how are these being applied? Are there common standards in research on permanent access? Or is it still too early to speak of standards, and is it only possible to distinguish best practices?

In 2004−2005, *Koninklijke Bibliotheek* conducted a survey for the IFLA-CDNL Alliance for Bibliographic Standards (ICABS) on the use and development of standards in digital archiving within the international library world. The survey resulted in this overview of the current state of affairs in 15 libraries (baseline July 2005). The libraries involved are the national libraries of Australia, Austria, Canada, China, Denmark, France, Germany, Japan, the Netherlands, New Zealand, Portugal, Sweden, Switzerland, the United Kingdom and the United States of America.

The survey addresses both operational and R&D activities aimed at digital preservation. The main focal points are the use of standards in operational safe place environments and the state of affairs on permanent access strategies, such as migration and emulation. Apart from providing information on the status, function and organisational embedding of digital repositories in the library organisations, the survey also gives an overview of the current national and international R&D projects.

A comparison with two earlier surveys on developments in digital preservation (Neil Beagrie for CLIR and the Library of Congress in 2003 and the survey report of the PREMIS Working Group in 2004) leads to the conclusion that digital preservation in the year 2005 is becoming increasingly integrated in the day-to-day library activities. A growing number of national libraries consider it their mission to safeguard not only the paper cultural heritage, but the digital cultural heritage as well. Some of them are still focussing on the national heritage, whereas others have a more international perspective. Digital preservation is a rapidly developing discipline and although there are not many official standards for it yet, emerging best practices might develop into standards within a few years.

Since 2003, considerable progress has been made in building digital repositories, or networks of interconnected computer systems. A second development is the emergence of large networks for national and international cooperation. At first, the focus of these was limited to the cultural heritage sector, but now the perspective is broadening, and cooperation between the cultural heritage sector and the science sector is becoming increasingly important. Although such cooperative projects or platforms often started with a focus on knowledge dissemination and knowledge sharing, it is likely that they could be used in the near future for the joint development of tools for permanent access as well.

In 2004−2005 the National Library of Australia (NLA) also carried out a survey for ICABS on digital preservation. This survey focused on the availability of suitable guidance documents for preserving digital materials. During the ICABS Session of the IFLA World Library and Information Congress 2005 in Oslo, the KB and NLA reports were presented together in one lecture. The NLA report will only be available online. However, since both surveys emerged from the ICABS Alliance, a summary of the NLA report can be found in the Appendices of this volume.

In 2003, *Koninklijke Bibliotheek*, together with five other national libraries, has been one of the co-founding partners of ICABS, the strategic IFLA-CDNL alliance on bibliographic standards. Now that we are more than half way through the first term period of ICABS and the first evaluative thoughts on continuation and enhancement of the alliance emerge, it is important to stress the advantages and opportunities an international cooperative like ICABS offers to knowledge sharing within the library world.
Within its mission, ICABS forms a framework to stimulate the development of new strategies and to promote different aspects of the long-term preservation of electronic resources. With this survey on best practices in digital preservation *Koninklijke Bibliotheek* hopes to provide a worthwile contribution to ICABS.

Ingeborg Verheul prepared both the survey and the report on behalf of *Koninklijke Bibliotheek.* I am extremely grateful to her for providing us with this broad international perspective on digital preservation.

Dr Wim van Drimmelen

Director General
Koninklijke Bibliotheek

December 2005

Foreword

The IFLA-CDNL Alliance for Bibliographic Standards (ICABS) − an alliance founded jointly by the International Federation of Library Associations (IFLA), the Conference of Directors of National Libraries (CDNL) and the national libraries of Australia, Germany, the Netherlands, Portugal, the United Kingdom, and the United States of America − is a continuation of the late UBCIM Core Activity (with respect to Bibliographic Standards), parts of the Universal Dataflow and Telecommunications Core Activity, and the CDNL digital initiatives which involved preservation and digital resource management, access mechanisms, interoperability and much more.

The alliance has a strategic focus and offers a practical vehicle for improving international coordination and steering developments in these key areas. The alliance aims to maintain, promote, and harmonise existing standards and concepts related to bibliographic and resource control, to develop strategies for these, and to advance the understanding of issues related to the long-term archiving of electronic resources, including the promotion of new and recommended conventions for such archiving.

Within ICABS, *Koninklijke Bibliotheek* (KB), the national library of the Netherlands, and the National Library of Australia have been exploring the requirements and conditions for the long-term archiving of electronic resources. Moreover both libraries have been exploring and promoting strategies, methods, and standards for migration and emulation.

In this context KB conducted an international survey on recent developments in digital preservation in 15 national libraries. This report presents the outcomes of this study based on recent publications, information about ongoing projects and survey results.

We hope that this report will serve as a useful guide for other libraries and cultural heritage institutions as they face the future challenges of long-term archiving and preservation.

Renate Gömpel

Chair of ICABS Advisory Board
Die Deutsche Bibliothek

Table of contents

Websites: ICABS (http://www.ifla.org/VI/7/icabs.htm)
PADI (http://www.nla.gov.au/padi/)

All URLs in this publication were valid as of December 1, 2005

Acknowledgements

I would like to thank a number of colleagues of *Koninlijke Bibliotheek* who helped with this survey. Thank you, Jeffrey van der Hoeven, Ingrid Dillo, Hans Jansen, Marco de Niet, Erik Oltmans, Judith Rog, Dennis Schouten, Johan Stapel, Johan Steenbakkers, Astrid Verheussen, Caroline van Wijk and, last but not least, Hilde van Wijngaarden, for providing ideas, input and comments, and in some cases draft text for some chapters.

Special acknowledgements are also in place for all digital preservation colleagues of the national libraries involved for their cooperation, additional information, critical review and friendly comments. Thank you Bettina Kann and Max Kaiser of the Austrian National Library; Peter Rochon and Deane Zeeman of the Library and Archives of Canada; Wang Zhigeng, Sun Wei and Qi Xin of the National Library of China; Birte Christensen-Dalsgaard of *Statsbiblioteket* (Denmark); Birgit Henriksen and Grede Jacobsen of *Det Kongelige Bibliotek* (Denmark); Catherine Lupovici of *Bibliothèque nationale de France*; Reinard Altenhöner, Kathrin Ansorge, Hans Liegmann, Susanne Oehlschlaeger, Thomas Wollschlaeger of *Die Deutsche Bibliothek* (Germany), Hisayoshi Harada and staff of the National Diet Library (Japan), Steve Knight and Sudha Rao of the National Library of New Zealand; José Borbinha and Paulo Leitao of *Biblioteca Nacional* (Portugal); Gunilla Jonsson and Johan Mannerheim of *Kungliga Biblioteket* (Sweden); Hansueli Locher and Barbara Signori of *Schweizerische Landesbibliothek* (Switzerland); Adam Farquhar, Roderic Parker and Helen Shenton of British Library (UK) and William Lefurgy and staff of Library of Congress (USA).

A special word of thanks is due to Neil Beagrie of JISC, Rebecca Guenther and Priscilla Caplan of PREMIS for their additional comments on the analysis section, to Barbara Sierman of *Koninklijke Bibliotheek* for mind mapping. And of course thank you, Colin Webb of the National Library of Australia for the red line of cooperation through time zones.

While gratefully acknowledging all of the inputs for this survey, any errors, ambiguities, misinterpretations or misconceptions are considered entirely the author's view.

Ingeborg Verheul

Research & Development Division
Digital Preservation Department
Koninklijke Bibliotheek

Introduction

Context

The research has been carried out within the framework of the IFLA-CDNL Alliance for Bibliographic Standards (ICABS).[1] It has been funded by structural funding from *Koninklijke Bibliotheek*, the National Library of the Netherlands, (KB) has received from the Dutch Ministry of Education, Culture and Science (OCW) for cooperation between KB, the National Archives, and the Netherlands Institute for Cultural Heritage (ICN) in the field of 'preservation' in the broadest sense. This cooperation concerns both paper and digital preservation. In 2004 part of this funding was earmarked for KB research on recent international developments on standards and best practices in digital preservation within the library sector, which resulted in this ICABS survey.

ICABS
ICABS is a strategic alliance of national libraries that focuses on improving the international coordination of bibliographic standards. The mission of ICABS is to maintain, promote and harmonise current standards on bibliographic and resource control. The efforts undertaken within the framework of ICABS have to stimulate the development of new strategies and as such promote different aspects of the long-term preservation of electronic resources and the promotion of standards.

ICABS was founded in 2003 to continue and expand the coordination work formerly done by the IFLA Universal Bibliographic Control and International Marc (UBCIM) and Universal Dataflow and Telecommunications Core Activity (UDT). Since the 1970s, UBCIM has focused on coordinating activities to develop systems and standards for bibliographic control at a national level and on the international exchange of bibliographic data. UBCIM cooperated closely with UDT. The IFLA Core Activity UDT mainly focused on analysis and research of technologies and standards for interoperability, information retrieval, digitising and metadata. The activities of UBCIM and UDT have been continued within ICABS. UBCIM and UDT received structural financial support from the Conference of Directors of National Libraries (CDNL). After the disbanding of these two Core Activities, the Committee on Digital Issues (CDI) was founded within CDNL. The committee's work on bibliographic standards and digital preservation is now being incorporated into the ICABS mission, while the Committee's work on deposit agreements will be continued separately by the National Library of Australia (NLA).

[1] ICABS: See: http://www.ifla.org/VI/7/icabs.htm

ICABS and digital preservation

Cultural heritage institutions are now becoming increasingly aware of the urgency of digital preservation: the long-term preservation of and permanent access to electronic objects. Several institutes have taken action to design and build trustworthy and permanent electronic repositories (sometimes referred to as 'safe places'). International projects such as NEDLIB (Networked European Deposit Library)[2] have improved cooperation, knowledge dissemination and standardisation in this field.

ICABS aims to promote knowledge dissemination on the long-term preservation of and permanent access to digital objects. Within ICABS, KB has assumed the responsibility for this task and is carrying it out in close cooperation with NLA.

[2] NEDLIB: See: http://www.kb.nl/coop/nedlib/

Aim, scope and methodology

Aim

This survey is the result of a research project KB carried out for ICABS in 2004−2005 on the use and development of standards in digital archiving within the international library world. This has resulted in an overview of current practice in digital preservation in 15 national libraries.

The main focal points in the survey are the use of standards in operational safe place environments (the current state of affairs of initiatives on electronic repositories), and the current state of affairs with respect to, and the need for permanent access strategies (such as migration and emulation). The survey also provides an overview of current national and international projects on digital preservation. The study made use of desk research, interviews and an analysis of new developments.

Scope

This KB survey can be seen as a continuation of two previous surveys: 1) the outcome of the research of Neil Beagrie on national digital preservation initiatives in 2003 at the request of the Library of Congress and the Council on Library and Information Resources[3] and 2) the outcome of the research by the OCLC/RLG PREMIS Working Group Preservation Metadata: Implementation Strategy (PREMIS) in 2003, the results of which were published in 2004.[4] The design of this survey combined subjects from both publications.
The report of Neil Beagrie was written to provide NDIIPP, the national digital preservation programme in the USA, with input and reference material to outline specific points of interest in an international context. It described current practice in digital preservation in Australia, France, the Netherlands and the UK in March 2002. The report provides a good perspective for designing a topical study. It also provided a good comparison of what had changed in three years, not only due to its international perspective but also the set of forecasts it provided about developments in digital preservation in the near future. Beagrie's report formed the basis for the sections on legal deposit, funding and current activities in this ICABS survey.

[3] Beagrie, Neil. 2003. National Digital Preservation Initiatives: An overview of the developments in Australia, France, the Netherlands, and the United Kingdom and of related international activity. See: Appendices – References.

[4] OCLC/RLG PREMIS Working Group. 2004. *Implementing Preservation Repositories For Digital Materials: Current Practice And Emerging Trends In The Cultural Heritage Community*. See: Appendices – References.

The PREMIS survey was carried out to provide a background study to examine and evaluate strategies for managing and exchanging preservation metadata. The examination would provide the context for the development of a preservation metadata scheme and a data dictionary. These were published in May 2005.[5] The PREMIS survey offers the perfect layout for describing the current developments in the field of digital preservation. Its set of questions and choice of subjects was the result of intensive international cooperation and detailed discussions between experts in the field. Therefore it provides a thorough coverage of all aspects involved in current digital preservation activities. The sections on digital repositories and preservation strategies in this ICABS survey are modelled on the PREMIS approach. As this survey is based on these two previous reports it provides an adequate opportunity to visualise the recent changes and progress within the field of digital preservation.

Whereas Beagrie and PREMIS both focus on national libraries and other cultural heritage institutions, the ICABS survey only considers national libraries. This is because ICABS primarily serves this type of library. The emphasis is on the common activities of all national libraries: the handling of the electronic equivalent of print and born-digital materials. The handling of audio or visual equivalents, which is also an important task of some national libraries, falls outside of the scope of this survey.

Methodology

This report contains two parts. Part one provides an analysis of the current state of affairs at 15 national libraries. Part two contains the detailed overviews per library. Each overview is divided into four sections: General; Digital repository; Preservation strategies; and Current activities.

Within the overviews, the general section contains information on recent developments with respect to deposit laws for digital objects (online and offline); the embedding of digital preservation activities (building of the repository and research) in the library organisation, and how digital preservation activities are financed.

The second section, on digital repositories, contains information on: the status of the digital repository; services provided; depositing; software and OAIS; materials; metadata and metadata schemes; and access.

[5] OCLC/RLG PREMIS Working Group. 2005. *Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group.* Dublin, O.: OCLC Online Computer Library Center Inc. See: http://www.oclc.org/ research/projects/pmwg/premis-final.pdf

The third section, on preservation strategies, describes which strategies are currently applied in the processing of digital objects and the plans for the future. The fourth section, on current initiatives, gives an overview of current projects and working groups in which the national libraries are involved or in which they participate (national and international projects) and a general impression of institutes active on digital preservation in other cultural heritage sectors (museums, archives and audiovisual institutions).

The fifteen libraries involved in the study are (in alphabetical order of country names) the national libraries of: Australia, Austria, Canada, China, Denmark, France, Germany, Japan, the Netherlands, New Zealand, Portugal, Sweden, Switzerland, UK, and USA.

Initially a selection of the national libraries to survey was made on the following basis: i) ICABS partners; ii) national libraries known to be active in the area of digital preservation. Obviously a selection inevitably leads to a somewhat subjective interpretation. The survey does not pretend to be complete and from a practical point of view is limited in size. A limit of fifteen libraries seemed feasible within the framework and time-span of this research.[6] In the next stage an overview was made of all libraries involved, based on the websites of the national libraries, the information provided by the research of Beagrie and the survey of PREMIS, and additional literature. All of the topics were then completed and updated using information gathered by phone or e-mail. The overviews were then sent to the contact persons in the respective libraries for commenting, correction and addition. Lastly the final version of the report was sent for review to the ICABS partners, the contact partners in the national libraries surveyed, and the authors of the CLIR and PREMIS reports.

The overview represents the current state of affairs in July 2005. In the analysis part, the libraries mentioned to illustrate the general observations are used as examples. This has been done to avoid endless enumerations of the information that can be found in the separate overviews per library.

Within the framework of ICABS, NLA also carried out a survey in 2004−2005. This concerned the availability of guidelines for digital preservation activities and various digital objects. This report will only be available online and a summary of it is included in the Appendices of this report.

[6] Other national libraries presently known to be active in the field of digital preservation but not considered in this study due to scalability and other reasons are, for example, the National Libraries of Finland, Norway, South Korea, Italy, Scotland, and Slovakia.

Practical definitions

In 2005, a wide range of concepts and definitions are still used in the field of digital preservation. Despite the many sources available, a standard definitive glossary has yet to be provided. This might impede communication and knowledge sharing. To illustrate how specific terms and concepts are used in this survey, the main basic concepts are explained below. These definitions are partly based on daily practice in the Digital Preservation Department of KB, and have been compared with and complemented by the definitions used in a selection of the existing glossaries.
The list only serves to provide a standard reference point for this survey and does not pretend to be complete or to form the basis for a final glossary. For practical reasons the list is subdivided in three themes: the discipline, the material and the system.

The discipline
Digital preservation or *long-term preservation* is the general term for all activities concerning the maintenance and care for/curation of digital or electronic objects, in relation to both storage and access. *Long-term* means five years or more; *short-term* is less than five years. Within digital preservation the main activities can be divided into *digital archiving* and *permanent access*.

Digital archiving means the process of backup and ongoing maintenance of digital objects and the associated software and hardware, as opposed to strategies for digital preservation. (Source: *DPC Handbook*).

Permanent access is usually paired with the term *digital preservation,* indicating that preservation is only half the battle. Within the digital environment, providing permanent access and adequate rendering of the digital object will be one of the greatest challenges, given the technological changes that have and will continue to occur. (Source: *CENDI report*).

Day-to-day activities are those activities concerned with the daily operational workflow of handling the digital objects in the framework of digital preservation within the library.

Preservation strategies (such as migration and emulation) are methods for keeping stored material permanently accessible. It refers to all techniques that provide more than would be obtained by merely storing the digital objects and never looking at them again.

Developing **tools** means, among other things, developing a device that provides a mechanical or mental advantage in accomplishing a task. Examples are: UVC, JHOVE, PANIC, Preservation Manager. (Source: Wikipedia).

The material

Digital objects is a general term used for the body of digital material that is subject to digital preservation, both *digital publications* and *digital records*; both *online* and *offline*; both *born-digital* and *digitised.*

Digital publications (or e-publications) are digital materials which have been released for public access (publication) and made available free of charge or for a fee.

Digital records refer to the type of electronic or digital materials the archival sector usually has to deal with. Digital records are created digitally in the day-to-day business of the organisation and assigned formal status by the organisation. They may include, for example, word-processing documents, e-mails, databases, or intranet web pages. (Source: *DPC Handbook*).

Online refers to digital objects that are connected to some larger network or system.

Offline An offline publication is not connected to or accessible through a network, but stored on a stand-alone carrier (CD, DVD, tape, optical, hard disk). It means that a digital object can be held in the hand and put on a shelf, without transforming it into any other form (paper or microfilm, etc.).

Born-digital refers to materials which are not intended to have an analogue equivalent of the object, either as the originating source or as a result of conversion to analogue form. (Source: *DPC Handbook*).

Digitised implicates the transformation of the information of the original physical, analogue carrier into a digital form. Digitised refers to the mode of production.

The system

Digital repository or *electronic repository* is the system (or combination of systems) that provides long-term storage and preservation of and permanent access to digital objects.

There are four stages in the development of a digital repository and these can be more or less overlapping: *design*, *development*, *implementation* and *production.*

Design phase is the stage in which plans, models and workflows are worked out and written down.

Development phase is the stage in which the repository is built.

Implementation phase is the stage in which the system is implemented in the library, and linked to other systems and processes in the organisation.

Production phase is the stage in which the repository is able to perform the day-to-day operational processes for handling the digital objects, both now and in the future.[7]

[7] Useful definitions can be found in the *Handbook* on the website of the Digital Preservation Coalition (UK): Jones, Maggie and Neil Beagrie. 2001. *Preservation Management of digital materials. A handbook*, London: The British Library. Updated frequently at the DPC-website. See: http://www.dpconline.org/graphics/intro/definitions.htm and in the NEDLIB glossary.

Borbinha, José Luis, Fernando Cardoso and Nuno Freire. 2000. *NEDLIB Glossary.* Webpublication: see: http://www.kb.nl/coop/nedlib/glossary.pdf and Clavel-Merrin, Genevieve. 2000. *The NEDLIB list of terms.* Study jointly funded by the European Commission's Telematics for Libraries*,* Den Haag: Koninklijke Bibliotheek/ NEDLIB Consortium. The NEDLIB Report Series, nr. 7. See: http://www.kb.nl/coop/nedlib/results/NEDLIBterms.pdf.

Wikipedia also gives a very up-to-date explanation of a lot of definitions in digital preservation, but is broader than digital preservation only. See: http://en.wikipedia.org/wiki/Main_Page. The PADI website Thesaurus is useful for more general descriptions of terms, combined with links and further information per subject. See: http://www.nla.gov.au/padi/topics/thesaurus.html. The Cornell Tutorial of Digital Preservation gives a good overview of existing glossaries. See: http://www.library.cornell.edu/iris/tutorial/dpm/terminology/g_resources.html. In: CENDI. 2004: Hodge, Gail and Evelyn Frangakis. 2004. *Digital Preservation and Permanent Access to Scientific Information: The State of the Practice* the often problematic use of terms and concepts in digital preservation is made clear in Chapter 4.1 ('Archiving Concepts and Definitions').

I. Analysis

1. General

Legal deposit legislation

Legal depositing can strongly influence various aspects of digital preservation. For example, legal deposit legislation can provide libraries with a firm basis for making agreements with depositors, or for setting up guidelines for the deposition of materials. The PADI website presents a clear case for legal deposit, especially in relation to digital preservation:

'Legal deposit is a statutory provision which obliges publishers to deposit copies of their publications in libraries in the country in which they are published. The principle of legal deposit is established in international conventions and in the national legislation of many countries, and aims to ensure that access to a nation's published cultural material in libraries and archives is preserved.

Increasingly, material is being published in digital form: this material also needs to be collected and preserved to ensure a complete record of a nation's published cultural material. Legal deposit legislation therefore requires a new legal framework in order to encompass digital publications. The complications associated with the collection and control of electronic materials, together with the lack of a comprehensive legal model, have made drafting appropriate legislation problematic and slow.

Because online information can easily be distributed and copied and is often multiple accessible, and because of the lack of a physical item to "deposit", new major issues occur, in coherence with the renewal of the legal deposit legislation: copyright, preservation requirements, public access, scope of coverage, method of collection, protection of publishers' rights, penalties, and implementation of revised legislation.' [8]

For many of the fifteen countries covered by this survey, 2005 seems to be a significant year with respect to developments in legal deposit.

Of the 15 countries surveyed, 14 have a legal instrument of some kind for the deposit of publications concerning their specific country and people. This can be a legislation based on legal deposit laws, on copyright acts or on national library acts. The Netherlands is the only country without any legislation on depositing. *Koninklijke Bibliotheek* (The National Library of the Netherlands) collects publications on a voluntary basis.

[8] PADI website: See: http://www.nla.gov.au/padi/topics/67.html

In most countries the national library is the national institute appointed by law to receive deposit material. In the USA and Japan, the parliamentary library has been given this task. In some cases several libraries have a deposit role (for instance in France, Switzerland, Australia, the UK and Germany). This is often due to the federal structure of the countries involved. In most of these cases however, the national library is considered to bear the main responsibility. In France there is one exception: here the legal deposit of dissertations and theses is a task of the university libraries and not the national library.

In some countries more than one library or institute is appointed as the main deposit institute. For example, Denmark has two deposit libraries which focus on different types of material. In Canada the national library and the national archive have been merged into a single institution that has been awarded the deposit role.

If a country has legislation covering deposit, the national library is always heavily involved, together with other institutions and organisations, in preparing new amendments to the law. Such discussions often take place in special working groups and can be quite lengthy, especially where digital objects are involved (see PADI citation earlier). Once the legislation has been approved, several more years can elapse before it is fully effected.

Legislation for digital objects is often divided into that covering offline publications and that covering online publications. The initial legislative amendments often concern offline publications only. This is probably because these type of publications are closer to analogue publications than their online counterparts. An exception is China, which has one law that is covering digital objects in general.

Offline publications
In the 1988−2000 period, most of the countries surveyed introduced new legislation or amendments to existing legislation to cover offline or physical electronic publications: USA (1988), Germany (1990), Switzerland (1992), Sweden and France (1993) and Austria and Japan in 2000. Both Australia and the UK have voluntary agreements for offline publications. In the UK this concerns an interim agreement under the Voluntary Code of Practice effective since 2000. This will continue until new legislation becomes effective (expected in 2006).
Portugal is still developing legislation for offline publications.

Online publications

Up until now only China and Canada have a legal deposit legislation that also includes online publications and digital records. In China this has been regulated since 1996 in a notice on the deposit of electronic publications by the National Copyright Administration. In Canada the Library and Archives of Canada Act provided for this in 2004.

In 2002, Sweden established a special government agreement for the harvesting of Swedish web pages and online databases. Online publications, however, have been deposited in Sweden on the basis of voluntary agreements up until now. The other countries are currently preparing new legislation for online publications (Austria, Switzerland) or waiting for it to become effective soon. Three countries expect legislation for online publications to come into force in 2005: Denmark and France on 1 July, and New Zealand by the end of the year. Sweden expects the law for online publications to come in force by the end of 2005 or early 2006. Japan expects this to happen in 2006, as do the UK and Germany. Although Canada already has an effective legislation, it expects the 2004 law to be extended in the near future to include other types of electronic publications that have not been subject to deposit until now, for example maps.

In 2007 legal deposit legislation on digital objects (both offline and online) will be achieved in almost half of the countries surveyed (7 out of 15). Their experience and approach might help to accelerate developments in the other countries. Portugal expects their future law on depositing to cover the legal deposit of offline publications and a selective deposit of online publications.

The National Library of Australia has signed voluntary agreements with publishers of online publications, and has thus negotiated the right to harvest online publications. *Koninklijke Bibliotheek* has signed special agreements with individual international publishers since 2002, so that it can collect and preserve their electronic journals in its digital repository. The first agreement was concluded with Elsevier Science Publishers. Since then ten more contracts have been signed.

Copyright legislation

Long-term preservation is impossible without copying original material. Very often electronic material is protected against copying by copyright law or even by technical restrictions that prevent copying. Three countries are actively preparing future amendments to their copyright legislation, so that national libraries are allowed to make copies for preservation and long-term access purposes, under special circumstances. In 2004 *Die Deutsche Bibliothek* in Germany signed a special agreement on this with the German national music union and the German national publishers union. The *Bibliothèque nationale de France* has also formulated plans in this area. In the USA the Library of Congress recently formed a study group to look at the Copyright Act and the consequences of this for electronic media (May 2005).

The appearance and rapid growth of digital publications in general, and Internet as a publication and communication medium in recent years, has had a considerable influence on the development of the legal deposit obligation for national libraries, whose perspective on safeguarding the written cultural heritage of their country has expanded enormously. This has led to a lot of activities in the field of legal deposit legislation.
Countries actively involved in web-harvesting and web-archiving activities at an early stage (for instance Sweden, Denmark and France) were the first to expand their deposit legislation to cover digital publications. An exception is the Netherlands. Despite not having a legal deposit legislation it also started activities on depositing digital objects at a very early stage. This might have been helped by the good contacts between *Koninklijke Bibliotheek* and several individual publishers. However, as little experience on web archiving has been gained in the Netherlands to date, the possible consequences of not having a Dutch legal deposit legislation of some kind on these specific digital preservation activities cannot be predicted yet.

Organisational embedding

Most of the current policy plans of the institutes surveyed, stress the importance of developing or expanding the scope of digital activities as a major issue over the next 2 to 5 years.
This is also reflected in the fact that all 15 libraries have at least one department, unit or division that refers to digital objects in some way.[9]

For three libraries the digital aspect is so important that they have a director of E-strategies amongst their board members: the national libraries of the UK, New Zealand and the Netherlands. Three other libraries have an e-oriented Division in the top level of their organisational structure: the British Library (e-strategies and IT Directorate), the National Library of New Zealand (Electronic Services Directorate) and the *Statsbiblioteket* in Denmark (Digital and Web Resources

[9] The Digital Archiving Unit, Digital Collections Management Branch and the Digital Preservation Unit in Australia; the Digital Preservation Department in Austria; Digital Collection Technology and Digital Preservation Technology in Canada; The Electronic Information Services Section in China; the Digital Objects Section and the Digital and Web resources Section in Denmark; the Digital Library Department in France; the Digital Information Planning Office and the Digital Library Division in Japan; the e-Depot Unit and the Digital Preservation Department in the Netherlands; the Electronic Services Department in New Zealand; the Electronic Publications Unit in Portugal; the E-Helvetica Unit in Switzerland; the E-Strategies Directorate in the UK; the Digital Resources Management and Planning Division in the USA.

Division). As digital preservation also involves routine library tasks (acquisition, cataloguing, collection care, IT), none of the libraries have placed all activities on digital preservation strictly within one unit.

The position of the digital-oriented units in the libraries varies. They are positioned under collection-oriented divisions, process-oriented divisions, IT-oriented divisions or strategy-oriented divisions.[10]

The term 'digital-oriented' is used deliberately here, since the digitising and digital preservation activities within one unit can be highly interrelated. In many cases the libraries started with digital library projects or programmes which focussed on digitising materials for access. The preservation aspect of the digitised materials arose later almost as a matter of course. Others became involved in digital preservation by carrying out projects that focussed mainly on infrastructural aspects. These projects resulted in the construction of repository systems suitable for long-term preservation and access (kopal in Germany, Digital Services Projects in Australia, National Digital Heritage Archive in New Zealand). Of course the boundaries between these two types of projects are not always that strict. In Japan for instance, the Digital Library Project focuses on both producing digitised images for access and building a repository system. The E-Helvetica Project in Switzerland is built along similar lines.

Responsibilities

In practice, digital preservation is a subject for all units and falls under the responsibility of the library as a whole. Moreover, it will have to be embedded in the normal workflow activities of the library in the future. Eight libraries, however, currently have one single department which bears the main responsibility for digital preservation. In Austria this is the Digital Preservation Department, in Canada the Documentary Heritage Sector, in Denmark (*Det Kongelige Bibliotek*) the Digital Objects Section, in Germany the IT Department, in New Zealand the Electronic Services, in Switzerland the E-Helvetica Unit, in France the Conservation Department and in the UK Collection Care.

[10] *Collection-oriented*: Collection Management (Australia), Collection Development and Processing (Austria), Documentary Heritage Collections (Canada), Special Collections (Sweden), Collections Section (Switzerland). *Process-oriented*: Professional Department (Switzerland); Administrative Department (Japan), Acquisition and Processing Department (Netherlands), Department of Administration (Sweden). *IT oriented*: Information Technology (Austria, Germany); Information Technologies Services (Canada); *Strategy-oriented*: Research & Development (Netherlands), Innovation & Development (Portugal), Office of Strategic Initiatives (USA).

The choice of France, the UK and Canada for 'Collection Care' reflects their anticipation/expectation of a shifting or broadening perspective of the traditional preservation department's working area to include the care of digital objects. It stresses what the possible opportunities for exchanging existing knowledge on preservation strategies between paper preservation and digital preservation might provide for the preservation policy of a library in general.[11]

Activities

Having one specific department as the main unit responsible for digital preservation, does not mean that all of the activities for digital preservation are carried out within this single department. Digital preservation always implies cooperative activity between at least two or more units within the library.

The status of the digital repository (see also page 35) often seems to influence the number of units involved. In libraries focussing on the development and construction of the system, the activities can be more centralised. In this phase, system development and R&D often take place within one department (for instance in Japan, Germany, Austria, Switzerland and the UK). Libraries who are already working with a more or less operational system, have often split up the digital preservation activities into day-to-day workflow care activities and R&D activities (for instance in Australia and the Netherlands). *Koninklijke Bibliotheek* is one of the few libraries with separate units that specifically focus on one of these specific tasks (e-Depot Unit for day-to-day workflow and Digital Preservation Department for R&D). Australia makes a distinction between conceptual R&D (responsibility of the Digital Preservation Unit) and implementation R&D (responsibility of IT).

IT contributes to digital preservation in all libraries. In most cases it has the technical responsibility for the repository system, but it can also assume other roles: for instance overall coordination (Germany) or responsibility for day-to-day activities (alone or in cooperation with other departments, for example, Australia and Denmark (*Statsbiblioteket*)).

Cooperation

Cooperation between units or departments is often formalised within library working groups involving staff from all departments. These cross-divisional working groups focus on practical aspects of digital preservation or on strategic issues. They can operate on a temporary or longer-term (structural) basis.

[11] Merging between digital preservation and other library tasks: See also: Friedlander, Amy and Deanne Marcum, 2003. 'Keepers of the Crumbling Culture. What digital preservation can learn from library history', in: *D-Lib Magazine*, Vol. 9, nr. 5 (may 2003) http://www.dlib.org/dlib/may03/friedlander/05friedlander.html

Temporary working groups are, or have been, generally active in the start-up phase, for instance in Austria. Structural working groups are generally active in the construction and development phase: for instance in France for all aspects of digital preservation, or in the USA for strategy issues.

There is also a tendency to form cross-sectoral working groups which will continue their activities during the operational phase. These working groups have to establish a continuous workflow, sustainable development and a firm embedding of digital preservation in the day-to-day activities of the library in the future, based on cooperation and sharing knowledge. The British Library is currently establishing such a group. In addition to these specific working groups, most libraries have an intensive consultative structure at all levels (meetings).

Staffing levels
The number of staff currently involved in digital preservation activities is highly variable and fluctuating. Most libraries find it difficult to state how many people are actually working on digital preservation, because digital preservation activities also involve the activities of regular staff from other sections: acquisition, IT, digital content creating, cataloguing, web harvesting. Moreover, digital preservation is often part of a task, and not a full-time job. The number of staff active in digital preservation on a full-time basis varies from 3–15 Full Time Equivalents. An exception to this are for instance the large library organisations in the USA and in Canada.

The organisational chart of a library is subject to almost constant change.[12] Whatever the current embedding of digital preservation is, the majority of the libraries state that there will be changes in the future: in Japan this will concern day-to-day responsibilities; in the UK, France and New Zealand the implementation in normal library routines and operational services; in Germany and Canada a complete restructuring of the organisational embedding of digital preservation; in Denmark the involvement of the preservation department; and in Switzerland the possible separation of services for day-to-day activities and for R&D.

Once again many of these plans are related to the library's position with respect to digital preservation. If the library is still in the development phase (building the repository), the positioning of digital preservation activities is not a main priority within the library organisation. Once the production phase becomes

[12] An organisational chart has been added to every overview, on which the units currently involved in digital preservation are marked. See: Overviews Part II-2: Organisational Charts.

more concrete, the embedding of activities becomes a real issue and the organisation will be subject to change.

Funding

National libraries can draw on resources from their own daily budget or from external funding to fund digital preservation activities (building the digital repository and R&D activities). External funding can be supplied by the government or a third party, on a national or international level, but is mostly incidental. The Netherlands is the only country where the library receives structural extra funding for digital preservation activities.

At present six national libraries depend solely on funding from their internal budget for digital preservation activities. These are the national libraries of Canada, China, Austria, Japan, Sweden and Switzerland. Most libraries, however, use a part of their own budget (and that of their partner institutions; for instance in Germany) and, at the same time, receive incidental external funding for specific projects or activities on top of this.

External funding
External funding is generally used for building the repository (Denmark, France, Germany, New Zealand) or for specific research and development projects (Australia, the Netherlands, Portugal). Most libraries expect that extra funding might be needed to continue or broaden digital preservation activities in the future.

External funding is mostly supplied by the government (Ministry of Culture or Ministry of Education) and is often limited to initial development activities related to the construction of a digital repository. Library of Congress receives congressionally-appropriated funds, and although it houses the National Digital Information Infrastructure and Preservation Program (NDIIPP),[13] it only receives some of the funds through this programme.

France, New Zealand and Germany have recently received external funding from their Ministries of Culture or Education, earmarked for the building of the repository. In France the funding is currently limited to the storage procurement procedure. New Zealand expects that the funding earmarked for the National Digital Heritage Archive (NDHA) Programme, will lead to a future increase in baseline funding to help ensure the scalability and sustainability of the digital preservation activities in New Zealand. In Germany part of the government

[13] NDIIPP, see: http://www.digitalpreservation.gov/. Further information on NDIIPP can also be found in section II-1 of this survey.

funding is specifically for the kopal project. Both projects (NDHA and kopal) aim to build a digital repository system.

National level

In some cases a third party provides extra funding for special projects or activities. For example, in the UK specific digital preservation projects in which the British Library is involved are funded by JISC (the Joint Information Systems Committee)[14] and the Higher Education Funding Council of England. The national libraries of Denmark have received a one-off funding from DEFF[15] (the Danish Electronic Research Library) for the initial phase of developing the Danish repository. And in Portugal there has been some funding for digital preservation activities by FEDER[16] (*Fonds Européen de Dévelopement Regional*) and from PIDDAC,[17] the generic investment-funding programme of the Portuguese government for the central administration, which covers investment projects in all the sectors.

In 2000, the US Congress granted the Library of Congress considerable funding for the NDIIPP programme. NDIIPP is the National Digital Information Infrastructure and Preservation Programme of the USA, which focuses on the development of a national strategy for digital preservation. Within NDIIPP the US Congress puts up funding for digital preservation within the framework of a national programme. This funding is meant to support the development of a national strategy and to serve as a catalyst for preservation through targeted investments. This year, NDIIPP partnered with the National Science Foundation to fund ten peer-reviewed preservation research grants. Certain investments are matched with funding from partners. In 2004, NDIIPP used cost-matching to fund eight Digital Preservation Partnership projects. Each project supplied a match equal to NDIIPP-supplied funds. Selections were made on the basis of a competitive evaluation process.

Since 2004, national cultural heritage institutions in the Netherlands have received structured additional funding from the Ministry of Education, Culture and Science for research and development activities related to both paper preservation and digital preservation. This structural budget is expected to increase over the next three years. One of the conditions for this is that part of the money has to be spent on joint projects in which the archival and library sectors cooperate. *Koninklijke Bibliotheek* acts as treasurer.

[14] JISC: See: http://www.jisc.ac.uk/

[15] DEFF: See: http://www.deff.dk/default.aspx?lang=english

[16] FEDER: See: http://www.info-europe.fr/document.dir/fich.dir/QR000925.htm (French)

[17] PIDDAC: See: http://www.gep-mopth.pt/?id=5&MID=11&MTY=2 (Portuguese)

International level

The European Commission is one of the main funding organisations at a pan-European level. Projects funded are often joint projects in which institutions from several countries are cooperating. At the moment there are funding opportunities for digital preservation activities within the EUs Sixth Framework Programme (FP6).[18] This Programme focuses on Research and Technological Development within Europe: Some national libraries are currently active in preparing project proposals within this framework. Funding opportunities are also expected within the Seventh Framework Programme (FP7).[19]

Another novel way of funding digital preservation initiatives is the Digital Preservation Award.[20] This award worth £5000 was first granted in 2004, and is meant to reward projects worldwide which stand out due to recognised leadership and achievements in the developing field of digital preservation. The prize is sponsored by the Digital Preservation Coalition in the UK, under the banner of the UK Conservation Awards, and is managed in a partnership of key conservation, restoration and preservation management organisations in the UK. A call for nominations is issued each year and an independent jury evaluates these.

[18] FP6: See: http://europa.eu.int/comm/research/fp6/index_en.cfm

[19] FP7: See: http://europa.eu.int/comm/research/future/index_en.cfm

[20] Digital Preservation Award: See: http://www.dpconline.org/graphics/awards/

2. Digital repository

Status

At present all of the national libraries surveyed already use repository systems for the storage and/or access of digital material they are producing in-house or receiving because of their deposit role. Libraries involved in web harvesting and web archiving have a storage and access system for these kinds of materials as well. These current systems usually consist of several separate systems that are connected. Most of these are not considered suitable for long-term storage and permanent access yet, and the libraries are currently planning to build an overall system which can meet these requirements.

A repository system suitable for long-term storage and permanent access is capable of the safe storage of digital material both now and in the future. It allows current and future access to the stored material according to worked-out procedures that address technical changes and innovations in hardware or software. The system supports information on and management of permanent access. It also supports or provides functionality for this purpose. Discussions are currently taking place on the defining elements of a long-term storage system.

There are four stages in the development of a digital repository, which can be more or less overlapping: *design*, *development*, *implementation* and *production.* In the majority of cases, development and implementation happens in phases and per type of material.

Three libraries currently have digital preservation repository systems in production. These are Australia, the Netherlands and Austria. Australia was the first with trustworthy digital repository facilities. In 2001 it launched its DOSS for the storage of digital objects and PANDAS for the storage of websites in the PANDORA Archive. The Netherlands has had its e-Depot in production since 2003. Both libraries consider their systems operational, but state that a continuous R&D effort is needed to improve and broaden different aspects and tools, both now and in the future. In Austria the system is operational for storage and ingest.

At present, eight libraries are in the development phase (Canada, China, Japan, New Zealand, Portugal, Switzerland, Sweden and USA) and five libraries are in the implementation phase (Austria – for the access part –, Denmark, France, Germany and the UK). Japan and France are in the process of accepting tenders

for building the storage part of their digital repository and Switzerland for the ingest part.

Austria and the UK expect their systems to be fully operational in 2005, Denmark in 2005/2006, Germany in 2006, France in 2007, Canada in 2007/2008, New Zealand in 2008, and Japan and Switzerland in 2009. The USA, Sweden, Portugal and China do not have such tangible plans yet.

Location and cooperation in building
Most systems will be housed on site. In Denmark the two national libraries will build a joint repository system (referred to as 'the national repository') with geographically divided redundancies in the system. *Die Deutsche Bibliothek* will get an offsite repository. The German system will be housed on the site of the *Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen (GWDG)*, one of the kopal partners. Sweden is currently discussing a partnership for building a digital archive together with the National Archives of Sweden and the Swedish Archive for Recorded Sound and Moving Images. The digital repository in Switzerland, which will be used by both the National Library and the National Archives, will be housed at the site of the IT Service Centre of the Federal Department of Home Affairs. The UK system will be housed on the multiple sites of the British Library to provide a high level of resilience, and for reasons of disaster management. For web archiving BL uses a system developed by the UK Web-archiving Consortium. This system is housed by third parties. At present the Library of Congress has an off-site repository for dissertations in digital form, which is owned and run by a private company, UMI, by agreement with the Library. New Zealand has not yet decided whether to opt for one central system, or several decentralised – but connected – systems. Austria has a backup system on location and a second backup facility at a distance in a governmental high-security data centre.

As the majority of the libraries surveyed are currently in the middle of the decision or development stage, the various aspects of an electronic repository are being widely discussed, but very little practical experience has been gained to date. The following section therefore focuses more on theoretical ideas, as opposed to practical joint achievements.

Services provided

All of the libraries involved are building their digital repository with the purpose of retaining the digital objects collected in perpetuity, in a structured, scalable and secure environment. This objective is translated in the central mission of the repository, namely, providing long-term preservation of and access to digital material. The importance of safeguarding the integrity of the digital object as part of the mission is an issue for all libraries involved in digital preservation. Both Denmark and Sweden still have to develop an adequate mission for the digital repository.

Differentiation of the general mission
Sometimes long-term preservation is referred to as long-term management (Australia), and sometimes the general mission is broadened to include long-term storage (Switzerland, New Zealand) or formulated as: providing storage, description and access (USA).

Four libraries state that their mission is also to serve as a safe place repository for digital objects of other institutions, on a national or an international level (the Netherlands, New Zealand, Portugal, Germany). In the UK this is still a point of discussion. For this aspect the terms 'trusted repository' and 'third party repository' are also used.

Under the auspices of the Research Libraries Group (RLG) and the National Archives and Records Administration (NARA), an international cross-sectoral task force is addressing the issue of trustworthy digital repositories. It is generally accepted that trustworthy digital repositories are fundamental to the international foundation of digital preservation. The Digital Repository Certification Task Force is currently developing a certification process for assessing which repositories can be trusted with the responsibility of long-term preservation and permanent access, and on what criteria. This year the Task Force will publish Guidelines that can serve as a handbook and audit instrument.[21] By the end of 2005 the digital repository system of *Koninklijke Bibliotheek* will be one of the systems tested in accordance with these guidelines.

The libraries sometimes choose a gradual approach to digital preservation in order to achieve the mission of their digital repository. Germany is first of all

[21] Digital Repository Certification Task Force. See: http://www.rlg.org/en/page.php?Page_ID=580 and also: Guidelines 2005. *Guidelines for the Certification of Trustworthy Digital Repositories and Archives. Handbook and Audit Instrument*. (Draft version 1 July 2005. Review version to be published by the end of 2005).

focussing on long-term accessibility and then on preservation. Portugal is currently limiting its efforts to short-term preservation and will consider long-term preservation and access in a second phase. Japan has stated that it might not be possible to preserve everything due to technical difficulties, and that access might be limited for copyright reasons, even though free access for all users is the primary aim. Copyright restrictions apply in all other libraries as well and are more or less inherent to digital preservation.

Expected services

The libraries are fairly similar in their expectations about the services the repository will provide, once operational.[22] A distinction can be made between archiving services (including storage and preservation) and access services. In archiving the following services are foreseen: (in decreasing frequency): secure storage; preservation treatments; and data management, which is split into metadata management and digital object management in New Zealand and Switzerland. On the access part services are foreseen for: search and discovery in controlled online access to archive copies and service copies (or preservation copies and access copies), with some limited access only to service copies (Austria and Switzerland), or user copies (Germany).

These services are in fact services that all are necessary for the functionality of a permanent storage and access system. A choice can be made as to whether something should be allowed or not. This touches the field of access management and DRM techniques, and falls outside of the scope of this survey.

A rather new type of service emerging at present is the delivery of stored items on behalf of, or for, publishers. There are two aspects to this: assistance with user services in the event of a calamity and a business-to-business service between repository and publisher. Business-to-business services go beyond the situation of a calamity as these involve the repository storing (part of) a publisher's holdings and delivering copies at the publisher's request. By doing this the publisher hopes to safeguard its holdings. *Koninklijke Bibliotheek* is currently discussing new business models with national and international publishers to examine new possibilities.

[22] The analysis of 'Expected Services' is based on information from: OCLC/RLG PREMIS Working Group. 2004.

Depositing

Since all of the libraries surveyed play a role as deposit libraries, the digital repositories are to a great extent meant to safeguard deposited material deposited because of the existing deposit legislations or voluntary agreements.

At present, national and international publishers have the largest share in depositing digital objects to the libraries (China, France, Germany, the Netherlands and New Zealand). Governmental institutions, research institutions and other cultural heritage institutions also contribute to depositing (Austria, Germany, Japan, Portugal and USA). And if this is currently not the case, it will be in future (the Netherlands). Some libraries have temporarily limited the deposit in the digital repository to digital objects created within their own institution (and that of their partner institute). This is for instance the case in the UK, Switzerland, and Germany. In most cases this will change once the repository is considered fully operational. Both Canada and Denmark first need to reformulate procedures and/or workflow for the deposit in the digital repository before expectations can be given.

Agreements with depositors
For some libraries the legal deposit legislation is considered the main agreement on which depositing is based (China, France, Austria, Denmark). As in most countries legal deposit for digital publications is not yet fully effective, several libraries have signed agreements with individual depositors. Some of these agreements concern contracts or arrangements on the voluntary deposit of certain types of material (in Australia on archiving web material and disk-based material; in Austria on online publications and e-theses/research papers; in Denmark on statistics databases; in France within Gallica, the digital library of BnF; in Germany on online theses and publications; in Switzerland on online material and university output; in the USA on dissertations; and in the UK on online publications as well).

In some cases the agreements are made within the framework of pilot projects on depositing (in the Netherlands with cultural heritage institutions and universities in the DARE project on digital academic repositories). An interesting case in this context is the current voluntary deposit experiment of the *Bibliothèque nationale de France* with regional newspaper publishers, to test digital depositing as an alternative to printed depositing. Agreements can also be signed with umbrella organisations (Germany and the Netherlands with the German and Dutch National Booksellers & Publishers organisations respectively; Sweden with an umbrella publisher and with the Digital Scientific

Archive (DiVA) on university output; and Switzerland with the Conference of University Libraries (KUB) on output of university libraries).

Some libraries do not have individually signed agreements yet (China, Denmark, France, Japan, New Zealand, UK), but all libraries expect that there will be either a full deposit legislation or individually signed agreements in the future.

Processing digital objects
At present most libraries process their digital objects both manually and automatically. The digital objects are obtained by both submission and harvesting. This often depends on the kind of material involved. Websites are harvested; digital images from in-house digitising project and CD-ROMS are mostly submitted manually. Automated processing is sometimes embedded in experimental projects.

The general opinion is that the workflow, particularly the processing and the quality control, must become as automated as possible in the future if digital preservation is to become a stable process. However, it is generally recognised that residual manual processing and checking may still be required in the future.

Software and OAIS

The generally held view is that for the design, construction, development and extension of a trustworthy digital repository, the physical storage and retrieval of the digital objects both now and in the future must be independent of the technical properties and characteristics of the physical storage. A design that incorporates this principle should be able to accommodate many generations of physical storage implementations.

Since only a few libraries currently have a fully operational digital repository (Australia, the Netherlands, Austria for the storage part), the experience in building (and thus software and system use) is still very poor.

Analysis
Australia and the Netherlands both built their own system using products developed on demand or software developed in-house, and adding existing commercial or open source hardware and software where necessary to broaden the features of the system. The main reason for this approach was that commercial products available at that time did not meet the specialised needs of the libraries. In the Netherlands this resulted in the development of the Digital

Information Archival System (DIAS) by IBM,[23] now available as a commercial product. Austria however used an existing product for the storage part of their system (DigiTool from Exlibris),[24] which it adapted to its needs.

Germany, France and the UK are currently building their own overall systems. The German system is based on DIAS and the French system on the archival system of the *Centre National d'Etudes Spatiales* (CNES).[25] The UK will develop its own system.

The other libraries have yet to reach a final decision on the basic system and/or software that will be used in their overall repository, although a lot of assessing and testing is going on. The most tested systems are FEDORA[26] and DSpace.[27] The Library of Congress and the libraries involved in NDIIPP are also currently testing systems such as aDORe,[28] the repository at Los Alamos National Library, LOCKSS[29] and the OCLC Digital Archive System (USA).[30]

Although DIAS, DSpace and FEDORA seem the most standard systems currently in use or under consideration, there still is some hesitance about choosing one solution or the other.
When a library starts developing its own system, tuned to its individual preferences, the development of the existing systems continues to be closely monitored. This is also the case when a library is still in the decision phase. A particular point of concern is whether the existing systems have sufficient reliability to be scalable enough for future long-term preservation needs. Libraries often develop an individual system to ensure that their own needs and standards can be optimally met. Integration of the existing library systems with the new system is also a very important aspect.

In all cases the general solution is not being sought in one single system. Although one basic system might be used for the central part of the repository, additional tools will always be needed. Web archiving in particular needs specific system and software solutions within the overall repository system. Most of the libraries expect to use a combination of existing off-the-shelf products, both commercial and open source, and solutions developed in-house

[23] DIAS of IBM: See: http://www-5.ibm.com/nl/dias/
[24] DigiTool of Exlibris: See: http://www.exlibrisgroup.com/digitool.htm
[25] CNES: See: http://www.cnes.fr/html/_455_.php
[26] FEDORA: See: http://www.fedora.info/
[27] DSpace: See: http://www.dspace.org/
[28] aDORe: See: http://african.lanl.gov/aDORe/projects/adoreArchive/
[29] LOCKSS: See: http://lockss.stanford.edu/
[30] OCLC Digital Archive System: http://www.oclc.org/digitalarchive/default.htm

for specific applications. There is a slight preference for the use of open source software, as this is expected to meet the previously stated vital conditions for the repository.

OAIS

The only official standard in digital preservation at present is the Open Archival Information System (OAIS) Reference Model, which originates from the CCSDS (Council of the Consultative Committee for Space Data System). The OAIS standard was certified by ISO in 2003 (ISO 14721:2003). An OAIS is an archive that contains an organisation of people and systems which has accepted the responsibility to preserve information and make it available for a designated community. It provides a model or general framework to build and maintain a repository for the long-term preservation of and access to digital material. The model offers guidance, but is not meant as a blueprint for the design of an archive.

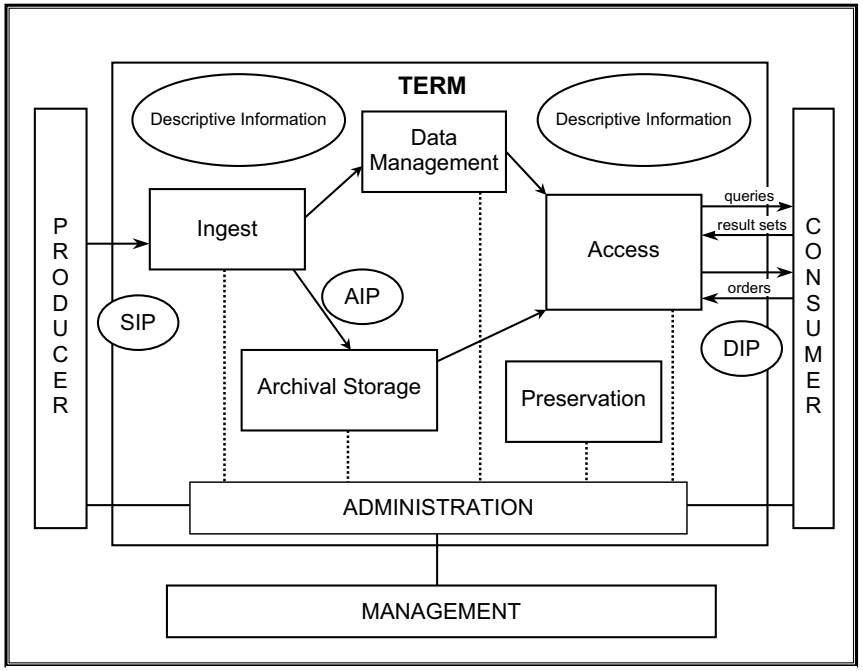


Fig. 1. Major functions of the OAIS Reference Model from Consultative Committee for Space Data Systems (CCSDS), CCSDS 650.0-W-1, Producer-Archive Interface Methodology Abstract Standard, (OAIS), White Book, Issue 1, Draft Recommendation for Space Data System Standards.

The OAIS-model. (after http://www.Dlib.org)

The most valuable aspect of the OAIS model is that it provides a shared vocabulary, detached from the traditional terms used in the different types of institutes or sectors. This general glossary has proven to be very useful for worldwide knowledge dissemination.[31]

All libraries state their digital repository is or will be OAIS compliant, if not on all aspects then at least on specific requirements which still need to be determined. This means that at least all libraries are familiar with the only standard for digital preservation at present.

The Netherlands has based the design of its e-Depot and DIAS on OAIS, particularly for the processes, the functional model, the construction of information packages within DIAS and the distinction between submission, archival and dissemination information packages (the SIPs, AIPs and DIPs). The workflow is based on the interpretation of OAIS within the NEDLIB project. The NEDLIB project contributed to the development of OAIS by suggesting the implementation of a preservation-planning module in it.[32]

Both the Netherlands and Australia have mainly used OAIS as a reference model. In the Netherlands it was useful in choosing the scope and approach, but less helpful for the technical design. In Australia OAIS is mainly being used for the conceptual check on developing and managing the digital repository. When process models do not fully comply, this is identified and monitored. OAIS has been the key concept for the archival information packages. The system in Austria is partially OAIS compliant (for the ingest, archival storage, access and data management).

In France the current systems are OAIS compliant for the web archiving part and for the ingest, management and preservation of digital objects. Both the delivery system and the communications/access system are not compliant. When the overall system is built, OAIS will mainly be used to make all current systems compliant. The CNES System, on which the French system will be based, is OAIS compliant for the storage part. In the USA the OAIS model is

[31] For a thorough description of OAIS see: Consultative Committee for Space Data Systems. 2002. http://ssdoo.gsfc.nasa. gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf

For a visual picture of OAIS see the Cornell Tutorial: http://www.library.cornell.edu/iris/tutorial/dpm/terminology/ oais.html

[32] Diessen van, Raymond J. and Johan F. Steenbakkers, *The long-term preservation study of the DNEP project: an overview of results*. IBM/KB Long-Term Preservation Study Report Series, Nr. 1. Den Haag December 2002. See: http://www-.ibm.com/nl/dias/preservation.html

mainly being used as a guide to identify requirements and to build prototypes. The architectural framework currently being developed within the NDIIPP will be OAIS compliant. In China OAIS is being used to modify the processing model and to make a distinction between the different information packages; in Switzerland it will serve as a basis for planning and building the archival system; in Japan to ensure permanent access and to prevent loss of information. To date Portugal has mainly used the OAIS model as a tool to raise public awareness of the problem of digital preservation.

Materials

All libraries expect a great variety of digital objects to be stored in their digital repositories in due course. Due to their deposit role, all countries safeguard the national cultural heritage in digital form, whether this be purchased or deposited, offline or online, published (e-journals, digital publications), semi-published or non published materials. In some cases the focus broadens beyond the limits of the national boundaries and is international, for example, when libraries want their repositories to act as safe-place repositories.

Frequently mentioned categories of digital objects to safeguard, now and in future, are: deposited digital materials, digitised objects derived from in-house digitisation projects; digital objects generated by web archiving; digital surrogates of images, sound and text from non-digital originals made for preservation purposes, and more complex, packaged digital materials such as databases, spreadsheets, CAD Drawings and GIS output. The majority of the libraries are also considering the storage of electronic objects from other institutions in the future. The storage of software is not that common yet. Some libraries do store software with specific restrictions (Denmark), and others do not store software applications yet (Austria).

Both the Netherlands and the UK have a clear step-by-step approach. In the Netherlands the e-Depot mainly stores e-journals at present and to a lesser extent deposited offline materials. In the near future this repository will also contain other materials such as theses and dissertations, web material and digitised objects from other cultural institutions. In the UK the variety in material types will expand as the system develops. At present the focus is on voluntary deposited materials, which will be followed by e-journals and then by other materials. The expansion will be based on the technical implementation of the new legal deposit law for digital publications. Japan and Canada also collect a large quantity of archival objects, due to their constitution or their role.

Formats
At present the libraries accept all types of file formats. The most common formats at the moment are: TIFF for images; PDF and XML for text; HTML for websites and WAV for audio. Less frequently mentioned formats that are currently stored are ZIP, PNG, ISO, JPEG, GIF, TeX, PS, EPS, AI, RTF, TXT and ASCII. Although there are preferences, these have not led to restrictions yet.

Nevertheless, most libraries are aware of the importance of regulating or limiting the file types in the future. Several initiatives are underway to develop guidelines or defined recommendations for preferred or accepted file formats. This is the case in Austria, Canada, France, the Netherlands and the USA.
Austria has already officially formulated recommendations on PDF file formats, based on the ISO PDF/A Standard.[33] France not only defines the type of preferred file formats, but also the quality level. New guidelines with a special focus on the preservation purpose of the digital repository are expected in September 2005. The research on file formats in the USA is broad, and focuses mainly on the preparation of future preservation strategies. In Canada the focus is not only on library material but also on governmental or archival publications. Canada makes a distinction between recommended and acceptable file types and interchange formats. Recommended in this context means that the specific file format is promoted as a standard for computer generated information from a technical point of view.

Preservation and access copies
In paper preservation, the use of more than one-generation surrogate copies has been widespread for several years (preservation master, duplicate copy, user's copy). In digital preservation two practices occur. Libraries either choose to generate and store a preservation and an access copy at the same time, or they choose to generate just a preservation copy and make access copies on the fly. At present the latter practice is more common. In most cases only preservation copies get the full preservation treatment.
Australia and New Zealand both use the three generations variant for digital preservation as well. In Australia this is currently limited to web objects. The preservation master is the object as harvested, without any alteration. The access master is checked for quality and if necessary adjusted to restore functionality. The access copy is made from the access master on demand and is used to give the public access. Both the preservation and access masters are stored.

Selection
The growing complexity of the digital objects to be stored is generally recognised by most institutions. This aspect, combined with the ever-growing

[33] The ISO PDF/A Standard is expected in October 2005.

flow of digital objects means that there will be an increasing emphasis on developing selection methods, not only due to the possible technical limits there might be for future storage, but also because of the costs involved in storage and long-term preservation care in general.[34]

The care of digital objects has added a new dimension to the selection procedure used by libraries. Paper material and digital material will be selected separately. Yet it is quite plausible that a choice might have to be made between paper and digital: will the paper object be safe kept or the digital object, or both? Additionally the question might arise as to whether digital objects can replace paper objects in general, once digital preservation is a stable factor. Both themes are currently being discussed and no final pronouncements have been made yet. In this context the thoughts of New Zealand on choosing the preferred preservation copies are interesting: 'The additional functionality associated with digital versions of analogue objects might even suggest that it is worthwhile acquiring both and deciding at a later date, maybe after reflecting on how the material is actually used, which will be the preservation copy.'

Another aspect to selection is found in a recent survey in the UK on trends in electronic publishing. This estimated that by the year 2020, 40% of UK research monographs will be available in electronic format only, while a further 50% will be produced in both print and digital. A mere 10% of new titles will be available in print alone. These figures point to a seismic shift for the library, it's partners in publishing and for the information sector in general. This shift will have significant implications for selection.[35]

Metadata and metadata schemes

Every stored digital object has to be supplied with metadata to describe it in a structured way. Metadata generally contain information on bibliographical, structural, administrative and technical aspects of an object. All of these

[34] Digital versus Print: See: Shenton, Helen. 2003. *Digital versus print as a preservation format – expert views from international comparator libraries*. http://www.bl.uk/cgi-bin/print.cgi?url=/about/collectioncare/digpres1.html

[35] Powell, David J. 2004. *Publishing output to 2020*. London: Electronic Publishing Services Ltd/The British Library. See: http://www.bl.uk/about/articles/pdf/epsreport.pdf . This survey was carried out by the end of 2004 by Electronic Publishing Services Ltd, which was commissioned by BL to research publishing output and trends in the UK and overseas to 2020, to inform BL's longer term projections of collection growth and establish its physical storage need to 2020.

metadata categories can contain information that might be useful for preservation purposes.

'Preservation metadata' is usually not seen as another category of metadata, but as a combination of existing metadata sets that provide the information needed for long-term preservation of and permanent access to digital objects. Preservation metadata have to contain technical details on the format, structure and use of the digital content; the history of all actions performed on the resource including changes and decisions; authenticity information such as technical features or custody history; and the responsibilities and rights information applicable to preservation actions.[36] Administrative metadata and technical metadata are generally considered to be the most important for preservation.

Since the OAIS model provides only a general framework for preservation metadata, several initiatives have been adopted to identify and specify metadata and to develop metadata schemes for digital preservation.[37] As each institution used its own approach and interpretation of the OAIS framework this resulted in various specifications and models (for instance NEDLIB, National Library of New Zealand Metadata Scheme, LMER). In May 2005 the PREservation Metadata: Implementation Strategies Working Group (PREMIS), sponsored by OCLC/RLG, released its data dictionary on preservation metadata, which includes many of the results of previously designed schemes. There is a PREMIS maintenance activity hosted at the Library of Congress as well as PREMIS implementers' Group.[38] At present also the International Internet Preservation Consortium (IIPC)[39] is developing an archival metadata scheme, with a special focus on web archiving. This will be available by the end of 2005. The use of metadata and metadata schemes is still under development, and to date none of the libraries have taken any final decisions. This is also the case with defining what preservation metadata are supposed to contain. However the current activities do show some common practices.

Submission and processing

Even though the libraries do not yet use requirements or agreements on metadata supply, it is generally expected that the depositors should submit metadata with the digital objects. If this is not the case, the libraries will generate the metadata. It is expected that in future the processing of metadata

[36] The analysis of 'Preservation metadata' is based on: PADI-website: http://www.nla.gov.au/padi/topics/32.html

[37] The analysis of 'OAIS and metadata schemes' is based on: Cornell Tutorial: http://www.library.cornell.edu/iris/tutorial/ dpm/foundation/metadata/index.html

[38] PREMIS maintenance activity: See: http://www.loc.gov/standards/premis

[39] IIPC: See: http://www.netpreserve.org/about/index.php

will be done as automatically as possible, although some manual work will still remain.

Storage
The libraries are currently storing metadata in three different ways. The most commonly used is the separate storage of metadata in an XML-related database (the Netherlands, Australia, Canada, Germany, New Zealand). In those cases the link between metadata and the original objects is assured by using persistent identifiers such as NBN or URN. Another storage method is separate storage within the repository (Austria, Denmark). It is also possible to bundle metadata with the related content files (France). How the metadata is stored, depends in part on the category of metadata concerned.

Metadata schemes
All libraries currently use (elements from) more than one metadata scheme. METS is used by all libraries. METS is a scheme designed to be used as a transmission standard. However it has a highly flexible design which incorporates the use of other 'extension' schemes for certain forms of metadata.[40] It is therefore normally used in conjunction with other schemes, for example, the NEDLIB Schema, the metadata schema of the National Library of New Zealand, and the NISO Z 39.87 Data Dictionary. Libraries also frequently refer to the research of the OCLC/RLG Preservation Metadata Working Group 2001.

At this stage four of the libraries surveyed explicitly mention the use of the PREMIS data dictionary, which can be implemented using a set of 5 XML schemes (Australia, Austria, New Zealand and the USA). Others are considering its use in the near future. The PREMIS scheme has the potential to become an emerging best practice.

Access

Since libraries started digitising (parts of) their collections about 10 years ago, and made a start on building their digital libraries, the main target has been to improve access to the collections. One of the main advantages of digitisation was the ability to protect vulnerable items within the collection from the decay and damage often caused by use.

[40] METS: see: http://www.loc.gov/standards/mets/

A large percentage of the paper collection suffered from internal decay and had to be preserved. This was initially done by microfilming. However, the advantages of using a digitised substitute over an analogue substitute are obvious: more comfort, more detail, scalability, accessible everywhere and possible multiple access at the same time. From a protection viewpoint, digitisation also had advantages for the use of unique and valuable special collection items. Alongside these in-house digitised objects, the libraries were increasingly collecting various digital objects through depositing, purchase or web harvesting.

Whereas libraries in general wish to give free access to all collected digital material, giving such access is not as self-evident as it might seem at first sight. Access is often limited by copyright regulations and agreements with depositors.

The extent to which access is given, can depend on several factors such as type of material, public, place, policies on rights, permission and restrictions. At present most libraries only provide access to their digital material on site, often under special conditions. Due to the process of development the library is in, Switzerland provides no direct access to the digital material yet. In Germany the accessibility is extending as the development of the repository progresses; at first only the depositing libraries will have access to the stored objects. Public access procedures will be worked out in a second phase.

Current access

A few countries currently provide Internet access for a limited set of material, under strict conditions: in Denmark Internet access is provided for, if the material (Internet publications, broadcasts and movies) is not commercially available, and limited to research purposes only. France provides open access for material without copyright restrictions, but only in low resolution. When material has copyright restrictions or when high-resolution copies are requested, only onsite access is possible. In Canada electronic publications are available online via the LAC website. The Netherlands will have open access for open access journals as of the summer of 2005. Sweden makes a distinction between commercial (only on site accessible) and non-commercial material. Japan has a similar distinction between governmental and private publications. When libraries also start storing digital objects of other cultural heritage institutions in their repository, it is expected that these objects will only be accessible for the institution involved.

Access is only given through access copies, which most libraries will generate on the fly, or on demand (New Zealand). Preservation copies can only be accessed for very limited purposes, mostly restricted to staff only, and in principle only to improve preservation (for example with migration). Paid

access is not being considered at present, except in France, where paid Internet access to digital images is currently a subject of internal discussion.

Access is often provided through the library catalogue, which in most cases forms a separate system, linked with the digital archive. Most libraries also provide a link between the paper originals (if existing) and the digital copies.

In most libraries future access still has to be determined, and for now all options are being kept open. Limited access on site, limited access within all depositing libraries (Sweden), access to one item at a time (UK) and no limitations at all, are some of the options currently stated. The outcome will heavily depend on the new deposit laws and on negotiations with the depositors. Most libraries see the development of access procedures as a step-by-step process, which requires careful negotiation within the given practical limits. Copyright questions are often the issue, and all libraries are trying to achieve the best possible way of preserving the material and making it accessible, without adversely affecting the producers' commercial interests. Denmark for instance, is currently considering the development of an archive for Internet material that will be partly open and partly closed, in order to tackle problems with access to personal data. In the UK, the UK Web Archiving Consortium is currently clearing permissions individually with rights holders of web materials, to allow materials to be publicly available in an online archive.

The general intention of the libraries surveyed is to provide enhanced access, consistent with copyright and other restrictions, as the technical infrastructure continues to develop.

3. Preservation strategies

Current strategies

In general the libraries feel that there is no one single strategy to achieve long-term preservation and access for all different types of digital objects. The development of different strategies should be encouraged, together with research on evaluating strategies and methods for preservation planning. At the same time, all libraries state that the basis for long-term storage requires proper procedures for media refreshment and a good backup regime.

Bit-level preservation, normalisation and conversion
The most frequently used preservation strategy is preservation of the bits and bytes, or 'bit-level preservation' as it is described by PREMIS and others. This means providing secure storage, with proper procedures for backup and refreshment. Restrictions on submission (not accepting all formats) and normalisation (conversion of formats into one acceptable format) are the second most used techniques, after secure storage. It is difficult for a library to influence the production-side of digital publication. Publishers choose their format, and libraries just have to accept what is deposited with them.

However, there is a growing tendency towards promoting specific 'preferred' formats and publishers seem to be interested in cooperating. The technical metadata and how they are included are an important issue as well. This might help to achieve a more or less standardised form for use within the complete lifecycle process of the digital objects, from creation to access. Although it does not yet generally lead to excluding certain formats, it might help to simplify the handling of digital objects in general, from creation to access.

Increasingly attention is being paid to the quality of the digital objects deposited (France), and the development of risk identification and shared-representation information tools (Australia with PANIC,[41] the Netherlands with the Preservation Manager,[42] the UK with PRONOM[43] and the proposed Global Format Registry).[44] This could exert a positive effect on the lifespan of digital material.

[41] PANIC: See: http://metadata.net/newmedia/

[42] Preservation Manager: See: http://www.kb.nl/hrd/dd/dd_onderzoek/preservation_subsystem-en.html

[43] PRONOM: See: http://www.nationalarchives.gov.uk/pronom/

[44] Global Format Registry: See: http://hul.harvard.edu/gdfr/

Migration and emulation
Migration and emulation are considered to be the most promising preservation strategies. Migration in the sense of format migration is the translation of data from one (version of a) format to another. Emulation is the process of a software package mimicking a piece of hardware or software so that other processes think the original equipment/function is still available in its original form.[45]

Migration is aimed at the digital object itself. It changes or updates the format of an object (and by doing this, the object itself) to adapt it to a new environment. Emulation does not focus on the digital object itself, but on the environment with which the object is rendered. It aims at (re)creating an environment in which the digital item can be rendered in its authentic form. Migration and emulation represent two specific ways of creating permanent access to digital items. In future, new strategies may be developed that choose an innovative approach or use a combination of the basic migration and emulation techniques.

There is a general awareness of the fact that new versions of hardware and software follow one another frequently, and that the − often irrevocable − decay of digital objects can cause a huge loss of information. Several libraries have in fact already been confronted with the decay of offline digital publications recently (for example China). This mainly resulted in enhancing backup procedures and bit-level preservation, as these are the most feasible strategies for most libraries at this time.

Currently, migration and emulation are only applied on a very small scale. The overall attitude towards the application of these strategies, however, is one of caution and maintaining a low profile with respect to its advocacy. This is due to several factors. Both migration and emulation demand a considerable effort in both financial and technical terms. Practical tools to implement these preservation strategies have not been evaluated, developed or generally made available yet. Tools that do exist have not been developed for the specific purpose of long-term preservation. The existing emulation tools, for example, are often made for nostalgic reasons, such as rebuilding platforms for computer games. Another type of current emulation tool has been developed from a business-optimisation point of view, so as to make it possible to run several operational systems on one server.

Research is needed to assess emulation and migration tools with respect to their suitability for long-term preservation purposes. The majority of the libraries involved are not yet in the full operational phase of digital preservation and are

[45] Preservation Strategies: See also: http://www.nla.gov.au/padi/topics/18.html

therefore currently giving priority to the establishment of the digital repository and the mechanical processes of archiving and access.

Although the libraries are aware of the risks, it appears that in in some libraries there is little desire for immediate efforts in the development of tools for strategies such as migration and emulation. Some libraries state that their digital collection does not require migration or emulation yet. In other libraries, file format migration has only had to be applied once over the past ten years. Some libraries hope to postpone the need for migration for another five- year period by using specific storage formats and metadata that will hopefully provide a longer shelf life.

Keeping a low profile does not, however, prevent most libraries from following the developments on preservation strategies with great interest. And of course there is still the silent hope that the computer industry will take action in time and provide suitable long-term solutions that can be used by the libraries. But even if this happens, libraries still have to evaluate these tools and fit them in their working process.

First steps in developing tools for long-term preservation
Both France and the Netherlands have taken the first steps in the development of tools for emulation. In France an emulation project is carried out within the audiovisual department, focussing in the first instance on audiovisual material.

In 2004 a prototype of the Universal Virtual Computer (UVC)[46] for digital still images was developed in the Netherlands. This UVC can be used to restore images (in JPEG or GIF format) such that they will look exactly the same as the original, while making use of a translation to a logical, humanly-readable description. In a way, the UVC is a combination of emulation and migration and it serves as an example of the innovative possibilities that can be explored.

Recently *Koninklijke Bibliotheek* and the national archive of the Netherlands started a project on the development of an emulator for preservation purposes. This project is the result of a study that was carried out by *Koninklijke Bibliotheek* in 2004. The first version of the emulator is based on a modular design that enables incremental development and the reuse of existing developments of earlier emulation experiments, and which will be capable of dealing with complex compound objects as well. It is expected to be operational by the end of 2006.

[46] Universal Virtual Computer: See: http://www.alphaworks.ibm.com/tech/uvc and http://www.kb.nl/hrd/dd/dd_onderzoek/uvc_voor_images-en.html

In the autumn of 2005 several libraries and other cultural heritage institutions coordinated by the British Library, plan to submit a project proposal to the European Commission, called PLANETS (Preservation and Long-term Access through NETworked Services). One of the goals of this project is the joint development and distributed use of various tools for preservation strategies. National joint initiatives, for instance in the USA with NDIIPP, or in the UK by the Digital Curation Centre,[47] are also promising for the development of tools for preservation strategies.

Future strategies

In general most libraries expect to continue with their current strategies. The majority also expect to apply migration and emulation in the future. Future strategies are expected to be as flexible as possible, not precluding any future developments. They must also be affordable from both a technical and financial viewpoint. The choice of the future strategy depends heavily on the development of the new infrastructure and policies within the library.

The moment the archiving systems are taken into production, the implementation of preservation strategies will become relevant. Since emulation is expected to be an expensive strategy, the level of funding a library has, will also be important. At this time all options are being kept open, and most libraries expect that whatever will be necessary and appropriate for the range of materials will be done, as long as this falls within the limits of common sense.

A number of national libraries are also engaged in discussions with other institutions on developing registries of file format and representation information for preservation planning. This is an area where there is likely to be significant cost reductions and preservation benefits from collaborative activity, but where complex governance and business model issues may need to be resolved before new services can be launched.

Focus
Preserving the digital object in its most authentic form is an objective most libraries subscribe to, at least in theory. When starting with the practical development of tools, concepts such as authenticity suddenly become very ambiguous. An object's authenticity concerns both the content and the original 'look and feel'. A choice for authenticity means ensuring that five aspects of a digital object remain intact. These are: the content, the context, the structure, the

[47] Digital Curation Center: See: http://www.dcc.ac.uk/

appearance and the behaviour. Priorities need to be made when choosing a future preservation strategy: is the emphasis on preserving the content, or preserving the 'original look and feel'. Moreover hybrid strategies need to be considered, as neither concept is clear-cut. Some libraries expect to focus on the content, some on appearance. The majority feel, however, that content is too interwoven with appearance to be treated separately, and that the focus cannot therefore be fixed. Sometimes context can lose its meaning without the original look and feel.

Several questions arise once strategies start to be developed and usually these can only be answered through trial and error. What exactly is the 'look and feel'? Does it also include the representation of the original computer screens, or the speed of the computer, for example? If the focus will be on the 'original look and feel', does this automatically mean emulation is the only way to go? And what exactly is 'original'?

For most libraries it is not yet clear where the priority will be. If the focus is to be on the content, does this mean that the library is also allowed to make derivatives, and will this automatically lead to conversion strategies such as migration? At present migration clearly has several disadvantages: it is not always possible to make an exact digital copy or replica of the digital object (as hardware and software change) whilst still maintaining the compatibility of the object with the new generation of technology. Further migration is a strategy that requires continuous efforts over time and it also bears the risk of error propagation.

The type of material is generally expected to be an important defining factor. Yet another new issue is increasingly coming to the fore: what do we want to offer our future users? The question is not only what will we be capable of developing technically, but also what we wish to offer the library audience in the future: an authentic looking document that nobody knows how to operate, or a documents that future users are familiar with, but that contains errors and gives a misrepresentation of the original content?

4. Current activities

National activities

Alongside the numerous activities that are going on within the individual organisations, all national libraries surveyed are involved in digital preservation activities on a national and/or international level as well. These activities can be temporary or structural. They are carried out within the framework of projects or working groups, in cooperation with other libraries, or in cooperation with other (cultural heritage) institutions.

The listing of national and international activities given in the Overviews in the second part of this survey, seems to give a very random picture. Its value lies in underlining which activities the national libraries in each country consider to be the most important at this point in time. A picture of the existing communication pattern can be found in the cooperation scheme attached to Section II-3 of this survey.

At present six national libraries state that building and implementing their digital repository is the main activity on digital preservation for the time being (Austria, Canada, France, Germany, New Zealand, British Library). Yet at the same time, most of them are also active on a national level.

Framework
Half of the libraries are part of a national framework. Examples are JISC and the DPC in the UK, NDIIPP in the USA, DEFF in Denmark, nestor in Germany, E-Helvetica[48] in Switzerland or PIN[49] in France. In the Netherlands, the structural funding provided for preservation by the Ministry of Education, Culture and Science also serves as a framework. These kind of frameworks can be a programme, an overall project or a consortium of some kind. The activities carried out within the frameworks are not just limited to temporary projects, but can also have a structural character. Sometimes the framework primarily provides funding and sometimes more practical facilities are offered to improve cooperation, such as coordinating offices, embedding within project organisation, websites, facilitating meetings and seminars, etcetera.

Focus
The activities in which the national libraries are involved on a national level can have various focal points. The purpose can be to improve their own (national)

[48] E-Helvetica: See: http://www.e-helvetica.admin.ch/
[49] PIN: See: http://vds.cnes.fr/pin/ (French)

digital repository (for instance the project with regional newspapers in France), or to develop tools or functionality (such as the emulation project in the Netherlands, or the project to preserve metadata in Australia). Once the tool or functionality has reached the status of an acceptable prototype form, or has become fully operational, it is normally meant to serve digital preservation outside the developing institution as well. For example the UVC, developed in the Netherlands, has been made freely available as a prototype. Improving facilities for other types of institutions can also be an important objective of national cooperation.

At present, the most intensive national cooperation takes place around subjects such as the development of (trusted) institutional repositories and web archiving. The target group for the repositories can be the academic world (for example Denmark, Switzerland, Sweden) or higher education in general (Canada, Australia, UK). Web-archiving and web-harvesting projects are being carried out, for example, in Japan, China, Sweden, Denmark and France.

Knowledge dissemination and improving national cooperation, not only within the library field but in the cultural heritage sector as a whole (archives, libraries and museums), is also a very important issue. At least five libraries that have yet to establish a solid national cooperation framework, state this will get top priority in the near future (for instance Austria, Japan, The Netherlands, New Zealand, Portugal).

The cultural heritage sector and national information structures
National libraries often play an important role in (initiating) cooperative activities within the cultural heritage sector and national information structures.
In 2004 Sweden established a special office within the National Library to improve cooperation between the archiving, library and museum (ALM) institutions. Austria has recently signed a resolution, based on the UNESCO Charter on Preservation of Digital Heritage, to set up a national strategy for digital preservation in all cultural heritage sectors. This was the outcome of a conference of the National Library of Austria and the Austrian Commission of UNESCO in March 2005. Future national cooperation in Austria will focus on forming a cooperative platform, copyright and legal deposit actions, cost calculations, and the setting up of a research programme. More detailed plans will be published in the near future.

Australia has taken the initiative to improve collaboration in all cultural heritage sectors in Australia, particularly in the areas of risk assessment and digital preservation management, as a follow up to the international web-archiving conference in 2004.

Portugal, whilst stating that the current focus is on digitising and not that strongly on digital preservation, has recently written a project proposal, called PREDICA, to form a national centre for digitisation and preservation in Portugal. The assessment procedure on this proposal is still pending.

Other institutions
National libraries are generally not the only institutions involved in digital preservation. In all cultural heritage and scientific research sectors, leading national institutes are involved in digital preservation activities as well, tailored of course to the type of material they primarily collect. Heavy involvement of national institutions is, for example, most common (though not limited to national institutions only) in the archival and data sectors. In countries with a federal government structure, the institutions at state or regional level also play an important part in digital preservation activities. This is particularly the case in the library and the archival sector (France, Germany).
A comparison of the cooperation between the different heritage sectors reveals that the cooperation between the library and archival sectors is the most established to date.

The university libraries are also important players within the library sector, usually in close cooperation with the national library. In the USA in particular, the role university libraries play in digital preservation is considerable.

Only the National Libraries of Germany and Austria mention current activities on digital preservation within the arts sector in their country. More activity is undoubtedly going on in this field, but the contacts between the arts sector and the library sector are not very intensive at present.

Within the audio-visual sector digital preservation is primarily a subject for the national institutions for sound, audio and film and for the broadcasting companies (the BBC in the UK). Sometimes the national library is involved in the care of parts of the national audiovisual collections as well (Australia, France USA, Switzerland, UK). At present the most important international project on preserving audiovisual materials is the PRESTO SPACE project, which is funded by the European Commission and coordinated by the BBC.[50] PRESTO SPACE focuses on technical solutions and integrated systems for complete digitisation and digital preservation of all kinds of audio-visual collections.

To date digital preservation has transcended the boundaries of the cultural heritage sector, and also involves (technical) universities, research institutes,

[50] PRESTO SPACE: See: http://www.prestospace.org/

governmental institutions and private or commercial companies (France and China) who offer their own digital archiving systems.

Subject data centres play an important role in science, the social sciences, and some areas of the arts and humanities. Collaboration with data centres has been particularly growing in areas such as the development of the OAIS standard or the preservation of the scientific record.

International activities

Two main activities in the digital preservation field connect the national libraries surveyed. Firstly the IIPC (International Internet Preservation Consortium): this web-archiving consortium connects more than half of the libraries surveyed (Australia, Canada, Denmark, France, New Zealand, Switzerland, UK and USA). Secondly – on a smaller scale – ICABS (the IFLA CDNL Alliance on Bibliographical Standards) connects six libraries: Australia, Germany, the Netherlands, Portugal, UK and USA). Four national libraries have also been involved in the PREMIS (PREservation Metadata Implementation Strategies) working group: New Zealand, the Netherlands, UK and USA.

On a European level the sixth and seventh framework programmes will also provide opportunities for future cooperation. Not only within the cultural heritage section but also between the cultural heritage sector and the science sector. Activities currently been undertaken to prepare a project proposal for the fifth call of the Sixth Framework Programme of the EU unite Austria, Denmark, the Netherlands and the UK. This project proposal called PLANETS (Preservation and Long-term Access through NETworked Services) is being coordinated by the British Library and involves not only libraries but also archives, universities and commercial firms. PLANETS proposes to conduct R&D on preservation planning, file format characterisation and emulation/migration services.

On the same subject Australia plans to work more closely with the British Library, *Koninklijke Bibliotheek* and the Digital Curation Centre (UK) to improve cooperation on research and in particular to test and improve the PANIC Service. This service provides format risk identification and networked access to preservation tools. The PANIC (Preservation web services Architecture for New media and Interactive Collections) project is one of the current research activities being undertaken by the MAENAD group at Distributed Systems Technology Centre (DSTC).

The National Archives in Sweden is also coordinating a project proposal for the fifth call of the Sixth Framework Programme, in which – amongst other institutions – also *Kungliga Biblioteket* and the Archive of Sound and Moving images of Sweden are participating. Under the working title 'PROTEAN' a project proposal will be submitted. The aim of PROTEAN (Preservation Over Time by Electronic Archiving and Networking) is to develop and demonstrate strategic models and methods for ensuring the continuous existence and accessibility of digital information over time, focusing on authenticity, reliability and IPR, and based on the OAIS model.

As a result of the EU Conference 'Permanent Access to the Records of Science' organised within the framework of the European Presidency of the Netherlands in 2004, an international Task Force is trying to improve cooperation on digital preservation between the cultural heritage institutions and the scientific organisations. This Task Force, with representatives from both sectors, is currently developing an R&D programme that can be incorporated in the outline for the Seventh Framework Programme of the European Union. The Task Force aims to improve cooperation at a strategic level. *Koninklijke Bibliotheek* is coordinating this action. The national libraries of France, Germany and the UK were amongst the initiators and are closely involved in the Task Force as well.

Other European Union related projects that create strong networks are for instance Delos (Network of Excellence on Digital Libraries),[51] Erpanet,[52] Minerva Europe[53] and the Firenze Working Group.[54]

Cooperation within regions
The German-speaking countries (Germany, Austria and Switzerland) have a strong mutual working relationship in several working groups and projects. The Scandinavian countries are strongly connected through their pioneering web-archiving activities, and often represent each other in the several activities on digital preservation. System related networks are emerging as well: such as the DIAS cooperation between Germany and the Netherlands or the participation of Austria in the DigiTool Exlibris working group.

Informal networking
Alongside these formal contacts the informal network of experts also plays an extremely important role in knowledge sharing and knowledge dissemination.

[51] Delos: See: http://www.delos.info/
[52] ERPANET: See: http://www.erpanet.org/
[53] MINERVA EUROPE: See: http://www.minervaeurope.org/
[54] Firenze Working Group: See: http://www.minervaeurope.org/structure/nrg/documents/firenzeagenda031017draft.htm

Informal contacts can reveal themselves in lectures, e-mail lists, informal visits for information exchange, intentional cooperation contracts or trainee posts. The relatively experimental nature of research within digital preservation and the relatively small field of experts, means that the communication lines between the experts are very direct and not difficult to establish. International workshops and conferences have proven their value in this context, for instance the European-Chinese Workshop on Digital Preservation in Beijing (China) in 2004. A follow up of this workshop is planned in Göttingen (Germany) in September 2005.[55]

Role models for cooperation

Easily accessible knowledge centres or centres of expertise for sharing information are equally important. These 'platforms' aim at knowledge sharing on digital preservation in general, and are usually not limited to solely the cultural heritage sector, but have a broad scope. The centre of expertise can serve as a guide that provides both theoretical and practical tools (glossaries, links to literature and websites, or organising meetings, etc.) via the Internet, as well as in every day practice. National libraries generally play an important role in coordinating (the initial) efforts.

PADI

The PADI website of the National Library of Australia (Preserving Access to Digital Preservation) serves as an international 'platform' for information dissemination.[56] To improve and ensure the management of digital preservation, PADI maintains a website for information supply (glossaries, reports, website links) and the promotion of relevant activities (information on research, conferences and workshops). PADI also provides an associated discussion list called *padiforum-l* for the exchange of news and ideas about digital preservation issues. PADI began in 1996 and can be seen as the subject gateway to digital preservation resources. An international advisory group which represents the whole cultural heritage field, has been established to provide advice and guidance for the PADI initiative. To be as up to date as possible, PADI relies on regular information provision from leading institutes in the field of digital preservation worldwide. One of the main advantages of this approach is that PADI has an undisputed role as the main international source of information, in part, because everyone has a share in it. There is generally no need to copy the PADI effort on a national level. Similar initiatives in non-English speaking

[55] European-Chinese Workshop on Digital Preservation 2004: http://www.csdl.ac.cn/meeting/cedp/index_en.html

[56] PADI: See: http://www.nla.gov.au/padi/

countries do exist (nestor in Germany), but always with a close link to PADI for the international aspect.

DPC, nestor, NDIIPP

Another initiative on knowledge sharing and dissemination is the Digital Preservation Coalition (DPC) in the UK.[57] In a similar vein to PADI, DPC also provides a website with current information on digital preservation activities worldwide, but DPC's added value lies in its achievements in building an (inter)national network of expertise within the digital preservation community at the same time.

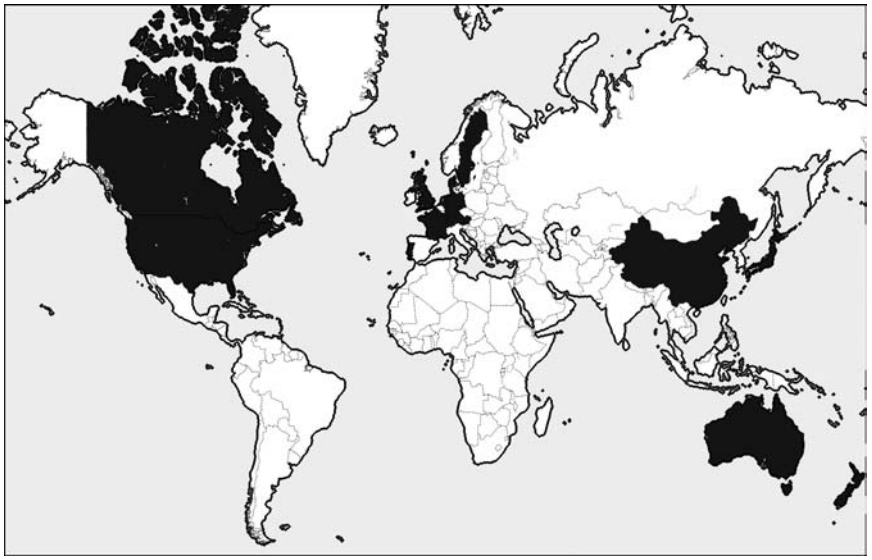
DPC was established in 2001 to foster joint action to address the urgent challenges of securing the preservation of digital resources in the UK and to work with others internationally to secure the global digital memory and knowledge base. This is not only done by providing information on digital preservation worldwide, but also by organising conferences and workshops, by initiating and participating in projects and by providing extra communication tools such as a handbook and the digital preservation award. DPC is built on a membership structure, current membership 27 including several coordinating bodies, and is not limited to the cultural heritage sector only. DPC functions as an independent body and has strong relationships with other similar organisation structures outside the UK, for instance NDIIPP (USA). It has become a company limited by guarantee and is an independent legal entity.

DPC is generally seen as an exemplar for establishing networks of excellence at a national level. In 2003 Germany established a comparable forum with nestor. Nestor (Network of Expertise in Long-term Storage and Long-term Availability of Digital Resources),[58] although focusing primarily on the German speaking digital preservation world and having a project status till 2006, has grown into a valuable source with an expanding international focus as well. After 2006 nestor hopes to continue as a structural facility. The NDIIPP programme in the USA also has the potential to grow to a level comparable to the DPC. More initiatives are currently emerging (for instance in Austria). As soon as there is a number of countries with it's own national coalition, the need for new coordinating initiatives on an international level might emerge.

[57] DPC: See: http://www.dpconline.org/graphics/index.html

[58] nestor: See: http://www.langzeitarchivierung.de/

All national libraries involved:



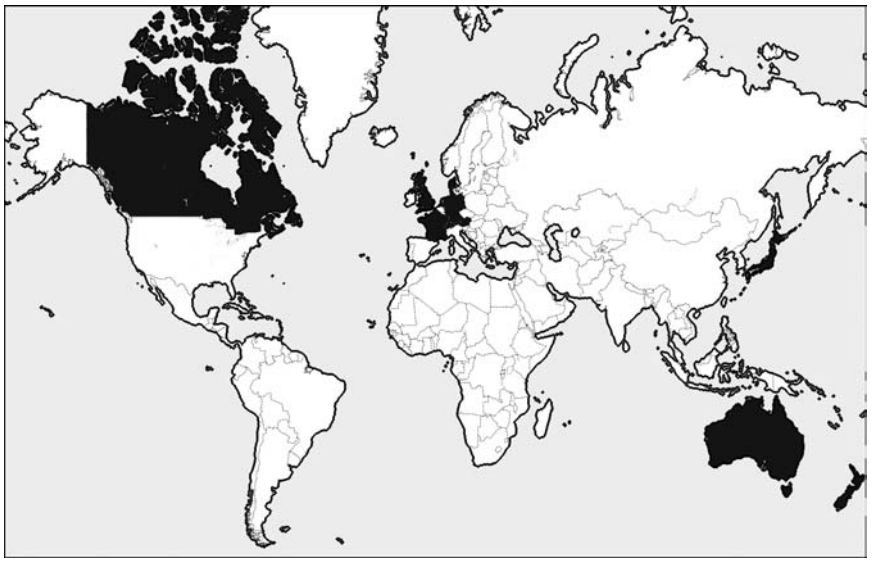
June 2005:

Operational repositories in Australia and The Netherlands
Illustration courtesy: Robert Gillesse, KB

December 2005:

Operational digital repositories in Australia, Austria, The Netherlands, The UK

2010:



Operational digital repositories in Australia, Austria, Canada, Denmark, France, Japan, The Netherlands, New Zealand, The UK

5. Conclusion

The situation at present

In mid-2005, a lot of activities are being undertaken in the field of digital preservation. This relatively new working area has obtained a central position within the national libraries and many developments are currently being realised. Several initiatives have been set out on which cover many different aspects of digital preservation, and these are now yielding tangible results.

Two of the national libraries surveyed, already have a fully-operational digital repository at their disposal right now. By the end of 2005 there will be two more, and within the next five years nine out of fifteen national libraries will operate such a facility. At least two countries with legal deposit legislation, expect a legislation covering digital objects to become effective this year. In 2007 seven of the fifteen countries involved will have a legal deposit for digital material. The first tangible measures to deal with copyright restrictions have been established. The first guidelines on preferred file formats, types of material and metadata for depositing will be published this year. The embedding of digital preservation within the organisational structure is gaining attention. International cooperation and knowledge dissemination structures have been set up and are already proving their worth. Plans for improvement are being developed right now. In fact all countries involved can be seen as pioneers, putting special efforts into specific aspects of digital preservation.

Broadening the scope

The emergence of digital archiving and preservation is having a growing impact on traditional day-to-day library routines. Activities on digital preservation often start off as a project. Once the project has been completed and the end result can be implemented in the organisation, things change. As digital preservation influences all aspects of traditional library tasks, all departments and units of the library get involved. A growing number of libraries are setting up a new working structure to ensure that the necessary measures can be implemented as smoothly as possible. They realise this by not only according digital preservation (day-to-day management of the repository and R&D) a central position within the organisation, but also by determining who will take the main responsibility for the subject. Several libraries have taken up initiatives to create cross-divisional working groups in order to secure a solid base for the future.
In this context it is interesting to note the emerging role of preservation and conservation departments with regard to digital preservation. Strengthening the

link between digital preservation and paper preservation in the near future, might benefit the general preservation strategy of the library. Since digital preservation inevitably leads to a blurring of national borders, libraries are widening their perspective on the outside world. Due to the growing complexity of digital objects – a document now scarcely contains just text – the distinction between library material and archival material is fading. And since the rise of Internet, publications are no longer strictly limited to one specific location. This is also leading to a blurring of the boarders between national and international publications. All things being considered, it is not just the imprint of a publication that is becoming unclear, but also the very nature of the publication. Can a document only be termed a publication when it is published by an official publisher? Or can everything communicated on the web be considered a publication?

Worldwide communication networks are gradually expanding. There is growing cooperation in the library sector, at both a national and international level. The same is true for cooperation between different types of heritage institutions (libraries, museums and archives) within the cultural heritage sector. Recently the first steps have also been taken towards establishing a closer relationship between the cultural heritage sector and the science sector.

Practices, best practices or standards?

To date there is only one globally accepted standard for digital preservation, the Open Archival Information System (OAIS). This standard, which has in fact been developed by the science sector in close cooperation with the cultural heritage sector, provides the digital preservation world with a reference model that offers guidance on the building and maintenance of a repository for the long-term preservation of and access to digital material. If nothing else, OAIS has provided the digital preservation community with a shared vocabulary and is thus facilitating and improving international cooperation.

At present various practices on digital preservation occur. These emerging practices are being closely monitored within the library community. The experts know where to find each other to discuss these practices. Yet as the scene is constantly changing and developing, it is still too early to talk about ‘best practices’.

Stating which aspects of digital preservation would benefit from standards development is not easy. Since digital preservation deals with cross-organisational themes, working out models that only address one specific subject often seems impossible.

Measuring progress

A picture of what has changed recently in the digital preservation landscape can be gained from comparing the outcomes of this report with the reports of Neil Beagrie[59] and PREMIS.[60]

Beagrie describes digital preservation in five countries (Australia, France, New Zealand, the Netherlands and the UK) in spring 2002. The report highlights focal points which emerged at that time on the following topics: national initiatives and funding; mission; collaboration and partnership; staff training and development; R&D and dissemination.

A comparison shows that progress has been made on several topics, but some things still need to be done. For example, progress has been made on (inter)national cooperation, funding perspectives, new deposit agreements (business models), awareness amongst stakeholders and government, globalisation of activities, development of a metadata framework (PREMIS), and certification of digital archives. The national libraries have made considerable progress in profiling their digital preservation activities on their websites, but it appears that maintaining efforts in this area is still difficult. In particularly, keeping the website information up-to-date, seems easier said than done. Since 2002, little progress has been made with respect to the development of persistent identifiers and selection procedures.

It is interesting to notice that since 2002 digital preservation has increasingly reached an operational phase. Consequently, the focus of R&D has shifted from archival aspects to permanent access issues. R&D in 2005 focuses more on the evaluation of operational aspects, the definition of practices and the evaluation of experiences. Another major change is reflected in the digital objects, which are becoming increasingly complex.

In Beagrie's report the national libraries involved gave four suggestions for international cooperation. Two of these suggestions still require further improvement: 1) the development of a preservation technology watch for file formats and new technologies, emulators and migration routines, and information on and repositories for obsolete software; 2) the development of shared services and central support for digital preservation in a distributed network of digital archives within larger national programmes. Several countries are carrying out projects on the third topic, the long-term requirements of the academic sector and scholarly communication in any national collaborative

[59] Beagrie, Neil. 2003.
[60] OCLC/RLG Premis Working Group. 2004.

scheme. The fourth area, fostering research on long-term preservation by developing standards and good practices can be viewed as an ongoing and still highly relevant concern.

The main focus of the PREMIS Survey, conducted in 2003−2004, was the management of preservation metadata in digital repositories in a broad context. A total of 48 cultural heritage institutions responded to the survey. As well as national libraries, the respondents included state libraries, university and research libraries, library consortia, archives, museums, and a few other organisations. As the scope of this ICABS report is limited to just national libraries, the practices of this type of institution can be compared with those of the broader group surveyed by PREMIS.

The PREMIS survey report gives a picture of a digital heritage community that has common practices, but which is not yet in a position to agree on standards. These common practices occur in six topics: the storage of metadata; the metadata scheme, the OAIS model, the maintenance of multiple versions of an object, preservation strategies, and the distinction between types of objects.

The PREMIS survey report made the following conclusions. Metadata are generally stored in two ways: redundantly in an XML or relational database or with the content data objects itself. The first practice allows fast access for use and flexible reporting, the second practice makes the object self-defining outside the context of the repository. For structural, descriptive and administrative metadata the METS metadata scheme is the most frequently used container, and for technical metadata for still images the Z39.87/MIX. All cultural heritage institutions use the OAIS model as a framework and starting point for designing the preservation repository. At the same time there is little flexibility to add functions and services that go beyond OAIS. With respect to the maintenance of multiple versions, there was a general habit of maintaining multiple versions (originals and at least some normalised or migrated versions) in the repository. For all versions the complete metadata sets are stored as well. Most institutions stick to a founded choice of multiple strategies for digital preservation. In general a distinction was made between different types of objects. All repositories record metadata pertaining to many different types of things: collections, logical objects, files, and bit streams. It was concluded that repository systems might make more granular distinctions and explicitly related the different metadata elements they record to the appropriate types of entities.

PREMIS concludes by summing up three best practices on storing metadata: the widespread use of METS, the practice of storing metadata redundantly in separate databases and with the stored objects and the idea of keeping all options open in the choice of preservation strategies.

When looking at the conclusions of the PREMIS Survey Report it becomes clear that the national libraries conform to the general picture when it comes to the choice of metadata formats and OAIS. A more detailed examination of the libraries, as provided in this report, reveals considerable movement and progress since the publication of the PREMIS report a year ago, particularly on the development of digital repositories. It is worth noting that the issuing to national libraries of the PREMIS data dictionary for metadata to support digital preservation, developed in 2005, could form a basis for interoperable preservation repositories and is well on its way to becoming a best practice for encoding preservation-related metadata.

Compared to both previous studies, this ICABS report not only gives an update of the achievements in digital preservation, but it also provides new information on workflow and organisational embedding. This might help our understanding of digital preservation in practice. A discussion of these issues was possible due to the focus of this ICABS survey on one specific type of institution and because the implementation of digital preservation within the organisational structure of the national libraries recently has become an important issue.

Future perspectives

In recent years considerable efforts have been undertaken to set the stage for digital preservation. Developments to build networks are well underway. In this context networks refer on the one hand to the technical networks that underpin the digital repository systems within the libraries and on the other hand to the emerging (inter)national cooperative networks for R&D and knowledge dissemination.

Despite the progress made, future challenges remain. Expanding the existing cooperation from being a network to building a network, will be a major achievement that can be used to develop concrete tools. Research and development has to be directed towards developing test beds, tools for preservation planning and procedures, tools for risk analysis, tools to evaluate solutions, and tools for the invocation of solutions. As the libraries that have an operational digital archive in place are also starting to develop permanent access solutions, a shift in R&D from archiving towards permanent access can be expected to take place on a larger scale over the next few years.

Digital preservation is still under construction. Yet one thing is clear from all of the achievements to date: the promising future of networking is about to start today.

(July 2005)

Imagine the challenge of digital preservation

Long-term storage

Photo courtesy: Reinier Deinum, Hans van Dormolen, Jacqueline van der Kort, Tom Thijs, Jos Uljee (KB)

Permanent access

II. Overviews national libraries

1. Overviews

Australia
National Library of Australia – Canberra

1. General

Legal deposit legislation
Legal deposit in Australia is mandated by the Commonwealth Government and the various State governments within the country. Under the Copyright Act of 1968 (and various State Acts), one copy of every library material published in Australia must be deposited with the National Library of Australia (NLA) and the appropriate State library, (and in some States to other designated libraries as well). These provisions extend to publications of private individuals, private and community organisations, as well as commercial publications.

Current Commonwealth legislation does not require the deposit of digital publications, although some of the States do extend their legal deposit provisions to various kinds of digital publications. NLA continues to seek the extension of legal deposit provisions for digital publications so as to ensure that a greater range is preserved for ongoing access.

NLA collects digital publications under voluntary arrangements, either by voluntary deposit of offline materials or by negotiated right to harvest online publications. The Library's collecting of digital publications is guided by publicly available acquisition guidelines. Commercially-produced audio and moving-image materials are collected by the National Film and Sound Archive, without legal deposit provisions.

Digital preservation in NLA

Organisational embedding
Since 1998 NLA has operated a Digital Services Project. This project forms NLA's key infrastructure strategy to support digital preservation activities. The aim of the project is to provide a technical infrastructure for the long-term management of digital material (both born-digital and digitised, both offline and online) in order to provide long-term preservation and permanent access. The project encompasses a wide set of IT development and procurement activities to support the overall framework and systems architecture for NLA's digital repository. The project is driven by the key library business processes of selection, acquisition, storage, resource discovery, delivery, access control and preservation.

NLA is organised into six main divisions and two other executive appointments, under a Director General. The two divisions primarily involved in digital preservation activities are Collections Management and IT.

Collections Management consists of six branches, of which three − Digital Archiving, Digital Collections Management and Preservation Services – play a major role in digital preservation. IT has five branches, each responsible for a particular part of the current digital repository and other library information technology systems: Business Analysis, Business Systems Support, Applications, Website Services, and Client Services. NLA's Digital Services Project is managed by the IT Division, but it is driven by the business needs of these branches, as well as those of others in the Library.

Day-to-day management of digital preservation is split between IT, Digital Preservation and Digital Archiving. The system is the responsibility of IT; research is split between conceptual research for which Digital Preservation is responsible, and implementation research for which IT is responsible. The Digital Preservation Unit is a part of Preservation Services.
The Innovation Division has some involvement in digital preservation at a policy level. See also Section II.2 for the organisational chart.

At present 26 people are involved in digital preservation: 4 in the Digital Preservation function; 5 in acquiring and archiving online materials; 2 in acquiring and cataloguing offline digital materials; 14 in creating digital content that is ingested for preservation; 4 (IT staff) in managing digital collection materials; and 2 (IT staff) in developing applications directly related to digital preservation functionality.

Developments in managing digital collections and services are guided at a strategic level by the Library's 'Directions' statements (updated every 3 years).[61] At an operational level, the key planning documents are the Library's IT Plan[62] and the Digital Preservation Policy.[63] There is a backup system of the digital repository both on location and at various locations of site.

Funding
Digital preservation activities (development and maintenance of the digital repository and research) are mainly funded from NLA's operational budget. A small number of research grants have helped fund specific research projects. For example, in 1996 the Australian Vice Chancellors Committee Working Group

[61] NLA Direction Statements: See: http://www.nla.gov.au/library/directions.html
[62] NLA IT Plan: See: http://www.nla.gov.au/policy/itplan.html[0]
[63] NLA Digital Preservation Policy: See: http://www.nla.gov.au/policy/digpres.html

on Electronic Publishing provided some funding to NLA for a 'proof of concept' pilot for the PANDORA Archive. And currently NLA is involved in the Australian Partnership for Sustainable Repositories (APSR), a project funded by the Australian Government's Systemic Infrastructure Initiative. Further information on APSR can be found in Section 4 (Current activities).

2. Digital repository

Status
NLA's digital repository consists of several interacting parts, including a Digital Collections Manager (DCM) (which includes a metadata repository and search system), a Digital Object Storage System (DOSS), and a Digital Archiving System for the PANDORA Archive (PANDAS).

DOSS is the digital mass store (disks and tape system) on which most of the digital collections are stored. Operational since 2001 and upgraded in 2002, DOSS is the responsibility of the Business Systems Support Branch.
PANDAS, DCM and DOSS are all digital object collection and management systems. PANDAS is the system for archiving and managing archived websites. Operational since 2001 and upgraded in 2002, PANDAS is being developed and maintained by the Applications Branch.

DCM has been developed to record management and technical information (metadata) about digital objects and surrogates. It is designed to support long-term management of the digital objects. It currently manages the digital image objects created through NLA's digitisation programmes. It is being developed and maintained by Digital Services, Applications and Web Services Branches.
Offline digital materials are managed by the Digital Preservation section using manual systems.

The overall system is operational, but depending on the digital material type the repository is in various stages of development: Storage and management of disk-based publications are in a planning phase. The preservation metadata management, the appropriate management systems, and management of digital audio are currently under development. The facilities for managing digital images and archived web materials are in full production, using separate management systems.

Services
The mission of the digital repository is to support access and to provide long-term management of digital collection material in the storage system. The repository not only stores material for preservation, it also holds derivative

copies for online delivery and systems for collection management of the materials and metadata, and for resource discovery. Once the digital repository is fully operational, it will provide services for search & discovery, controlled online access to service & archive copies; secure storage; data management and preservation treatments.

Depositing
Only NLA can deposit materials in the DCM. In PANDAS, PANDORA partners can also deposit the web publications they have selected for archiving and preservation. For the archiving of web-based material, the Digital Archiving Section negotiates permission agreements with publishers.[64] Disk-based materials are accepted under voluntary deposit agreements.[65]

Online publications are processed both manually and automatically; disk-based materials are processed manually. Websites and online resources are harvested by the PANDAS System. Image files, generated from digitisation programs are submitted to the repository as part of the workflow via the DCM. NLA will probably automate as much of the processing and quality control work as possible, but recognises that some residual manual checking may be required. The process will only be sustainable if most of it is automated.

Software and OAIS
Since 1999 the Digital Services Project has acquired or developed infrastructure components to support the management of NLA's digital collections. Although NLA prefers to buy software components, and only develops in-house components when off-the-shelf products do not meet their requirements, commercial products were unable to meet the Library's specialised needs. An in-house capability to build elements for the infrastructure therefore had to be developed and the management systems software for DCM was developed in-house. Accordingly DCM is not based on a more or less standard system like DIAS or FEDORA. However, NLA expects that these kinds of systems will probably be used in the future, when repository management software has developed a greater capacity to manage preservation functions.

The web-archiving management systems were developed with WebObjects Development software and incorporated an open source harvester (HTT Track). DOSS uses Application Storage Management (ASM) software for the storage processes. Online delivery is provided through an Apache web server. Metadata is held in Oracle databases.

[64] NLA Permission agreements with publishers: See: http://pandora.nla.gov.au/manual/general_procedures.html#formlet[0]
[65] NLA Voluntary deposit agreements: See: http://www.nla.gov.au/policy/vdelec.html[0]

NLA uses the OAIS Reference Model as a conceptual check for developing and managing its own digital repository. When process models do not fully comply with OIAS, this is both identified and monitored. NLA uses the key concept of OAIS with respect to archival information packages.

Materials
The digital collection of NLA includes digital publications acquired through purchase or voluntary deposit, digitised files generated by the Library (mostly in image and audio formats), and deposited original digital materials such as digital photographs and manuscripts.

The DCM manages objects created through NLA's digitisation programme (pictures, maps, printed music, manuscripts, books, etc.) and sound preservation programme (digital audio). Methods to store and manage disk-based or offline publications within the system still need to be developed. The PANDAS system manages websites and online publications.

For material that has been created within NLA digitisation programmes and managed through the DCM, standard formats are used to support preservation: currently, TIFF for images and BWF for audio. Items collected through the PANDAS system may be selected for ingest regardless of file format, although some items may not by currently ingestible due to technical constraints. Texts created by NLA to support delivery or to provide contextual metadata for items within the repository are encoded in XML, whereas NLA's website is in HTML.

For every digital object a preservation copy and an access copy are made. Both preservation and access copies are kept in the same repository, but are stored and delivered separately. Three copies of every harvested publication are kept in the web archive: the preservation master is a copy of the publication as it has been harvested (without changes); the access master is checked for quality and adjusted to restore functionality when necessary; an access copy is made from the access master on demand and used to give the public access to the object. Only the preservation master and the access master are stored.

Metadata and metadata schemes
For submission of certain types of material (e.g., databases) NLA expects depositors to supply relevant metadata. If this is not the case, metadata is automatically extracted (technical and structural metadata). Most of the administrative and descriptive metadata are submitted manually, although NLA will try to minimise this or to make this process semi-automatic.

In DCM the administrative metadata are submitted manually. The bibliographic metadata are automatically obtained from the NLA catalogue. Technical

metadata are obtained from created digital files or from creation processes administered by the system. For the digital audio files, the administrative and descriptive metadata are obtained from existing management database for materials. Technical metadata for this kind of material is automatically created by digital audio workstations. In PANDAS the administrative and bibliographical metadata are manually submitted. MIME data are extracted from the harvested files. The remote contributors who deposit in PANDAS are expected to submit a complete metadata set as required by the PANDAS management system.

NLA develops methods to accept or harvest batch metadata from publishers. Materials collected via other management systems are generally created and deposited in-house.

Most metadata are stored and updated in ODBC-compliant relational databases that are associated with and accessed through appropriate management systems. For these systems, which are not fully integrated yet, the metadata are held in flat files (associated XML-metadata files for digital audio files). For specific file types specific metadata may also be automatically embedded in individual file headers during file creation (TIFF for image, BWF for digital audio).

Metadata and files are generally stored separately. Persistent Identifiers are used to connect the metadata with the files. In the DCM the image masters (for preservation) and the derivatives (for duplicating and for access) are stored as individual files. Some system administration metadata are also stored with the file objects.

The stored metadata will contain information on rights and permissions, provenance, technical and structural aspects, administrative and management aspects and bibliographic/descriptive aspects. Information will be stored on collections, logical objects, non-digital source objects, files, bit streams and metadata. NLA is creating (or expects to create) metadata at all of these levels, although not all levels will be relevant for all materials. The required details have yet to be fully determined at each level.

Four types of metadata are considered important for preservation purposes: metadata to identify and characterise the object (including technical metadata), metadata about relationships with other objects, change history metadata to record object (and metadata) provenance, and any rights which must be managed. In future new metadata records will automatically be created for every new version of an object. In these records relevant change history metadata will be kept, in order to store information on transformations and to keep the relations between the historical versions.

NLA uses elements of the following metadata schemes: CEDARS, METS, MIX or Z39.87, NEDLIB, the NLA Metadata scheme and the NLNZ Metadata scheme. NLA is currently in a process of determining further specific preservation metadata requirements, so that a final model can be developed and approaches found for implementing a collection of the required preservation metadata within the repository management systems. This is likely to be based on the emerging PREMIS implementation model. In the future NLA does not expect to make exclusive use of one system, but will instead adapt recommendations and elements to fit determined requirements for particular materials, actions to be managed, business processes and system capabilities.

Access
Access to digital objects depends on the kind of material, the rights, permissions and restrictions. Only access copies are available for online access or distribution. Preservation copies are restricted and secured. They only will be made available through library staff and under special conditions. The library catalogue (for access) is not integrated in the repository management system, but both systems are linked. In the same way non-digital versions of digital material are linked with the digital object.
At present digital audio material is only accessible on site, and under special restrictions. In future, the digital audio material will be accessible for the public under the same conditions as the objects in the web archive, subject to access restrictions applicable to the original recording.

3. Preservation strategies

Preservation strategies actively applied at the moment are limited to the restriction on submission and bit-level preservation. Downstream strategies such as migration and emulation have not yet been implemented. For the in-house created digital objects NLA has chosen storage formats that are expected to have a long lifespan, before migration will be necessary. For digital objects created outside the Library's control, some archived formats do not have clear preservation paths, but they will be kept in storage. Research is continuing on how access may be reliably and effectively maintained.

In the future, restrictions on submission will remain for the in-house created digital objects. In that way NLA hopes to ensure standard formats and quality. No strategy will be excluded yet (migration, migration on demand, emulation and future strategies). Whatever is necessary and appropriate for the range of materials in the repository will be applied. In its choice of future strategies NLA is committed to maintaining the ability to present the look and feel of the digital objects it archives. Yet it also recognises that there will be constraints and for some materials representing the original look and feel will not be possible.

4. Current activities

National activities
NLA is currently involved in the following national digital preservation projects and programmes:

- PANDORA: (Preserving and Accessing Networked DOcumentary Resources of Australia). The PANDORA Archive of Australian online publications was established by NLA in 1996 and has subsequently been developed in collaboration with other Australian libraries and cultural institutions. A whole domain harvest of Australian websites is being undertaken in 2005, to complement the highly selective approach adopted for PANDORA. The Pandora website gives access to the collections and also provides information on the history, development, policy, statistics on size and growth, and the technical infrastructure. Partners in PANDORA are: Australian Institute of Aboriginal and Torres Strait Islanders Studies, Australian War Memorial, Northern Territory Library, ScreenSound Australia, State Libraries of New South Wales, Queensland, South Australia, Victoria and Western Australia. Term: 1996 – ongoing as a permanent commitment. For further information:
http://pandora.nla.gov.au/index.html

- ARROW: (Australian Research Repositories Online to the World). ARROW is a digital repository project with the purpose to identify and test software solutions to support best practice in institutional digital repositories comprising e-prints, digital theses and digital publishing. NLA will develop a repository and associated metadata for ARROW to support independent scholars (those not associated with institutions). A complementary activity of ARROW is the development and testing of national resource discovery services (developed by NLA) using metadata harvested from the institutional repositories, and the exposing of metadata to provide services via protocols and toolkits. ARROW is funded by the Commonwealth Department of Education, Science and Training. Partners: Monash University, Swinburne University of Technology, University of New South Wales Term: 2004−2006. For further information:
http://eprint.monash.edu.au/archive/00000046/01/ARROW_Public_Final_v1.3.pdf

- APSR (Australian Partnership for Sustainable Repositories): APSR is a digital preservation project with the purpose of developing demonstrator repositories and supporting the continuity and sustainability of digital collections. This also includes an investigation of the ramifications of accessing and managing research data produces for and generated by Australia's grid infrastructure and an exploration of DSpace. NLA will scope and provide advisory services to higher education institutions with respect to digital continuity through long-term preservation and collection management strategies. NLA will lead an

investigation of preservation metadata requirements to manage risks associated with the repositories, and a study of mechanisms to provide integrated access to available digital preservation tools. Partners in APSR are: the Australian National University (project leader), Australian Partnership for Advanced Computing (APAC), Universities of Sydney and Queensland.
Term: 2004−2006. For further information: http://www.apsr.edu.au/

Other institutions
Other institutes in Australia, active in digital preservation, include:
on library materials: a number of higher education sector libraries
on audio preservation: National Film and Sound Archive
on records archiving: Monash University, the Public Record Office of Victoria, The National Archives of Australia
on data archives: GeoScience Australia, the Australian Bureau of Statistics

International activities
NLA provides an international service for knowledge sharing, called PADI:

- PADI (Preserving Access to Digital Information): Partners: DPC (UK) and Erpanet. PADI is an information programme to support archiving and preservation efforts, by making knowledge available. It aims to provide mechanisms to ensure that information in digital form is managed with appropriate consideration for preservation and future access; to facilitate the development of strategies and guidelines for the preservation of access to digital information; to develop and maintain a website for information and promotion purposes; to actively identify and promote relevant activities; and to provide a forum for cross-sectoral cooperation on activities promoting the preservation of access to digital information. The PADI website is a subject gateway to digital preservation resources. It has an associated discussion list (padiforum-l) for the exchange of news and ideas about digital preservation issues. For further information: http://www.nla.gov.au/padi/index.html

NLA also participates in the following consortia and working groups:

- IIPC (International Internet Preservation Consortium) is a consortium of eleven national libraries and the Internet consortium. NLA is the leader of, and an active contributor to, the Deep Web Working Group. This is developing methods and tools for the identification of websites that are inaccessible to crawlers, the collection and storage of these sites, and the provision of access to them. NLA is also a member of the Framework, Researcher Requirements, Content Management, and Metrics and Testbed Working Groups. For further information: http://netpreserve.org/about/index.php

- ICABS (IFLA CDNL Alliance of Bibliographic Standards). ICABS is an alliance of six national libraries: the British Library, the Library of Congress, *Die Deutsche Bibliothek*, *Koninklijke Bibliotheek*, the National Library of Portugal and NLA. Within ICABS the NLA's activities focus on advancing the understanding of issues related to the long-term archiving of digital resources**.** There is a special focus on web harvesting (explore and promote methods to archive web-based publications collected by web harvesting) and on the preservation of digital materials (carrying out a survey of existing standards, guidelines and codes for preservation of digital materials (in cooperation with IFLA's Preservation and Conservation Section)). Within ICABS NLA and *Koninklijke Bibliotheek* are cooperating closely on various digital preservation subjects. Term: 2003− ; For further information: http://www.ifla.org/VI/7/icabs.htm

NLA is also participating in the Commonwealth Metadata Pilot Project, the purpose of which is to provide automated deposit and archiving of online government resources. This project will contribute to the archiving effort of PANDORA. The outcomes of the pilot project will be further refined and once that has been completed the project will resume. Term: 2003− ongoing. For further information: http://www.nla.gov.au/kinetica/metadata.html

A conference on web archiving was held in Australia at the end of 2004. This highlighted the need for action at a national level to identify digital materials requiring preservation, and for a more widespread acceptance of responsibility. A number of national agencies have expressed interest in working together to build a more coherent and comprehensive approach to managing digital information resources. NLA is currently paving the way for such cooperation by discussing the plans with the British Library, *Koninklijke Bibliotheek*, and the Digital Curation Centre. One focus of proposed cooperation is research to test and improve the PANIC service (Preservation webservices Architecture for Newmedia and Interactive Collections) being developed by the Distributed Systems Technology Centre (DSTC) in Australia, for format-risk identification and networked access to preservation tools.

Website National Library of Australia:
http://www.nla.gov/au/

Austria
Österreichische Nationalbibliothek (Austrian National Library) – Vienna

1. General

Legal deposit legislation
Legal deposit in Austria goes back to the sixteenth century. Deposit regulations for printed or published publications were reinforced in the early nineteenth century, and again in 1922 (Austrian Press Law). Since 1981 legal deposit has been regulated by the Austrian Media Law.
Österreichische Nationalbibliothek (ÖNB) receives and preserves four copies of every periodical Austrian publication, two copies of all other Austrian publications on paper, and one copy of every university thesis. In addition, the ÖNB also collects foreign literature that specifically refers to Austria and humanities literature relevant for the collection.
In 2000 the deposit law was extended with the legal deposit obligation for digital publications. This only concerns offline publications (except for music/audio and film media). ÖNB receives one copy of every offline publication. ÖNB is participating in a working group, coordinated by the Federal Chancellery of Austria, charged with drafting a legal deposit law for online publications.

Digital preservation in ÖNB

Organisational embedding
In 2002−2003 digital preservation issues were the responsibility of a special task force, comprising staff from the IT department, system librarians, cataloguing experts, a digital library expert and a legal expert. A special Digital Preservation Department was created in 2004. This department is a subdivision of the Collection Development & Processing Department (one of the three main departments reporting directly to the Director General). At present, the Digital Preservation Department has a staff of three, who is mainly concerned with strategic planning and research. The Digital Preservation Department is responsible for developing and implementing the digital repository and for coordinating digital preservation issues within the library, and at a national and international level.
The Department for Information Technology Services maintains the technical basis of the digital repository. Ingest and technical cataloguing currently take place in the Digital Preservation Department (technical issues, data conversion, etc.), whereas processes like acquisition and cataloguing are mainly assigned to other units of the Collection Development & Processing Department (Acquisition and Processing Department, Subject Cataloguing Department,

Central Bibliographical Services). See also Section II.2 for the organisational chart.

Funding
Digital preservation activities (developing the repository and doing research) are funded from ÖNB's institutional budget. No external funding is available at present.

2. Digital repository

In 2003–2004 ÖNB tested several software solutions for implementing and developing a digital repository. At the end of 2004 an implementation plan was accepted, and ÖNB started building and developing a repository. The current system version is considered to be an interim solution. Although the system is operational (ingest and storage), access is still limited to internal use and preservation functions still need to be added. The digital repository is expected to be fully operational by the end of 2005.

Services
The mission of the digital repository is to take care of ingest, data management, dissemination and long-term preservation of digital objects. Once the digital repository is fully operational, it will provide the following services: deposit and ingest of objects, object validation and transformation, search and recovery, online real-time access to service copies, secure storage of digital objects and data management. Preservation treatments are planned for the next stage.

Depositing
At present materials can be deposited by the general public, by the research community, by ÖNB itself, and by other (cultural heritage) institutes. There are several agreements on obtaining materials: the deposit of offline publications and certain governmental online publications is regulated by law, for example the Federal law gazette. There are also some special agreements with publishers for the voluntary deposit of online publications, as well as special agreements with universities for the deposit of e-theses and research papers.
Digital documents are submitted both automatically and manually. Certain publications are obtained by selective harvesting. Automatic harvesting of the .at domain is planned for 2006/2007.

It is expected that in future the workflow (submission, ingest, storage and access) will also be carried out both manually and automatically.

Software and OAIS

The digital repository is based on DigiTool from ExLibris (currently version 2.4). A new version of DigiTool (3.0) will be available in the summer of 2005. This version will be implemented in the repository to provide it with archiving functions. The final version of the repository will contain different modules, and will be based partly on commercial software (DigiTool as repository software and Oracle for database management), and partly on open source (Linux as operating system, tools like JHOVE for file format validation and metadata extraction, and a NAS system for storage).

The system is expected to be partially OAIS compliant for the following functions: ingest, archival storage, access and data management.

Material

As a legal deposit library, ÖNB has to accept and manage all types of files: offline media on various carriers, online publications and digital material from in-house digitisation projects. A pilot project on web harvesting took place in 2000−2002.[66] ÖNB expects to continue web harvesting in 2006/2007.

ÖNB examines the publication status of digital documents on the basis of four different categories: deposited digital publications, semi-published materials (theses and e-print papers, held in document servers and deposited by universities), non-published materials (such as personal records, digital manuscripts of the special departments of the ÖNB, e-mail correspondence) and digital surrogates of images, sound and text from non-digital originals held in the ÖNB. Software applications are not currently stored.

Various formats are accepted for the different file types (text, still image and audio). The preferred format for texts is XML, for still images uncompressed TIFF, for audio WAV, for websites ZIP archives (PNG Images, ISO disk images for whole CDs; PDF and HTML only with restrictions). ÖNB will publish recommendations for accepted file formats. A recommendation for PDF files based on the forthcoming ISO Standard for archivable PDF (PDF/A) is already available.

Metadata and metadata schemes

Metadata will be used to describe collections, logical objects and files. There will be special schemes for images, audio, video and text files. All metadata will be stored in XML containers within the digital repository. There will be a link between the library system of the ÖNB, which contains descriptions in MAB2 and the digital repository that contains (qualified Dublin Core) descriptive, technical, preservation, history, rights and structural metadata. The bibliographical metadata in the library system and the digital object in the

[66] Pilot project web harvesting: See: http://www.ifs.tuwien.ac.at~aola

repository will be linked through the URNs, based on the scheme urn:nbn:at. This uses MIX for images, AMD and VMD (LoC schemes) for audio and video, and a METS extension (textmetadata) for text. For rights metadata ÖNB uses the METS Rights Declaration, and preservation and history metadata are now based on the recently released PREMIS scheme (ÖNB previously used its own preservation metadata scheme, based on the scheme of the National Library of New Zealand). For automated metadata extraction and quality control ÖNB uses NLNZ metadata harvester and JHOVE.

Access
Access to the material will depend on the type of object and the permission given. There will not be any paid access.
The planned access workflow will be as follows: the user will use either the OPAC of the ÖNB library system, the OPAC of the Austrian Union Catalogue or the search interface of the digital repository. Metadata record from the library system will be linked to the digital objects in the repository. Via the search interface of the repository, features such as full text search, advanced search using Boolean operators, search in different metadata fields, search in desired media types and/or file formats, and browse through collections will be available. Access rights will be checked and if necessary conversion-on-the-fly will take place. The user will be able to access the resource online and use personalisation services such as saving records or search results. There will be different viewers for different types of objects (e.g., a multipage viewer for digital books). Some of the material will be available for restricted in-house use only and other material will be available free-of-charge. An API for metadata harvesting via OAI-PMH will also be available.

3. Preservation strategies

Currently the digital resources are stored in archivable formats. Regular backups and integrity checks take place. Backups are made on locally separated backup servers and on tapes. Additionally backup copies are stored in a governmental high-security data centre located several hundred kilometres away. If necessary damaged files will be refreshed. Obsolete formats will be migrated.
Preservation strategies (such as format conversion) are only applied on the preservation copies. Bit-level preservation is applied on all objects in the repository. Submission restrictions apply for e-theses and research papers. For the digital objects that result from preservation projects with hybrid filming and scanning (e.g., the historical newspaper project) additional microfilm copies are made for the preservation backup. In the future, ÖNB will focus on restriction on submission, normalisation, migration and migration on demand.

When choosing future strategies, ÖNB will primarily focus on the content, but keeping close to the 'original look and feel' will also be important factors. The strategy chosen will also depend on the type of object to be preserved.

4. Current activities

National activities
Most of ÖNB's current activities on digital preservation within Austria are focussing on the building and implementation of its own internal digital repository.

As a result of a conference on digital preservation organised by the ÖNB and the Austrian Commission of UNESCO in March 2005, a resolution has been signed (based on the UNESCO Charter on the Preservation of Digital Heritage) to set up a national strategy for long-term preservation in Austria.[67] A national cooperation platform, modelled on the DPC (UK) and nestor (Germany) will be set up, and the scope of the digital heritage that needs to be preserved and the responsibility of the different cultural heritage institutions will be defined. Efforts must be undertaken to adjust the copyright and legal deposit legislation. The costs involved need to be calculated and a research programme needs to be drawn up.

Other institutions
Other cultural heritage institutions in Austria, active in the field of digital preservation are:
on library materials: University Library of Innsbruck
on web archiving and preservation metrics: Technical University of Vienna
on digital art preservation: Ludwig Boltzmann Institute for Digital Culture and Media Science in Linz (research programme starting in summer 2005)
on audio preservation: *Phonogrammarchiv* of the Austrian Academy of Sciences and the *Österreichische Mediathek*
on e-archiving: Federal Chancellery

International activities
On an international level, ÖNB is currently participating in two projects that focus on digital preservation: the reUSE project and the Digital Preservation Cluster of the Delos Project.

[67] Resolution for National Strategy: See:
http://www.onb.ac.at/about/lza/veranstaltungen/unesco/

The reUSE project is funded by the European Commission (eContent Programme). Project partners are the University Library Innsbruck (coordinator), the national libraries of Germany, Estonia and Slovenia, the Humboldt University in Berlin, the University Library Graz and the Universities of Linz and Ljubljana. The purpose of reUSE is to set up trusted digital repositories maintained by national and university libraries in order to collect, preserve and make available digital documents that are currently used for paper-based publishing only, based on the OAIS principles. Term: 2004–2006. For further information: http://www2.uibk.ac.at/reuse/

In 2005 ÖNB started to participate in the Digital Preservation Cluster of the Delos Network of Excellence, funded by the European Commission (6th Framework Programme). ÖNB's role is to evaluate tools and strategies for file format validation and characterisation, and to work on a preservation evaluation metric. For further information: http://www.dpc.delos.info/

Further ÖNB is involved in two different Working Groups: the D-A-CH Group, and the Ex Libris Focus Group of Digitool software users. The D-A-CH Group is a trilateral working group that is working on strategies, metadata, formats, management of digital objects, persistent identifiers and workflows. Project partners are the national libraries of Germany, Switzerland and Austria. For further information see: http://www.onb.ac.at/about/lza/index.htm

At present the ÖNB is also participating in the preparations of the PLANETS project proposal for the Fifth Call of the Sixth Framework of the European Commission.

Website National Library of Austria:
http://www.onb.ac.at

Canada
Library and Archives of Canada − Ottawa

1. General

Legal deposit legislation
Canada has had legal deposit legislation in place since 1953, when the National Library of Canada was created. This law was amended in 2004, when the National Library and the National Archives of Canada were merged into a single institution, Library and Archives of Canada (LAC). The Act to establish the Library and Archives of Canada came into force by order of the Governor in Council on 21 May 2004.

According to the Library and Archives of Canada Act, Canadian publishers are required to deposit two copies of their publications with LAC: one for preservation purposes and one for the general collection. There are also regulations which support the legal deposit provisions of the Act. These are in the process of being revised. The Library and Archives of Canada Act introduces a new legal concept 'documentary heritage', which includes publications and records in all media related to Canada. The Act strengthens the mandate of LAC to preserve this documentary heritage, by providing for online publications and future new media to be included in legal deposit, for archiving websites of interest to Canada, and for the transfer of government records of archival or historical value, deemed to be at risk.

Legal deposit applies to all types of publications in all types of formats. The original legislation, which applied primarily to books, was extended to include serial publications in 1965, sound recordings in 1969, multimedia kits in 1978, microforms in 1988, CD-ROMs and video recordings in 1993, and digital publications on all types of physical formats in 1995. Legal deposit covers all individuals, associations, federal government departments and agencies, trade and periodical publishers, and publishers of audio, video, multimedia, microforms and digital publications issued in physical formats. One of the future goals of LAC is to extend the breadth of coverage of the legal deposit to digital publications and maps not currently acquired, by finalising the regulatory process to this extent.

Digital preservation in LAC Canada

Organisational embedding
A fundamental challenge for LAC is to balance the careful preservation of Canadian cultural heritage with the requirement to provide appropriate access,

both now and for future generations. Long-term preservation and access of digital objects is one of the focal points. At present LAC is closely monitoring international developments and has implemented solutions for digital publications, graduate theses, and some e-mail systems. Over the next three years solutions will be proposed for other digital materials, including Canadian websites, research data sets and the digital records.

LAC is one of the few national cultural heritage institutions in the world that covers and combines digital preservation activities in the library field and the archives field in one organisation. Since the practice of long-term preservation and access in both fields has different focal points, objectives, approaches and solutions, LAC is in the position to combine the practices of both worlds within one institution, and to possibly serve as an example for other countries.

In the period prior to 2004, when the National Library of Canada and the National Archives of Canada functioned as independent cultural heritage institutions, they were both active in their own field of digital preservation. However, since the merger in 2004 a review process has been taking place within LAC which covers all activities, work processes and practices, and involves rethinking approaches, relationships, services and programmes. In the 2005−2008 period LAC will reallocate resources to support the development of a Canadian Digital Information Strategy with other Canadian partners.

Recently LAC published the 'Report on plans and priorities 2005−2006'.[68] This contains the overall strategy of LAC for the 2005−2008 period.
Three strategic outcomes have been formulated: 1. Canada's documentary heritage is safeguarded and organised for future generations; 2. Canada's documentary heritage is known, accessible and used; 3. Information and knowledge are effectively managed within the Government of Canada.

LAC has three main operational sectors: Documentary Heritage Collection, Programmes & Services and the Government Information Management Office.

The Documentary Heritage Collection (DHC) Sector ensures that Canada's documentary heritage is collected, organised and safeguarded for current and future generations. The sector has three branches for the acquisition (including web archiving) and description of four types of material: publications, government archival records and private archival records, multimedia and special collections. The Intellectual Management Office, part of the DHC Sector, supports innovation and standards (including metadata expertise).

[68] Frulla, Liza. 2005. *Library and Archives Canada. 2005-2006. Report on Plans and Priorities*: http://www.tbs-sct.gc.ca/est-pre/20052006/LAC-BAC/pdf/LAC-BAC_e.pdf

The Care of Collections Branch, part of the DHC Sector, extends across all collections and takes care of maintenance, conservation, digitisation and preservation of the objects.

LAC also has four cross-sectoral branches: Strategic Office, Communications, Information Technology, Corporate Management. The Strategic Office takes care of strategic policy (national and international), strategic planning and rights policy and the coordination of initiatives regarding Multicultural and Aboriginal heritage. The IT Services Branch provides IT services, develops and maintains applications and provides infrastructure management. All sectors and branches within LAC have shared responsibility for digital preservation.

The IT Services Branch has approximately 110 employees who are responsible for LAC's IT infrastructure. The total number of staff involved in digital preservation throughout the organisation is difficult to assess.

The Documentary Heritage Collection bears the main responsibility for activities related to the first strategic objective ('safeguard and organise documentary heritage'). To achieve this the LAC will extend the legal deposit to digital publications and maps; develop an infrastructure and policy framework for managing and preserving digital content; develop processes for web archiving; establish guidelines and recommendations for protecting government records of archival and historical values in federal institutions; develop a framework for metadata; and develop a risk management framework for the care of documentary heritage collections.

Programmes & Services bears the main responsibility for activities related to the second strategic objective ('making the documentary heritage known, accessible and used'). It will realise this by transforming its service delivery to ensure easy access to one collection, both published and unpublished; plan and implement e-transactions and improve processes for providing access to government records.

Governmental Information Management Office bears the main responsibility for activities concerning the third strategic objective ('effective management of information and knowledge within the Government of Canada'). To achieve this LAC will, for example, develop records-management metadata and a new model for storing Government records of business value to all media. See also Section II.2 for the organisational chart.

At present the sectors and branches are housed in a number of buildings. In the autumn of 2004 all of the staff, with the exception of Program and Services, started to move to LAC's new facility at Place de la Cité, located in Gatineau.

This new facility neighbours the Gatineau Preservation Centre. Moves to the new facility will be completed within a year.

Funding
The digital preservation activities (development of a repository system and research) are funded from LAC's daily operational budget. Significant additional investments will be required to enhance the IT structure of LAC, to safeguard the cultural heritage of Canada and to ensure that this remains accessible for future generations.

The Canadian Initiative on Digital Libraries (CIDL) promotes, coordinates and facilitates the development of Canadian digital collections and services in order to optimise national interoperability and long-term access to Canadian digital library resources. CIDL is not directly involved in digital preservation but plays an in important role in the coordination and promotion of digital initiatives. Although CIDL does not provide direct funding for digital preservation, it does provide bursaries for learning opportunities in the field of digital preservation.

2. Digital repository

Status
One of the ways to achieve the first strategic theme for 2005−2006 is to advance the long-term accommodation infrastructure strategy. The collection is geographically dispersed; most buildings (with the exception of the Preservation Centre in Gatineau, Quebec) do not meet the standards for preserving documents. IT is a critical enabler for LAC to manage and provide integrated access to Canada's documentary heritage collection. A secure and reliable IT infrastructure is being designed to provide the flexibility and scalability required to handle the increasing complexity of collection information and to continue the trend towards digitisation and digital preservation.

In developing and implementing a new infrastructure and policy framework for the management and preservation of digital content, new methodologies, systems, tools, policies and procedures will be put in place to enable the organisation to streamline how it acquires, manages, preserves and provides access to Canada's digital cultural heritage. Many aspects of this have still not been decided upon. Following the completion of pilot projects in 2005−2006, LAC will develop an infrastructure for the online contribution and processing of digital archival records and digital publications, along with metadata for loading into corporate digital and metadata repositories. A digital content management system will be developed to facilitate the acquisition, management, preservation and accessibility of the digital documentary heritage. This new capacity will

complement other related acquisition activities such as the legal deposit of digital publications and the archiving of a sample of websites of interest to Canada. Secondly a streamlined process to capture, describe and preserve websites of significance to Canada will be developed, to make these available for long-term access. LAC aims to implement a technical infrastructure, guidelines and a policy framework for web harvesting and web archiving.

Services
The digital repository will ensure the long-term preservation of LAC's holdings and make these available to the public through circulation, loans and copying. The repository will therefore be built to provide long-term preservation and permanent access services.
Once the digital repository has become fully operational, it will provide long-term preservation, migration and access; search and recovery; online, real-time access to service copies and archival copies; secure storage; data management; storage and management of non-digital versions; and preservation treatment and formal distribution of archival copies on request. Prior to the creation of LAC, there was no repository for digital archives at institutional level. Work is underway to establish such an institutional repository.

Depositing
As a result of the current changes and reformulations, there is no information available at present with respect to current and future depositing.

Software and OAIS
It has yet to be decided which types of software LAC will use to build the institutional digital repository. How the new system will be OAIS-compliant is also still under discussion.

Materials
The digital repository system will contain all digital (machine-readable) objects within the LAC collection, including published and archival materials of national significance in all media. The new concept of documentary heritage encompasses publications, public and private archival records in all media, the legal deposit of online publications and a sample of websites of interest to Canada. It will also include computer-generated digital audio, digital still imagery, digital video, documents (text, e-mail, geospatial data, structured data), databases and spreadsheets, and CAD drawings. The Government Information Management Office is currently developing guidelines on recommended and acceptable file types and interchange formats. Recommended for use means LAC promotes these standards for the creation of computer-generated information from a purely technical rationale. These file types and interchange formats are also those preferred by the LAC for the transfer of digital

information to its control, after its operational business, value to an organisation, has ceased. These guidelines are still under development.[69]

Metadata and metadata schemes
Metadata are obtained in various ways. Thesis metadata are provided by the author. For e-collections, metadata may be provided by either the author or the publisher. Metadata related to video is auto-tagged, and metadata for online publications are extracted automatically. In the future, efforts will be made to obtain as much metadata as possible through automated processes. LAC requests depositors to deliver technical content metadata. However, at present there is no requirement to provide metadata.

Metadata are stored in many ways: in relational, XML, object-oriented, proprietary database or format, in flat files or bundled with related content files. Unzipped metadata are inserted in the header of tapes; metadata for video is captured in separate databases and linked with the repository. Preservation processes, such as migration information and reports, are maintained in a separate database. For e-publications there is a hot link from the bibliographic record in the LAC online catalogue, to the e-publication.

The stored metadata will contain information on rights and permission, provenance, technical and structural aspects, administrative and management aspects and bibliographic/descriptive aspects. The repository stores information at a collection level and a logical-object level, on non-digital source objects, files, bit streams and metadata. For normalised and migrated versions of digital material, information will be stored on the new format, the administrative metadata and the preservation metadata.

LAC uses metadata elements from different metadata schemes: CEDARS, METS, MPEG21, the NEDLIB scheme, the scheme of the National Library of Australia, OCLC Digital Archive Metadata, and schemes such as Dublin Core and RAD (Record of Archival Description) and AACRII (Anglo-American Catalogue Rules). To realise a new approach for enhancing user access, LAC will develop a strategic framework for using metadata. Metadata are currently produced as part of the description process of physical and digital media. LAC will contribute to the development of common international metadata standards. The aim is to develop methods of automatically accepting and normalising metadata that are submitted from outside LAC, for example, metadata from

[69] Brown, David L. 2004. *Library and Archives Canada: Guidelines on Computer File Types, Interchange Formats and Information Standards, Ottawa*: Library and Archives Canada. See:
http://www.collectionscanada.ca/information-management/0612/061204_e.html

publishers or metadata attached to government records. Ideally metadata will be created once, preferably at the time of content creation and/or publication, and then enhanced and reused in a variety of ways.

Access
All use of the digital objects within LAC is subject to legal restrictions, including copyright laws. How objects are made accessible (or will be accessible in future) will depend on policies concerning rules, place and public, the type of material and the permissions given. All kinds of access can be possible. There is no paid access to materials at present.

Access and preservation copies are served from the same source. For library material, LAC provides relationship links between access and preservation copies from Amicus (the national union catalogue of Canada) for specific digital files. For archival material there are links with the relational database for access purposes. The planned future integrated archival description access system, called Amican, will include both published and unpublished materials. LAC provides online access to electronic publications through the LAC website.
LAC aims to provide a seamless, efficient and multichannel access to Canada's documentary heritage in the future, subject to legal restrictions.

3. Preservation strategies

At present LAC is focussing on standardising formats for the creation, use and transfer of digital information, as an essential element of the long-term preservation of digital objects. This should allow reliable access to digital records for a period of five years, before the information must be migrated to a new format. The choice of future preservation strategies will be part of the development of the new infrastructure and policy for the long-term preservation of, and access to, the digital heritage. Whether the focus of future preservation strategies will be on safeguarding the content of the object or on safeguarding the original look and feel is not clear yet. This is still under discussion.

4. Current activities

National activities
In addition to the internal strategic project on the development of the infrastructure and policy framework to safeguard the Canadian heritage, LAC is at present actively involved in the following national projects on digital preservation:
- Canadian Initiative on Digital Libraries: E-Preservation: Preserving Digital Information in Canada. E-preservation was developed through a cooperative

effort between the National Library of Canada and the Canadian Initiative on Digital Libraries (CIDL). E-preservation is intended to provide Canadians with easy access to policies and research on the creation, use and preservation of digital collections. This project includes digital resources from libraries, archives and museums. The site focuses mainly on Canadian initiatives, and complements the PADI website (National Library of Australia).

- E-collection: Preserving Canadian e-publications for Long-Term Access. Purpose: Since 1994, LAC has been collecting, organising, and providing access to digital publications created by federal departments. During this time, LAC has established partnerships with 37 federal organisations, has collected over 1500 titles, and has created a guide to best practices for Canadian publishers.

- National Research Data Archive Consultation: In partnership with the Social Sciences and Humanities Research Council, LAC undertook a consultation and investigation with respect to the management, preservation of, and access to, social science and humanities research data in Canada. This consultation was conducted by a working group of experts in the fields of social science and humanities research and data archiving.

Other institutions
LAC has a key leadership mandate in partnership with central agencies and the information management community to develop processes, standards, training and practical tools for government users and federal libraries.

International activities
On the international level LAC is participating in IIPC, the International Internet Preservation Consortium. Partners are the Library of Congress, the British Library, the national libraries of Australia, France (coordinator), Denmark, Finland, Iceland, Italy, Norway, Sweden, the Internet Archive and LAC. One of IIPC's main objectives is the development of tools for web archiving. IIPC is currently working on finalising an IIPC toolkit for acquisition, selection and storage, which will incorporate IIPC standards (Arc 3.0, Metadata and API). The toolkit will be available in June 2006 as open source. For further information: http://netpreserve.org

Website Library and Archives Canada:
http://www.collectionscanada.ca/

China
Zhong Guo Guo Jia Tu Shu Guan – (National Library of China) − Beijing

1. General

Legal deposit legislation
Depositing of publications in the People's Republic of China is regulated in the National Copyright Act. Since 1916, the Metropolitan Library (the predecessor of the National Library of China) has received deposit copies of Chinese publications. This marks the library's function as a National Library.
The National Library China (NLC) has undergone several name changes since 1916. The first was in 1928 when it became the National Library (of Peiping). In 1998 the State Council approved a change to the Chinese name of the library and it became Zhong Guo Guo Jia Tu Shu Guan (National Library of China).

In 1996, the National Copyright Administration of China issued a 'Notice' on deposit of electronic publications. This Notice stipulated that after 1 January 1996, all electronic publications should be deposited at the National Copyright Administration of China, the Copyright Library of China (located in Beijing) and NLC. In January 1997, The State Council of the People's Republic of China established 'Publication Administration Byelaw No. 210', which regulated legal deposit to NLC.

NLC has the largest collection of Chinese publications in the world, and it collects all books published in China. NLC also puts an emphasis on the acquisition of unpublished materials in China and it has a large collection of publications in library and information science, a centre of Chinese yearbooks and a special collection of publications from the Hong Kong, Taiwan and Macao regions. NLC also has a collection of publications in foreign languages dating from the 1920s, the largest of its kind in China.

As one of the depository libraries of UN publications, NLC also collects publications of international organisations and foreign governments. The Doctoral Dissertation Centre (a subdivision of NLC) is the only organisation in China authorised (by the Committee of Academic Degrees, the State Council) to collect doctoral dissertations, post-doctoral reports and overseas Chinese dissertations since 1981.

Digital preservation in NLC

Organisational embedding
NLC is a comprehensive research library, a national repository of publications issued in China, a national bibliographic centre, a national centre of library

information networks, and the library research and development centre. NLC provides services for the central legislature, government, key research institutions, academia, education, business and the general public. NLC is also the national centre of bibliographical records, digital libraries and R&D in library science. It applies modern technologies and plays a leading role in standardisation, digitisation and networking in the library community in China.[70]

As the National Library, NLC plays a critical role in the long-term preservation of both Chinese and Foreign languages. NLC plans to preserve digital resources from Chinese main Internet content providers (ICPs) in its buildings. NLC is also the key library in the cross-preservation of basic Chinese digital libraries. NLC also plans to preserve purchased digital resources from foreign ICPs. NLC is cooperating with educational and science libraries in the development of digital preservation strategies.

In China, the digital library (including its infrastructure system design) is established as a government system. On 1 July 1998 NLC applied to the Ministry of Culture to implement the China Digital Library Programme. This project was established in 2000 by a joint meeting which was initiated by the Ministry of Education. In the same year, the Ministry of Education and Ministry of Science and Technology started to implement a digital library to serve their system.

In 2001 the State Council approved the start of the National Library of China Phase II (which means developing new buildings and new tasks of the library) and the National Digital Library of China Project. This is part of the China Digital Library Programme. All digital libraries are cooperating in standard research. The China Digital Library Programme focuses on composing a digital library of reproductions of primary source materials to ensure better preservation and to improve the study of Chinese cultural heritage. Another achievement of this program will be to set up a national platform to enhance the access to digital repositories which arise from this project. Main goals of the program are the creation of digital repositories, the construction of digital library soft- and hardware infrastructure, the development of application systems, setting standards and specifications structures, and training of professionals.[71]

---

[70] Zhizhong, Li. [et.al.] 2003. *National Library of China,* Beijing: National Library of China. See: http://bgu61.nease.net/homepage/nlc/nlc_introduction.htm#chart

[71] Liu, Wei. 2004. 'The new development in digital libraries in China'. Presentation on the International Symposium on Digital Libraries and Knowledge Communities in Networked Information Society. DLKC '04. March 2–5, 2004 Kasuga Campus, University of Tsukuba. (Tsukuba, Ibaraki, Japan). In cooperation with NSF, DELOS and the Japan Society of Library and Information Science: See: http://www.kc.tsukuba.ac.jp/dlkc/e-proceedings/papers/dlkc04pp120.pdf

The China Digital Library Programme (including the National Library of China Phase II and the National Digital Library of China Project) are being carried out within the framework of China's Five-Year Plan. The Five-Year Plan is a part of China's economic plan, which mainly focuses on the national critical construction project and the instruction of China's economy. To date, China has completed 10 Five-Year Plans.

NLC acts under the responsibility of the Ministry of Culture. Under the Library Director the library is divided in two parts: an Executive part and a part called Professional Department. The Executive part consists of the Directors Office; the Coordination Division; the International Cooperation Division; the Financial Department; The New Building Construction Division; and the Digital Library Administration Division.

The Professional Department consists of: the Acquisition & Cataloguing Department; the Serials Department; the Rare Books Department (including the Preservation and Conservation Section); the Stack Management & Reading Services Department; the Automation Department; the Reference Research Department; the Microfilm& Microform Centre and The Branch Library.
See also Section II.2 for the organisational chart.

Five sections and divisions within NLC are responsible for digital preservation. The Electronic Information Service Section (part of the Serials Department) is responsible for deposited and purchased digital resources. The Audio and Video Service Section (also part of the Serials Department) is responsible for deposited and purchased audio and video resources. The Rare Book Department and The Branch Library of the NLC are responsible for digitised rare books. The Stack Management & Reading Services Department is responsible for electronic resources with books.

Ten staff members are currently responsible for the maintenance of these resources. Digital resources are stored in the repository system and in the online reading system. Two staff members have a full time responsibility for these systems. Digital resources are preserved in four divisions: the Serials Department of the Audio and Video Section (audio and video); the Serials Department of the Electronic Information Service Section (purchased and deposited resources); the Rare Book Department (offline resources and resources that derive from the International Dunhuang Project, an international project on digitising manuscripts, paintings and artefacts from Dunhuang and making these accessible on the Internet); and the Branch Library of NLC (offline disks of the digital Gazetteer, a digital collection on the local history of villages and cities in China, produced by NLC).

Further the buildings of the NLC house 11 specialist organisations, companies, offices, etc. which are closely associated with NLC. These are The Branch Library, the NLC Publishing House (Beijing Library Press), the National Microfilming Centre for Library Resources; the Secretariat of the China Society for Library Science; the NLC R&D Institute; the National Centre of National Cultural Information Resources Sharing Project; the Office of Leading Group for NLC Digital Library Development; the Construction Engineering Office; the Division for the Supervision & Management of NLC State-Owned Assets; the China Digital Library Corporation Ltd and the Tuxin Sci-Tech Development Corporation. The Wenjin IT Research Centre is an independent subcompany of the China Digital Library Corp. Ltd, which specialises in the development of integrated systems for the digital library. The centre has a group of more than 30 employees. It is involved in the national and Beijing municipal projects and is responsible for software development and the maintenance of the computer systems of NLC.

Once the National Digital Library of China Project has been completed, a special Digital Resource Repository and Management Centre will be responsible for all the digital resources in NLC. This is expected to happen from 2007 onwards.

The electronic repository system will mainly be housed in the new building. After the completion of the new building, the digital resources will be stored in online, near-line, and offline systems. Some offline resources will remain in the storage facilities in the old building.
A data recovery system will be built to store all the electronic resources of NLC. A repository management centre will be built for the maintenance of electronic resources, including electronic account, preservation, examination and stimulation. A system-development centre will be built that will carry out research for the digital repository.

Funding
The National Digital Library of China Project is part of the China Digital Library Project, but it does not receive funding from the government.
At the moment the main investments are focussed on research and development of digital library technology to enhance the functionality and technology of the digital library. Long-term digital preservation forms the main topic in digital library research of NLC.
The government has granted 10 million Renminbi for the digitisation of NLC collections before 1949. NLC will apply to increase the funding to 30−50 million. There is no special funding for digital preservation at present.

2. Digital repository

Status
NLC began its library automation in the middle of the 1970s and started to use a mainframe integrated library management system in 1989 for the development and application of library automation.

NLC developed the Wenjin Integrated Library Management System, which is used for acquisition and cataloguing of Chinese books and circulation desks. Since 1995, NLC has been keeping track of the latest developments in digital libraries throughout the world. NLC is also responsible for several projects related to digital libraries and has completed a pilot demonstration system of the Digital Library, which has been one of the basic tools for the China National Digital Library project.

To make the China Digital Library Project work, NLC founded the China Digital Library Corporation (CDLC) which launched the China Digital Library Project in 2000. The CDLC aims to set up high-quality multimedia repositories that provide digital information services and develop e-commerce.

With the China National Digital Library project, NLC will realise the digitisation of document resources and the networking of service patterns. The Wenjin IT Research Centre has been devoted to the development of library automation and computerisation. Since 1996 the centre has been focussing on research and development for the digital library. To date it has achieved the following: a Digital Resource Processing and Service System and the Digital Library LAN Service System.

As mentioned above, a new building is currently being built for NLC, which is due to open in 2007. This includes not only a new library building, with improved space for users and materials, but also an opportunity to introduce new services such as virtual reference, web archiving and improved access to digital collections. It will also offer new opportunities for the digital electronic repository and digital preservation activities in general, as described above.

Services
The mission of the digital repository is to provide long-term preservation and access. Once the repository is fully operational it will be providing the following services: secure storage; preservation treatments; online, real-time access to service and archival copies; and formal distribution of archival copies on request.

Depositing
Currently more than 120 publishers deposit audio and video publications to NLC; more than 70 publisher deposit electronic publications. There are no signed agreements with depositors, since the deposit is regulated through the Notice on the Deposit of Electronic Publications.

At the moment the deposit workflow is operated manually. The publisher deposits both the publication and the list of deposit (bibliographic description). Once the National Digital Library Project of China has been established, NLC will be able to handle three forms of deposited material: material that is deposited manually; online deposition by a single person or institution, or batch deposition by institutions like universities.

Software and OAIS
In 2003, the integrated library management system Aleph500 became fully operational in NLC. This system is also used for digital preservation and access. The new digital repository system will be based on FRBR and will be OAIS-compliant. The standard model will be combined with the status of the NLC collection. It will be adjusted such that all NLC collections can be safeguarded. The physical and digital resources of NLC will be unified in the FRBR model. The OAIS model will be used for building the AIP, the DIP and the digital warehouse, including electronic account, copyright registration and data conversion, NLC expects to develop software in-house, depending on the internal wishes and demands, but commercial software will also be purchased.

Materials
The digital collection of NLC contains audio and video; electronic publications; foreign language databases; Chinese databases; Deptal databases; in-house generated bibliographic data; indexing data and full-text digitised images. The main focus of NLC concerns digital preservation of valuable text and images; websites issued through Cathay; and audio and video digitised in-house.

NLC makes a distinction between preservation and access copies. High precision digital resources are stored in the repository; frequently used and valuable research resources which are suitable for the Internet, are used for access services.

In future the repository will also function as a repository for digital output of other heritage institutions.

With the development of various information media and digital network services, NLC has been collecting numerous items of microforms and AV

materials, databases on CD-ROM and digital publications, both in Chinese and in foreign languages (including e-journals, thesis and dissertations in Western languages), the GAIE and UMI web-databases in science, technology and business and a websites on law and regulations. Once operational, the digital repository system of NLC will contain several types of electronic information resources: electronic publications collected by NLC; digital archives created in-house by digitising printed matter; information resources available on the internet and electronic information provided by external institutions.

Metadata and metadata schemes
NLC will use the Audio-Visual Metadata Standard Framework of the China State Administration for Radio, Film and Television material. There also will be a reference to the metadata framework, which has been developed in the Digital Audio-Visual Preservation Prototyping Project at the Library of Congress (USA).

Multiple models of metadata schemes will be used to store metadata and materials, some separately in MARC, DC and special audio and video schemes, and some stored within the repository with METS. METS will be used in the SIP (according to the OAIS model). Search systems will support multi-types of metadata, corresponding with DC to be compliant with OAI and also corresponding with MARC, to be compliant with Z39.50. Defined fields in MARC and DC will be comprised in METS; fields of audio and video metadata will be used. If necessary, new fields will be created to characterise the future digital resources.

Metadata for the digital repository will also be stored in the digital warehouse for backup. Metadata for distribution and service will be centrally and independently stored. Some types will be dispersed and stored in the same boot as the electronic resources, for instance DigiproMD metadata and administrative metadata.

It is expected that most metadata will be collected in an as automated a manner as possible. Part of the online resources metadata can be collected automatically, such as name, author and time. However some of this will need to be processed manually, such as class, subject, and abstract.

The Metadata Framework of China will include information on descriptive, administrative and structural aspects. For the long-term repository, preservation metadata is also being considered.
In future the metadata for digital resources will be obtained from the union metadata catalogue, which currently includes holdings of more than 900 libraries,

but will grow to 1500 libraries. They will be obtained through online capture; semi-automatic cataloguing and online deposit.

The digital repository will contain information on the following types of entities: for born-digital resources: information on background, bibliography, abstract and full-text; for digitised resources: information on born format, location, bibliography, abstract, colour management and original structure.

For digital preservation multi type metadata storage, encapsulated in METS, is considered the most important.

Access
In May 2005, a digital portal to collections of NLC was launched, providing online search and access to digitised resources of the library. This D-portal combines 37 Chinese language databanks, 77 foreign language databanks, some 16,000 periodicals in both Chinese and foreign languages, as well as special (digital) resources including local records, Dunhuang documents, periodicals of the Republic of China (1912−1949), doctoral dissertations and master's theses, all purchased or established by the National Library.

Currently the depositing publishers require NLC to protect the copyright. Therefore all the deposited resources only can be used in the library building without remote access.

3. Preservation strategies

At present NLC is applying the following preservation strategies for the digital library materials: secure storage, backing up, specification of naked data structure (preparing for migration) and normalisation. Since NLC discovered the disfigurements of CD-ROM, DVD-ROM and diskette in life circle and reuse, secure storage and backing up have become very important. NLC requires the ICP to specify the data architecture in detail for NLC so that the database can be reverted if necessary.
NLC expects to implement preservation strategies which are to be considered suitable for the long term, and that can be managed with low costs and simple technology. Emulation and migration might be applicable for resources considered valuable for research, for commercial purposes or for high utilisation.

When choosing a future strategy: both safeguarding the content of the object, and safeguarding the original look and feel are considered to be important.

4. Current activities

National activities
On a national level NLC is participating in the China Info Mall project, which focuses on preserving Internet Resources. Within this project NLC cooperates with the Peking University. The Web InfoMall, launched in 2001, is the China Web archiving project. Currently the Web InfoMall holds about 1 billion pages (15 terabyte). All archived data are freely accessible to the public. For further information see: http://www.infomall.cn/index-eng.htm

The Sci-tech retrieval centre is located in the National Science and Technology Library (NSTL). NSTL is a virtual institution which contains the Library of the Chinese Academy of Sciences, the National Engineering and Technology Library, the Library of Chinese Academy of Agricultural Sciences and the Library of the Chinese Academy of Medical Sciences. NSTL aims to: collect and exploit documentation and information in the natural sciences, engineering and technology, the agricultural sciences, and the medical sciences as required by national development; provide documentation and information services to the national science and technology community; and provide other information-based services to wider public. NLC is one of the partners.

Other institutions
Other institutions which are involved in digital preservation in China are for example:
- The China Academic Library and Information System (CALIS). Funded primarily by the Chinese Government, under the leadership of the Ministry of Education, the mission of CALIS is to promote, maintain and improve library resource-sharing between Chinese universities, and between academic libraries and other libraries and information institutions. As a nationwide academic library consortium, CALIS has been charged with building the China Academic Digital Library, so as to create and extend access to digital resources at all of the 1000 Chinese universities and to the public at large. The CALIS National Administrative Centre is located in Peking University Library.
- The Computer Networked and Distributed Systems Laboratory (CDSL) is also active in digital preservation, web archiving and a network grid. CDSL also gives training courses on several information technologies.
- The Peking University; this university is one of the partners of the NLC in the project China Info mall.
- The Shanghai Library is active in carrying out various digital preservation projects;
- China National Knowledge Infrastructure (CNKI) is a digital library which contains full text databases. These are multimirrored and function as backups for each other;

- The Chongquing VIP Information services company is based on the same principle as the CNKI.
- SSReader is a Chinese e-book provider which works with the multimirror principle. It preserves the material as TIFF format, and distributes it as PDF format. It contains 1.8 million digital resources;
- China Central Television (CCTV) archives digitised TV programmes in the form of disk cassette near line preservation. The CCTV archive contains more than 0.4 million hours.
- The Palace Museum: is active in cultural relic digitising and provides online disk and/or DVD-ROM preservation.
- The Digital Innovation Technology Co., Ltd. (Wenjin IT Research Centre) is developing state-of-the-art digital library software which includes a digital resource repository and management system based on OAIS.

International activities

On an international level, NLC is participating in the IDP project for the preservation manuscripts and paintings in DunHuan. Partner in this project is the British Library. For further information see: http://idp.bl.uk/

Website National Library of China:
http://www.nlc.gov.cn/

Denmark

Det Kongelige Bibliotek – Copenhagen & the Statsbiblioteket – Århus
The Royal Library – Copenhagen & the State and University Library − Århus (National Libraries of Denmark – Copenhagen / Århus)

1. General

Legal deposit legislation
Denmark's first legal deposit law was passed in 1697. The current law, the Act on Legal Deposit of Published Works, was passed by the Danish Parliament in 1997 and became effective in 1998. This law covers not only printed matter but all works published in Denmark, regardless of the medium used for the production of copies. This definition also covers published works on the Internet that form a final and independent unit and are produced for a Danish audience. A new version of the Legal Deposit law which also allows for the harvesting of the Danish part of the Internet was passed in 2004 and will be effective from 1 July 2005.

Publications are increasingly published in printed form, on the web and sometimes on CD-ROM as well. All three forms of publication are subject to legal deposit. Only one copy of each edition is placed in the national collection, which means that duplicates and unaltered editions are discarded. One copy is preserved in The Royal Library, *Det Kongelige Bibliotek* (DKB) in Copenhagen and one copy in the State and University Library, *Statsbiblioteket* (SB) in Århus. Originally (before 1902), book and art printers in Denmark and on the Faeroe Islands (except Greenland) were required to deposit between two and five copies of everything they printed with DKB. Between 1902 and 1997 one copy of everything printed was to be deposited with both libraries (one in the DKB and one in the SB).

This changed in 1997. Two copies of works in print, (except newspapers), in microform, combined works and works in digital form, Braille and photographs have to be sent to DKB. One copy will subsequently be sent by DKB to SB. Conversely, SB will receive two copies of newspapers, phonograms and videos. SB will send one copy of these materials to DKB, except for videos, a copy of which is sent to the Film Institute. Works in databases are also subject to deposit, if they are static. Dynamic databases were not included in 1997, due to the need for further clarification with respect to the legal aspects and technical problems associated with the law on collection, preservation and use. Computer programmes are excluded from deposit. Feature films and videos are covered by a more recent law requiring deposit with the Danish Film Museum. Radio and TV broadcasts are deposited on a voluntary basis at SB which houses the

national Media Archive. A form of legal deposit applies solely to the Danish Broadcasting Corporation. Admission to radio and television broadcasts is only permitted for research purposes in accordance with copyright law. The new law of 2004 (effective 1 July 2005) also includes Radio and TV programmes broadcast in Denmark and/or for a Danish audience. The National Media Archive (which is part of the SB) will digitise the programs as they are broadcast.

The so-called digital legal deposit[72] is currently based on producers reporting to DKB/SB that a digital publication is ready for deposit. The library then harvests the publication via Internet and ensures that it is stored. From 1 July 2005 the digital legal deposit will increasingly be independent of the producers because DKB and SB will harvest the Danish Internet on a regular basis. DKB bears the main responsibility for collecting and preserving books and serials, and SB has the main responsibility for newspapers and audiovisual material (movies excluded). The collection and preservation of Internet material will be a joint responsibility. The development of new legal deposit forms has taken place in a close collaboration between DKB and SB.

Digital preservation in DKB/SB

Organisational embedding
In Denmark two relatively recent 'initiatives' have been important for setting the stage for the current digital preservation activities. These are the DEFF (*Danmarks Elektroniske Fag- of Forskningsbibliothek* or Denmark's Electronic Research Library) and a survey by the Ministry of Culture's committee on the preservation of the cultural heritage, which lead to a report in 2003.

DEFF
DEFF was initiated in 1996 by three Danish ministries and several research libraries. Since 2003 it has received structural funding. DKB participates in the Steering Committee of DEFF. Both DKB and SB participate in several programme committees. One of the main reasons for initiating DEFF was the need to improve the cooperation on development of digital services and on digital content in the Danish library world. The mission of DEFF is defined as follows: 'To contribute to an optimal exploration of research-based information resources. This will be achieved through cooperation between library partners, common development projects and the establishment of a technical infrastructure.' The overall aim of DEFF is to improve the application of IT with a view to supporting research and education. DEFF supports joint projects where several libraries cooperate on testing systems and are prepared to pass on their

[72] Danish Digital legal deposit: See http://www.pligtaflevering.dk (Danish).

experiences to the rest of the library sector. Currently the main focus of DEFF is on digitisation activities, but DEFF also stimulates initiatives on digital preservation. DEFF has formulated three strategic lines of action. Within each line of action seven programme areas are defined: Technical frame and web environment (containing: system architecture, portals/tools and user facilities); Content (containing: licences); and New Applications (containing: e-learning and e-publishing). An action plan has been developed for each programme area, which contains several project proposals. These action plans form the basis for the concrete implementation of the strategy and for budgeting.

The system architecture to be implemented in DEFF will comply with common standards and will be based on the common (3-layer) model. Within the system architecture programme area, two projects have been carried out in 2004: on AAI (the Authentication and Authorisation Infrastructure) and on XML web services and the 3-layer architecture (XWS).

DEFF supported the activities which led to the development of strategies and methods for preserving the Danish part of the Internet with an international conference on relevant strategies in the summer of 2001. In the period 2001−2002, DEFF also supported a trial to identify the Danish strategies. The Ministry of Culture subsequently provided support for defining the required activities and supplied the relevant budget for developing the required software.

Within the framework of DEFF, the subject of institutional repositories has received a lot of attention from all universities and their libraries. Most are engaged in implementing such repositories, or will shortly make a start on this. This activity is related to the existing work on registering all research activities and the idea is to submit relevant material together with the metadata. To this end, a private company has developed a software programme for the registration part which will be supplemented with either DSpace or Fedora, to provide an archive for the material.[73] This work involves both DKB and SB.

Ministry of Culture – Committee on preservation of cultural heritage
In spring 2003, the Ministry of Culture's Committee on the preservation of the cultural heritage published a report that dealt with preservation issues in the seven major national cultural institutions of Denmark from an overall perspective. The seven institutions involved were museums and archives, DKB and SB. The report gives a number of progressive recommendations concerning the collection of the digital cultural heritage, within the framework of the new law on legal deposit, which will be effective as of 1 July 2005. This law revokes the difference between static and dynamic material on the Internet and central har-

[73] Pure Archive: See: http://www.pure.atira.dk/PURE

vesting of all forms of material from the Internet will be recommended. This means that unlike the physical offline cultural heritage, anyone who publishes material on the Internet will not be obliged to legally deposit it. The area of preservation is only dealt with on a very high level and the main recommendation is to form a group that will start to address questions concerning digital preservation. It is expected that this group will be formed in 2005. This report has been important for digital preservation activities in DKB/SB, since further funding was based on the assumptions in the report.

Responsibilities
Denmark has two libraries that function as a National Library. The Royal Library, *Det Kongelige Bibliotek* (DKB), in Copenhagen and the State and University Library, *Statsbiblioteket* (SB), in Århus. The content and profiles of the two libraries are very different, each reflecting their own mission, with DKB focusing on printed material and photographs and SB focusing on video and sound.

DKB and SB are collaborating on the development of a joint repository, where they will act as a backup for each other. The first example is the bit-archive holding of the collection generated by the web-archiving activities. Here, the storage systems of the two libraries are used to create one storage system, with geographically divided redundancies in the system.

Digital preservation in DKB
With a simultaneous focus on building up, mediating and preserving both conventional and digital collections, DKB is a hybrid library. There are two main focus areas: digitisation for access and digital preservation, both supported by the Ministry of Culture, and DKB's active involvement in DEFF.

As a National Library, DKB administers the national cultural heritage in terms of published works in conventional and digital form. DKB must provide the best possible access to the collections under present-day conditions for the purpose of research, studies and experiences, while at the same time making sure that the collections are preserved, secured and will be handed on to posterity.

DKB is divided in three administrative areas, which are divided into 17 function-oriented departments. Together the departments must fulfil the four main objectives of the library: University Library, National Library, Culture and Research. Administrative Area A (Område A)[74] contains the Department of Legal Deposit (with a special section for networked publications). Administra-

[74] Since the chart of the Danish libraries is in Danish, the Danish names of the departments are added here.

tive Area B (Område B) contains the Department of Documentation and Digitisation (DDA), which is responsible for building and developing the digitisation of DKB's collections. This department has four sections: Digital Objects, Subject Specialists, Web Editorial Section, Web Technology. DDA is responsible for the processing, dissemination and storage of digital documents and books as well as the continuous, ongoing development of web-based library services. It also takes the responsibility for the development and operation of systems for storage of net publications and digital collections, including collaboration with the SB regarding preservation of the Internet DDA also participates in various national and international projects and networks to develop and promote virtual information systems, for example Denmark's Electronic Research Library and CultureNet Denmark. See also Section II-2 for the organisational chart.

Currently digital preservation in DKB is the responsibility of the Digital Objects Section. At present a staff of 5 is working on digital preservation, all within the Digital Objects Section. The day-to-day operation is split between the Digital Objects Section and the IT department (Teknik) (sub department of Area C). Research & Development mainly falls under the Digital Objects Section . In the long term the Preservation Section will also be involved, but at present they do not have any experience in the digital field. One of the tasks of the Digital Objects Section is to ensure that the digital cultural heritage is stored for the future. The section has the technical responsibility for harvesting, storing and preserving Danish Internet publications.

The Subject Specialists Section is responsible for selecting books and periodicals in conventional and digital forms. It is also responsible for obtaining licences and giving access to licensed digital resources through the database application Elektra. The Web Editorial Section handles various development and digitisation projects in order to be able to continuously offer new digital resources to the users, more user-friendly web services, and online-services and guidance. The Web Technology Section is responsible for the operation of DKB's web servers and the operation of the increasing amount of digital resources in Elektra, a registry of the library's collection of digital periodicals, databases, e-books, etc.

Digital preservation in SB

SB in Århus serves three communities: the Århus University, the Public Libraries in Denmark and the general public. The library has a number of functions: it hosts the State media archive with films, sound and music collected since the start of the twentieth century; it hosts the National Newspaper Collection; and it has various other obligations such as a nationwide responsibility as a central agency for the interlibrary loan service.

SB has six main departments which are directly accountable to the Director General. Three of them concern the main functions of SB (University Library Department, National Library Department and a service for public libraries). The other three concern more overall tasks (the Administrative Department, the IT Department (IT&Proces), and the Department for Digital and Web Resources). The Digital and Web Resources Department (Digitale Ressourcer &Web) is divided into a Digital Resources Section and a Web Resources Section.

The work on digital collections takes place in three departments: the IT department bears the main responsibility for bit preservation, the Digital and Web Resources department is responsible for the software development of the relevant databases and ingest, and the National Library Department (Nationalbiblioteket) bears the main responsibility for audio, video and sound. The general policy adopted is that the work related to video will be outsourced and SB will acquire expertise needed to handle sound and music. The IT Department has the responsibility for the day-to-day operations of the digital repository system. A total of 10−15 people are involved in digital collections and in digital preservation. See also Section II.2 for the organisational chart.

The digital repository system and the backup systems are housed in the IT Department of DKB, with the exception of the Internet preservation project. Here there are two copies stored online, one in DKB and one in SB, and a third copy stored offline in SB. This current strategy will continue once the repository has become operational.

Funding
DKB and SB have received external funding from the Ministry of Culture for building the digital repository system and other digital preservation activities. In the initial phase some funding was also received from the DEFF Programme. In 2003−2004 DKB and SB received extra funding to work on technical aspects of web harvesting and to build a harvesting system, which will be ready to operate in July 2005.

2. Digital repository

Status
The joint repository system on which DKB and SB are currently working is expected to be operational in 2005−2006 (DKB part). The bit-archive, holding the collection generated by the web archive activities, is already operational. Here the storage systems of the two libraries have been used to create one storage system with geographically divided redundancies in the system.

Services
Once the digital repository is fully operational it will provide preservation and access to digital objects. The access will be limited, since the legal deposit law only allows access for research purposes. DKB/SB expect to formulate a mission for the digital repository in due course.

Depositing
DKB/SB have yet to formulate the deposit procedure and/or workflow for the national repository system in detail. For the statistics databases the requirement is fulfilled by obliging the depositor to inform the designated institutions about the publication of the database and at the same time ensuring that the institutions are provided with the necessary access codes and other relevant information needed to access the works. Agreements have been reached with the government, based on the legal deposit law, but there are currently no other signed agreements with depositors on digital materials.

Software and OAIS
At present DKB/SB are testing Fedora for possible future use in the digital repository. A decision has yet to be taken concerning the system and the type of software that will be used for the repository. The national system is expected to be OAIS compliant, where relevant.
In the *Internetbevaringsprojektet*, a project on web archiving in which DKB/SB are closely cooperating, the ARC format designed by the Internet Archive for its archival systems has been used. This format has been extended to serve the storage of converted files as well. Within the Net archive project, DKB's special focus has been on snapshot harvesting, whereas SB has concentrated on selective and event-based harvesting and delivered material. The two libraries are working closely together on the establishment of the entire archive and are cooperating to develop strategies and software for the collection, archiving, preservation and access of material. Archiving in institutional repositories is one of the areas already included.

Material
DKB/SB collects national cultural heritage in terms of published works, in conventional and digital form. The digital repository will contain all kinds of digital objects: online and offline publications, databases, web publications and so forth. The repository will also contain material from in-house digitising projects. Computer programs in the categories system programs, utility programs and tool programs are not subject to deposit. However, if a system program forms part of a work of another kind and is being published with it, it is covered by the law. Multimedia, educational computer programs and digitised reference books are examples of types included. At present the majority of the legally-deposited materials from the web are in HTML, JPEG or GIF.

DKB/SB believe that any file received for archiving must be preserved in its original form, in addition to any conversions that may take place, to allow for higher-quality conversion or emulation at a later stage. A distinction will be made between preservation copies and access copies. At present it is not known yet whether the digital repository will also function as a repository for digital output of other heritage institutions in the future.

Metadata and metadata schemes
Within the repository, metadata and materials are stored separately, but within the same format for harvested Internet sources. The stored metadata will contain information on technical, descriptive and preservation aspects, and on rights management. The repository will contain information on collections, logical objects, non-digital source objects and metadata. In the future, most metadata will probably be collected and stored by using automated processes wherever possible. However, submission by both internal and external depositors will also take place. For digital preservation, DKB/SB believes it is important to keep metadata on the original format type and format transformations, authenticity aspects, access limitations and of course content information such as the event and time, the registration time, the author, person(s) concerned, the location, copyright and description.

The focus with respect to the choice of a metadata scheme is very practical. At the moment DKB/SB are using DC qualified and the forthcoming IIPC standard WARC. Some experiments are being carried out with METS.

Access
DKB/SB subscribes to a large number of e-journals, e-books and databases. However, the licences limit the use of these digital resources and so access is only possible from within the premises of DKB/SB. At present the two libraries are only allowed to give online access to Internet material, broadcasts and movies, if the material concerned is not commercially available, and is solely for research purposes under very strict conditions. To give people access to archived Internet material the archive might be divided into an open and a closed archive (this would also tackle the problem of personal data).[75] Future access procedures still need to be worked out.

[75] Larsen, Svend. 2005. 'Preserving the Digital Heritage: New Legal Deposit Act in Denmark.'. To be published in: *Alexandria.*

3. Preservation strategies

Up until now, conversion and migration are the two preservation strategies that have been carried out. The choice of future strategies, for example instance emulation, will also depend on the level of funding the libraries receive.
When choosing a future strategy DKB/SB feel that a perfect copy of a digital object would preserve not only the appearance and functionality of the original, but the entire look and feel, i.e. the design and operational quirks of GUI elements, the resolution of the monitor, even the speed of the machine. While this may be overkill for most preservation, this is quite common amongst game enthusiasts. The main focus will however be on the content, although there are some digital objects (e.g., Flash) where the content has no meaning without the look and feel.

4. Current activities

National activities
In 2003 DKB established the basis for a database that will contain Danish scholarly periodicals in full text. This was the result of a project called Archive for Danish Periodicals, which was financed by DEFF. It is based on the idea that all research libraries share their resources. Via a digital periodicals archive with interdisciplinary search facilities, everybody interested is offered a user-friendly, free, and independent of time and place, access to a digital version. The master files created in this project will be stored for long-term preservation.

Currently DKB/SB are involved in a project for the preservation of the institutional research repositories. This project is being carried out within the framework of DEFF. The project will start in the autumn of 2005 and will end in 2006. There will be an intensive cooperation on information exchange with the repository initiatives that are currently underway in the UK.

In connection with the activities on institutional repositories, DEFF has given support to a project that will use the OAi-PMH protocol to harvest the relevant metadata and URLs – and subsequently harvest the material. This will ensure that the preservation of material in institutional repositories will follow the same procedures as web-harvested material. Work will start in October 2005 in collaboration with a similar UK-funded project Sherpa-DP, and the possibility to store the material in a structural form will be investigated.

Another national project on digital preservation in which DKB/SB are currently involved is Netarchive.dk, a project to develop a fundamental strategy for archiving Danish Internet material. For further information:
http://www.netarchive.dk

The aforementioned activities will begin as projects and all will subsequently be implemented in the day-to-day activities of the libraries.

Other institutions
At present the main activities on digital preservation in Denmark are carried out within the framework of DEFF. The main institutions active in the field are libraries.

International activities
On an international level DKB has been participating in the International Internet Preservation Consortium (IIPC) since 2003. One of the major goals of IIPC is the development of tools for web archiving. At present IIPC is working on finalising an IIPC toolkit for acquisition, selection and storage, which will incorporate the IIPC standards (Arc 3.0, Metadata and API). The toolkit will be available in June 2006 as open source. Partners in IIPC are: the Library of Congress, the British Library, the National libraries of Australia, Canada, Denmark, Finland, Iceland, Italy, Norway, Sweden and the Internet Archive. For further information: http://netpreserve.org

Together with a group of European national libraries, archives and universities, SB is preparing a project proposal to be submitted to the European commission in September 2005. The British Library is coordinating this proposal under the working title PLANETS (Preservation and Long-term Access through NETworked Services). This project will conduct research and development on preservation planning, file format characterisation and emulation/migration services.

Websites national libraries of Denmark:
http://www.kb.dk/ (Det Kongelige Bibliotek) and
http://www.statsbiblioteket.dk/ (Statsbiblioteket)

France
Bibliothèque nationale de France – (National Library of France) − Paris

1. General

Legal deposit legislation
France has had legal deposit legislation since 1537. This first deposit law only concerned printed books. Over the years the legislation has been extended to cover new forms of publications: in 1648 printed material like maps and broadsheets, in 1793 music scores, in 1925 photographs and LP Records, in 1975 videos and publications on various carriers, and in 1992 digital publications (CD-ROMS, etc.). The legal deposit legislation for offline publications became effective in 1993.

The *Bibliothèque nationale de France* (BnF) collects and preserves two copies of every publication, published, produced or distributed in France. This concerns all kinds of publications in every media. For paper publications France also has 19 regional libraries with a regional legal deposit function. These libraries do not collect digital publications. Since 1992/1993 the *Institut National de l'Audiovisuel* (INA) has been the legal deposit institution for radio and television and the *Centre National de la Cinématographie* (CNC) for movies authorised for public viewing in cinemas. Movies on DVD and promotional material for films are deposited in BnF.

BnF has no legal deposit for university papers and dissertations on paper or in a digital form. The depositing of this type of material is the responsibility of the university libraries. In 2000 there was a proposal to broaden the legal deposit legislation to cover digital materials on the web and online publications. This should have been enacted in 2003/2004, but legislation covering this kind of material is not in force yet. This is expected in July 2005. In anticipation of this law, BnF started to collect French websites through periodic snapshots, complemented by the voluntary deposition of deep web contents. In the near future new amendments will be necessary to regulate copying for preservation purposes.

Digital preservation in BnF

Organisational embedding
One of the main objectives stated in BnFs policy plan for the 2004−2007 period, is the strengthening of the fundamental mission. Reinforcing preservation and safeguarding the collections is one of the action points to achieve this. Preservation activities will not only focus on paper preservation, but also on the long-

term preservation of digital material that has been deposited (audiovisual, multimedia and digitised documents). A second action point is the building of a digital repository for the preservation of and access to all of BnF's digital resources. These efforts are supported by the Ministry of Culture (*Ministre de la culture et de la communication*).

BnF is housed at six different locations: three in Paris, one in Avignon (each with their own collections and reading rooms) and two locations for conservation, storage and technical workshops in Marne la Vallée (Paris Suburbs) and Sablé sur Sarthe. The current digital repository systems are housed in the François Mitterand Buildings in Paris. The new repository system will also be housed in Paris with a backup in Marne de la Vallée (Bussy St-Georges).

Within BnF there are three main departments that report directly to the Director General: Collections (*Direction des collections);*[76] Services & Networks (*Direction des services et des réseaux*) and Administration & Staff (*Direction de l'administration & du personnel*). Following an internal reorganisation of BnF in 2004, the decision was taken to divide the digital preservation tasks over several departments: the Conservation Department (*Département de la conservation*), the Information Systems Department *(Département des systèmes d'information*), the Digital Library Department (*Département de la bibliothèque numérique*), the Audiovisual Unit (*Services Audiovisuel*) and the Records Management Unit (*Gestion de la production documentaire et des archives*).

The Conservation, Information Systems and Digital Library departments are all subdivisions of Services & Networks, which is also responsible for the Legal Deposit Department. The Audiovisual Department and Records Management Department are part of the Collections Division (*Direction des collections*). Records Management reports directly to the Director General. The Conservation Department bears overall responsibility for digital preservation. This department is responsible for managing the digital repository system and, in cooperation with the Information Systems Department, is also responsible for the system development, exploitation and services. It also closely cooperates with the other collection departments already in charge of digital collections: the Audiovisual Unit, the Digital Library Department and the Records Management Unit.

One of the reasons why the Conservation Department was chosen to be in charge of digital preservation, was the decision that BnF would increasingly focus on digitisation as the sole preservation method. At present digitisation for preservation is already applied to colour material and high-quality black and

[76] Since the chart of the BnF is in French, the French names of the departments are added here.

white images. Some microfilming still takes place, but this is mainly for economic reasons, and it is expected that microfilming activities will be phased out within a few years. A very practical reason for this switch from microfilming to digitising is the expectation that it will be difficult to buy microfilms in the near future, as fewer firms will be able to produce films of an acceptable quality for a reasonable price.

The Audiovisual Unit has been involved in digital preservation right from the start, as this is the Unit that receives all offline digital publications. One of the copies is kept there and the other copy is sent to the BnF conservation building. One of the Audiovisual Unit's current activities is the development of an emulation strategy. The Digital Library Department is responsible for the access module of the digital repository system, including dissemination and the Deptalue with digital system applications. This department is also responsible for the entire web archiving process (from ingest to access). Once the new digital repository system is stable embedded in the organisation, these tasks will be integrated in the classical organisation of BnF. This is expected to occur within less then 5 years. See Section II, 2 for the organisational chart.

At present there are 26 people in BnF (including three full-timers) working on digital preservation, in several working groups: on metadata formats, on functional requirements, on specifications, on integration of guidelines and on preservation policy.

Funding
The digital preservation activities of BnF (development of and research for the repository system) are partly funded from the library's daily operational budget and partly from some specific funding from the Ministry of Culture (extra funding for the first part of the storage procurement procedure).

2. Digital repository

Status
BnF has two different repository systems at present: one for material that has been digitised for Gallica and needs to be preserved and one for audiovisual material. The repository for the audiovisual material has its own system for communication and storage. Part of the existing repository system is a so-called 'dark archive'. This is the part of the system where the preservation copies are stored.

In 2003 a coordinated working group within BnF drew up an implementation plan for building a digital repository. In 2004 BnF started working on the realisation of an integral digital repository system, in which the current systems will

be integrated. This repository, which is called SPAR (*Système pour preserver, archiver et répartir l'accès aux données*, or: System for Preserving, Archiving and Distributing Access to Data), must be operational in 2007. A procurement process is currently being undertaken for the storage part of the system, with a call for tender. Tests have been carried out and a decision on the storage part will be taken by the end of 2005. It is expected that the system will grow to 2 Petabyte within 7 years. One-third will be for the storage of web material, one-third for audio, and the rest for digitised material and voluntarily deposited material.

Services
The mission of the digital repository is to provide long-term preservation and access. There will be a separate repository for access services. Once the repository is fully operational, the system will provide the following services: search & discovery; secure storage, preservation treatments, and reporting. The access repository will provide online, real-time access to service and archival copies and the formal distribution of archival copies on request.

Digital objects are deposited at BnF because of the existing legal deposit law for offline publications, or because they come from digitisation projects within BnF or from research communities under cooperative agreements as part of the Gallica repository. Due to the legal deposit legislation there is a governmental deposit agreement with the French Ministry of Culture for offline publications. With the exception of this agreement and the cooperative agreements for Gallica, there are however no specific contracts with clients or depositors yet. However this will change once the legal deposit law on online publications has come into force.

Both harvesting and submission are used for web material at present. Websites are deposited through ftp of the physical carrier or through harvesting. Other digital objects are submitted. Until the legislation for online publications is passed, these types of publications are being harvested on an experimental basis. BnF is implementing a robot for the comprehensive capture of French websites for archiving purposes (harvesting). This is being done within the framework of the International Internet Preservation Consortium (IIPC). There also will be a voluntary deposit by the website producers on some deep web parts, to ensure that the collection is as complete as possible. In 2004 BnF realised the harvesting of the Regional and European Collections and the harvesting of the .fr domain in mid-December for a 1-month period. At present BnF is also conducting a voluntary deposit experiment with two regional newspaper publishers to test digital deposition as an alternative to printed deposition for all the local versions. This concerns the deposition of pre-production files.

Software and OAIS

The current archival repositories for the material from web archiving and digitising projects are OAIS compliant for the ingest, management and preservation stages, but not for delivery. The current communications/access system is not OAIS compliant. In the design phase the OAIS model has mainly been used to make the current systems compliant. SPAR will be based on the CNES (*Centre National d'Etudes Spatiales*) system which is OAIS compliant and has operated for more then 10 years. Currently BnF and CNES are setting up a partnership to implement the storage part of the system. BnF will not integrate the entire CNES system, but will modify existing CNES software for its own use. BnF will also develop additional software and tools in-house. A decision on the storage part will probably be taken by the end of 2005.

Most offline publications are currently processed manually by the Audiovisual Unit. Some tests on automated processing have taken place, for example to assess the physical status of objects and their quality. Digitised materials are processed automatically. Websites or portions of sites are ingested automatically after the harvesting process. The future digital records ingest will be as fully automated as possible.

Materials

At the moment BnF receives offline digital publications by legal deposit, web archives, institutional records, from its own collections and from digital surrogates of items it produces for preservation purposes, and for the online digital library (prestigious editions, dictionaries, image databases and periodicals). No materials from other libraries are stored. SPAR will store digital resources that are currently disseminated through Gallica, resources which are progressively constituted for conservation purposes (especially audiovisual data), the archives of BnF and resources generated by archiving the web.

File formats and quality level are defined by special BnF guidelines for digital material. BnF is currently working on new guidelines, which should be completed by September 2005 and will be suitable for the specific preservation purposes of SPAR. These new guidelines will also contain information on preferred formats for ingest. BnF currently accepts all formats offered.

At present preservation and access copies do not have the same format and are not linked. In the new system there will be one preservation format. The communication (access) format will be made on the fly. The preservation copy of objects from the web archiving projects and the digitising programme are currently stored in the dark archive only. The final decisions concerning the treatment of preservation and access copies have yet to be detailed.

A project is now being conducted with the ingest of the official Federal law journal. These journals have a permanent signature (due to authentication rules for e-publications from the government, which are expected to become part of the official deposit law in future). BnF unwraps the document at the ingest, and only stores the content and not the full envelope. It will be the task of the National Archive to study, develop and keep track of the digital archiving of the permanent signatures.

Metadata and metadata schemes
If possible, metadata will be submitted by the depositors. If not, it will be generated automatically by the system. There are currently no requirements for the provision of metadata. The stored metadata will contain information on rights and permissions, provenance (document history), technical and structural aspects, administrative and management aspects, and bibliographical/descriptive aspects. At present, some metadata are stored in the document file (in file headers and structure maps), and other metadata are stored externally in specific digitising documentation. For web archiving, metadata are associated with the documents through METS. It is expected that in SPAR most metadata will be bundled with related content files. A solution for updating the current dark archive has yet to be found. Information will be stored on collections, logical objects, non-digital source objects, files and bit streams. A decision on the storage of normalised or migrated versions has yet to be taken (the same applies for the metadata of these versions). Technical metadata is considered important for rendering and management metadata is important for information on the document's life cycle.

The metadata schemes used will be a combination of METS, the NEDLIB Scheme, and the OCLC Digital Archive Metadata. Whether or not METS-Rights.xsd will be used for rights declarations has not been decided yet. Recently the IIPC consortium started discussing metadata schemes for web archiving. A decision on which metadata scheme can be used for ingest and preservation is expected by the end of this year.

Access
At present, visitors can only have on-site access to the digital deposit material in one of the Paris locations. In future, access will depend on the access rights and permissions. For the material without copyright restrictions there will be open access for all end users, through the web, at a low resolution. All the material with copyright restrictions, and/or high-resolution images will only be accessible on site. Preservation copies are stored in the dark archive part of the current system and will only be accessible to BnF's Reprographic Department for reproduction on demand (under specific conditions). In the future system, access and preservation copies will be derived from the same original. An internal

discussion is currently taking place within BnF as to whether digital images should be sold or made freely available on the Internet. A decision has not yet been taken.

3. Preservation strategies

Digital publications are currently supplied with restrictions on submission for specified formats or quality, bit-level preservation (secure storage, backup, refreshing, etc.), normalisation, migration, migration on demand and emulation for offline publications. The same strategies will be applied once SPAR has become operational. When choosing future strategies BnF will pay attention to both preserving and giving access to the 'original look and feel' of the objects and to the content. For preservation purposes efforts will be made to retain the original look and feel without damaging the content. However, these two aspects require different technical decisions and so discussions are still taking place as to which aspect should from the most important focus.

4. Current activities

SPAR is BnF's main project on digital preservation. This is an internal project, aimed at realising a long-term storage system, the definition of a preservation metadata set, and a persistent identifier system. At the moment the technical requirements for the ingest of digitised material are being modified to adapt them for preservation, and to address the issue of the automatic generation of metadata. BnF is working closely with CNES to realise the construction of the storage part of the system. Term: 2004−2007.

National activities
On a national level BnF is participating in the PIN working group (*Perennisation des Information Numeriques*). This is an informal working group coordinated by the CNES, whose task is to work out an OAIS reference model standard. Recent achievements are: a French translation of the OAIS standard which will be submitted to ISO; a training session on digital preservation for librarians and archivists and a training session for the ENSSIB (the National School of Librarianship). BnF is responsible for these training sessions which are intended to raise awareness of and provide practical approaches to digital preservation activities, such as in-depth teaching on metadata and functional requirements. A national conference on legal aspects of digital archives, with a special focus on preserving digital signatures, was also organised. For further information: http://www.vds.cnes.fr/pin/ (French).

Other institutes
Other institutes in France, active in the field of digital preservation are for example: on library materials: the University Libraries of Lyon and Marne de la Vallée; on audiovisual materials: the *Institut National de L'Audiovisuel*; on e-archiving: the National Archives, and several regional, departmental and municipal archive agencies. The National Archives has developed guidelines for digital archives and is also involved in building an digital repository for long-term preservation; others: CDC Zantaz[77] is a private company, offering digital archiving facilities in France. Digital art preservation has yet to be fully developed in France.

International activities
On an international level BnF is participating in IIPC, the International Internet Preservation Consortium. BnF has a coordinating role within this consortium. Partners are the LoC, the BL, the National libraries of Australia, Canada, Denmark, Finland, Iceland, Italy, Norway, Sweden and the Internet Archive. The development of tools for web archiving is one of the major goals of IIPC. At the moment IIPC is working on finalising an IIPC toolkit for acquisition, selection and storage, which will incorporate the IIPC standards (Arc 3.0, Metadata and API). The toolkit will be available in June 2006 as open source. BnF and the BL are currently conducting a smart archiving crawler project, to implement large-scale, automatic focus crawls. The priority will be based on citation linking and thematic assessment. The first prototype is expected mid-2005. For further information: http://www.netpreserve.org.

BnF also was one of the initating libraries for the FP 7 Task Force, an initiative that arose from the Netherlands' presidency of the European Commission in 2004. For further information: http://www.kb.nl/coop/euconference/frame-conclusions.html.

Website Bibliothèque nationale de France:
http://www.bnf.fr/

[77] CDC Zantaz: See: http://www.cdc-zantaz.com

Germany
Die Deutsche Bibliothek – (National Library of Germany) − Frankfurt/Leipzig/Berlin

1. General

Legal deposit legislation
Germany has had legal deposit legislation since 1969. *Die Deutsche Bibliothek* was established in 1990 on the basis of the Treaty of Unification in a merger of the existing institutions the *Deutsche Bücherei* Leipzig (founded in 1912) and the *Deutsche Bibliothek* Frankfurt am Main (founded in 1947), of which the *Deutsches Musikarchiv* Berlin (DMA) has been an integral part since 1970. The Treaty of Unification enabled the institutions to join together in fulfilling the legally-specified objectives of the national library. *Die Deutsche Bibliothek* is responsible for the collection, processing and bibliographic indexing of all German and German-language publications issued since 1913.

Every commercial and non-commercial publisher in Germany is required to submit two deposit copies of each publication to *Die Deutsche Bibliothek* (DDB). Germany consists of 16 federal states. The location in Frankfurt receives two copies of each publication from 9 states; the location in Leipzig receives two copies of each publication from the other 7 states and one copy of all publications from abroad. Once described, one copy of the duplicate set is forwarded to the partner library for archiving and use. The single copies of publications from abroad are only archived in Leipzig. Printed and recorded music are only processed at the DMA in Berlin. After processing DMA forwards second copies of these to the *Deutsche Bücherei* Leipzig for archiving. This requirement applies to traditional publications on paper, microforms, sound recordings and offline digital publications.

Online publications (digital periodicals, digital full texts and websites) are not yet covered by law. For this category of publications DDB has formed a working group, Digital Deposit Library, together with publishers and producers to test procedures and to make agreements. Based on these agreements DDB already collects networked digital publications on a voluntary basis and this includes university dissertations and theses (1998). A new law, *Gesetz über die Deutsche Nationalbibliothek* (Law concerning the German National Library), which also covers the legal deposit of this type of material is expected to come into force in 2006. Online digital publications are only deposited in Frankfurt.

To prevent the illegal copying of music, CD-ROMs and e-books, the rights of the artists and authors are protected by special copyright law. Often material is

also technically protected against copying. However, the long-term preservation of these materials cannot take place without copying. In November 2004 DDB, the *Bundesverband der Phonographischen Wirtschaft* and the *Börsenverein des Deutschen Buchhandels* signed an agreement on the use (and duplication) of publications that are under restricted publication laws, so that DDB will be able to fulfil its task on preservation and access. This means DDB has been permitted to provide users with copies of the deposited publications, when they are used as a DDB access copy, as a working copy for educational and scientific purposes, as a copy for a school collection, or when it has been deposited for two years for own use. DDB is responsible for personalised digital watermarks on the copied material and signed restricted use agreements with the customers, and it has to provide an annual report on how copyrighted works are used. The depositors agree to deposit online publications which are as free from technical protection as possible. The agreement will be reviewed by all three partners on an annual basis.

Digital preservation in DDB

Organisational embedding
Digital preservation activities are coordinated by the Information Technology Department. This Department is one of the two main departments that report directly to the Director General of DDB and do not fall under the responsibility of one of the two location directors. The IT Department is mainly housed in Frankfurt, with branches in Leipzig and Berlin. See also Section II.2 for the organisational chart.

For the development and construction of a digital repository that meets the standards of a trusted repository for long-term preservation and access, DDB has formed a partnership with the *Niedersächische Staats- und Universitätsbibliothek Göttingen*, the *Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen* (GWDG) and IBM Germany. The GWDG will be responsible for the technical operation of the long-term archive. IBM *Deutschland* will ensure a professional adaptation of software components and provide long-term support. This partnership falls within the kopal project (see below).

At present approximately 15 people from all four partner institutions are involved in building the repository (5 of them in the IT Department of DDB). The system will be physically housed at the GWDG in Göttingen and a backup system will also be housed in Göttingen. Another backup system to be located in Munich is under consideration. In June 2005 a coordinated task force, in which all three DDB locations are participating, started working out a renewed internal organisational structure for digital preservation activities (research and

day-to-day handling of the repository). In the future digital preservation will not only involve the IT Department, but the other main departments as well.

Funding
In 2004 DDB started the kopal project to build a long-term archive for digital data (kopal = *Kooperativer Aufbau eines Langzeitarchivs digitaler Informationen* or cooperative development of a long-term digital information archive). The Federal Ministry for Education and Research (*Bundesministerium für Bildung und Forschung*) is funding the project for three years. Further, the four partners are to support the development and building of the digital repository from the daily operational budgets of their own institutions.

2. Digital repository

Status
DDB has processed digital publications since 1998. However, the system used does not yet meet the requirements of a repository for long-term preservation and access. Early in 2004 a start was made on the design and development of a trusted digital repository. The digital repository is expected to be operational in 2006. The kopal project will end in 2007.

Services
The mission of the digital repository is to ensure the long-term accessibility of digital objects. In the future it will also provide an opportunity for other institutions (academic, business and administrative) to make data available on a long-term basis. Once the digital repository is fully operational, it will provide the following services: long-term preservation of and access to the authentic form, in so far as the technical solutions allow this, online real-time access to archival copies, secure storage of digital materials, preservation treatments (e.g., migration) and reporting.

Depositing
DDB currently receives digital publications from German publishers (Springer and other publishers affiliated with the German Booksellers and Publishers Association (*Börsenverein*)) and digital theses from university libraries. Agreements on the voluntary submission of online publications have been reached with both universities and publishers. Due to the legal deposit legislation there is also a governmental deposit agreement with the German Federal Ministry of Education and Research.

At first, only DDB and the University Library of Göttingen will be allowed to make deposits in the digital repository. However, at a later stage it must also be possible for other institutions to deposit materials. At present, digital publica-

tions are processed both manually and automatically. The method used depends on the kind of material and how it has been supplied. In the future most of the processing will be done automatically. Both submission and harvesting will be used once the repository has become operational.

Software and OAIS
The digital repository will be based on the Digital Information Archival System (DIAS) developed by IBM for the *Koninklijke Bibliotheek* (KB) in the Netherlands. It will be built according to the international standards for long-term archiving and metadata as described in OAIS. Where possible the digital system must be integrated with existing library and information systems used in Germany. The use of object formats through metadata schemes such as METS and LMER will require special attention.

A special DDB/KB User Group has been set up (together with IBM Germany and IBM the Netherlands) to make the two DIAS systems (in the Netherlands and in Germany) as compliant as possible. For both countries the basic system will be extended with additional tools, based on individual wishes. The DIAS 2.0 version developed will only be available to the DIAS Partners. The local workflow software that the four German partners develop for kopal will become available as open source.

Materials
The digital repository will contain all of the material that DDB is legally required to safeguard. Göttingen will supply a lot of digital objects derived from the SSG activities (*Sondersammelgebiete*). This is a German-wide project funded by the *Deutsche Forschungsgemeinschaft* for the collection and indexing (including the long-term preservation) of international publications on various disciplines, in both paper and digital form. The repository will also function as an academic repository for the University of Göttingen.

It will contain all kinds of digital documents (offline and online publications, including digital periodicals, digital full texts, websites) in PDF, TIFF, TeX and similar formats, including complex objects such as digital videos. In future the system will also contain data from other institutions (libraries, universities, large companies and governmental institutions). At present, most of the files are stored in TIFF, PDF and WAF, but in future, the repository will accept any type of digital file.

Metadata and metadata schemes
Most metadata will be supplied by the depositor. The stored metadata will contain information on provenance (document history), technical and structural aspects. Some technical metadata will be generated in the digital repository, in a

relational database, together with related content files. Descriptive metadata will be stored in a separate cataloguing system (outside the repository system). Information will be stored on logical object and file level. The storage of the different types of metadata in the DIAS system has been defined. When migration takes place, all different and pre-versions (and metadata) will be safeguarded, so that the migration history of the objects can be restored. Metadata for specific identification of the file formats are considered to be the most important metadata (meaningful in its own right and as a reference to a detailed format repository).

In future, automatic extraction will have to be used as much as possible. There are preferences with respect to the supply of metadata, but METS will be a guideline. Other metadata schemes that will be used are: the scheme developed in NEDLIB and LMER. LMER is a metadata scheme for technical metadata, developed by DDB for use in long-term preservation aspects. This scheme is based on the model of the National Library of New Zealand.

Access
Initially, only the main suppliers (DDB and SUB Göttingen) will get access to their own material. Access to the material will depend on the access rights and permission. In the future it is expected that on-site access and open access will both be possible for specific types of material.

3. Preservation strategies

At present the digital publications are supplied with bit-level preservation (secure storage, backing up, refreshing, etc.), migration, and reformatting to more current versions of the format once the source format becomes obsolete. Once the repository has become operational migration, migration on demand and emulation will be applied to maximise the safeguarding of long-term preservation and access. Initially the thinking on future strategies was lead by the idea of giving priority to safeguarding the content of a digital publication. However keeping the 'original look and feel' of a document is now also considered to be an important aspect. Therefore, both approaches will be taken into account when drawing up preservation strategies.

4. Current activities

National activities
DDB is currently involved (as a leading partner) in two main projects on long-term preservation: kopal and nestor.

- kopal: is an operational project focussing on the cooperative development of a depot system for digital resources. It is considered to be both a system (project result) and a project. Project partners are DDB, the State and University Library of Göttingen, IBM Germany GmbH and GWDG (*Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen*). The kopal Project must be completed in 2007. For further information:
http://www.kopal.langzeitarchivierung.de/ (German)

- nestor: (Network of Expertise in Long-Term Storage of Digital Resources). The nestor project aims to bring together existing knowledge, people and expertise on the long-term storage of digital resources as a starting point for a future alliance for Germany's digital memory. Nestor aims to develop a network for information and communication for current and future long-term preservation activities in Germany, to establish a cross-sectoral community to promote and support long-term preservation activities and to raise awareness in society, to trigger synergies between ongoing activities in Germany and to cooperate with international partners and projects, and to develop strategies for the coordination of long-term preservation activities in Germany. The project will end in 2006 with a proposal for a long-term organisational model to continue the service as a network of excellence, along the lines of the DPC in the UK.

Under DDB's leadership nestor's partners are libraries (Bavarian State Library, Lower Saxony State and University Library), a media centre (Computer and Media Service of Humboldt University, Berlin), an archive (Bavarian State Archive – Head Office) and a museum representative (Institute for Museum Studies, Berlin).

The nestor Advisory Board consists of publishers, representatives of science & technology, museums, archives, libraries and universities and as well as members of culture & politics and research institutions/computing centres. Nestor currently provides a website for knowledge dissemination on long-term preservation, built using PADI as an example. Nestor also publishes research results on several topics in digital preservation and coordinates the activities of several national working groups. Two nestor working groups are active at present: on Trusted Repository Certification and on Multimedia Archiving. A working group on Preservation Policies and Selection Criteria is currently being set up. For further information: http://www.digitalpreservation.de.

DDB is also participating in the EUBAM working group. EUBAM (Portal zu EUropäischen Angelegenheiten für Bibliotheken, Archive, Museen und Denkmalpflege, the German digital library forum) which aims to institutionalise the coordination of national efforts with respect to a wide range of activities such as the digitisation of content, metadata, copyright, preservation of and access to

digital cultural heritage at a strategic level. The experts in EUBAM are asked for their personal expertise and do not act on behalf of their institute. EUBAM's activities started in 2002. For further information:
http://www.eubam.de/ (German)

Other institutions
Due to the federal structure of Germany a lot of institutions are working on digital preservation in the different federal states. An overview of all institutions is available on the nestor website under Projects. This overview is to get an impression of the (kind of) institutions which are very active at present:

- on library materials: University Library Göttingen, Bavarian State Library Munich, The *Hochschul Bibliothekszentrum* Cologne (web archiving), *Rheinischen Landesbibliothek* Koblenz, *Landesbibliothek* Stuttgart, *Landesbibliothek Karlsruhe*; *Regionalbibliotheken Bundesland Rheinland Westfalen*, Humbold University Berlin (ReUSE), Max Planck Institutes (ECHO), *Universität der Bundeswehr* (*Institut für Software Technology*) Munich
- on digital art preservation: *Datenbank Virtuelle Kunst*, Berlin; *Institut für Museumskunde*, Berlin
- on archival records: *Staatliche Archive*, German Federal Archives, Archives School of Marburg, Archives at the level of the *Bundesländer*
- funding: *Deutsche Forschungsgemeinschaft*, *Bundesministeriums für Forschung und Bildung*

International activities
At an international level DDB is participates in the ReUse project. ReUSE is a cooperative project of libraries and universities from Austria, Estonia, Germany and Slovenia funded through the European Commission's eContent Programme. Its objective is to set up trusted digital repositories maintained by national and university libraries in order to collect, preserve and render available (in accordance with OAIS principles) digital documents currently used for paper-based publishing only. It focuses on the re-use of digital masters of grey (governmental) literature. Term: 2004–2006. For further information:
http://www2.uibk.ac.at/reuse/service/.

DDB is also involved in several international working groups: D-A-CH, ICABS (IFLA CDNL Alliance of Bibliographic Standards), the Firenze Agenda and the FP 7 Task Force.

- D-A-CH (*Deutschland* – Austria – *SCHweiz*) is a working group with the National Libraries of Austria and Switzerland, which is working on strategies, metadata, formats, management of digital objects, and persistent identifiers.
D-A-CH is cooperating closely with EPICUR on the subject of persistent identifiers in order to build a mutual resolving service for Uniform Resource Names

(URNs). For further information:
http://www.onb.ac.at/about/lza/kooperationen.htm (German).

- In ICABS, DDB is not currently focussing on digital preservation. For further information: http://www.ifla.org/VI/7/icabs.htm.

- The Firenze Agenda is a follow-up of the Italian Presidency of the European Commission in 2003. A group of experts proposed an agenda with a few, focussed objectives in response to the challenges of preserving digital memory. The focus addresses creation, preservation and access issues, including both digitised and born-digital objects. It also respects the interests of museum, libraries and archives, and the differences between media format. The agenda covers a short period (12−18 months) and aims to identify concrete and realistic actions. DDB is participating in this working group. For further information:
http://www.minervaeurope.org/structure/nrg/documents/firenzeagenda031017draft.htm.

- FP 7 Task Force is an initiative that arose from the Netherlands' presidency of the European Commission in 2004. It calls upon the growing importance of an infrastructure for permanent access to scientific knowledge within the framework of the creation of a European Research Area (FP7). In particular, it focuses on promoting cooperation and resource sharing within Europe in the field of long-term preservation and permanent access, fostering research and development in the field of long-term preservation and permanent access by using the instruments for the promotion of science and technology within the European Union, and promoting training and capacity building and the dissemination of knowledge and information in the field of long-term preservation and permanent access. Under the leadership of the *Koninklijke Bibliotheek* the FP 7 Task Force (of which DDB was one of the initiating libraries) will define a research agenda and develop scenarios for a European networked infrastructure for long-term preservation and permanent access. For further information:
http://www.kb.nl/coop/euconference/frame-conclusions.html.

Website Die Deutsche Bibliothek:
http://www.ddb.de/

Japan
National Diet Library – (National Library of Japan) – Tokyo

1. General

Legal deposit legislation
The legal deposit system of Japan mandates that copies of all new publications published in Japan must be sent to the National Diet Library (NDL) in accordance with the National Diet Library Law (Law No. 5, 1948). This covers publications such as books, pamphlets, serial publications, musical scores, maps, phonographic records and so on. The National Diet Library Law was amended in 2000 in order to start a new legal deposit system which includes CD-ROMs and other packaged digital publications. The next policy target is the deposit legislation for online publications. NDL recently investigated .jp-domain websites with respect to the technical and operational challenges that might occur when archiving. This research has been one of the preparations for extending the legal deposit legislation to online digital information. This deposit legislation is expected to be approved in 2005, and to be effective in 2006.

Legal deposit in Japan bears no reference to copyright. The purposes of legal deposit are categorically defined in the National Diet Library Law for governmental publications and private publications: governmental publications should be deposited for official use by NDL and for the international exchange of governmental publications with other governments, while private publications have to be contributed for 'the accumulation and utility of cultural goods'. When NDL was established, legislators argued as to whether legal deposit legislation should be based on the copyright law. However, this idea was not realised. Under the legal deposit system, national or local government is required to deliver a prescribed number of copies of works published to NDL. Private publishers are required to deliver one copy of their new publications and, in return, are entitled to receive a sum of money from NDL corresponding to the normal cost of publication (in principle, half the list price).

NDL was established in 1948. Its primary objective was to support the Diet, the central government of Japan. Today NDL serves not only the Diet but also the administrative and judicial branches of the government and the Japanese people in general. Therefore NDL primarily functions as a parliamentary library. As NDL is also the National Library of Japan, it also fulfils a research function. As the only depository library in Japan, NDL acquires all materials published in Japan and preserves them as national cultural heritage. The legal deposit system applies not only to books and serials but to other types of publications as well.

Digital preservation in NDL

Organisational embedding
Digital preservation activities in NDL are carried out within the framework of the Digital Library Project. This project is based on the Digital Medium-Term Plan 2004, written in February 2004. This plan specifies the two main objectives of NDL digital library services for the next five years: 1. Digital Archives: Building a Digital Repository; Web Archiving; Digital Deposit (E- Journals); Digitisation of Books, etc. 2. Digital Archive Portal functions: Portal site of digital archives throughout Japan.

NDL uses the term 'Digital Archives' to refer to the digital repository system for long-term preservation. In addition to the Digital Library Project, NDL is carrying out a Digital Library Development Project, which involves the whole workflow in NDL. This project focuses on the Electronic Library Information System (ELIS). Between 1998 and 2004, NDL constructed ELIS as its mission-critical system on a project basis. The main objective of ELIS is to integrate the services of the three main locations of NDL, and in particular to develop a system which enables the Tokyo Main Library and the Kansai-kan (see below) to operate as a single unit. To realise these objectives, NDL planned to rationalise procedures and standardise rules by developing an integrated system to replace the previously separate systems for each task. Another goal was to provide digital library services such as a digitised materials database and networked information resources collection. However, ELIS and the Digital Archives system are two completely different systems in both a physical and logistical sense. NDL is currently discussing the future possibilities for integrating the contents of the Digital Library Subsystem of ELIS into the preservation system of the Digital Archives system. For further information: http://www.ndl.go.jp/en/publication/ndl_newsletter/133/333.html.

The holdings of NDL are collected through the legal deposit system, by purchase, exchange and donation. These collections are held separately in the Tokyo Main Library, the Kansai-kan and the International Library of Children's Literature, the three subsidiary libraries of NDL.

NDL consists of the following libraries:
(1) The Main Library comprising the Administrative Department, Research and Legislative Reference Bureau, Acquisitions Department, Bibliography Department, Public Services Department, Reference and Special Collections Department, the Detached Library in the Diet, and the Kansai-kan;
(2) The International Library of Children's Literature;
(3) The Toyo Bunko (Oriental Library);

(4) Twenty-six Branch libraries in the executive and judicial agencies of the government. The Branch Libraries in the government ministries, agencies and the Supreme Court not only provide library services but also acquire government publications for NDL. This unique system of Branch Libraries has no counterpart in other countries.

- Main Library

Within the Main Library, the Acquisitions Department assumes the main responsibility for NDL's preservation work, through its Preservation Division (one of the four Divisions within this department). At the moment the activities of the Preservation Division are limited to paper preservation only. The Administrative Department has a Digital Information Planning Office. The function of this Office is to plan the Digital Library Project and to negotiate with other departments in NDL as well as with the external institutions involved. Together with the Digital Library Division of the Kansai-kan, the Digital Information Planning Office is in charge of the Digital Library Development project. However, the efforts to construct the digital library are shared between all departments of NDL. The Information Systems Division of the Administrative Department bears the main responsibility for system development, except for the development of the Digital Library Subsystem (a part of ELIS). This is the responsibility of the Digital Library Division (see below). See also Section II.2 for the organisational chart.

- The Kansai-kan

The Kansai-kan consists of a Collections Department (in three divisions), a Projects Department (in two divisions), and an Administrative Division. The Digital Library Division, one of the two divisions of the Projects Department is responsible for: digitisation of materials and their supply via the Internet; the development, administration and examination of the digital library system; the collection of online resources; research on and development of digital libraries; the implementation of the Digital Library Project; and the digital preservation activities in NDL.

It is expected that the Digital Archives system will be housed in the same location as the department responsible for the day-to-day operations. The precise location will only be known when a decision has been taken as to which department will be responsible for the system. At present, the Digital Library Division is in charge of developing the systems and the R&D for digital preservation.

Within NDL 6 staff members are involved in digital preservation activities: 3 members of the Digital Information Planning Office and 3 members of the Digital Library Division.

Funding
The digital preservation activities are funded from NDL's daily operational budget. The total budget for 2005 is, approximately 9 hundred million Japanese Yen. However, this might change in 2006.

2. Digital repository

Status
The Kansai-kan, a facility in the Kansai Science City, has been especially developed for the mass storage of library materials and for digital library functions. The focus is on the development, maintenance and operation of the Digital Library Project. One of the objectives of this project is the development, administration and examination of the digital library system. Acquisition and preservation of domestic digital publications will be one of the focal points.

NDL has been developing technical, legal and systematic frameworks to implement the plan and is now preparing a requirement to build the Digital Archive systems. This will contain a comprehensive web-archiving system, a Digital Deposit system and a Digital Archive Portal manner as for conventional paper publications. Up until now, this has been the case for every system, which will be released on the web in 2009. This requirement will also include the preservation system designed for Digital Archives.[78]

Library staff are currently working on the construction of a database of books on microfilm and the information retrieval of reference materials. Tapes of the database are regularly exchanged between the library near the Diet building in Tokyo's Nagatacho and the Kansai library. Efforts are being made to establish a backup system. The backup tapes of ELIS's data are also shared and exchanged between both libraries for the purpose of risk hedging and disaster management. A decision on the backup of the Digital Archives system still has to be made.

Services
NDL was established as a result of the conviction that 'truth makes us free, and with the objective of contributing to international peace and the democratisation of Japan as promised in the Japanese constitution'. The mission of the Digital Archives project is to preserve digital materials as a sort of national heritage and to give users free access to these. However, NDL is also aware that the library might not be able to preserve certain materials due to technical problems, and that copyright restrictions will place limitations on user access.

[78] ELIS: See: http://www.ndl.go.jp/en/aboutus/elib_plan2004.html

Depositing
The majority of books and publications come to NDL through the Legal Deposit Law. Depositing at NDL is done by central government, metropolitan and prefectural governments, city governments (including special districts), town governments and village governments, as well as by publishers of private publications. There are currently no signed agreements with depositors on digital materials, but it is envisioned that NDL will hold certain types of digital information through cooperative arrangements with publishers (including academic societies).

NDL has established criteria and methods for selecting the best editions of non-governmental packaged digital publications for preservation and utilisation purposes. These criteria are described in Article 25 of the National Diet Library Law. If there are different editions of a non-governmental packaged digital publication with an identical content, from the same publishers at about the same time, the best edition will be selected on the basis of the following six criteria: superior durability; in a container; no special facilities or equipment are needed; availability of a manual for usage; widely-used standard of the medium or the equipment for using it; any special function is included. However, special function that is designed for particular purposes is disregarded. The 'content' in these criteria, refers to the essential content recorded in the publication. Non-essential parts such as advertisements on videotape are not included. On the other hand, a difference in title or dubbing, etc. is regarded as a difference in the 'content'. A difference in the platform of the application or game software is regarded as a difference in 'content' also. Each publication with a different 'content' is treated as a different publication to be deposited. An upgraded edition is a different publication and has to be deposited as well. Publications with the same content but released by different publishers are also considered to be different publications and have to be deposited.

It is expected that once the Digital Archives System has become fully operational, the requirements for the producers to deposit material will be treated in the same packaged digital publication, such as CD-ROMs. How these materials will be migrated in the preservation system and how long-term access can be ensured has yet to be decided upon. NDL realizes that ensuring the long-term preservation of a vast amount of information will necessitate an automated workflow wherever possible. NDL is currently investigating how to migrate the packaged digital publication in the preservation system and how to ensure its longevity. At present much of this is ingested manually.

Software and OAIS
NDL's Digital Archive System, which is currently under development, will be a long-term preservation system. It will consist of a large-scale storage facility to

enable long-term preservation, a preservation system to ensure access to bit-stream with implementation of migration and emulation strategies, XML-based metadata schemes for AIP and a subsystem for acquisition, cataloguing and dissemination. Identifiers for long-term data storage and for the preservation of uniformity will be provided, as well as metadata for access or storage.

Web information on the Internet will be collected in the web archives, and the structure of each site as it was transmitted will be retained as much as possible. The information that has been collected will be stored and provided in a format that permits time-series recognition. Information resources provided over networks such as the Internet which cannot be collected by mechanical means (for example the deep web) or which should be treated as separate intellectual property, will be collected, organised, stored and provided individually.

In 2002, NDL started to store information on the Internet on a structural basis. The scope of information gathering is however limited due to the copyrights involved. NDL is currently making an inventory of all collection objects, including digital objects such as digital intellectual materials and digital governmental records, which cannot be collected by robots, but should be collected because they count as intellectual property. This inventory will provide the basis for a selection policy with respect to such materials. NDL would like to collect as many of these materials as possible. However, the construction of such a collecting system will probably encounter many problems, especially if copyright issues are considered.

NDL will use the OAIS Reference Model to construct the Digital Archives system so as to ensure permanent access and to prevent the loss of information. Although DIAS and DSpace appear to be the most standard systems currently in use, NDL will nevertheless consider other similar systems which foreign national libraries are employing (or will employ). Which system will be used as basic system has not been decided yet.

NDL has yet to specify the type of software to be used for the construction of the system. However it feels that it is important to use open source software, or to adopt a de facto standard wherever possible.

Materials

The digital repository system of NDL contains several types of digital information resources: digital publications collected by NDL; digital archives created by digitising printed matter; information resources available on the Internet; and navigation resources to locate digital information provided by external institutions. A distinction can be made between primary information and secondary information. Primary information contains: full-text database

systems of the Minutes of the Diet (provisions of Diet session proceedings through cooperation with the House of Councillors and the House of Representatives), digitised image files and rare books image database from NDL special collections (consisting mostly of colour picture materials of the Edo Period), and digitised rare books (mainly from the Meiji Period).

Secondary information contains: NDL-OPAC, the list of titles of the Japanese Periodicals Index, the report on Braille books & tape-recorded books in Japan, the directory of Japanese Scientific Periodicals (covers serials on science and technology edited and published by organisations operating in Japan), information on all types of catalogues compiled by NDL, edited content such as digital exhibitions, digital reference materials such as subject bibliographies and explanatory notes to reference books and publications by NDL.

The system also contains web pages harvested through the Web Archiving Project (WARP) and more specifically selected information from the web that has been systematically collected since 2002 and which contains digital magazines and homepages of government agencies and cities, towns and villages, including reports on the homepages of international sports and cultural events such as the 2002 FIFA World Cup. Since 2000, a 3-year project of digitising approximately 170,000 volumes of Japanese books published between 1868 and 1911 has been underway. This material will also be stored and safeguarded in the digital repository system. The system also will contain access to databases provided by external institutions and navigation to information resources available on the Internet.

NDL has not yet formulated preferred formats for text, still images, audio and websites to be store in the Digital Archives System. It is expected that every deposited format will need to be preserved. Within the preservation system a distinction will be made between preservation copies and access copies. There will be AIPs for preservation copies and DIPs for access copies.

Metadata and metadata schemes
In 2001 NDL formulated the National Diet Library Metadata Element Set, a Japanese version of the Dublin Core Metadata Element Set. It is expected that once the system is operational, the collection of metadata will be automated wherever possible. However, NDL has not yet decided how to store metadata and materials within the repository. The stored metadata will contain information on descriptive aspects, technical aspects, rights management issues, preservation aspects and administrative aspects. The types of entities on which the Digital Archives system will contain information has yet to be decided.
In the future NDL would like to realise automatic metadata generation. A tool will be needed that can extract and harvest metadata from web-based resources

to create catalogue records, and can detect and report changes in resource content and bibliographic data, so that these records can be maintained. NDL has yet to decide which type of metadata should be considered the most important for digital preservation. METS and MPEG-21 are being considered as the main metadata scheme for the AIP, but a final decision has yet to be taken.

Access
There are several possibilities for accessing the material: it may be limited to the facility where the server is currently located, or to the facilities where the Main Library, the Kansai-kan and the International Library of Children's Literature are located. It may be accessible in the branch libraries, the public libraries and similar institutions, or it may not be limited in any way whatsoever.

For private-sector publications that are out of print or otherwise difficult to obtain commercially, and government publications digitised by consultation with the copyright holders, the extent of access will be subject to negotiation. Access to other private-sector publications will be limited to the extent allowed by agreement with the copyright holder. For the time being, the materials preserved through the WARP project, are accessible to all users through the Internet. There are some specific materials that can only be accessed within the premises, due to agreements that have been made with the depositors. In general NDL would like to permit free access for the digital material preserved by the library. However the issue of copyright has a limiting effect on preservation activities. If material is deposited in agreement with the depositors, there might be limited access for users, based on these agreements.

NDL plans to actively promote the digital library as a new library service in the future. Through consultations with all those concerned, step-by-step progress will be made within practical limits, especially with respect to giving access to the digital objects. For recent paper publications that are used very frequently, the possibilities for digitising journals, and materials on science, technology, etcetera, will be investigated at the stage where the copyrights clearance system can be executed faithfully.

3. Preservation strategies

NDL is currently in the planning and designing phase for digital preservation strategies. Migration and emulation are considered to be inevitable strategies for long-term preservation and long-term access. Migration will be approached in two manners: data migration will be required in view of the lifespan of the digital media. The Digital Archive to be developed, with its massive storage facility, will ensure bit-level preservation and periodic data migration with due consideration to the HDD lifespan. To ensure long-term preservation of the

various contents (for example, HTML, PDF and PPT) file format conversion will also be necessary. NDL is considering an automatic conversion method to interpret technical metadata that are contained in the AIP.

As NDL has many different types of application programs for the library material, it will be necessary to keep these applications working. NDL is considering a rendering system that interprets technical metadata in the AIP, creates an emulation environment and executes application programs. Whether NDL will choose to keep the 'original look and feel' of an object or to keep the content, the information within the object, will be decided after the future preservation strategies have been determined.

4. Current activities

National activities
At present NDL has an internal project, called WARP (Web ARchiving Project). WARP is a web-archiving project aimed at preserving information on the Internet in Japan (including digital magazines). This project started in 2002. For further information: http://warp.ndl.go.jp/pamph-e.pdf

NDL is not yet involved in any specific working group concerning digital preservation on a national level. Since 2002 NDL has been researching the preservation of digital materials. The research performed by the Digital Library Division has resulted in the following:
- Research on packaged digital publications: A number of packaged digital publications held in the Electronic Resource Room were examined for their renderability and performance. The outcome was that almost 70% of the 200 samples of digital resources acquired before 2000 have problems with the latest computer environments. It also transpired that commercial emulator and data conversion software do not have sufficient capability to assure long-term access. Under the emulated environment, only 29 of the 100 samples could be read. The commercial data conversion software was able to convert only 2 of the 100 samples.
- Guidelines: NDL has started to develop draft guidelines for digital preservation, based on the UNESCO Guidelines for the preservation of digital heritage and other resources.
- State of the art investigation: NDL has investigated current trends in digital preservation worldwide, and has translated some documents into Japanese and published these on the Internet.
- Survey on preserving digital information in Japan: NDL has distributed a questionnaire to companies, ministries, government offices and so forth to investigate current practice with respect to the preservation of digital information in Japan. According to the survey results, almost 60 % of the

respondents preserve digital information for five years or more. They recognise the importance of saving digital information, but the challenge lies in the obsolescence of software, operating systems and hardware.
Future activities on digital preservation will be: the building of the Digital Archives for long-term preservation and access; setting up guidelines to establish the long-term strategy for digital preservation in NDL; publishing the final report of the investigation on packaged digital publications, in Japanese and English; and organising a consortium of digital preservation in Japan.

Other institutions
To date NDL's cooperative activities have been mainly limited to cooperation among libraries. However, because the digital library goes beyond the scope of a library, NDL's cooperative activities in digital preservation have started to expand beyond the world of libraries. In Japan, digital preservation is a recent phenomenon. There are few institutions active in this field. The National Institute of Informatics (NII) is currently researching formats of metadata, based on OAI-PMH and on useful software for institutional repositories.

International activities
Over the last few years, pioneers among several foreign national libraries (e.g., the Library of Congress and *Koninklijke Bibliotheek*) were invited to give lectures at NDL and to discuss digital preservation activities. However, NDL is not involved in any regular international projects and working groups or programmes at present.

In 2005 NDL plans to sign an agreement with *Koninklijke Bibliotheek* in the Netherlands to strengthen their cooperation on strategic goals such as widening access to the catalogues and collections, and preserving the national digital and paper heritage. The aim of this agreement is to build a strong working relationship, to share information and experiences and to promote the joint development of solutions for the storage of and long-term access to digital documents. Both libraries will each nominate staff members in the key areas to develop their relationship with counterparts, and they will exchange key documents such as strategic plans and annual reports. NDL and *Koninklijke Bibliotheek* are currently formulating a draft agreement.

Website National Diet Library:
http://www.ndl.go.jp/

The Netherlands
Koninklijke Bibliotheek – (National Library of the Netherlands) – The Hague

1. General

Legal deposit legislation
The Netherlands does not have a legal deposit obligation. Voluntary deposit to the *Koninklijke Bibliotheek* (KB), the National Library of the Netherlands, has been in place since 1974, and is based on a general agreement with the umbrella organisations, representing the Dutch and Flemish book trade.

KB not only receives paper publications for deposit, but since 1996 digital publications as well. Since 2002 KB has had a fully operational digital repository for the long-term preservation and accessing digital publications, which is called the e-Depot. The deposit of e-publications is based on individual archiving agreements with publishers, and, since 2005, on a general agreement with the Dutch Publishers Association. In these agreements, the publishers commit themselves to the delivery of e-publications and KB commits itself to the long-term preservation of these.

Deposit of e-publications is not limited to Dutch publications or publishers or to publications about the Netherlands. Since 2003 several international publishers have signed individual archiving agreements with KB to store their publications for the long term.

Digital preservation in KB

Organisational embedding
KB has four main divisions: Acquisition & Processing, User Services, Expert Services & Collections and Research & Development. Two of these divisions, the Acquisition & Processing Division and the Research & Development Division are involved in digital preservation.

Day-to-day operation of the e-Depot is the responsibility of the e-Depot Department, within the Acquisition & Processing Division. This department currently has a staff of 5 fte. The e-Depot Department also employs trainees on a regular basis.

One of the subdepartments within Research & Development is the Digital Preservation Department. This Department has existed since January 2003 and has grown from 2 digital preservation officers in 2003 to a staff of 6 people (4.6 fte) in 2005. The Digital Preservation Department is responsible for digital

preservation research and projects. The main focus is on the development of digital preservation strategies and the development of additional functionality and new services for KB's digital repository system.

KB has a special Research & Development Division. This Division is responsible for coordinating and executing national and international programmes and projects on digital preservation, digitisation and digital services. One of KB's three directors specifically focuses on digital preservation matters, the Director e-Strategy & Property Management. KB's Research & Development Division reports to this Director.

The IT Department is responsible for the technical maintenance of the e-Depot, with two people involved in the daily activities for the systems operations.
See also Section II.2 for the organisational chart.

Decisions on digital preservation are taken in joint consultations between all departments involved (e-Depot; Digital Preservation and IT). For this a working group meets on a regular basis.

The e-Depot system is located within the KB building, where backup tapes are stored separately. Plans are currently underway for a second backup facility at another location.

Funding
Digital preservation activities are partly funded from the daily operational budget of KB and partly from external grants, which are mainly supplied by the Ministry of Education, Culture and Science (Ministry of OCW). IBM Netherlands who developed the technical heart of the e-Depot, also invested in the financing of the digital preservation system by carrying out extensive R&D during the development phase.

Since 2004 the Ministry of OCW has provided structural funding for research and development on digital preservation. This funding is earmarked for innovation. One of the stipulations for this additional funding is that part of the budget be spent on joint projects between KB and the National Archives. The total grant for digital preservation research was € 200,000 in 2004 and € 700,000 in 2005. KB also tries to obtain additional external funding for new developments (for instance from the European Union).

2. Digital repository

Status
The e-Depot was implemented in 2002 and has been fully operational since 2003. Since 2003 several features of the e-Depot have been under development to improve the preservation functionality, future access possibilities and batch delivery. The development of the preservation functionality is expected to require continuous R&D.

Services
The mission of the e-Depot is to provide long-term storage of and permanent access to the deposited digital publications. KB also wants to serve as a safe place for digital objects from other institutes. At the moment the e-Depot contains mainly e-journals in PDF format (scanned and born-digital documents). In future the e-Depot might also serve as a repository for digitised master images from KB and other Dutch cultural heritage institutions (museums, etc.) and as a repository for the storage of academic output from Dutch university libraries (see below: Current Activities).

The e-Depot provides for secure storage of digital material and is the technical base for preservation treatment, should this be needed. The Preservation Manager is linked to the e-Depot system. It stores technical file format information and tracks technology changes. Development, testing and prototyping of migration and emulation tools (e.g., Universal Virtual Computer) will offer permanent access capabilities within the next few years. Additional services for publishers are also under development, such as the delivery of stored items or assistance with user services, in the event of calamity at the publisher's site.

Depositing
At present both Dutch and international publishers deposit in the e-Depot. They deposit their publications on tape, DVD or through FTP connection. In future it will also be possible to deposit digital publications online. This facility will not only be available to publishers, but also for the research community (DARE project) and Dutch cultural heritage institutions (pilot TIFF archive project).

KB has archiving agreements with individual publishers and the Dutch Publishers Association concerning the deposition of publications. In the summer of 2005 these are: Elsevier, Springer, *Nederlands Tijdschrift voor Geneeskunde*, BioMed Central, Oxford University Press, Blackwell Publishing, Taylor & Francis, International Union for Crystallography, NLR, Sage, and Brill. Negotiations with other publishers are currently taking place. Most of these publishers focus on STM publishing: Science, Technology and Medicine. This has

resulted in the e-Depot being the world's largest digital archive for STM publishing.

Software and OAIS

The technical heart of the e-Depot is the Digital Information Archiving System (DIAS). DIAS was developed for KB by IBM Nederland. After the joint development, IBM shaped DIAS into a commercially available product. The system is a combination of components, including Tivoli Storage Manager, Content Manager, Web Sphere Application Server and DB2. DIAS can be implemented in any organisation and is dedicated to long-term archiving. Together with specific library modules the KB implementation forms the e-Depot system. In 2005 the partners in the German kopal project, in which a digital repository system will be built, chose to implement DIAS as well. Both the Netherlands and Germany will, together with future users of the system, work with IBM to enhance the system and add preservation functionality.

The design of the e-Depot and DIAS is based on the OAIS reference model. Processes, the functional model and the construction of the information packages within DIAS are based on OAIS. A distinction is made between submission, archival, and dissemination information packages (SIPs, AIPs and DIPs). KB considers the OAIS model to be a very useful, though high-level, reference model for the design of a digital archive. OAIS does not provide clear guidelines for the technical design of a digital repository, but it was very helpful for choosing the scope and approach during the development of DIAS and the e-Depot.

For the workflow, especially the separation of storage from other services, the interpretation of OAIS in the NEDLIB project (Networked European Digital LIBraries) has been followed. KB was coordinator of this European project that ran from 1998 to 2000. The NEDLIB partners became actively involved in the design of the OAIS reference model by adding the preservation-planning module. The process model DSEP that was developed within NEDLIB provided a firm basis for the design of KB's digital archive and DIAS.

Materials

The majority of publications in the e-Depot consist of electronic journals, or so-called online journals. Until 2005 these were mainly publications in PDF format. Publishers, like Elsevier, are also digitising their back issues, resulting in the delivery of TIFF images or PDF to the e-Depot. The e-journals in PDF sometimes include annexes in a different file format (e.g., Excel, XML, JPEG, MPEG). CD-ROM publications are stored in the e-Depot according to a different procedure: they are fully installed on a dedicated workstation and are subse-

quently stored in the e-Depot in a package with a full disk image of the installed publication.

KB accepts all current file formats, but is working on a list of recommendations for publishers to improve the usage of file formats in a durable manner. Three new projects focus on the future storage of new types of material, such as academic output, digitised images and web resources.

In 2005 KB will start a project on web archiving. Selected Dutch websites will be harvested and stored in the e-Depot. Just as for the long-term storage of e-journals, substantial efforts will be made to maintain the accessibility of the web pages.

Deposited material is submitted by publishers on tape, DVD/CD-ROM or through FTP connections. In the near future the e-Depot will also have a harvesting capability. The first results are expected within the DARE project (see below under Current Initiatives). However future harvesting will not be limited to academic output only.

Metadata and metadata schemes
Bibliographical, structural and administrative metadata are submitted by the publisher. Technical metadata are partly added by the system and will be manually registered and enhanced in the technical metadata registry, Preservation Manager, and stored according to a structure of Preservation Layer Models and View Paths. Technical metadata will be stored on a format level. Every file has a unique identifier (NBN) that links the object to the format type number, the bibliographical metadata and document history.

KB stores bibliographical and technical metadata in separate databases, to enhance flexibility and to improve durability. The AIP contains some metadata such as a unique number and some structural information as well as the complete bibliographical description. The bibliographical metadata are stored in *KB Titel*, the cataloguing database of KB.
This uses a KB-specific DTD, but is based on Dublin Core and can be translated into other formats.

Depositing publishers provide bibliographical and structural metadata, preferably according to KB specific format in XML. If not already delivered in KB preferred format, a script will transform the submission such that it can be stored in the KB catalogue. The connection with the stored packages is guaranteed through the automatically generated NBN. The submission of e-journals and connected metadata is fully automated. Offline publications like CD-ROMS are described manually.

The technical metadata stored in the Preservation Manager are created manually at KB. These metadata are structured such that they can be used to serve future purposes including the rendering of stored digital objects. As well as storing technical metadata, the Preservation Manager is a tool to analyse the consequences of technical changes and to register alternative ways of viewing different kinds of file formats. Import and export facilities are currently under construction.

Research is currently underway on possible improvements to the data model of the e-Depot. In response to the purchase of DIAS by the German kopal project, a new version of DIAS is under development that offers possibilities for improvements. KB will look at the recently published PREMIS model to learn about the possibility of including additional preservation metadata. The use of METS might also be considered. When preservation strategies like migration and emulation are applied, provisions have to be added for the storage of information on preservation treatment.

Access
KB does not distinguish between access and preservation copies. Access copies are generated on the fly from preservation copies. At the moment public access to stored e-publications is only possible on site, and in accordance with the agreements made with the publishers. From August 2005 onwards, remote access will be available for open access publications. In future, access will depend on the type of material. The publisher will define the manner of access and the restrictions. When the e-Depot stores material from other institutions as well, like academic output or TIFF files, only the institutions involved will have access to their own material.

3. Preservation strategies

KB is dedicated to preserving digital documents in their original format and aims to keep them accessible in a way that includes all original functionality. This is the main reason why KB is working with IBM on the development of the Universal Virtual Computer (UVC). In 2004 a prototype of an UVC for JPEG has been developed. This prototype proves that the approach of such a tool is viable for restoring the original looks of a digital image. More research is needed to develop this strategy for other formats and to test the robustness of the UVC approach as a whole. The prototype for JPEG is available as open source.[79]

In 2005 KB, together with the National Archives of the Netherlands, started a project to develop an emulator for preservation purposes. Emulation is often

[79] UVC: See: http://www.alphaworks.ibm.com/tech/uvc

seen as the primary solution for maintaining the 'look and feel' of digital objects in the future. An increasing number of digital publications are becoming complex; compound objects that cannot be migrated as a whole at once. Emulation has to offer a solution for this new problem as well. Yet despite this clear need for an operational emulator, projects to develop one are scarce.
In 2004, KB carried out a feasibility study for the development of an emulator. The results of this study are very promising: if an emulator is built according to a modular design, a first version could be ready by the end of 2006. The modular design enables incremental development and the reuse of existing developments.

In addition to emulation, KB is also considering migration or format conversion as well. KB realises that future users might want to use versions of software that they are used to, or might be unable to operate documents running under emulation. For relatively simple digital objects like text documents or images, migration can work, although errors may appear. Methods for determining how high-quality conversions can be performed will be investigated.

The R&D developments described above have not yet reached an operational phase. Until then, the e-Depot system provides a basis for future implementations. Electronic publications are currently supplied with bit-level preservation (secure storage, backing up, refreshing, etc.).

4. Current activities

National activities
At a national level KB is currently working on two projects, in close cooperation with other institutions: the DARE project and the pilot project TIFF archive.
- DARE (Digital Academic Repositories). DARE is a cooperative project with the Dutch university libraries and SURF (the 'higher education and research partnership organisation for network services and information and communications technology') to make their academic output jointly accessible. Apart from storage in linked repositories, all digital objects will also be stored in the e-Depot for long-term accessibility. Term: 2003−2006. For further information: http://www.kb.nl/hrd/dd/dd_projecten/projecten_dare-en.html.

- TIFF archive: This pilot project is a joint project with some Dutch cultural heritage institutions to explore the feasibility of storing master images resulting from digitisation projects in the e-Depot. An important factor within this project is not only the development of a pilot system for storage, but also the development of a good prototype business model to expand the service in the future. Term: January 2004−December 2005. For further information:
http://www.kb.nl/hrd/dd/dd_projecten/projecten_tiffarchief-en.html.

Another current national project is the above-mentioned emulation project with the National Archive. The emulation project is funded out of the structural funding for 2005 from the Ministry of OCW.

KB is continuing to conduct its own research on the requirements for long-term access. This includes the work on file format registration with the Preservation Manager and the development of additional preservation functionality for the e-Depot. Starting in the summer of 2005, a study will be conducted on recommendations and risk analysis with respect to the durability of file formats.

Early 2006, KB will take the initiative of organising preliminary discussions between different Dutch institutions to see whether a stronger national cooperation on digital preservation can be established.

Other institutions

In the Netherlands KB is the most active institution involved in digital preservation. Other institutions active in digital preservation include:

- on library materials: several university libraries are involved in the DARE project
- on archival materials: The National Archive has considerable experience in the field, through the Dutch Digital Preservation Test bed project and as one of the leading partners of Erpanet. The National Archive is currently conducting a study on building/acquiring an archiving system for archival materials.
- on audiovisual materials: The Netherlands Institute for Sound & Vision is responsible for video and audio preservation in the Netherlands.[80] They are working on preservation issues on a limited scale, for instance by cooperating in the European project Prestospace: http://www.prestospace.org.
- The National Film Museum focuses on the digital preservation of movies.[81]
- other:
- NIWI (Netherlands Institute for Scientific Information) used to be an organisation involved in preservation research, but has recently been dissolved.[82] Part of NIWI now resides under a new initiative, called DANS (Data Archiving and Networked Services).[83]

[80] Netherlands Institute for Sound & Vision: See: http://www.beeldengeluid.nl

[81] National Film Museum: See: http://www.filmmuseum.nl/website/exec/frontpageread/deddigeghji?id=66696c6d6d757365756d-722-6e6c2e66696c6d6d757365756d2e50616765

[82] NIWI: See: http://www.niwi.knaw.nl/en/

[83] DANS: See: http://www.dans.knaw.nl/nl/

- European Commission of Preservation and Access (ECPA).[84] ECPA is part of the KNAW (the Royal Netherlands Academy of Arts and Sciences). ECPA is also focussing on digital preservation activities at an international level, and is currently coordinating the TAPE project (Training for Audiovisual Preservation in Europe). For further information: http://www.knaw.nl/ecpa/TAPE/.

International activities
On an international level, KB is an active participant in various preservation networks.
- IFLA CDNL Alliance on Bibliographic Standards (ICABS): within the ICABS group, KB is working together with the National Library of Australia in 2004−2005 to promote digital preservation with two surveys and the organisation of a session on digital preservation at IFLA 2005. In 2005−2006 it will undertake more practical developments in the area of preservation planning developments.

- *PLANETS*: together with a group of European national libraries, archives and universities, KB is preparing a project proposal to be submitted to the European commission in September 2005. The British Library is coordinating this proposal under the working title PLANETS (Preservation and Long-term Access NETworked Services). This project will conduct research and development on preservation planning, file format characterisation and emulation or migration services.

- Records of Science: at the EU Conference 'Permanent Access to the Records of Science' (organised by KB and the Dutch government on 1 November, 2004 in The Hague) the participants called for an international task force, with representatives from the cultural heritage sector and from international scientific organisations, to lead a more comprehensive action to bring together the main players at a strategic level. For too long the issue of long-term preservation and permanent access has been separately championed by memory institutions and scientific organisations. The taskforce will join forces to develop an R&D programme that can, in part, be incorporated in the proposals for FP7 (Seventh Framework Programme of the European Commission). KB is coordinating this initiative. For further information: http://www.kb.nl/coop/euconference.

Improving international cooperation
Letters of intent to improve cooperation on several subjects, including digital preservation, have been drawn up between KB and several other libraries worldwide, for example with the British Library. KB is currently also discussing

[84] ECPA: See: http://www.knaw.nl/ecpa/

with the National Diet Library of Japan on intensifying contacts. KB staff also have been involved in PREMIS (International Working group on Preservation Metadata: Implementation Strategies) and the RLG/NARA Taskforce on Trusted Digital Repositories. A pre-existing close working relationship with *Die Deutsche Bibliothek*, and the State and University Library Göttingen has now assumed a more formal nature in the DIAS User Group.

Website Koninklijke Bibliotheek:
http://www.kb.nl

New Zealand
National Library of New Zealand Te Puna Mātauranga o Aotearoa – Wellington

1. General

Legal deposit legislation
Legal deposit legislation in New Zealand is administered by the National Library of New Zealand (NLNZ) through the National Library Act 1965. This Act requires that three copies of every book or publication published in New Zealand are deposited at the National Library of New Zealand.

In May 2003, the 1965 Act was revised with the passing of the National Library of New Zealand (Te Puna Mātauranga o Aotearoa) Act 2003. Legal deposit as applied to printed copies now requires 'a publisher of a public document to give the National Librarian, at the publisher's own expense a specified number of copies (not exceeding 3) …'. The National Library of New Zealand Act 2003 has mandated NLNZ to collect, preserve, and make available not only New Zealand's traditional paper documents, but also material in digital form (including websites and digital journals).

The legal deposit provisions apply to any person, group or organisation that publishes books, magazines, newsletters or any other work, for sale or free of charge, to any section of the public. One copy is made accessible through the general collections of NLNZ, and the other is preserved for use by future generations as part of the Alexander Turnbull Library, New Zealand's premier research library and a part of the National Library of New Zealand.
Digital legal deposit is expected to come into force in late 2005.

Digital preservation in NLNZ

Organisational embedding
NLNZ consists of seven units: three of these units: Electronic Services, Collection Services and the Alexander Turnbull Library (ATL) are involved in digital preservation. Each unit is managed by a director. The directors form the Senior Leadership Team and report to the Chief Executive/National Librarian.

Electronic Services provides national access to library and information resources by providing application and systems environment support to clients, and new ventures in digitisation and digital preservation. Collection Services provides access to the general collections, is responsible for collection development and collection management, and produces NLNZ's bibliographic services.

The ATL focuses on collecting and providing access to its research collections. Policy and Strategic Development supports NLNZ in the development of research and policy advice in its role as a key advisor to the Government on information management and delivery.

The Library's Electronic Services business unit is responsible for the development, establishment and implementation of the National Digital Heritage Archive (NDHA) Programme (see below under funding). There is a core team of approximately eight people, including staff seconded from other areas of the Library and contractors, working through the issues related to the NDHA.

At this stage it looks as though NLNZ will have distributed responsibilities for the range of activities associated with digital objects and their preservation. Once NDHA is implemented, a business unit with ongoing responsibility for its management will probably be set up. This will probably be within Electronic Services. The whole-of-domain or broad crawling of the .nz domain will probably take place within Electronic Services. Selection, acquisition and cataloguing/indexing of published digital material will probably be the responsibility of Collection Services. Appraisal, acquisition and arrangement/description of unpublished digital material will probably be the responsibility of the Alexander Turnbull Library. In Collection Services and in the Alexander Turnbull Library some environmental functions, for example, virus checking will to be undertaken prior to those functions. Selective and event web-harvesting will probably be undertaken within the Alexander Turnbull Library. It is likely that decisions on particular preservation strategies and their impacts (e.g., notions of acceptable loss) will be a shared activity between Electronic Services and the Alexander Turnbull Library. These activities have still to become fully embedded practice. See also Section II.2 for the organisational chart.

Funding
NLNZ is a government department and therefore all of its activities are funded by the New Zealand government. In May 2004, as part of its annual funding bid to government, the National Library received government funding to build a trusted digital repository now known as the National Digital Heritage Archive (NDHA) programme. An important component of this funding is that at the end of the NDHA project, NLNZ will receive an increase in baseline funding to help ensure scalability and sustainability of its digital preservation activities.

2. Digital repository

Status
With the passing of the National Library of New Zealand (Te Puna Mātauranga o Aotearoa) Act 2003, NLNZ took a strategic approach to the management of digital material through the development of a digital library for both born-digital objects and digital objects created through the Library's digitisation programme. The preservation of digital materials therefore became a significant new business requirement within the Library.

Following this, a review of NLNZ digital library activities was undertaken by Seamus Ross, Director of HATII and Erpanet, in 2003. The purpose of the review was to either validate current progress or highlight areas of deficiency, and to provide a base document from which NLNZ could proceed in its efforts to incorporate digital material into its core business, including long-term storage and preservation. The review concluded with a few specific action points for digital preservation: the creation of a digital service team; the establishment of long-term preservation services, including a digital repository; and the establishment of a Digital Library Delivery Service with the responsibility for strategic developments in the areas of selection, acquisition, cataloguing, providing access to and preserving of digital materials.[85]

At the same time NLNZ was developing its first digital strategy. The key issues flagged in the strategy were:
- the growing demand for digital information including digital originals and digital objects created through digitisation, which will provide both increased access to the Library's collections and increased recognition of the Library in the nation's information infrastructure;
- the need for a heightened awareness of the users' requirements in a digital environment and being able to respond effectively and efficiently to these;
- the requirement for open, interoperable standards based on resource sharing on a national and international basis.

The strategy also noted that success is dependent on how NLNZ builds the capability and capacity of its staffing resources to meet the digital challenge and how well NLNZ is able to implement a full range of activities and services related to the provision of digital information. In essence success will be measured by the extent to which NLNZ is able to:

[85] Ross, Seamus. 2003: National Library of New Zealand Te Puna Mātauranga o Aotearoa. Digital Library Development Review. Final Report. See: http://www.natlib.govt.nz/files/ross_report.pdf

- provide enhanced access to digital information for New Zealanders, e.g., online databases, digital journals, and especially New Zealand content;
- collect digital resources, especially those relating to New Zealand and New Zealanders;
- ensure the long-term storage and preservation of New Zealand's online heritage;
- provide enhanced access to the Library's collections through digitisation.[86]

The Library has chosen a broad definition of the 'digital library' to encompass all the services and resources delivered in a digital environment along the lines of the Digital Library Federation definition of digital libraries as 'organisations that provide the resources, including the specialised staff, to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that they are readily and economically available for use by a defined community or set of communities'.

A Digital Strategy Implementation Committee has been established to monitor the implementation of the Digital Strategy and the Library's annual programme of work related to the Digital Strategy.

Recently, a Business Requirements Specification has been developed along with an interim Object Management System for the handling of digital material while the digital repository, the NDHA, is being developed. This includes issues of redundancy in a country where issues of geological stability are key. A Functional Requirements Specification will be developed in the second half of 2005. The current activity is a logical extension of previous work on a range of Digital Library activities including digital preservation (e.g., preservation metadata schemes/data model, persistent identifiers, structural and rights metadata, OpenURL etc).

Services
The mission of the NDHA will be to retain the digital objects that are collected by NLNZ in perpetuity in a structured, expansible and secure environment. The repository will provide long-term storage, preservation and access. It will also provide metadata management and digital object management. The repository is expected to be fully operational by 2008.

When operational, the repository must be fully accepted as a trusted repository for digital objects. It is expected that this will include not only material that reaches NLNZ through legal deposit or from donations of unpublished materi-

[86] Ross, Seamus. 2003.

als, but also material hosted by NLNZ as a third party archive, for example, for the high-resolution objects created by other projects or as a preservation archive for research data sets.

Once the digital repository is fully operational, it will provide services for search and discovery; secure storage; data management; preservation treatments; online, real-time access to service copies (depending on usage restrictions that may apply) and access copies (on a restricted basis) and reporting. There will not be any formal distribution of archival copies on request, since dissemination copies are derivatives of the preservation master.

A number of systems and processes within NLNZ will be dedicated to the long-term preservation of digital material. These will work collectively to ensure the smooth and efficient processing of digital material from acquisition through to integrity checking of the file repository. This is a key component of the approach of NLNZ. In order to ensure the integrity of the object, NLNZ aims to preserve whatever is necessary to adopt a life-cycle approach, so that the object can be tracked from ingest into the preservation archive layer and in ongoing management.

Depositing
NLNZ has a legal mandate to collect and preserve published and unpublished material relating to New Zealand, New Zealanders and the Pacific. Material is acquired by legal deposit, purchase or donation. Unpublished material is acquired mainly by donation or purchase; online published material is mostly harvested and sought directly from producers; offline published material comes in through legal deposit or purchase.

Interim digital deposit systems have been put in place as part of the NDHA programme, which allow publishers to deposit digital materials. This process includes the development of a cover sheet describing the digital publication and the files associated with it, automatic authentication processes and automatic uploading to a 'drop-box' at the Library from where virus checking, cataloguing, other metadata processes and loading into the file repository are undertaken.

The Library will be developing formal permission/contractual processes to acquire born-digital material as part of the implementation of the Electronic Legal Deposit Requirement, resulting from the National Library of New Zealand Act 2003. This is scheduled for late 2005.

For the deposition of online publications, NLNZ is seeking the support of a range of publishing organisations within New Zealand, with a view to forming a Publishers' Reference Group to discuss issues such as setting up submission

agreements and defining the most suitable methods for processing digital objects for deposit at NLNZ.

Software and OAIS
For the purposes of digital preservation and the management of 'distributed' Archival Information Packages (in OAIS terms) the design of the NDHA is unlikely to be delivered in terms of one single solution. The library has core resource discovery and digital application software both of which run on an Oracle backend and are not currently designed for the extra functionality required to support digital preservation. Components of preservation metadata are currently in an SQL Server repository and the current object storage strategy uses standard UNIX file management tools in a SAN environment.

Preservation metadata is extracted programmatically from the objects, using a locally developed stand-alone application. This application provides data for the metadata repository and has been successfully integrated with the Library's Object Management System. The Library hopes to undertake a process shortly, which will align the Harvard University file format characterisation tool JHOVE with the Library's preservation metadata extract tool.

The Library is currently using the National Library of Australia's PANDAS web-harvesting tool for selective web-harvesting. It is hoped that a 'next generation' application for web harvesting will arise out of the International Internet Preservation Consortium activity and that a first iteration of this will be developed in 2006. The increasing use of Content Management Systems (CMS) and Digital Asset Management Systems (DAMS) for storing digital objects that institutions deliver dynamically to the web, will make it more difficult to capture web publications in the not too distant future. This will require significant investment in new technologies and necessitate the establishment of formal deposit arrangements with information creators and providers to help facilitate the collection process.

For the new system, NLNZ assessed the potential use of key emerging digital repository software, including DSpace and Fedora. However, it was uncertain whether these would scale to provide an enterprise class solution to a National Library's requirements to provide long-term storage and preservation 'in perpetuity'. As a result of this assessment NLNZ decided to go out to RFI for a preferred software supplier to develop a commercially viable and sustainable digital preservation and management software application. This process is still underway.

While NLNZ has yet to decide on the mechanics of being a third party host for the preservation of other agencies' digital materials, there is another wider

discussion still to be had as to the nature of digital preservation within New Zealand. This discussion is centred around choosing between developing one centrally shared preservation repository that can be used by other organisations, and a system of distributed preservation repositories in which organisations manage their own repository but benefit from common/shared preservation and/or infrastructure services.

Materials [87]

In developing its digital holdings NLNZ will continue to emphasise its collection of material produced within and related to New Zealand and the Pacific. NLNZ is unlikely to store large volumes of non-New Zealand material, although it does collect overseas material in certain areas, some of which may more commonly arrive in the future in digital form (e.g., on CD-ROMs). NLNZ will collect websites, unpublished digital products (such as the e-mails of key figures in New Zealand life and letters and other electronically stored papers), packaged digital publications, and other types of digital objects as they emerge. The repository will contain four classes of digital objects: packaged objects, products from website harvesting, unpublished digital materials and output of NLNZ's digitisation programmes.

Packaged objects will include CD-ROMs, tapes, solid state devices, and other portable media that house publications ranging from databases, e-books, games, and image collections to software. Objects harvested from the web will be derived from targeted selections and 'automated trawls of the web', and will include a variety of file formats and document structures. Unpublished digital materials will be mainly documents (e.g., drafts of publications, e-mails) of authors, politicians, and other New Zealand icons. All four categories are unlikely to be homogeneous in the types of media they contain. Increasingly they will be composed of complex entities consisting of text, images, moving images, audio, and virtual reality. They will have interactive qualities. Software, both bespoke and off-the-shelf applications, will be integral to their performance. No particular file formats, application types and so forth will be explicitly excluded.

The additional functionality associated with digital versions of analogue objects might even suggest that it is worthwhile acquiring both and deciding at a later date, maybe after reflecting on how the material is actually used, which will be the preservation copy.[88]

[87] Ross, Seamus. 2003.
[88] Ross, Seamus. 2003.

Metadata and metadata schemes
Metadata collection will largely be automated to make the process as consistent, reliable and auditable as possible. As far as possible metadata relevant to preservation is programmatically extracted from objects as XML, and output according to NLNZ preservation metadata schemes developed for the preservation metadata repository. However, it is still likely that some preservation metadata will need to be produced manually, although it has not yet been formalised which will be obtained automatically and which manually. At this stage NLNZ asks contributors to provide only minimal metadata with their contributions as part of the digital legal deposit process.

NLNZ's current thinking allows for preservation, descriptive and other metadata to be stored separately from the digital objects, currently in standard XML-enabled relational databases. NLNZ underlines the importance of having clear definitions as to what preservation metadata is and what it comprises, as opposed to other sorts of metadata and where the demarcation lines are to be drawn. Rights and permission metadata, bibliographic/descriptive metadata, and structural metadata (including METS) for example, are not considered to be preservation metadata. The approach NLNZ schemes use is fairly rigorous in defining what constitutes preservation metadata. The metadata considered necessary for preserving objects, is only a part of the total metadata known about an object.

The repository will store information on logical object level, file level, bit stream level and metadata (mainly to document changes to the metadata record for an object). No information will be stored on collections (except to relate collection objects to each other) and on non-digital source objects (which will be held as part of the descriptive metadata).

Preservation as an activity will rely on all metadata and not that deemed to be unique to preservation metadata. All metadata will have to be available in a unified and meaningful manner for all processes (resource recovery as well as preservation). The whole metadata record of an object is considered important, not just for supporting preservation, but also for providing information for continual and ongoing resource discovery and administrative management.

NLNZ has developed its own preservation metadata scheme, data model and XML scheme definition. This work has been mapped to the OCLC/RLG Preservation Metadata Working Group[89] and the NISO Z39.87 Data Dictionary,

[89] OCLC/RLG Preservation Metadata Working Group : See: Research Libraries Group. 2001. Preservation metadata for digital objects: A review of the state of the art. A

Technical Metadata for Digital Still Images.[90] The Library expects to mainly develop the preservation metadata repository in line with this, although a comparative analysis of NLNZ and the recently released PREMIS models will be undertaken to determine the best approach for NLNZ.

The minimum set of metadata that NLNZ requires for preservation of digital objects is prescribed in NLNZ preservation metadata schemes and data models. These are flexible, within certain boundaries, and can be expanded when metadata becomes more accessible in the future.

NLNZ preservation metadata schemes cater for the existence of an original object, a preservation master (there is only ever one preservation master at any point in time) and previous preservation objects that were formerly masters but have been transformed, for example, through migration from an obsolete format.

As previous preservation objects and preservation masters are discrete objects, they can have their own set of preservation metadata. Metadata records are cumulative and the fullest record is always with the preservation master. Any changes or migrations made to an object are recorded as a process in the preservation metadata. Thus any preservation master will carry with it the record of its own creation and of any processes carried out on it. The relationship between the objects is maintained through a naming structure (_og = original; _pm = preservation master; _pp1/pp2 = previous preservation objects superseded over time, e.g., through migration from an obsolete format).

Access

The maximum preservation effort will be focused on the preservation master. Access to digital originals and preservation masters will therefore be restricted to specific staff for management purposes such as transformations and migrations. Derivatives are derived from the Preservation Master, in the required format, as and when required. This will limit the need to access the Preservation Master. Access copies of the material are not considered to be preservation objects and will receive no preservation treatments. Different access copies may be required for different purposes and in different formats/versions, and will be produced accordingly. Generating access copies on the fly and on-demand access are also being looked at. The way in which objects are accessible (or will be accessible in future) depends on policies concerning rules, place and public,

---

White Paper by the OCLC/RLG Working Group on Preservation Metadata. See: http://www.oclc.org/research/projects/pmwg/presmeta_wp.pdf

[90] NISO Z39.87-2002 : Data Dictionary, Technical Metadata for Digital Still Images. http://www.niso.org/standards/standard_detail.cfm?std_id=731

the type of material and the permission given. All kinds of access might be possible in future, although paid access is not being considered at this stage.

Under the new legal deposit legislation NLNZ may provide a specified number of copies but no more than three of a deposited document for use by the public (within or outside the library) but is not allowed to make the document available on the Internet without the express permission of the publisher/copyright owner. However, if a publication has been made accessible on the Internet without any authentication or commercial considerations NLNZ can make that publication available for access and use for the public via the Internet.

3. Preservation strategies

NLNZ is developing a long-term strategy for preserving its digital material in line with the requirements of its Act to preserve in perpetuity. NLNZ is aware that there are, as yet, no existing practical tools for the long-term management of digital objects, such as tools for emulation. Any given strategy depends upon reliable tools that can be used to implement that strategy. Without tools, the minimum preservation strategy for an object is to hold some form of agreed preservation master in anticipation of the future availability of tools, and/or have an original iteration of a digitally-born object to act as a reference copy for use with the tools.

NLNZ's current position is that the first priority must be given to 'the mechanics of possession' whereby NLNZ must be able to ensure the capture and storage, including some level of curatorial endorsement of the look and feel of an object. By their very nature preservation strategy issues must be secondary to the mechanical process at this stage.

Future strategies of NLNZ will be based upon automated processes carried out on bulk types of objects. It may be possible to restrict submission formats in order to normalise objects, but this cannot cover all of NLNZ's material. Normalising and migration have their limitations in a preservation sense, in that an object may ultimately become so removed from it's original self as to lose all relevant, or desirable look and feel. In order to respond to a dynamic and changing external environment any chosen preservation strategy must in itself be flexible and dynamic, if it is to be an adequate and enduring response. Possible strategies for NLNZ will be: restriction on submission (under consideration, because it is not yet clear what the implications might be for NLNZ and the donors); normalisation (under consideration, but only for text materials); migration and emulation. Migration on demand will only be applied as the implications become clearer. Current strategies are to a certain extent based on what is currently appropriate for implementation while allowing NLNZ to engage with

digital material and their preservation issues in a meaningful way. Only by engaging with digital material and by learning from that engagement can NLNZ assume a position from which it can develop and/or implement more comprehensive strategies and policies that cater for the long-term.

4. Current activities

National activities
NLNZ is currently the only New Zealand agency looking at digital preservation in depth. This has clearly arisen from the change in the legal deposit legislation giving NLNZ the mandate to collect and preserve the nation's digital heritage, and is primarily reflected in the work of the NDHA programme. The recent passing of the Public Records Act means that Archives New Zealand will probably become more active in this area as well.

Other institutions
A range of other institutions are likely to have requirements for archiving their materials without necessarily having the same scope as NLNZ, particularly organisations with specialist digital materials, for example, government departments with large collections of GIS materials such as Land Information New Zealand or those with large data sets such as Statistics New Zealand. Nearly all of these organisations need access to the technical, organisational, and economic expertise to fulfil their respective obligations. It is hoped that successful digital preservation strategies and practices at NLNZ will give the Library the capability to provide leadership for New Zealand in this area and to promote widespread adoption of standards and coherence of practice as these emerge.

Similar requirements are arising with regard to audiovisual materials housed by the National Film Archive, the Sound Archive/Nga Taonga Korero and the New Zealand Television Archive. While these institutions have their own ownership, funding and governance structures, it is clear that the problems they face are the same as those faced by NLNZ. Consequently, NLNZ hopes to enter into discussions with these and other related agencies to try and ensure that responses to the preservation needs of these materials are undertaken as cohesively as possible. Conversely, there is also a lot of interest from these other institutions in how NLNZ deals with and provides answers to the challenge of digital preservation 'in perpetuity'.

International activities
NLNZ is also engaged in a range of international digital preservation activities including:
- representation on the Advisory Board for the recently completed PREMIS work on preservation metadata;

- discussing the viability of integrating the JHOVE file format characterisation tool with NLNZ's preservation metadata extraction tool;
- working as observers with the International Internet Preservation Consortium on the development of a Curator Tool for the capture of websites using both selective and trawling methods.

Website National Library of New Zealand:
http://www.natlib.govt.nz/

Portugal
Biblioteca Nacional (National Library of Portugal) – Lisbon

1. General

Legal deposit legislation
Portugal has had legal deposit legislation since 1805, and this mandates the Bibliotheca Nacional (BN) to function as the legal deposit library. Since 1986, theses and dissertations have also been covered by the legal deposit law and must therefore also be deposited at BN. The current legal deposit law does not cover the deposition of digital objects, but BN is involved in developing models for the voluntary deposition of offline materials (digital theses and digital publications). BN is compiling a strategy to address the problem of depositing digital publications. These efforts are concentrating on three main components: an overall analysis of the structure of the problem; the development of the concept of publication genre; and the development of the concept of deposit by scenarios.

A possible future law on depositing will cover the legal deposition of offline publications and the selective deposition of online publications. This law was discussed several years ago, but because it comprises other issues besides digital deposit, it is not possible to estimate if or when will it be enacted.

Digital preservation in BN

Organisational embedding
BN has five main service departments. The Department of Services for Innovation and Development (*DSID −Direcção de Serviços de Inovação e Desenvolvimento)* is responsible for activities in the Digital Library Initiative (BND − *Biblioteca Nacional Digital*), a project comprising digital preservation.[91] The DSID Department has two formal subdivisions: The Porbase Division (*Divisão da Porbase*) and the Informatics Division (DINF − *Divisão de Informática*). The Informatics Division has three working areas: one on development and services, one on infrastructure, and a helpdesk. This Department also comprises two offices, the Office for Project Management (GAP − *Gabinete de Gestão de Projectos*) and the Office for Digital Publishing (GEDE − *Gabinete de Edições Electrónicas*).

The activities of the Digital Library Initiative span the entire Departments. The cataloguing activities receive support from Porbase, while DINF maintains the

[91] Since the chart of the BN is in Portugese, the Portugese names of the departments are added here

stable services and gives support to the development activities GEDE hosts the actions for digitisation and digital publication, and finally GAP is responsible for the human and financial resources management of the projects. The developments are assured by a team of researchers and engineers from INESC-ID, a research laboratory with which BN has been working for many years.
See also Section II.2 for the organisational chart.

Long-term preservation (*preservação a longo prazo das suas obras*) is one of BN's objectives. Focal points are the development of online deposition and access for voluntarily deposited 'simple' digital materials (monographs, theses and sites), establishing cooperation with publishers and the setting up of formal arrangements on the deposition of digital magazines, and the development of a technical solution for web harvesting. Up until now, however, the activities for digital resources (including preservation) have only been developed within the scope of the Digital Library Initiative, with no relevant roles for other departments. These actions are still considered to be in a 'project phase', with no outside formal services having been delivered yet. Up until now the subject of digital preservation has been addressed within the scope of the storage problem, with the equivalent of one full-time person dedicated to it. All of the infrastructure is currently housed at BN (including backups), but in the end BN expects to have it replicated at other physical locations in the country.

Funding
Digital preservation activities, as part of the Digital Library Initiative activities, have been externally funded since 2001 by FEDER (under the national programmes POC − *Programa Operacional da Cultura* and POSI − *Programa Operacional para a Sociedade da Informação*). Internal funding has been provided by PIDDAC (the programme for investment from the central administration). However, the main funding so far has been for the development of services for acquisition, registry, discovery and access. Digital preservation has been always a known issue, but long-term digital preservation has never been addressed by any specific action. That is expected to occur in the next phase of Digital Library Initiative, starting in 2006.

2. Digital repository

Status
Since 2000 BN has had a fully-operational system for offline and online digital publications. At the moment BN is building an internal digital repository for preservation and access. These plans arise from the activities BN has developed in the National Digital Library Initiative.[92] At present the Digital Library

[92] Portugal National Digital Library Initative: See: http://bnd.bn.pt (Portugese).

Initiative (BND) is still in a development phase. A first version of an operational repository has been available since the end of 2004, but this mainly covers the requirements for operational storage and short-term preservation.

Services
The digital repository of BN has to serve as generic repository for digitised and born-digital publications, which are created in or submitted to BN. Once the development phase has been completed, long-term preservation aspects (long-term preservation and long-term access) will be developed. However, the first version of the system will be adequate for short-term preservation storage.

The first prototype will also be used to disseminate results of actions as examples of good practice for other organisations in Portugal. In future it is expected to be the national digital repository for all digital objects/publications Portugal produces. A persistent identification service will also be added.
Once operational, the system will provide services for search & discovery; online, real-time access to service and archival copies (restricted or unrestricted); secure storage; preservation treatments, formal distribution of archival copies on request (real-time or batch) and reporting.

Depositing
At BN digital material can be deposited by the general public, the research community and other companies or institutions on the basis of voluntary deposition. The digital repository will also contain the digital material originating from the digital library project of BN.

The main actual actions for the digital depositing are developing in two directions: external and internal.
Externally there is an effective service within BND for the digital deposition of theses and dissertations (DiTeD − *Depósito de Teses e Dissertações*). Two other services are planned for the digital depositing of government resources (DIMAC − *Depósito Digital de Publicações da Administração Central*) and for depositing of digital copies of printed works by publishers (DIMIC − *Depósito Digital de Publicações Impressas Comerciais*). These services use a software framework develop by BN, called Deptal, which can be configured for several scenarios. Services using Deptal can interoperate with other services and with 'web services', export the deposited object structured in METS, and produce descriptive metadata in several formats (mainly in UNIMARC and Dublin Core).

Internally BND has digitised nearly one million images of multiple genres of originals. Most of these digitisations were done in high quality, which resulted in nearly 40 Tbyte of data. These images have also been used in several digital

publications created by GEDE. The preservation of these resources is a very important objective.

Software and OAIS
BND is developing an infrastructure for mass storage, where preservation will be a major issue. The actual infrastructure has nearly 50 Tbyte of storage, but cannot yet be termed preservation storage. This infrastructure has been built with normal hardware and open source software. The preservation storage is an extension of this, which will also be assured by open source software under development with the help of local researchers.

Research is underway to develop a final model to manage preservation and access according to the appropriate standards. Both commercial available software and open source software will be used for the development (Lustre, Linux). Commercial software will only be applied for specific purposes if needed. The new digital repository naturally can be considered OAIS compatible. Up until now, there were no special concerns with OAIS. The OAIS model has been particularly useful for making the general public/outside world aware of the problem of digital preservation. However a potential danger of the OAIS system is that it might provide an oversimplified picture of the digital preservation problem.

Materials
The digital repository will contain all digital objects that are deposited. Up until now there has been no limitation to special formats or subject areas; the system therefore contains various materials such as images, sound, websites, theses and dissertations. However more stringent requirements will be applied in the future. At present only the preservation copies are stored in the current system. Preservation and access copies will not be stored in the same repository. There will be two different spaces for that (storage for access, already developed and in use; and preservation for storage, supported by a non-definitive solution, that will be redeveloped after 2006).

Duplication and references to the bibliographical record are kept for all the digital works, including the digitised copies of printed works that are already described in the national union catalogue, Porbase. Other materials can have other descriptive metadata, such as special archiving material in EAD. Dublin Core crossings are also available.

Metadata and metadata schemes
At present most metadata are submitted by the depositors, or retrieved from external systems (bibliographical metadata from the Porbase). Structural and technical metadata are generated automatically or manually. Apart from

supplying bibliographical information there are no specific requirements regarding the delivery of metadata. Information will be stored on collections, logical objects, non-digital source objects, files, bit streams and metadata.

At present all metadata are stored in XML flat files, and reused in databases, active services, and so forth. Therefore, all metadata is stored in the repository along with the contents files. The stored metadata contains information on rights and permissions, provenance (document history), technical and structural aspects, administrative and management aspects, and bibliographic and descriptive aspects. BN considers all types of metadata to be equally important for future preservation and access.

The METS scheme is used as structural metadata scheme for all the copies (access and preservation copies). METS RIghts.xsd is used for rights declarations.

Access
In the current system access-copies are available from HTTP servers. The access policy for digital material varies, depending on restrictions, rights and the type of material. Online paid access once the system has become operational is not envisaged. BN wants to provide access copies whenever needed. Access copies are only submitted under normal security measures (backup or for special purposes). The repository access copies are previously generated from the preservation copies (masters).

3. Preservation strategies

At present two preservation strategies are applied: bit-level preservation (secure storage, backing up, refreshing) and normalisation. Once the system has become operational, normalisation, migration and migration on demand will be applied to the material in the digital repository. This strategic choice is mainly based on common sense. The actual priority is for the physical and logical preservation (the ability to read the medium, identify the files that are part of a resource, and understand its logical structure). Special requirements for intellectual preservation (preserve the 'look and feel', the original interaction behaviour, etc.) have not been considered yet.

4. Current activities

BN is not involved in any straightforward digital preservation projects at present. There are however several actions and activities related to preservation. This is not because digital preservation is seen as the ultimate solution, but

because BN is always concerned with the risks of decisions for long-term preservation, and how much BN can afford to eventually lose, redo, etc. BN feels that at present, not enough information is available to address the problem in a sustainable manner. It is therefore hoped that at some stage in the future, the current efforts focused on physical and logical storage will make a contribution to the problem of digital preservation, thereby defraying the need to answer the problem in detail now.

National activities
On a national level BN is currently active in digital deposition (internally and externally). See Section 2 Depositing.
BN is involved in a proposal for a national initiative (with other public bodies, universities, research laboratories and local private companies) for the creation of 'PREDICA' a Centre of Excellence on Preservation and Digitisation in Advanced Environments. This is an action within the framework of a highly competitive national programme for the generic creation of Centres of Excellence to help universities to cooperate with private companies and public bodies. As a part of the activities, links are proposed to external networks, projects and bodies with similar missions, and in this other National Libraries are important references. This proposal was put forward in May 2005. A funding decision has yet to be made.

Other institutions
No other institutions in Portugal are currently active in the field of digital preservation. Other libraries, museums and archives are developing digital services and spaces, but digital preservation as a specific issue is still not being addressed.

International activities
BN is not currently involved in international digital preservation activities. BN is participating in ICABS, the IFLA CDNL Alliance on Bibliographic Standards. However, within ICABS the focus of BN is on the promotion and development of UNIMARC according to the IFLA-UNIMARC Programme, and not on digital preservation.

Website Biblioteca Nacional:
http://www.bn.pt/

Sweden
Kungliga Biblioteket – National Library of Sweden − Stockholm

1. General

Legal deposit legislation
The first Swedish law on legal deposit, issued in 1661, was a government office regulation. All existing printing presses were ordered to send two examples of each printed document to the *Kungliga Biblioteket* (KB). Over the centuries, the law has undergone a number of revisions. In 1979 it was extended to include so-called combined material, for example, printed documents together with audio cassettes.

KB is the National library of Sweden. It also serves as a research library, primarily in the humanities. The task of KB is to collect, describe, preserve, and provide access to all materials published in Sweden, publications pertaining to Sweden but issued abroad, and a representative collection of foreign literature. This not only concerns publications on paper, but also digital documents. KB thereby fulfils the tasks prescribed by the Legal Deposit Act (1993:1392) and the Ordinance of Legal Deposit (1993:1439). In accordance with this Legal Deposit Act a document is subject to deposit if it is duplicated and published and if the intent is to make it available for the general public or a large private gathering. The document can be produced in Sweden or abroad, but primary dissemination in Sweden must be intended.

The Legal Deposit Act of 1993 (and the 1995 amendment) requires legal deposit of digital documents available in physical format, such as optical disks, to KB. Online digital documents, such as Internet publications, websites and databases, are not yet covered by this legislation. In Sweden there are seven more libraries with a deposit function.

In 1998 a report on securing the digital heritage and e-deposit was published. This report, *E-plikt : Att säkra det elektroniska kulturarvet* (1998:111), written by the Director of the Archive for Recorded Sound and Moving Images, and initiated by the Ministry of Education, contained the results of a review on e-publications. In 2000 amendments to legal deposit provisions were discussed as part of a research and development bill (*Forskning och förnyelse*). This mainly concerned Swedish web pages, but also some online databases. As a result of the *Kulturarw* Project (see below) a special decree by the Swedish government, approved in 2002, authorised KB to collect Swedish websites on the Internet and to allow public access to the archive within the library premises.

A legal deposit legislation for digital publications is expected by the end of 2005 or early 2006. The preservation of audiovisual media (broadcast radio and TV, film, video, records, CDs, multimedia, etc.) has been subject to legal deposit law since 1979. These materials have to be deposited to the National Archives of Recorded Sound and Moving Images.

Digital preservation in KB

Organisational embedding

In the area of digital preservation, KB was one of the first national libraries to start web harvesting on a structural basis. In recent years, a lot of digitisation activities have also been taking place in KB, mainly to protect fragile, unique and frequently used originals from getting damaged by everyday usage during intense library routines. However, KB feels that digitisation does not exempt it of its responsibility to preserve the originals, both on paper and in digitised form.

Digitisation implies the entire process of transferring physical objects from the collections to produce their digital equivalent. This involves the depositing of material in digital image archives and catalogues, efforts to secure the long-term preservation of digital material and permanent access, and the delivery of packaging for users. The focus is now on the digitisation of objects from the collections.

KB has six main departments that report directly to the National Librarian. Four of these departments (Collection Development & Documentation, Special Collections, Preservation & Access Department and the Department for Administration) are involved in digital preservation activities.

The Department of Collection Development & Documentation is responsible for the legal deposit. The Preservation & Access Department is responsible for preservation activities and for the local library system (Regina). The Libris Department is responsible for the national library systems. National coordination and development of the information infrastructure within the research library network is the responsibility of the Bibsam Department.

Coordination of digitisation activities is the responsibility of the Special Collections Department. However, each of the four main departments and the Libris Department have special units participating in digitisation activities. Special Collections also contains a small ALM centre, a joint initiative of KB, the National Museum, the National Heritage Board, the Swedish National Archives and the Council for Cultural Affairs. It was established as a project on 1 May 2004 and over the next few years it will concentrate on collaborative

digitisation projects. Web harvesting is carried out within the framework of the Kulturarw3 Project, which has been running since 1997. The Kulturarw3 Project falls under the responsibility of the division of Information Technology within the Department of Administration.

Digital preservation is considered the responsibility of the library as a whole. However, Collection Development & Documentation and Special Collections have been especially engaged at the strategic and practical levels. Research (currently only in projects with external partners) and development work are mainly carried out within these two departments, with support from the Information Technology division. KB has a large number of servers, which are controlled by the Information Technology division. There are currently 10−12 people involved in digital preservation activities on a regular basis.

Bibsam funds both research and development projects. The SVEP project is particularly relevant in this context (*Samordning av den Svenska Högskolans Elektroniska Publicering*, or Coordination of digital publishing in Swedish universities and colleges. See also Section II.2 for the organisational chart.

Funding
Up till now, digital preservation activities have mainly been funded from KB's own budget. From time to time external funding has been acquired for specific tasks.

2. Digital repository

Status
At the moment the archiving of web files, master files from digitisation and other material is carried out in an HSM system with a tape robot archive and a strong server. There are other systems closer to the staff and users working on their servers, which fetch files from the HSM system. There is also a system for looking at the web archive, a system for the photographic staff, web presentations and so forth. KB strives to establish direct links from records in the national Libris system and in the local library system Regina to the digital resources.

One of the principles of digitisation in KB is that all digitised documents are stored in an image bank and that the designations of all files are in accordance with specific guidelines. This is not a fully-fledged image bank, but a simpler solution. The digitised material will be stored according to a long-term preservation perspective and will be kept permanently available to the users. Conversions and other similar projects will be pursued together with the preservation of

other digital information. The digital objects that are deposited to the library are also stored on a server.

KB is currently discussing cooperation in the long-term archiving of digital information with the National Archives and the Archive for Recorded Sound and Moving images, and is also considering the possibility of forming one or several cooperating national digital repositories in the near future. However it is not expected that such facilities will be realised in 2005. The Technical University of Luleå might be involved in the technical design, and DiVA (the digital publishing unit of Uppsala University library) in designing solutions to technical problems. The current central archive facility of KB is already partly operational, but still lacks important functionality. KB hopes to be able to add some of this functionality during the course of this year. It is still not known when the system will become fully operational.

Although decisions have yet to be been taken, the technical part of the system for long-term preservation and access of digital material is currently under consideration. The KB server might be the basis for this system and function as the central archive. This central archive will have a tool called Resolution Service (developed by DiVA and KB), which determines where the digital document a client requests is stored in one of the many local digital archives (also trusted repositories) from any of the institutions using the system. It is expected that once the central digital archive in KB has been realised, there will still be digital objects in trusted repositories of other heritage institutes which have not yet been linked to the central archive. That is why this Resolution Service Tool, based on the use of URN:NBN, has been developed.

KB, the National Archives, the Archive for Recorded Sound and Moving Images and the Technical University of Luleå are currently working on a proposal to form a centre for research and development into long-term digital archiving. This will be located in Boden (close to Luleå). In Boden, the National Archives is currently undertaking a long-term digital preservation project.[93] This project is concentrating on the digital delivery of government institution records to the National Archives, by designing a new strategy for preserving digital content in archives, based on the OAIS model and SML-related standards. The plans for the joint centre in Boden can be seen as a continuation of this project, but on a much broader scale in terms of approach and participation.

[93] Long term preservation project Swedish National Archives: See: http://ldb.project.ltu.se/~Projekt_LDB (Swedish)

Services
Once the digital repository (or central archive) is operational, it will provide preservation of and access to digital objects. KB is currently formulating the mission of the new digital repository.

Depositing
In 1998 KB signed the first central agreement on depositing digital materials. KB currently has seven agreements, four with commercial or private producers and three with universities or central government bodies. Both monographic materials and periodicals, mostly serials, are covered. One of the commercial producers brings in 55 commercial publishers under one umbrella, and therefore covers a large part of the commercial sector. KB also has an agreement with DiVA that brings in some 10 or 11 other universities or colleges. At least two new agreements with other central government bodies are being negotiated. Up until now, online objects are being deposited on the basis of voluntary agreements.

Producers depositing in the central archive KB is developing, will be subject to requirements concerning the data delivered to the archive (so-called verifiable agreements). There will be a link between the Central Archive of KB and several trusted repositories from other Swedish producers (such as universities). Once the repository has become operational, the workflow will be as automated as possible.

Software and OAIS
KB expects to use mainly Open Source or CC-licensed software to build the digital archive. Some scripts that are in use now for the current system have been developed in-house, and this will probably also be the case in the future. Commercial software will probably have to be used for some services such as database management or search systems.A decision as to which system will form the basis of KB's central archive has yet to be taken. Investigations on this point have been started but a solution has not yet been found. The system will be OAIS compliant.

Materials
Once the system has become operational it will contain digital publications, periodicals, websites, digitised images that result from the digitisation project, and digital objects from other institutes. The currently preferred formats for texts are PDF and XML, for still images TIFF and for websites HTML. This might change in future.

Metadata and Metadata schemes
Up until now, KB has primarily focused on the capturing of digital materials. However, experience has also been gained in metadata standards and digital preservation strategies for digital publications. This mainly concerns the early engagement of KB in web harvesting and in the NBN/PI aspects.

XML will be used as a primary storage format for all metadata. The URN: NBN will be used as primary persistent identifier. For digital publications delivered by agreements, KB will require metadata to be delivered, preferably as separate files that are stored together with the publication in a 'package'. It is expected that most metadata will be acquired using automated processes.

When it becomes fully operational, the repository will contain metadata on technical aspects, descriptive aspects, preservation and rights management. At present the repository records information on logical object, file, bit stream and bibliographic metadata. However, it will be possible to include every type of entity in the future (also on collection, logical object, non-digital source object and others). Metadata is currently delivered by the depositors; some of it is automatically extracted by the depositor's own computer programme. KB is about to carry out an experiment with several government bodies to see whether they can integrate KB's metadata script so that KB can harvest their metadata automatically.

The bibliographic metadata is considered necessary for searching and retrieving files, and for knowing which publications the archive holds. The technical metadata are necessary for rendering the digital publications in the future.
METS and DIDL will be the main metadata schemes used by the system for the metadata elements.

Access
The digitisation concept of KB excludes born-digital resources, which are often part of a digital library alongside digitised material. By applying digitisation KB will meet the general public's right to access its collections as well as presenting the collections on a global scale. Digitisation will be used to contribute to the protection of the collections so that future generations can benefit from them.
A number of criteria must be complied with to enable access to digital collections. Different target groups need to be helped to access and navigate the collections. The venture should be pursued in a long-term perspective and involve future commitments.

At present, KB has different access agreements, depending on the producer. With the commercial producers there is an agreement that KB has the right to give access to the delivered files within the library building, but not through any kind of public network. The publishers are quite happy with this arrangement, and KB hopes to be able to convince the government that it is a satisfactory solution from the copyright owners' point of view. KB does not pay the commercial producers anything for their deliveries. So far, the publishers have found that the rapid inclusion of records for their resources in the national union catalogue, the Libris bibliographic database, is reward enough.

With the non-commercial producers there is an agreement that the link in the bibliographic record shall point to their own server, as long as the resource is kept publicly available. If and when it is removed, the producers must warn KB, and the link will be changed to go into the archive instead. KB hopes that if this occurs even the link into the archive will be allowed to be generally active, that is, not requiring anyone to come to KB. However, this is not clear yet. KB is planning to investigate the legal prerequisites and requirements and whether anything can be achieved using rights metadata to create these possibilities.

If the central archive is to be (as proposed) a joint venture between KB, the National Archives and the Archive for Recorded Sound and Moving Images, it would be logical for access to be permitted within these three institutions, and also at all seven deposit libraries of Sweden. No decisions have been made yet. The access procedure will need both a government and a parliamentary decision.he main target group KB needs to provide access for is the educational system in its widest sense: pupils, students, teachers, researchers and the general public. In accordance with a government ordinance, the national union catalogue has therefore been made freely available to the public on the Internet, as Libris Websearch, since 1997.

KB has collected Swedish web pages for six years. The search engine *Svesök* for Swedish web pages only automatically indexes the web pages most recently harvested by Kulturarw3, but the links point to the live web, not to the archive. KB plans to replace Svesök with a search engine that indexes the entire Kulturarw3 archive. However then the links can only be active within the KB building. To go through the archive, one must know the exact address of the page one wants to look at. It is also possible to browse by following the links and move between time layers. Manual search indexes will not be performed in the Kulturarw3 web collection, but KB hopes to use automated techniques to improve searching and usability.

3. Preservation strategies

The only preservation strategy KB has implemented to date is the decision not to accept all formats for delivered materials. However in harvesting, all formats are of course collected.

KB will continue to require certain formats for delivery (e.g., XML for metadata). Decisions about accepted formats have been guided by what KB considers to be sufficiently standardised and well documented. As for other preservation strategies, KB will look with great interest at the results of the emulation experiments being carried out by the *Koninklijke Bibliotheek* in The Hague, the Netherlands.

In the choice of new strategies, the choice between safeguarding the 'original look and feel' or the content is not that simple. A good example is a newspaper page where the 'original look and feel' is inextricably linked to the information content. It might even be the case that preserving the bit stream may be the best strategy that can be achieved, even if the preference is to be able to render a digital object in its original form. For some kind of publications, such as newspapers, not being able to render the original look would mean loosing essential information.

4. Current activities

Digitisation is an area well suited to coordinated ventures between KB and other libraries as well as other ALM (archives, libraries, museums) institutions with extensive cultural heritage collections. KB is trying to initiate a forum where issues related to digitisation and registering can be developed in cooperation with cultural heritage institutions. The same consideration applies to long-term preservation.

National activities
On a national level KB is participating in or carrying out the following activities:

- Kulturarw3
KB initiated the Kulturarw3 (Cultural Heritage Cubed) project in 1996. This project, which concerns the harvesting of Swedish web resources, is investigating the long-term preservation of published digital resources. There is some 'weeding' at the source, as generally collection is done only twice a year. Over the past two years more frequent (daily) collection has been done from web newspaper sites. The goal is to collect, preserve and make available Swedish

documents from the Internet. There are no partners involved in this activity, which started as a project but has now an ongoing status. For further information: http://www.kb.se/kw3/ENG/Default.aspx

- SVEP (*Samordning av den Svenska Högskolans Elektroniska Publicering*)
The SVEP Project is funded by KB (BIBSAM department). It aims to coordinate academic digital publishing in Sweden. One of the goals is the development and practical implementation of a generalised archiving workflow between a local repository and a national archive, in order to accommodate the variety of publishing platforms and systems currently used by Swedish universities. One of these publishing platforms (DiVA) bases its workflow on the use of URN:NBN as a unique identifier. A basic technical infrastructure will also be developed to support the transmission of dissemination and submission information packages (DIP/SIP in OAIS parlance) between university repositories and the National Library Archive. A primary objective of SVEP is to ensure long-term access to theses and dissertations produced at Swedish universities. Therefore, the project will also attempt to determine the minimum level of preservation metadata required to support archiving and subsequent discovery of these documents. Partners: KB, Lund University Library, Uppsala University Library. Term: September 2003−September 2005. For further information: http://www.svep-projekt.se/english/

- DiVA, Digital Scientific Archive (*Digitala Vetenskapliga Arkivet*)
DiVA is a comprehensive description of a system developed at Uppsala University Library, a repository where the documents are archived is one part of the system. This is why the name Digital Scientific Archive was chosen. The DiVA system has been developed within the framework of the DiVA project which started in September 2000. The system has been fully operational since January 2003. A common portal for the participating universities of the project has also been developed within the same project. This is sometimes referred to as the DiVA Portal. KB cooperates with DiVA to provide long-term preservation and access services. DiVA started as a project, but is now an ongoing activity. Some necessary functionalities are being developed with the financial support of KB within the SVEP project.
For further information: http://www.diva-portal.se/

KB is also the coordinator of the coordinated programme of activities for a Swedish digital library for higher education and research. Together with other libraries and other ALM institutions with extensive cultural heritage collections, KB is striving to initiate a forum where issues related to digitisation and registering can be developed in cooperation with cultural heritage institutions, so that the digitised material is cared for according to a long- term preservation perspective and made permanently available to the users. Conversions and other

projects pertaining to the above shall be pursued together with the preservation of other digital information. The main focus is more on access than preservation. KB has made efforts to obtain special government funding for this programme, but these have not been fruitful to date. Nevertheless, the coordination is still taking place. For further information:
http://www.kb.se/BIBSAM/utredn/natbibl/assignmentdescription.pdf

National initiatives for action
In October 2001, the Swedish Government instructed KB, the National Council for Cultural Affairs, and the National Archives to describe the steps taken with respect to cooperation in the area of ALM, as well as within and between the different sectors. The Government also asked for an evaluation of earlier achievements and lessons learnt, at both a national and international level. This initiative from the Government had its roots in a request from the Swedish Parliament's Committee on Culture Affairs for an analysis of the conditions for a higher degree of cooperation between the ALM institutions. A report on the ALM task was given to the Ministry of Culture in March 2002. In this report the cultural authorities mentioned a national plan for digitisation, a national plan for collecting digital material and a national plan for long-term preservation of digital material, limited to the context of digitisation.

In 2001, on behalf of the Ministry of Culture, an inquiry was also carried out on archival matters, especially on the long-term preservation of digital material. This resulted in a report which stressed the urgent need for actions to preserve digital information and proposed measures for this. It was also emphasised that the question is not only of fundamental significance for preserving today's cultural heritage but also for the introduction of e-government in the Swedish public administration. (http://www.regeringen.se/sb/d/108/a/1493 in Swedish). In 2004 a small coordination secretariat for the ALM sector was placed within KB. This office is financed by KB, the National Archive, the National Museum, the National Heritage Board and the National Council for Cultural Affairs.

During 2002 a government review of KB and its activities took place, the findings of which were presented in a report. (http://www.regeringen.se/sb/d/108/a/669 in Swedish). This report also made proposals concerning the legal deposit of digital publications. The Ministry of Education has stated that there will be a legal deposit legislation for digital publications, but this is still pending. Recently KB has had informal contacts with the Ministry on this issue. KB will work together with the Archive for Recorded Sound and Moving Images to support the Ministry with advice. Based on the outcomes of this report, KB forwarded proposals for a national digitisation programme and for the Net Library in the autumn of 2004.

At about the same time a government review of the Archive for Recorded Sound and Moving Images was taking place. This also resulted in a government report which contains several proposals for the digitisation of existing materials and the collection of future materials. (http://www.regeringen.se/sb/d/108/a/23549 in Swedish). The Archive for Recorded Sound and Moving Images has received special government funding to save major parts of its collections by means of digitisation.

Other institutions

There are no formally appointed competence centres on digitising and digital preservation in Sweden and no criteria for the selection of such centres. All institutions working with digitisation have gained a certain amount of special competences in digitising their own material, for example, KB: digitisation of printed material and preservation of web content; the National Archives: digitisation of archival material, long-term preservation of archival material in digital form, and databases on archival content; The Swedish Land Survey Authority: digitisation and presentation of maps on the web.

The following institutions in Sweden are also active in the field of digital preservation:

- Libraries: University Libraries of Uppsala and Lund
- Archives: National Archives
- Museums: Swedish National Collections of Music, National Museum of Cultural History (*Fotosekretariat*): photo preservation
- Audiovisual institutes: National Archives of Recorded Sound and Moving Images, Swedish Film Institute
- Others: Technical University of Luleå

International activities

KB and the Archive of Sound and Moving images are currently participating in a project proposal, coordinated by the National Archives, for the fifth call of the Sixth Framework Programme of the European Commission. Under the working title 'PROTEAN' the Swedish cultural heritage institutions, together with international partners from the cultural sector, the IT sector and commercial players in the field of digital storage, are preparing a project proposal that will be submitted to the European Commission in September 2005. The aim of the PROTEAN project (Preservation Over Time by Electronic Archiving and Networking) is to develop and demonstrate strategic models and methods for ensuring the continuous existence and accessibility of digital information over time, focusing on authenticity, reliability and IPR, and based on the OAIS model.

KB is participating in IIPC (International Internet Preservation Consortium). One of the major goals of IIPC is the development of tools for web archiving. At present IIPC is working on finalising an IIPC toolkit for acquisition, selection and storage, which will incorporate the IIPC standards (Arc 3.0, Metadata and API). The toolkit will be available in June 2006 as open source. Partners in IIPC are: the Library of Congress, the British Library, the National libraries of Australia, Canada, Denmark, Finland, Iceland, Italy, Norway, Sweden and the Internet Archive. For further information: http://netpreserve.org/about/index.php

Website Kungliga Biblioteket:
http://www.kb.se/

Switzerland
Schweizerische Landesbibliothek – (National Library of Switzerland) − Bern

1. General

Legal deposit legislation
Switzerland has no federal legal deposit legislation for any type of published material. Some cantons however, have separate legal deposit legislation, which generally only cover printed material. Since early 1900 the *Schweizerische Landesbibliothek* (SLB) has had a written agreement with the government to collect, describe, preserve and give access to so-called 'Helvetica': the entire literary production of Switzerland, all works published abroad with a bearing on Switzerland, its population, its past, present and future and works (including translations of them) by Swiss authors. The term Helvetica also includes all publications of associations and official bodies, music scores, daily newspapers, dissertations and telephone and address directories. This assignment has been made within the framework of a National Library law. In 1992 the governmental assignment was modernised to include not only printed publications, but also other information stored on information carriers, in pictorial form or as sound documents.

SLB has always conserved manuscripts of authors. In 1990 a special Swiss Literary Archive was founded, which forms a part of SLB. In 1995, when SLB celebrated its centenary, the Swiss ISSN centre was opened. In the same year *Memoriav*, an association for the preservation of the audiovisual heritage of Switzerland, was founded. SLB was a cofounder of this initiative.

Due to the lack of a national copyright act, SLB signed an agreement with the Swiss Booksellers and Publishers Association. Based on this agreement publishers deposit one copy of each of their publications with SLB, free of charge. In 1961 this agreement was renewed. There is no specific agreement for digital Helvetica at present.

Digital preservation in SLB

Organisational embedding
SLB functions under the responsibility of the Federal Office of Culture of the Federal Department of Home Affairs. There are three main sections, which report directly to the Head of the Swiss National Library. The Collections Section is responsible for acquisitions, cataloguing, conservation and storage. The Public Relations Section is responsible for public relations, research and

online services and lending. The Staff Section is responsible for, amongst other things, the computer systems and international relations.

Since 2001 all digital preservation activities in SLB have been concentrated in the e-Helvetica project. The project organisation of the e-Helvetica team is part of the Collections Section (Sektion Sammlungen).[94]
The e-Helvetica project is carried out in two subprojects: one on organisation and one on archiving.

The team consists of five project members: a project leader (also head of the Collections Section), a coordinator Organisation, a coordinator Archiving and two project members Archiving. The project is carried out under the auspices of a steering committee consisting of the director general and an international relations officer. See also Section II.2 for the organisational chart.

The Organisation subproject focuses on: acquiring and compiling the online and offline-Helvetica; drafting and testing new business models and workflows; copying offline publications for long-term archiving; defining collection guidelines for digital Helvetica and drafting guidelines for providers of digital publications. Another part of the Organisation subproject is to set up contacts with potential partner institutions and publishers as well as other producers of digital publications.

The goal for the Archiving subproject is to select and commission a final archival system. This includes compiling all library-related, technical and financial conditions and requirements, as well as establishing contacts with potential partner institutions. The Archiving subproject is the core of e-Helvetica and provides the technical basis for current work with test servers. The Archiving and Organisation subprojects cooperate closely, support each other and supplement each other's subproject mission.

SLB feels that long-term archiving can only be carried out in collaboration with other institutions. Therefore the cultivation and extension of partnerships at home and abroad are an important aspect of the project work.

The current storage system was jointly purchased by SLB and the Swiss National Archives. Since several institutions in Switzerland, for example, the National Archives, university libraries and cantonal libraries need solutions for long-term preservation, the choice was made for a coordinated approach, to save

[94] Since the Chart of SLB is in German, the Swiss name of the departments in charge is added here.

both money and manpower. SLB will therefore not develop an in-house archival system, but is working in close collaboration with the Swiss National Archive.

Both SLB and the Swiss National Archives fall under the responsibility of the Federal Department of Home Affairs. There is a central facility for IT services. The e-Helvetica project however, is a strategic project of SLB in which no other organisations are involved. The Swiss National Archive also has defined separate projects for long-term preservation.

The development of the new central digital repository system involves close cooperation between the two institutes. For each OAIS element, it is considered whether this can be developed and used by both institutes at the same time, or whether it is necessary to build separate systems. The storage part for example will be developed as a mutual system.

The future embedding of digital preservation activities in SLB has not yet been defined, but it will probably be a separate service. A clear distinction will probably not be made between the day-to-day processing and the research.

Funding
The e-Helvetica project is funded from a credit that is part of the annual budget of SLB.

2. Digital repository

Status
In 2001 SLB launched the e-Helvetica project, to develop a framework for the collection, cataloguing, retrieval and long-term preservation of digital *Helvetica* and to develop an archiving system for digital publications, together with several national partners. The duration of the project is estimated at eight years (2001−2008). In 2007 the ingest part of the repository will be operational, and probably the access part as well. The collecting, cataloguing, mediating and archiving of online and offline digital Helvetica should be an integral element of the day-to-day operations of SLB by then. The system will be fully operational by 2009.

The future digital repository system will be housed in the IT Service Centre of the Federal Department of Home Affairs in Berne. A second storage system that serves as a backup facility will be placed in Neuchâtel, where is will be situated at the Swiss Federal Statistical Office.

Together with the Swiss National Archives, SLB conducted a common WTO tender in 2003 for the storage subsector of a future archiving system. At the end of 2003 the acquisition of a joint mass-storage system was started. As a result of the WTO tender for the storage part of the system in 2003, two tape robots were purchased, with a Hierarchical Storage Management System (HSM). The tape robots are working with LTO-2 and AIT-3 tapes The storage infrastructure will serve as a joint technical basis for the long-term archiving of digital documents from both the National Archives and SLB. The system will be operational in November 2005.

With the exception of offline publications such as floppy disks, CD-ROMs and DVDs, there is no provision for the comprehensive collection of digital publications yet. The aim is to gain some initial experience by compiling well-organised sample collections instead, and to prepare the ground for a comprehensive collection procedure, which must be backed up by a high degree of automation.

In July 2005 a WTO tender for the ingest part was started. Outcomes of this tender are expected in October 2005. Then a pilot will start in January 2006. The ingest part is expected to be fully operational in August 2006.

Services
The digital archive that will be established within the e-Helvetica project will guarantee preservation of and access to the digital publications for future users. The mission of the repository is to store digital data, to preserve these and to make these accessible.

Once operational, the system will provide search and discovery services; online real-time access to service copies, secure storage and data management of digital materials, preservation treatments and formal distribution of archival copies on request and reporting. There will not be direct online access to archival copies.

Depositing
At the moment only SLB and the National Archives can deposit materials in the digital repository. Once operational, material will obtained by both harvesting and submission. In a test run during the last months of 2002, the automatic collection of websites from the .ch domain was tested with the help of a harvester.

No formal agreements have been signed with depositors yet, but some pilot projects have been carried out, which also touch on this subject. In 2004 SLB started with a pilot project called POP, Acquisition and Archiving of publishers' online publications. The aim of this is to set up an acquisition and processing

procedure for online publications of publishers in order to permit long-term archiving. This project is carried out with two Swiss publishers, Karger and Stämpfli.

A web form has been developed for the delivery of university publications (including metadata). At present delivery according to this method has been used by the university Libraries of Lugan, St. Gallen and Berne. Further enquiries are being made into an automated delivery of dissertations. An agreement with the members of the Conference of University Libraries of Switzerland (KUB) which aims to transfer on the basis of OAI-PMH was established in 2003. In order to guarantee the legitimacy of the delivery of digital dissertations for archiving purposes to SBL, the university libraries sign agreements with Ph.D. students or add a suitable clause to the order of conferral of a doctorate.

Software and OAIS
All IT projects relating to long-term archiving are being handled in the archiving subproject, in accordance with Federal Administration standards. The IT projects reproduce most processes in accordance with OAIS. The OAIS system will serve as a basis for the planning and building of the archiving system. In recent years recommendations from standards institutions (UNESCO, IFLA, CENL, etc.) have been analysed and implemented.

Procedures for various types of digital publications have been defined. The individual work stages, from the receipt of the publications to their archiving on the currently provisional long-term archiving storage system, are established in a test system.

Several parts of the archiving system are being developed in collaboration with the Federal Archive. Under the ASTOR project, the first things to be provided were the project archive storage facilities for the storage of digital publications (see also under Current Activities). Another project known as ADAM has also been planned in conjunction with the Federal Archives and will include data management (see also under Current Activities).

Digital online publications will be given a Uniform Resource Name (URN). This means that SLB can take advantage of the preliminary work and the technical infrastructure of the National Library in Germany, which is also working with URNs. For offline publications the persistent identifier will be allocated according to exactly the same criteria as for URNs.

Materials
The digital repository will contain digital publications about Switzerland in all forms (both offline and online publications). It will also contain a compilation of

specific websites concerning Switzerland and various digitised materials including pictorial material and sound documents, which are deposited to and digitised by SLB or the National Archives. All digital objects will be copied from the original carries to the archival storage. Non-digital versions of digital material keep their own management stream.

Since 2003 SLB has also has been collecting and cataloguing online documents from the University of Lugano and online theses from the University of St. Gallen and Berne. The digital copies are currently archived on a test server and will later be transferred to an SLB and SFA joint storage facility. The objective is to integrate the online theses of all Swiss universities in SLB collections.
Preferred formats will be PDF, GIF, JPG, PS/EPS/AI, RTF, TXT/ASCII, TIFF, XML. However as SLB has to take the publications in the published format, in reality all formats will have to be accepted.

Metadata and metadata schemes
Together with the National libraries of Germany and Austria, SLB agreed on a common metadata structure for online theses, in 2004. The members of KUB have undertaken to use the fields defined as obligatory in this metadata structure. The decision was taken to use the SLB library catalogue (Helviticat) as the only catalogue and to use it for the registration of both printed and digital publications. Bibliographic metadata are therefore stored and updated in Helveticat.

SLB has developed its own internal metadata structure. The MARC XML structure is filled into a METS container for bibliographic information. For the other metadata (technical, administrative, managerial aspects, rights and so on) the Preservation Metadata Structure from the National Library of New Zealand is used. This internal structure will be used for all types of digital material. It is expected that most of the metadata will be collected in automated processes.

Once operational, the repository will contain metadata on rights and permissions, technical, administrative and management aspects, and bibliographic and descriptive aspects. SLB intends to maintain a document history that starts at the time the document is delivered to SLB. All future migration processes will also be recorded. Structural metadata will be recorded for journals, as well as the detailed file structures of complex objects such as websites or offline publications on CD-ROM or floppy disks.

It is expected that the repository will record information on those types of entities that are similar to the entities used with paper publications. The librarians will define the logical units that have to be stored (e.g., a monograph or an issue of a journal). As the library system will offer the metadata for all kinds of publications, the information units for printed and digital publications have to be

similar. In future the producer of the digital publications will be expected to deliver a defined set of metadata. This metadata will be completed with automatically extracted information. In the worst case a librarian will also be involved in completing the bibliographic metadata.

SLB will use its own metadata scheme, which includes elements of METS, MARC XML and of the National Library of New Zealand Scheme.

Access
Negotiations with the producers concerning access to digital publications from the archive portfolios of SLB are still pending. Questions concerning copyright still need to be defined.

A special point of attention is preventing the producers' commercial interests from being adversely affected by SLB, so that publications normally paid for can be copied from SLB.

Direct access to the digital publications stored at SLB is not yet possible as the project is still very much in the early stages and the appropriate module does not yet exist. A query in Helveticat, the online catalogue, offers users an external link to the server of the relevant publisher or other provider. Current and future user groups of digital publications offered by SLB also still need to be identified.

3. Preservation strategies

No preservation strategies have been carried out yet, although there has been a pilot project on media migration, in which data from floppy disks were copied to hard disk for archiving in a file system. This project was been carried out in 2002−2004, in close cooperation with the German National Library.
In future migration and emulation will be implemented as preservation strategies.

When choosing future strategies SLB will pay attention to both preserving of and giving access to the content and to preserving of and giving access to the original look and feel. The content will receive the first priority, but a document often gains extra meaning from the presentation of the original look and feel. It is often difficult to split the content from the look and feel. It would be wrong to care for the content only, since this can always mean a loss of value. The accent can also depend on the type of material.

4. Current activities

National activities
The main project on digital preservation in SLB is e-Helvetica. This will be completed by the end of 2008. Partners are: the Swiss National Archives, the Commission for Swiss University Libraries (KUB), the Swiss cantonal libraries, Swiss publishers, the Swiss booksellers and publishers' association and the National libraries of Germany and Austria.

Within the e-Helvetica project four pilot projects concerning the building up of a digital collection are currently underway:

- POP (Pilot project on the 'Acquisition and archiving of Online commercial publishers' Publications'). The purpose of POP is to set up an acquisition and processing procedure for publishers' online publications so as to permit the long-term preservation. Partners: Karger, Stämpfli. There is no web link for the project at present.

- 'Online-Helvetica collection' is a pilot project to cooperate with the Cantonal libraries on building up a common collection of Swiss grey online literature. The purpose of this pilot project is to set up a realisable cooperative model for building up a common collection of online Helvetica. The aim is to share work when selecting, collecting, cataloguing and archiving relevant regional and cultural online Helvetica. A representative collection from the different cantons of Switzerland will be made available to current and future generations. Partners: Cantonal libraries of Switzerland. There is no web link for the project at present.

- 'Long-term archiving of the cantonal libraries' diskettes' is a pilot project to cooperate with the Cantonal libraries on the long-term archiving of floppy disks. Partners: Cantonal libraries of Switzerland. There is no web link for the project at present.

- 'Online-Theses' is a pilot project to cooperate with the Swiss university libraries on data management in order to fix the metadata for OAI-PMH harvesting on the university servers. Partners: Swiss university libraries. There is no web link for the project at present.
For further information: http://www.e-helvetica.admin.ch/

Other digital preservation projects in which SLB participates are:

- ASTOR: (Archival STOR). ASTOR aims to provide archive storage facilities for digital publications. Partner: the National Federal Archive. Term: 2003−2005. There is no web link for the project at present.

- ADAM (*Archivierung elektronischer digitaler Daten und Akten* (archiving digital data and files) – Data management). ADAM focuses on setting out guidelines for data management and implementing a data management system as required by the OAIS model). Partner: The National Federal Archive. This project has been planned, but has yet to be started. There is no web link for the project at present.

Other institutions
Memoriav is the association for the preservation of Swiss audiovisual cultural heritage (*Verein zur Erhaltung des audiovisuellen Kulturgutes der Schweiz*). This association was founded at the end of 1995, with the purpose of advancing the collection, preservation and dissemination of audiovisual materials within a framework of networked institutions. Participants within this association are both publishing and archival institutions.

The objective is to share information and services to improve and guarantee the access of audiovisual materials for science, schools and the general public. Pilot projects on new preservation technologies and strategies will be carried out within the *Memoriav* framework. SLB is one of the founding participants of this association. For further information: http://www.memoriav.ch/

The following institutions in Switzerland are also active in the field of digital preservation: on library materials: Canton libraries, university libraries; on archival materials: National Federal Archives, Cantonal Archives. (The Literature Archive is part of SLB); on audiovisual materials: Swiss National Sound Archives,[95] the Swiss Film Archive[96] and the Association for the preservation of the audiovisual heritage of Switzerland. (The National ISSN Centre is part of SLB). SLB knows of no ongoing activities in Switzerland for the digital preservation of art.

International activities
On an international level, SLB is participating in D-A-CH (*Deutschland*-Austria-*Schweiz*): an international working group that is exchanging information and experiences, and is working on strategies, metadata, formats, management

[95] Swiss National Sound Archive: See: http://www.fonoteca.ch
[96] Swiss Film archive: See: http://www.cinematheque.ch

of digital objects, persistent identifiers and workflows. Partners in D-A-CH are the national libraries of Germany, Austria and Switzerland. For further information: http://www.onb.ac.at/about/lza/index.htm (German)

Website Schweizerische Landesbibliothek:
http://www.snl.ch/

The United Kingdom
The British Library − National Library of the United Kingdom − London/Boston Spa

1. General

Legal deposit legislation
The United Kingdom has a legal deposit law which gives the British Library (BL) the right to receive one copy of every publication distributed in the United Kingdom or the Republic of Ireland (Legal Deposit Libraries Act 2003 and Irish Copyright Act 1963, which is currently being replaced by similar provisions in the Copyright and Related Rights Bill 1999).

The legal deposit regulations of the Copyright Acts 1911 and 1963, with slightly different conditions, also apply to the other five legal deposit libraries: the Bodleian Library in Oxford, the University Library of Cambridge, the National Libraries of Scotland and Wales and the Library of Trinity College in Dublin. The main difference for these five libraries is that they are entitled to claim (and, having claimed, to receive) one copy of any publication covered by the law, whereas BL is entitled to receive without having first claimed.

There is no active legal deposit legislation for digital documents in the UK yet. As an interim arrangement, the deposition of offline or hand-held digital publications is performed under the terms of the Voluntary Code of Practice of 2000. The Legal Deposit Libraries Act 2003 (which should begin to come into effect in 2006) will extend the present legislation to cover digital or e-publications and other non-print materials. It ensures that these publications can be saved as part of the published archive, and become an important resource for future generations of researchers and scholars.

The network of legal deposit libraries in the UK and the UK publishing community cooperate closely through the Joint Committee on Legal Deposit (JCLD). The JCLD will continue, in parallel with the Legal Deposit Advisory Panel, to develop the Regulations needed to effect the 2003 Act in daily practice (http://www.bl.uk/about/policies/legaldepositmembers.html).

Digital preservation in BL

Organisational embedding
Within the organisational structure of BL there are six Directorates that come directly under the Chief Executive. At the moment, two of these Directorates are

involved in digital preservation: The Scholarship & Collections Directorate (SC) and the E-strategy and Information Systems Directorate (eIS).

Since Collection Care has been responsible for preservation of the traditional material from the outset, and also has the expertise and broad perspective on preservation in general, BL has chosen just to broaden the scope of Collection Care with digital material, instead of setting up a new section for digital materials. Collection Care also played an important role in the initiating phase of digital preservation activities in BL. This was marked for instance with the appointment of a special digital preservation coordinator within Collection Care at a very early stage.

eIS is responsible for the development and implementation of e- and IT-strategies for digital media; services, projects and programmes; managing the delivery of the e-strategy and the IT change programme and leading the development of e-business models and tools. eIS has two main subdivisions (Architecture and Development, and Operations), and three distinct Programmes. One of these is the Digital Object Management (DOM) Programme. Within the DOM Programme, eIS, Collection Care and other sections within BL are working together to develop technical solutions to all the problems brought by digitised materials.

The DOM team has eleven full-time staff (software developers and programme management), but many other staff work full- or part-time on the Programme as required: such staff include system architects, business analysts, and technical operations specialists. The DOM Programme is managed by a Programme Board made up of senior-level staff from other Directorates.

Once the DOM system is operational, most of the services developed will be the responsibility of the public services staff, for instance in Scholarship & Collections (acquisition, metadata creation) and in Operations & Services (reading room services). Where the DOM team is setting up the infrastructure for digital materials, decisions on what to store, what metadata to create, or how to link different items and collections are not taken by the DOM team, but by the several library sections involved.

Collection care has at the moment two staff members working on digital preservation. One of them is the JISC-funded project manager of the LIFE project on digital life cycling (see also under Current Activities).
To make the combination of technical digital preservation knowledge and the traditional knowledge on preservation more profitable, BL will form a cross-directorate digital preservation team, which represents both Collection Care and E-Strategy. The team of three full time staff members will exist of two Digital

Preservation Managers (one for the Collection Care role and one for the Architecture and Development role) and a Technical Architect, and is expected to start with their activities in September 2005.
See also Section II.2 for the organisational chart.

Funding
BL is involved, with a number of UK academic partners, in several digital preservation projects, attracting over £750,000 of external funding from JISC. A further £3 million has been provided by the Higher Education Funding Council for England to fund Content Creation projects for sound and for newspaper archives. About £1.5 million has been allocated to the DOM Programme from the British Library operational budget for 2005/2006. For specific activities concerned with making the digital repository operational and with future developments, BL might apply for alternative external funding.

2. Digital repository

Status
In 2003 BL started the Digital Object Management Programme (DOM Programme) to develop the technical solution for the long-term preservation and access of digital items ('objects') as part of the Library's collections. Within this programme BL will develop the Digital Object Management System (DOM System). This system is designed to run on multiple sites with identical functionality (though not necessarily identical hardware). Initially there will be two sites, in BL buildings at Boston Spa and in London. BL has chosen to build a system with two peer clusters at different locations so as to be able provide a high level of resilience and for reasons of disaster management. There will also be a passive backup site, called the 'dark archive'.

At the moment BL is going from the planning and organisational stage into the development and production stage. Ingest of limited classes of material (items received under the Voluntary Deposit of Electronic Publications scheme, which includes 200,000 items occupying 1.5 Tbytes) is planned for mid-July 2005. The development of the system will be worked out in phases. For further information see the 'Materials' subsection below. The digital repository system of BL will be known as the Digital Object Management System. It is expected to be operational by the end of 2005 for the handling of voluntary deposited digital objects.

Services
The DOM Programme is developing a management system for digital objects that will store and preserve any type of digital material in perpetuity, provide access to this material to users with appropriate permissions, ensure that the material is easy to find, ensure that users can view the material with contemporary applications, and ensure that users can, where possible, experience material with the original look and feel. The integrity and authenticity of the digital material will be guaranteed with digital signature techniques.

Work done to date has concentrated on designing and developing a secure storage component. Later developments should lead to services including: one or more ingest processes for different types of material; full links with the Integrated Library System and BL OPAC; development of other resource discovery interfaces as needed; enforcement of digital rights management; active preservation processes (e.g., by migration) and providing appropriate copies of objects to system users (e.g., graphics at different resolutions). Search facilities are not provided yet.

BL will provide the DOM system with stringent security requirements. For this a hardware solution will be used, employing the SHA-1 algorithm: this is currently in use in the financial industries: NIST Federal Information Processing Standards.[97] The approach based on this standard must guarantee both authenticity and integrity of every digital object stored in the DOM system, and should not interfere with long-term preservation goals.

Depositing
At present BL is developing the system for its own use. It is discussing the use or adaptation of the system with the other Legal Deposit Libraries. The material currently being ingested is that received under the Voluntary Deposit scheme. The next block of material scheduled for ingest will probably arise from a pilot project handling e-journals. No formal signed contracts have been signed with depositors or users yet.

Software and OAIS
Present design and development work is concentrating on building a storage system that will endure for a long time. One main principle underlies this work: it is vital for the physical storage and retrieval of the digital objects to be independent of the technical properties and characteristics of the physical storage itself. A design that incorporates this principle should be able to accommodate many generations of physical storage implementations.

[97] NIST FIP Standards: FIPS 140-2, Security requirements for Cryptographic Modules. See: http://csrc.nist.gov/publications/fips/

The software for the DOM System will be developed by BL staff, using commercially available software when appropriate. There are currently no plans to use open source software. The system is intended to comply with the OAIS concept.

BL has been looking at existing systems, but has ruled them out so far, because in general these do not deal with the key issues of BL, or are not suitable at present. However, BL will continue to closely watch the development of the existing systems.

Materials
As the DOM System is developed, the range of digital materials stored in it will expand. As stated above, the material currently being ingested has been received under the Voluntary Deposit Scheme. The next block of material will probably arise from a pilot project handling e-journals. What happens after that will depend on the technical implementation of the 2003 legislation. It is likely that this will first be applied to physical (hand-held) formats, and then to digital journals received online. Where digital publications are also made available in non-digital forms, we expect that the metadata will describe the links to the non-digital versions of the digital material. The DOM System will be expanded in due course to include all of BL's digital collections.

Due to its role as a national library, BL is obliged to preserve all significant published output. This will include nearly all digital output sourced in the UK, such as journals and interactive multimedia on CD-ROM, disks accompanying print publications, networked/online publications sent by e-mail or file transfer protocol, and special collections material such as manuscript collections containing e-mail, disks, tapes of all types and even computers. BL will also preserve the images and structured products from digitisation and in-house projects (for instance the images of 'Turning the Pages', the newspaper collection and cartographic data), websites (mainly results of web archiving projects), audio (British Library Sound Archive) and some geographic information system data (GIS). However, sound and film are excluded from the coverage of the 2003 Act, as existing voluntary schemes have been working well and are expected to continue doing so.

BL has not expressed any preferred formats for texts/still images/audio files, though it does have an active interest in technical standards and may seek to influence producers from time to time.

BL anticipates that material to be stored in the DOM System will arrive in three distinct ways: Firstly: material that should be received under Legal Deposit. The situation is entirely analogous to the present situation with printed materials: producers should submit the items they publish (and be given an official receipt), and BL will also claim items that it believes should have been deposited but have not been. It is fully expected that a good understanding of mutual needs and benefits can be achieved with producers, making this a fairly efficient process, but considerable manual effort may be needed with the continual stream of new producers. Secondly: purchased and in-house generated material. Here there is also an analogy with arrangements for printed materials, with new digital items in effect being harvested. Thirdly: web archiving. There is no analogy to processes for non-digital materials. It is expected that all of the web pages archived will be done so with the full agreement of the producers. It should be possible to develop a fully automated process.

Metadata and metadata schemes
The DOM system will contain metadata to describe files and bit streams. It is expected that all of the metadata, whether liable to change or not, will be stored with the object. Metadata for collections, logical objects and source objects are expected to be held in a separate system.

The stored metadata will contain information on rights and permissions, provenance (document history), technical aspects, administrative and management aspects and structural aspects.

Automated methods for deriving metadata or extracting these from submissions are considered essential. It is to be expected that manual entry will only be possible for very select items and fields. There are no requirements on metadata provision yet, but this might be requested where possible. Although developments are still underway, it is expected that the metadata scheme used by the DOM system will be based on METS and MPEG 21.

Access
Access to the material will depend on the type of object and the permissions given. Digital rights may mean that access to some materials for some users will be restricted to access on a particular site, or that only one access to an item is allowed at any one time. Access may be limited to a restricted community, or it may be fully open and online. Wherever possible, access copies of material will be provided for users: these will be derived from preservation copies for security and ease of use.

3. Preservation strategies

At present, only bit-level preservation (secure storage, backing up, refreshing, etc.) is applied, because other strategies are not yet required for the current materials. BL states that no single strategy will achieve the objective of ensuring long-term access to all types of digital material. Any strategy should be as flexible as possible and not preclude future options and future developments (technical and managerial). A subset of strategies such as refreshing, transferring, migration, emulation and technology preservation should be encouraged with particular emphasis on transfer, migration and emulation. Refreshing will be a component of any good backup regime. BL will work with research projects on new developments in this area. Doing nothing and technology preservation are seen as impracticable and unacceptable approaches to digital preservation.

BL will aim to provide supporting preservation functions to accompany any of the above strategies, such as maintaining preservation metadata and negotiating with publishers to obtain a suitable digital version of material for preservation.

When choosing future strategies BL will firstly focus on safeguarding the original look and feel of the object. Keeping the original bit-stream has priority.

4. Current activities

National activities
The main project on digital preservation in which BL is involved is DOM, an internal activity.

On a national level BL is working in partnership with JISC (Joint Information Systems Committee). In June 2005 this partnership was formalised with a memorandum of understanding between the two organisations. BL and JISC are developing and collaborating in activities and projects, in the field of digital preservation; virtual research environments, resource discovery and content creation. Current projects in which BL and JISC are participating are:

- eTHos (e-Thesis National Service Pilot): eTHos is a digital repository project that aims at providing full text electronic access to theses produced in higher education institutions in the UK. eThos will develop and implement a hybrid IT infrastructure combining a central host based at BL and a single search interface for all e-theses repositories in the UK. Special attention will be paid to rights and permission procedures, dissemination and advocacy and business models.

This project is being coordinated by the University of Glasgow. Other participants are: the Universities of Birmingham, Edinburgh, Southampton and Warwick, Cranfield University, Robert Gordon University, SHERPA, the National Library of Wales and BL. Term: January 2005–June 2006. For further information:
http://www.jisc.ac.uk/index.cfm?name=project_ethos

- SHERPA (Securing a Hybrid Environment for Research, Preservation and Access): SHERPA is a digital repository project to establish 'e-print archives' compliant with the Open Archives Initiative (OAI) Protocol for Metadata Harvesting (OAI PMH) using eprints.org software. The creation, population and management of the repositories will be the core of the project. The project also aims to work in conjunction with publishers. Partners in this project are the Universities of Edinburgh, Glasgow, Nottingham, Oxford, the White Rose Partnerships (Universities of Leeds, Sheffield and York) the Arts and Humanities Data Service and BL. Term: 2002−October 2005. For further information: http://www.sherpa.ac.uk/

- PRESERV (PReservation E-print SERVices): This digital preservation project aims to implement an ingest service, based on the OAIS reference model for institutional archives, built using E-prints software. Working with The National Archives, the project will link E- prints through a web service to software for identification and verification of file formats. The project will emphasise automation, provide modular tools for capturing metadata and enable the identification and verification of file formats. The project will scope a technology watch service to populate and update services where full automation is not feasible for file format recognition. This ingest service will be integrated into the E-prints deposit process for two existing institutional archives at Southampton and Oxford Universities for evaluation, subject to prior satisfactory testing on pilot archives. BL and Southampton University will build and test an exemplary OAI-based preservation service. This service could be used with any OAI-compatible preservation archive to create a software-independent preservation archive. The project partners are: University of Southampton (project leader), The National Archives, Oxford University and BL. Term: October 2004−September 2006. For further information:
http://www.jisc.ac.uk/index.cfm?name=project_preserv&src=alpha

- LIFE (Lifecycle InFormation for E-literature): LIFE is a digital preservation project that will explore and develop a life cycle approach to costing digital archiving for e-journals. After reviewing the existing state of knowledge, it will implement a number of methodologies to selected e-materials for the study. An international conference will be held to evaluate and validate the findings. The outcomes of the project are intended to answer several key questions of costs for

long-term preservation for higher and further education. Partners are University College London and BL. Term: October 2004−September 2005. For further information http://www.ucl.ac.uk/ls/lifeproject/

- DAAT (Digital Asset Assessment Tool). The aim of the DAAT project is to develop a digital preservation assessment tool for use within the UK higher/further education and research, learning and teaching communities. The proposal will provide a tool for identifying the preservation needs of digital holdings. It will do so in a way which allows scarce resources to be focussed on those assets where the risk and cost of loss is greatest. Project leaders are the Arts and Humanities Data Service (AHDS) and the University of London Computer Centre (ULCC). Partners are: the National Preservation Office, BL, The National Archives, Kings College London, The School of Advanced Study, and the University of London. The aim is to encourage the deployment of the tool beyond the JISC community, contributing to the sustainability of the project's outputs. Term: 2004−2006. For further information: http://www.jisc.ac.uk/index.cfm?name=project_daat&src=alpha

- Digital Preservation Training Programme: This project aims to develop a modular training programme in digital preservation, with class-taught, online and offline components. It builds on existing examples of training and information provision, including the Cornell University digital preservation course, the DPC's travelling 1-day workshop, the 'Preservation Management of Digital Materials' handbook, and training from existing JISC-funded services such as AHDS. The training will be developed at multiple levels, to meet the needs of senior managers as well as practitioners and new staff. With the backing of the Digital Preservation Coalition (DPC), the course can be made available outside the JISC community. Partners in this project are the University of London Computer Centre (ULCC) and BL. Term: October 2004−October 2006. For further information: http://www.jisc.ac.uk/index.cfm?name=project_dptp

On a national level BL is also participating in the following working groups:
- The Common Information Environment Group: The aim of this group is to help realise a collaborative, cross-sectoral partnership to build a common online information environment, which will provide full access to the rich information and the exciting possibilities that the web has to offer to each and every one of us. Partners are several UK cultural heritage institutes such as, the National Library of Scotland, BL, The National Archives, the BBC, UKOLN and JISC. For further information: http://www.common-info.org.uk/index.shtml

The UK Web Archiving Consortium: This consortium is addressing the long-term preservation problems of web archiving in a project to develop a test-bed

for the selective archiving of UK websites. An initial 2-year project is currently underway to collect, archive and give access to some 6000 websites using PANDAS software. This project has to be completed by June 2006. Partners are: BL (project leader), The National Archives, the National Libraries of Scotland and Wales, JISC and the Wellcome Trust. For further information:
http://www.webarchive.org.uk/

- The Digital Preservation Coalition (DPC): Since 2001 the UK has had a national body for digital preservation activities: the Digital Preservation Coalition (DPC). BL is one of the full members of DPC. The aim of DPC is to secure the preservation of digital resources in the UK and to work with other international bodies to secure our global digital memory and knowledge base. DPC acts as a coordinating body, bringing together different sectors, exchanging know-how and raising awareness, and disseminating information on current research and practice. Since 2004 the DCC (Digital Curation Centre), part of DPC, has integrated digital preservation research. The DPC organises forums and meetings on all kinds of digital preservation issues, has its own survey line, and since 2004 has presented a Digital Preservation Award for innovative projects, achievements or research in the field of digital preservation worldwide. DPC works in close cooperation with PADI (Australia). The DPC is considered to be an example for many emerging activities on knowledge dissemination in other countries. For further information:
http://www.dpconline.org/

Other institutions
This rough overview is limited to DPC participants only:
- on library materials: the UK public library System; the five national deposit libraries; several university libraries, allied in CURL (the Consortium of University Research Libraries) Cambridge, Oxford, Southampton, Universities of London, Leeds, Sheffield, York; The University Library of Essex (UK Data Archive)
- on audio preservation: the BBC
- on records archiving: archival institutions such as The National Archives, the National Archives of Scotland, the Public Records Offices of Ireland; policy bodies such as the Central Laboratory of Research Councils, The Digital Curation Centre (DCC), CURL, JISC, The Council for Museums, Libraries and Archives (MLA), The Ministry of Defence

and various other institutes: such as OCLC, The National electronic Library for Health, The Arts and Humanities Data Service (AHDS), the University of London Computer Centre, Open University, Association of Learned and Professional Society Publishers (ALPSP), National History Museum, The Publishers Association, The Wellcome Trust Library.

The University of London Computer Centre is working together with The National Archives on preserving large-scale government data sets and on the development of a storage system, suitable for long-term preservation and access.

There are currently no plans for integrating this system with the DOM system. There will be some joint activities on digital preservation in the future, but the use of a common system is not being considered.

On an international level BL is participating in:

International activities
- LOCKSS (Lots of Copies Keep Stuff Safe): A large multinational effort developed by Stanford University, with participation from libraries and publishers, to distribute software and build communities of practice to enable the safe archiving and preservation of e-journals and other web-published content. Partners are: Cambridge University Library, Imperial College, and the Universities of Leeds, Edinburgh and Glasgow. For further information: http://www.lockss.org

- IIPC (International Internet Preservation Consortium): One of the major goals of IIPC is the development of tools for web archiving. At the moment IIPC is working on finalising an IIPC toolkit for acquisition, selection and storage, which will incorporate the IIPC standards (Arc 3.0, Metadata and API). The toolkit will be available in June 2006 as open source. Currently BL and BnF are carrying out a smart archiving crawler project, to implement large scale, automatic focus crawls. The priority will be based on citation linking and thematic assessment. The first prototype is expected mid-2005. Partners in IIPC are: the Library of Congress, BL, the National libraries of Australia, Canada, Denmark, Finland, Iceland, Italy, Norway and Sweden, and the Internet Archive. For further information: http://www.netpreserve.org

- ICABS: Although BL is participating in ICABS (IFLA CDNL Alliance on Bibliographical Standards) its current focus is not on digital preservation, as is the case for the *Koninklijke Bibliotheek* and the National Library of Australia.

- PLANETS: BL is currently coordinating a project proposal for the fifth call of the Sixth Framework Programme of the European Commission. Under the working title 'PLANETS' a group of European libraries, archives and universities is preparing a project proposal which will be submitted to the European Commission in September 2005. PLANETS stands for: Preservation and Long-term Access through NETworked services). This project will conduct research and development on preservation planning, file format characterisation and emulation/migration services. Partners in this project are: *Koninklijke Biblio-*

theek (the Netherlands), the *Österreichische Nationalbibliothek* (Austria), the *Statsbiblioteket* (Denmark), the National Archives of the UK and the Netherlands, and the Technical University of Vienna (Austria).

Website The British Library:
http://www.bl.uk

The United States of America
US Library of Congress (National Library of the USA) – Washington DC

1. General

Legal deposit legislation
The first copyright law of the United States of America goes back to 1790. In 1870, the American Congress passed a law that centralised the copyright system in the Library of Congress (LoC). In 1897 a Registrar of Copyrights was positioned within LoC.

Copyright is a form of protection provided by the laws of the United States (title 17, U.S. Code) to the authors of original works of authorship. Copyright was always given prior to publication up until 1909, when the copyright became effective upon publication. Copyright registration is still not required but since 1909 the Copyright Act has established a mandatory deposit requirement for works published in the US. The owner of the copyright or the owner of the exclusive right of publication of the work has a legal obligation to deposit, within three months of publication in the US, two copies (or, in the case of sound recordings, two phonorecords) of the work to LoC. There are special deposit requirements and exemptions for certain types of work, based on the collection policies of LoC.

For deposit purposes, LoC is also responsible for declaring what it considers the 'best edition' of a work published in more than one form. There are currently no regulations governing the best edition for deposit of works solely available online; automated databases solely available online in the US are expressly exempt from deposit under regulations. LoC's Copyright Office is testing voluntary and limited pilot programmes for electronic copyright registration and deposit.

Since 1988 magnetic tapes, CD-ROMS and microforms have also been deposited. The first e-books for electronic copyright legislation and deposit were received in 2001. Voluntary deposit agreements on CD-ROMS were established in 1993. There is no mandatory deposit law for online electronic publications yet.

Recently a study group has been formed to advise on possible amendments to Section 108 of the Copyright Act. Section 108 provides limited exceptions for libraries and archives. However it does not adequately address many of the issues unique to digital media, either from the perspective of rights owners or libraries and archives. LoC is sponsoring and participating in the Section 108

Study Group. This group will study how Section 108 of the Copyright Act could be amended to address the relevant issues and concerns of libraries and archives, as well as creators and other copyright holders. By mid-2006 the group will advise LoC on how to revise the copyright law with regard to the digital aspect, so as to ensure an appropriate balance among the interests of creators and other copyright holders, libraries and archives in a manner that best serves the national interest.

Digital preservation in LoC

Organisational embedding
In 1998 LoC started to plan for institutional changes associated with information technology. A study was commissioned by the Computer Science and Telecommunications Board (CSTB) of the National Research Council to provide strategic advice concerning the IT future of LoC, and in particular the integration of digital and analogue collections, and how to interact with emerging digital library technologies, systems and projects. The CSTB report was published in 2000.[98] In 2001, as a direct outcome of the survey, the Office of Strategic Initiatives (OSI) was set up and a new position of Associate Librarian for Strategic Initiatives was created.

Since 1994, LoC has been conducting the National Digital Library (NDL) Program. One of the major achievements of the NDL Program is the American Memory project, a large digitisation project that makes LoC collections accessible through the Internet. The practices for digitisation, persistent identification, and storage management that have been established by the NDL program have since been applied to other projects that provide access to digital reproductions of content from the LoC collections. Such projects include Global Gateway, the digital collections from the Music Division, the Prints and Photographs Online Catalog, and the Veterans History Project.

In December 2000, Congress directed LoC to develop and execute the National Digital Information Infrastructure and Preservation Program (NDIIPP) and provided US$ 99.8 million for this purpose. This funding was meant to set forth a strategy for LoC, in collaboration with other federal and non-federal entities, to identify a national network of libraries and other organisations with responsibilities for collecting digital materials, which will provide access to and maintain of those materials. Once the strategy is developed, the focus will be on developing the policies, protocols and strategies for the long-term preservation of such materials, including the technological infrastructure required at LoC.

[98] CSTB [a.o.]. 2000. *LC21: A Digital Strategy for the Library of Congress.* Washington D.C.: National Academy Press.

The Congress approved the NDIIPP plan in December 2002.[99] The plan outlined a process to develop a national digital preservation strategy, including the key components of the digital preservation infrastructure. OSI is the service unit that is responsible for managing NDIIPP.

The Law Library of Congress has also been actively involved in digital preservation through the Global Legal Information Network. Legal materials, particularly primary source materials, have unique preservation requirements; preservation of format is as critical as preservation of content. The Law Library is working with the Legal Information Preservation Alliance to define objectives, develop standards, explore models, create networks, and foster support for long-term sustainability.

LoC is the oldest federal cultural institution of the US and serves as the research arm of Congress. It is also the largest library in the world. The mission of LoC is to make its resources available and useful to Congress and the American people and to sustain and preserve a universal collection of knowledge and creativity for future generations. LOC's activities (including those relating to digital preservation) are currently housed in three buildings located on Capitol Hill in Washington DC.

LoC is made up of the following units: The Congressional Research Services, the Copyright Office, the Enabling Infrastructures, the Law Library of Congress, the Office of the Librarian, and the Office of Strategic Initiatives. See also Section II.2 for the organisational chart.

The Congressional Research Services (CRS) is committed to providing Congress legislative research, analysis, and information services throughout the legislative process that are timely, objective, non-partisan, and confidential, thereby contributing to an informed national legislature.

The mission of the Copyright Office is to promote creativity by administering and sustaining an effective national copyright system. The Copyright Office manages copyright registration and deposit activities, and also provides policy advice to Congress in the development of national and international policy.

The Office advises the Librarian of Congress on all copyright issues, including those related to digital preservation. It will also play an important role in implementing the strategic outcomes of NDIIPP. The Copyright Office's *Strategic*

[99] *Preserving Our Digital Heritage. Plan for the National Digital Information Infrastructure and Preservation Programme. A Collaborative Initiative of the Library of Congress.* October 2002. See: http://www.digitalpreservation.gov/repor/ndiipp_plan.pdf

Plan 2004–2008 includes the objective to increase the acquisition of digital objects in a variety of formats, through registration and mandatory deposit.[100]

The Law Library of Congress provides research and legal information to Congress, as well as to federal courts and executive agencies, and offers reference services to the public.

Library Services is responsible for the acquisition, organisation, accessibility, maintenance, security and preservation of LoC's collections. The Office of the Librarian directly comes under the Library Services and is the main responsible for the mission of the LoC.

The Office of Strategic Initiatives (OSI) is responsible for the overall strategic planning for the library and for the national program for long-term preservation of digital cultural assets. It leads a collaborative institution-wide effort to develop consolidated digital future plans and to integrate the delivery of information technology services. The Information Technology Services (ITS) directorate within OSI is responsible for LoC's computer and telecommunications infrastructure and supports the technology needs of the service units and their staffs.

All of these units have a stake in digital preservation activities within the Library. OSI is responsible for coordinating institution-wide technology initiatives and for integrating digital preservation best practices and other knowledge from NDIIPP projects. OSI convenes meetings of the Digital Executive Oversight Group (DEOG), which comprises the Chief of Staff in the Office of the Librarian, as well as the heads of Library Services, Law Library, Congressional Research Service, and Copyright Office. The DEOG, established in 2002, serves as the internal means for vetting, justifying, and allocating resources for the digital programs and information technology initiatives. Through its regularly scheduled meetings, the DEOG reviews existing work and sanctions new digital initiatives within the Library.

Currently there are 96 staff within OSI and 203 staff within ITS. Digital preservation activities in the other divisions occupy variable numbers of staff at any given time.

[100] United States Copyright Office, 2004-2005: *Strategic Plan 2004-2008. United States Copyright Office.*

Funding
LoC's digital preservation projects are funded through a combination of congressionally-appropriated funds and, for some external projects as part of NDIIPP, special matching financial arrangements.

2. Digital repository

Status
The responsibility for computer storage, telecommunications, security, backup, and all aspects of software and hardware operations is centralised within ITS. At the physical level (ensuring the ongoing integrity of digital bits on disk or tape), ITS is responsible for managing digital content, whether created within LoC or received from external sources. ITS maintains a robust storage area network, an alternative computing facility, plans and procedures for backup and media migration, and security policies. A major activity involves providing ongoing access via the website through American Memory, Global Gateway, and other services providing access to digital content. ITS also assists in developing prototypes for advanced digital preservation projects.

LoC currently manages over 80 terabytes of digital content, which consists of scanned images, web pages, photographs, audio, video, and other digital formats and types. The different divisions/sections with curatorial responsibility for digital content coordinate with ITS and are responsible for providing metadata and other information needed for retrieval and use.

In 1998, the David and Lucile Packard Foundation provided a grant to begin the development of a state-of-the-art National Audio-Visual Conservation Center on a site near Culpeper, Virginia. Support from the foundation and the allied Packard Humanities Institute has continued to the present day. The new centre has been authorised by Congress, and will enable LoC to develop a central storage and conservation facility for the audio-visual collections. It will feature specialised, newly-designed preservation laboratories for all of its audio-visual media, including digital content. The physical storage facility will open in 2005. The National Audio-Visual Conservation Center is expected to be fully operational in 2006. During the planning period for the new centre, LoC has undertaken a number of audio-visual prototyping projects to develop approaches for the digital reformatting of moving image and recorded sound collections as well as studying issues related to born-digital audio-visual content. The effort, coordinated by the Motion Picture, Broadcasting, and Recorded Sound Division and ITS, is working with outside specialists to develop detailed plans for digital systems within the facility and the extensive networking and storage systems that will be required.

In 2004, LoC announced a joint initiative with the National Endowment for the Humanities for the digitisation of newspapers from microfilm. The test bed phase of this project (2005−2007) will use Fedora as its core repository. For this content, a single repository will manage the archival master content and the production master files that will be used by the access application. Detailed specifications for image files, embedded metadata (including provenance and preservation metadata), and the representation of compound objects for newspaper issues have been established and will be validated before ingest into the repository.[101]

LoC expects that multiple repositories will be necessary to support the management, preservation, and access needs of its digital content. Currently, the focus is on gathering requirements for a repository framework, supporting focused internal and external prototypes, and sharing information with its partners about emerging technology, best practices, and lessons learned. LoC has a digital life cycle framework to explore requirements and plan projects for digital content. Significant additional information about preservation approaches is coming from work with NDIIPP partners. In particular, much knowledge is being acquired though the development and testing of a technology architecture to support multipartner efforts to ingest, preserve, and provide access to heterogeneous digital content. LoC expects to develop advanced digital preservation systems in an incremental manner by adopting technology that can scale to the massive size and complexity required to meet future needs.

Services

LoC now provides services to store, describe, and provide access to all digital content. The intent is to build on this solid foundation for the long-term preservation of heterogeneous digital objects and their associated metadata. Future systems would, for example, enable: multiple complex ingest procedures in accordance with specific agreements with diverse content providers; preservation management functions to enable appropriate migration, emulation, or other services to overcome technological obsolescence; detailed reports regarding the level of service for objects in terms of preservation and access; and seamless access to distributed collections of service and/or archival copies. In addition, LoC requires repository services that include secure storage, appropriate access to restricted or rights-protected content, and automated workflow management.

Depositing

With some exceptions, only LoC can now place content under digital management services. LoC does have an agreement with UMI, a Bell & Howell

[101] Digitisation of newspaper from microfilm: See: http://www.loc.gov/ndnp/

company, which enables electronic copyright registration and deposit of dissertations with the Copyright Office. In addition, the agreement designates UMI's ProQuest Digital Dissertations as the official off-site repository for a collection of more than 100,000 dissertations and theses converted to digital form since 1997, as well as those to be produced in the future.[102] The agreement is the first time LoC has designated an official off-site repository for deposited digital collections.

In accordance with the objectives stated in its Strategic Plan, the Copyright Office is developing a new system that will support both online registration and the deposit of electronic works.

Software and OAIS
The Open Archival Information System (OAIS) reference model is useful to LoC as a guide for identifying requirements and building prototypes. The architectural framework proposed for NDIIPP is also consistent with the OAIS reference model. In terms of systems development, the National Digital Newspaper Program will use Fedora as the basis for its test bed phase and for automated validation of digital files. LoC and its NDIIPP partners are testing multiple repository approaches, including DSpace, Fedora, aDORe, LOCKSS, and the OCLC Digital Archive, as well as customised local configurations that rely on commercial, in-house, and open source software.

*Material*s
LoC is establishing institution-wide practices that will ensure long-term retention of digital resources from all sources, whether created internally or acquired for its collection by purchase or through copyright deposit, whether born-digital or converted from analogue originals. The intent is to provide for preservation of digital content in a wide variety of formats, types, and media. In-depth analysis focused on the long-term sustainability of digital formats is now underway.

Metadata and metadata schemes
LoC is supporting the development and deployment of a variety of metadata schemes for use with digital objects. Examples include the Metadata Encoding and Transmission Standard (METS); Metadata Object Description Schema (MODS); Metadata Authority Description Schema (MADS); and Metadata for Images in XML (MIX). Library staff also are playing a key role in the advancement and trial use of the PREMIS data dictionary for preservation metadata. The Network Development and MARC Standards Office is hosting the PREMIS maintenance activity website, that will mainly focus on maintain-

[102] UMI's ProQuest Digital Dissertations: See: http://www.umi.com/umi/dissertations/

ing the PREMIS schemes and making further information about the PREMIS maintenance activity available.[103]

The NDL Program identified several categories of metadata to support access and management needs. Descriptive metadata supports discovery through search and browse functions. Structural metadata supports presentation of complex objects by representing relationships between components, such as sequences of images. In addition, administrative metadata is needed to support management tasks, such as access control, archiving, and migration. Individual metadata elements may support more than one function, but the categorisation of elements by function has proved useful.

In order to avoid duplicate catalogue records for digital and physical copies of the same item, LoC's catalogue will provide links, directly or indirectly (for example, through collection-level records), to all digital content to which LoC provides public access. A persistent identifier serves as the link between a descriptive record and a digital content object. LoC uses the Handle Server from CNRI to resolve its persistent identifiers to current locations. The Law Library has developed standards for descriptive metadata used by all nations that contribute legal documents to GLIN, the Global Legal Information Network.

For American Memory, Global Gateway, and other projects that provide access to digital reproductions, a limited set of compound digital object types is in use. These object types are represented by data structures incorporating structural metadata that relates component files (such as page image files or digitised audio tracks) to the parent object. For NDNP and the Veterans History Project, the METS standard is used to represent the compound objects. Tests are also underway in collaboration with the Los Alamos National Laboratory to use the MPEG-21 Digital Item Declaration Language to represent compound objects. LoC believes that by experimenting with different approaches it will come to understand the advantages and disadvantages of alternatives.

In terms of developmental and future systems, LoC is interested in tools and procedures to automatically generate and/or extract metadata during different stages of the digital life cycle, particularly creation and ingest. Another area of interest includes assisting creators and depositors to provide necessary metadata: robust methods to validate metadata, including technical and preservation-related elements.

[103] PREMIS standards maintenance: See: http://www.loc.gov/standards/premis

Access

Currently, the primary means of access to the digital content of LoC is through projects such as American Memory, Global Gateway, and the Veterans History Project. LoC also supports electronic access to US federal legislative information through the Thomas website and to international legislation through the GLIN website.

While most of the records in the main online catalogue describe physical items in the collections (books, print periodicals, sound recordings, and so forth), the catalogue includes a growing number of records for electronic resources, including digitised photographs and motion pictures, electronic journals and databases. To identify these electronic items in the catalogue, there are label links on the Brief Record display of the record. These links will take a patron to the electronic item described by the record.

The Prints and Photographs Online Catalog contains catalogue records and digital images representing a rich cross-section of still pictures held by the Prints and Photographs Division and other units of LoC. These materials have been made accessible for a broad public, to contribute to education and scholarship. The catalogue provides access through group or item records to about 65% of the Division's holdings. About 90% of the records are accompanied by one or more digital images. In some collections, only thumbnail images display to those searching outside LoC because of potential rights considerations.

Through subscription services, LoC also provides local access to electronic resources such as journals, newspapers, magazines, books, manuals, and other materials. Current subscriptions include over 200 databases, with new databases added regularly. These range from indexes and abstracts for magazines and scholarly journals to full text resources. Many databases cover specific disciplines or areas of study; others are general or multidisciplinary. Access to subscription databases is available from within LoC itself in any of the public workstations in the reading rooms, although there are a few databases that can be used only from a particular reading room.

Extending electronic access is a key long-term goal. As the technical infrastructure continues to develop, the intent is to provide enhanced access consistent with copyright and other appropriate restrictions

3. Preservation strategies

The current practice for retention of locally digitised materials is based primarily on ongoing refreshing of bits. Files on secure central servers are protected by standard backup and restore procedures, along with replication and frequent

refreshing. Steps are taken during digitisation to reduce the need for migration, including applying metadata that will support effective migration when necessary to avoid the future need for emulation or digital archaeology. It can be noted that in the 10 years since the establishment of the National Digital Library Program, only one file format has required migration. Currently, a careful investigation of the JPEG 2000 format may prompt some further consideration of migration of formats to sustain easy access.

Looking to the future, it is understood that optimal preservation of resources in digital form requires attention early in the life-cycle − preferably at the moment of creation, publication, or acquisition − as well as ongoing management to ensure continuing access and interpretability. A key consideration in this regard involves choosing file formats that are most favourable from a long-term preservation perspective. LoC is examining many different preservation approaches and is open to the potential use of multiple strategies. It is possible, for example, that certain digital object classes will require retaining some or all significant properties (such as formatting, presentation, and other 'look and feel' considerations), while others will require only ongoing retention of the information content.

4. Current activities

National activities
Among the major digital preservation-related activities of the Library within the US are:

- NDIIPP: see below
- Section 108 Study Group. See Section 1. General
- The National Audio-Visual Conservation Center near Culpeper, Virginia. See Section 2. Status.
- Standards for resource description formats, digital libraries, and information resource retrieval protocols. LoC is actively involved in helping to establish and implement a number of standards. For further information see:
http://www.loc.gov/standards/
- Web harvesting. In 2000, LoC began experimenting with web harvesting through the MINERVA project. A multidisciplinary team of Library staff studied methods to evaluate, select, collect, catalogue, access, and preserve these materials for future generations of researchers. In 2003, LoC expanded support for web content collection and entered into an international collaboration with 11 other national libraries and the Internet Archive to develop tools and articulate processes for archiving web content. For further information see:
http://lcweb2.loc.gov/cocoon/minerva/html/minerva-about.html

- Digital Formats. LoC staff are analysing digital content formats and making information available through a public website. The site is devoted to the analysis of the technical aspects of digital formats, with a particular focus on implications for policy matters, most significantly collection policies. For further information see: http://www.digitalpreservation.gov/formats/

Other institutions

A small sample of the organisations in the US active in digital preservation:

- National Archives and Records Administration;[104]
- National Library of Medicine;[105]
- National Institute of Standards and Technology;[106]
- RLG;
- OCLC;
- Council on Library and Information Resources;
- Legal Information Preservation Alliance.[107]

International activities

LoC is a charter member of the International Internet Preservation Consortium (IIPC), along with the National libraries of France, Australia, Canada, Denmark, Finland, Iceland, Italy, Norway, Sweden, the British Library (UK), and the Internet Archive (USA). LoC leads the Metrics and Testbed Working Group. This working group is organised to define metrics for web archiving and to define and characterise an evaluation process for the coverage and performance of web-archiving tools and processes. The main achievement identified a test bed that can be used for evaluating web crawlers and other archiving tools and it continues to define the set of metrics for the life cycle of archived web content. To date, the working group has developed an inventory and a taxonomy to address the methodological characteristics and issues of collecting, rendering, and preserving data from the web. The working group is also surveying each aspect and designing a working document that will serve as a guide to develop a common understanding of these issues. The metrics derived from this work will formalise the development of a testbed that will assist in measuring the quality and performance of tools and processes. For further information see: http://netpreserve.org/about/index.php.

LoC played a key role in the PREservation Metadata: Implementation Strategies (PREMIS) working group. The PREMIS working group, jointly sponsored by OCLC and RLG, was composed of international experts from institutions that

[104] NARA: See: http://www.archives.gov/
[105] NLM: See: http://www.nlm.nih.gov/
[106] NISO: See: http://www.nist.gov
[107] LIPA: See: http://www.aallnet.org/committee/lipa

had developed or were currently developing digital preservation capacity. The group compiled a final report, along with a core preservation metadata set, supported by a data dictionary, with broad applicability across the digital preservation community. For further information see: http://www.oclc.org/research/projects/pmwg/

NDIIPP is an allied organisation of the UK's Digital Preservation Coalition, which has a mission to secure the preservation of digital resources in the UK and to work with others internationally to secure our global digital memory and knowledge base. For further information see: http://www.dpconline.org/

Website Library of Congress:
http://www.loc.gov/

National Digital Information Infrastructure and Preservation Program: NDIIPP

Background
In December 2000, Congress authorised LoC to develop and execute a congressionally- approved plan for a National Digital Information Infrastructure and Preservation Program (NDIIPP). The NDIIPP plan, *Preserving Our Digital Heritage: Plan for the National Digital Information Infrastructure and Preservation Program*, outlines a process to implement the program.[108]

Purpose
The mission statement of NDIIPP is: 'Develop a national strategy to collect, archive and preserve the burgeoning amounts of digital content, especially materials that are created only in digital formats, for current and future generations.'

Initial emphasis
The initial emphasis of NDIIPP has focused on building a national network of partners, developing and testing a digital preservation architecture, and supporting advance digital preservation research. Communication with all stakeholders about NDIIPP activities is also a key emphasis.

Network of prartners
LoC is implementing NDIIPP through building partnerships with key federal agencies (including the National Archives and Records Administration, National Library of Medicine, and the National Agricultural Library), information technology companies, universities, non-profit entities (such as RLG, OCLC and CLIR), and other stakeholders. (See below for further information about specific partnerships.)

Main goals
The primary goals of NDIIPP are: to develop a national digital collection and preservation strategy; establish a network of partners committed to digital preservation; identify and preserve digital content that is significant and at risk; and support improved tools, models, and methods for digital preservation.

Achievements: Digital preservation partners
In 2004, LoC established a formal network of NDIIPP Digital Preservation Partners consisting of eight consortia, each of whom have multiple institutional members. Using a combination of NDIIPP investments and local matching resources, each consortium will identify and preserve at-risk digital content,

[108] The NDIIPP plan, along with extensive current information about NDIIPP, is available at http://www.digitalpreservation.gov.

develop scalable digital collection and preservation strategies, explore protocols and standards to support partnership operations, and support the development of tools, models and methods for preservation.

Projects will explore content selection and collection strategies, probe intellectual property issues related to preservation, consider economic sustainability for digital preservation activities, and address a broad range of technical architecture matters. All participants will also engage in collaborative interactions to identify and share best practices and develop effective partnership network policies and procedures. A summary of the projects is provided below.

Lead institution: California Digital Library at the University of California.
Partners: New York University, University of North Texas, The Libraries, and the Texas Center for Digital Knowledge. Collaborators: San Diego Supercomputer Center, Stanford University Computer Science Department and Sun Microsystems Inc.
Subject: This award is for a project to develop web-archiving tools that will be used by libraries to capture, curate and preserve collections of web-based government and political information. This literature is a critical element of our nation's heritage and is increasingly found exclusively online, putting it at greater risk of being lost. The collections will focus on local political activities and movements, such as the California gubernatorial recall election of 2003.

Lead institution: University of California at Santa Barbara.
Partner: Stanford University.
Subject: These institutions will lead the formation of a National Geospatial Federated Digital Repository to design an infrastructure and collect materials across the spectrum of geographic formats. The born-digital materials to be collected and preserved will range from Landsat imagery to other cartographic content from university, corporate and government resources as well as websites. The repository will preserve content vital for the study of history, science, environmental policy, urban and population studies, census construction and analysis, and other fields requiring US geospatial information.

Lead institution: Educational Broadcasting Corporation (EBC) (Thirteen/Wnet New York).
Partners: WGBH Educational Foundation, Boston, Mass.; Public Broadcasting Service (PBS), Alexandria, Va.; New York University (NYU), NY.
Subject: Partners in this project will collaborate to establish the initial procedures, structures and national standards necessary to preserve public television programmes produced in digital formats. EBC and WGBH are the two largest producers of public television content in the United States. Through PBS, their productions are made available to audiences from coast-to-coast. Together,

these three entities produce and distribute the majority of public television in the United States. NYU is home to one of America's most distinguished research libraries and has become a major player in the field of digital preservation of moving images. The four partners will focus on such influential series as 'Nature', 'American Masters' and 'Frontline', which are increasingly being produced only in digital formats, including the new high-definition standard (HDTV). The project will also examine issues associated with the preservation of important corollary content, such as websites that accompany broadcasts.

Lead institution: Emory University.
Partners: The University of Louisville Libraries, Virginia Polytechnic Institute and State University Libraries, Florida State University, Auburn University Libraries, Georgia Institute of Technology Library and Information Center.
Subject: This project will develop a MetaArchive of Southern Cultural Heritage (http://www.metaarchive.org) by creating a distributed digital preservation network for critical and at-risk content relative to Southern culture and history. The partners will select and preserve institutional digital archives, as well as ephemeral works such as online exhibitions and cultural history website displays. This body of digital content includes a wide variety of subjects complementary to LoC collections such as the Civil War, the civil rights movement, slave narratives, Southern music, handicrafts and church history.

Lead institution: University of Illinois at Urbana-Champaign Library, Graduate School of Library and Information Science and National Center for Supercomputing Applications. *Partners*: OCLC Online Computer Library Center, Tufts University Perseus Project, Michigan State University Library, and an alliance of state library agencies from Arizona, Connecticut, Illinois, North Carolina and Wisconsin.
Subject: This project will develop criteria for determining which digital materials to capture and preserve, as not all digital material can or should be preserved. These materials will include sound and video recordings, historical aerial photography, web-based government publications from the partner states, and primary and secondary historical materials made available by the Perseus Project.

Lead institution: University of Maryland Robert H. Smith School of Business.
*Partner*s: Center for History and New Media at George Mason University; Gallivan, Gallivan and O'Melia LLC; Snyder, Miller, Orton Lawyers LLP; and the Internet Archive.
Subject: This project will preserve at-risk digital materials from the American business culture during the early years of the commercialisation of the Internet − the 'Birth of the Dot Com Era,' specifically 1994−2001. The materials, collected through web portals at http://www.businessplanarchive.org and http:

www.dotcomarchive.org and through direct contact with former participants in the Dot Com Era, will be of incalculable historical value to Americans eager to make sense of this remarkable period of venture creation.

Lead institution: University of Michigan Inter-university Consortium for Political and Social Research.
Partners: The Roper Center for Public Opinion Research at the University of Connecticut, the Howard W. Odum Institute for Research in Social Science at the University of North Carolina-Chapel Hill, the Henry A. Murray Research Center at the Radcliffe Institute of Harvard, the Electronic and Special Media Records Service Division of the National Archives and Records Administration and the Harvard-MIT Data Center.
Subject: These institutions will create a partnership to identify, acquire and preserve data used in the study of social sciences to ensure that future generations of Americans have access to this vital digital material that will allow them to understand their nation, its social organisation and its policies and politics. Examples of data that will be preserved are opinion polls, voting records, large-scale surveys on family growth and income, and focused studies on the effects of events such as factory closures or the need to care for aging parents. Together the partners will build a shared catalogue, adopt a common standard for describing survey data and develop strategies for ensuring that the data remains available for analysis.

Lead institution: North Carolina State University Libraries.
Partner: North Carolina Center for Geographic Information & Analysis.
Subject: The project will collect and preserve digital geospatial data resources, including digitised maps, from state and local government agencies in North Carolina. Geospatial data are created by a wide range of state and local agencies for use in applications such as tax assessment, transportation planning, hazard analysis, health planning, political redistricting, homeland security and utilities management. Although this project will focus solely on North Carolina, it is expected to serve as a demonstration project for other states.

Achievements : technical architecture
During the initial planning period, NDIIPP convened four workshops with technologists from the private and public sectors. The workshops produced a Technical Architecture framework for preservation that is intended to guide the development of a national distributed network. The principles for the architecture state that it must support institutional relationships, separate preservation and access functions sufficiently to protect intellectual property, be constructed modularly and assembled over time, be able to upgrade parts without disruption of the whole, and use broadly-adoptable standards and protocols. Work has begun to explore strategies for the ingest and preservation of digital archives.

The current working version (0.2) of the Technical Architecture is available at http://www.digitalpreservation.gov. A one-year to challenge the architecture, the Archive Ingest and Handling Test (AIHT), was completed in April 2005. The test was designed around the transfer, ingest, management, migration, and export and exchange of a common test set of heterogeneous data contributed to LoC by the George Mason University. A full report of the work, which involved LoC and four universities (Harvard, Johns Hopkins, Old Dominion, and Stanford) will be published shortly. The Digital Preservation Partnerships will expand upon the aspects of interoperability and shared preservation among institutions.

Achievements: Digital preservation research
LoC and the National Science Foundation (NSF) have partnered to establish a Digital Archiving and Long-Term Preservation DIGARCH research program as part of NDIIPP, and were recently awarded US$ 3 million to undertake pioneering research. The projects will explore topics, such as preserving rich oceanographic data from hundreds of deep-sea submersible missions; automating methods to describe digital objects and place them in secure archival storage; testing how to preserve digital video when it is first created; and preserving complex three-dimensional digital content. All the projects are expected to produce study results in one year. The results of the projects will be integrated with the larger NDIIPP effort. All applications were subjected to a NSF peer-review process. A summary of the projects are listed below.

Institutions: University of California San Diego, Scripps Institute of Oceanography and San Diego Supercomputer Center; Woods Hole Oceanographic Institution.
Title: Multi-Institution Testbed for Scalable Digital Archiving.
Summary: These two institutions will develop a multiterabyte digital repository to preserve data from more than 1600 oceanographic research projects. The collaborating institutions will test processes for automatic archival ingest (acquisition), metadata extraction, validation and access control, and will also explore methods for management of rights-protected data.

Institution: University of Maryland.
Title: Robust Technologies for Automated Ingestion and Long-Term Preservation of Digital Information.
Summary: This project will explore automated ingest and verification for distributed digital collections. It will also develop and test a preservation architecture that can 'evolve gracefully' as technology changes and that is interoperable with different computer platforms.

Institution: Drexel University.
Title: Digital Engineering Archives.
Summary: This project will work with decades of three-dimensional Computer Assisted Design (CAD) engineering design and production data that currently have very limited preservation options. Researchers will use international standards to convert complex design data into more readily preservable content and will use the results to educate the engineering community about three-dimensional data preservation options.

Institution: University of California San Diego, San Diego Supercomputer Center.
Title: Digital Preservation Lifecycle Management: Building a Demonstration Prototype for the Preservation of Large-Scale Multimedia Collections.
Summary: The project will demonstrate a preservation life cycle management process for video content. Researchers will develop and document a practical preservation process for mixed collection of both legacy and born-digital video material.

Institution: University of Arizona.
Title: Investigating Data Provenance in the Context of New Product Design and Development. *Summary*: This undertaking will investigate ways to automate metadata capture through an innovative partnership with Raytheon, a commercial defence and aerospace systems supplier. Methods to develop 'self aware/self describing' production and design digital data will be explored.

Institution: University of Michigan.
Title: Incentives for Data Producers to Create Archive-Ready Data Sets.
Summary: The project will examine incentives for data producers to deposit 'archive-ready' data sets. Focus will be on collaboration between producers and archives, including identification of a process for archives to adjust their deposit requirements to better suit producer needs.

Institution: Old Dominion University.
Title: Shared Infrastructure Preservation Models.
Summary: This project will evaluate existing shared Internet infrastructure elements (such as Simple Mail Transfer Protocol or SMTP) to determine if they are suitable for digital preservation purposes. Researchers will explore options to reduce digital preservation costs through the use of cheap and widely deployed protocols.

Institution: University of Tennessee at Knoxville.
Title: Planning a Globally Accessible Archive of MODIS Data.

Summary: The aim of this project is to bring together leaders of the Moderate Resolution Imaging Spectroradiometer (MODIS) archive community with computer science researchers to discuss new distributed approaches to managing MODIS satellite data, which currently has a volume of about two petabytes.

Institution: University of North Carolina at Chapel Hill.
Title: Preserving Video Objects and Context: A Demonstration Project.
Summary: Development of rich descriptive terms and a process for applying them to digital objects is the focus of this study. Attention will also be given to demonstrating a cost-benefit methodology.

Institution: Johns Hopkins University.
Title: Securely Managing the Lifetime of Versions in Digital Archives.
Summary: This project will study technologies for secure deletion of information to protect personal privacy and provide a mechanism to ensure that no unwanted data is retained along with preserved data.

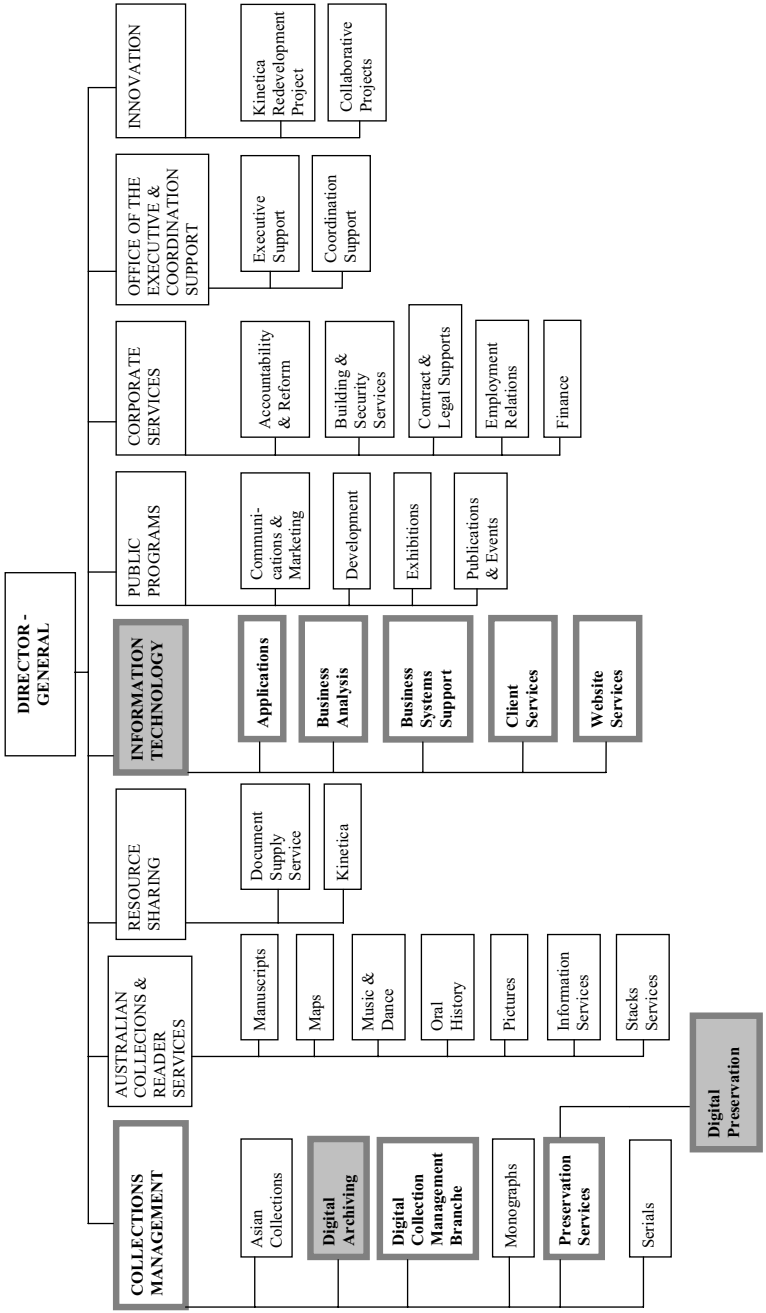
Website NDIIPP:
http://www.digitalpreservation.gov/

II. Overviews national libraries

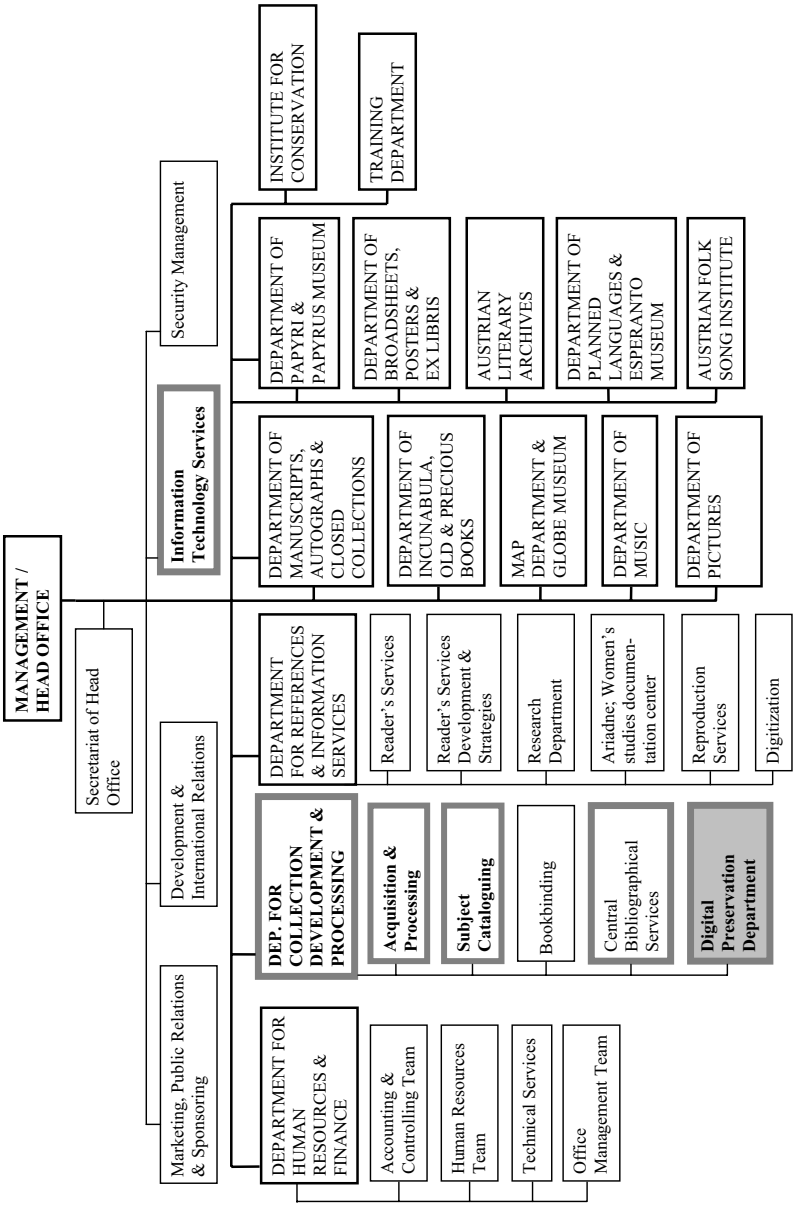
2. Organisational charts[109]

[109] The layout of the organisational charts is adapted to make them better comparable. The departments that are involved in digital preservation are marked in bold and grey. Organisational charts are subject to frequent change. Visit the website of the national library involved to see the current situation.

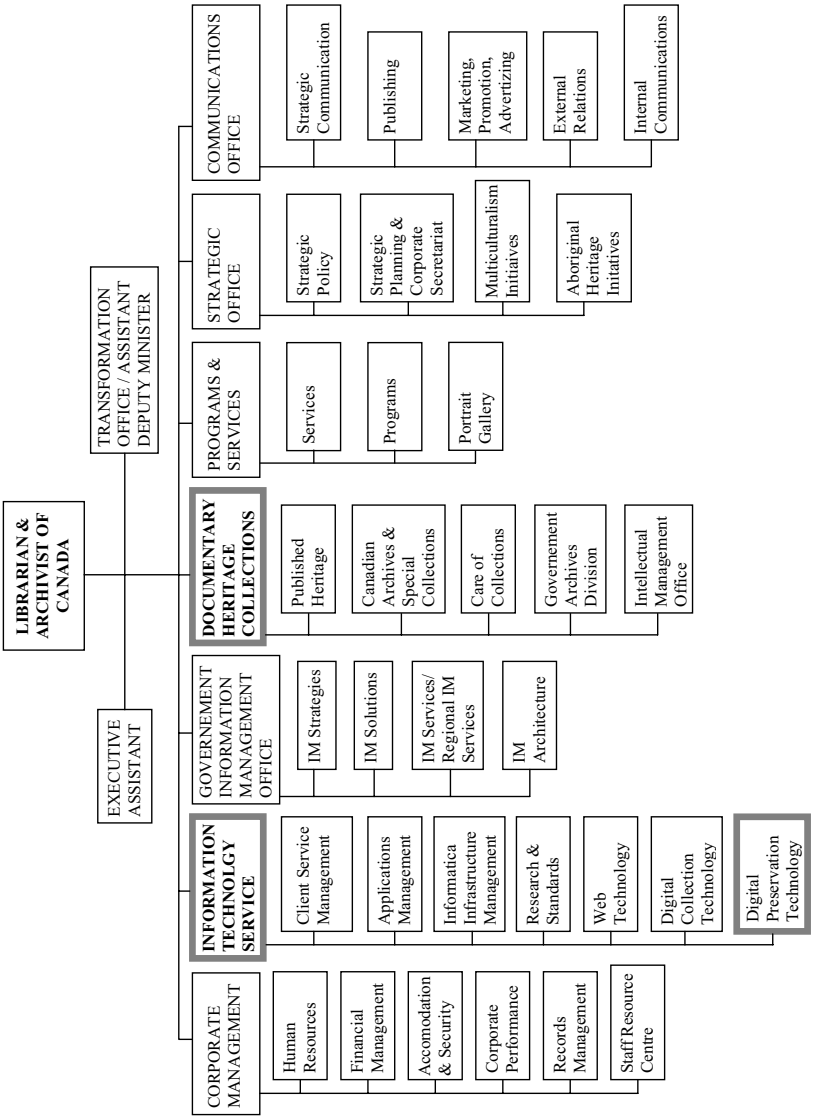
National Library of Australia – June 2005



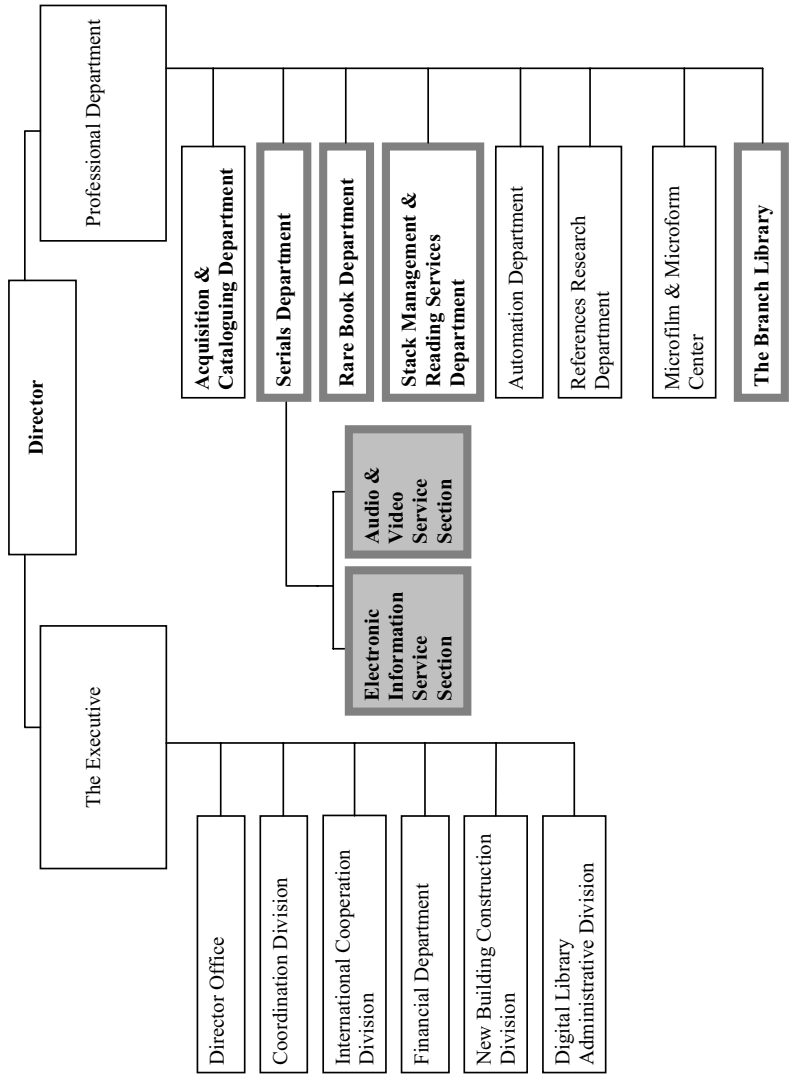
National library of Austria – January 2005



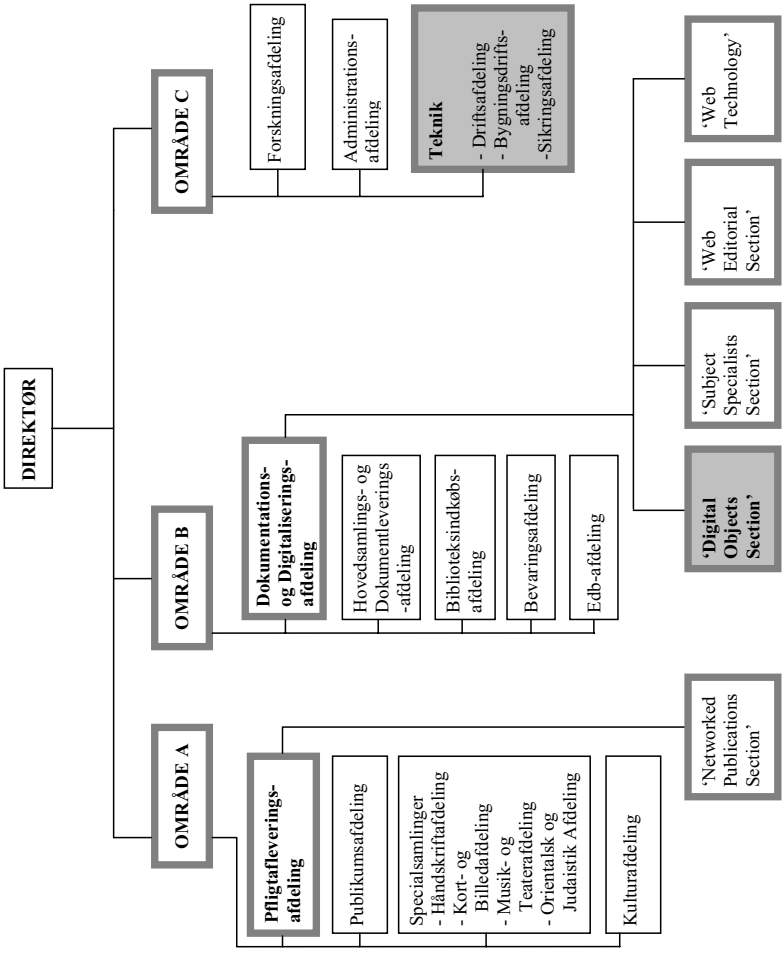
Library and Archives Canada – July 2005



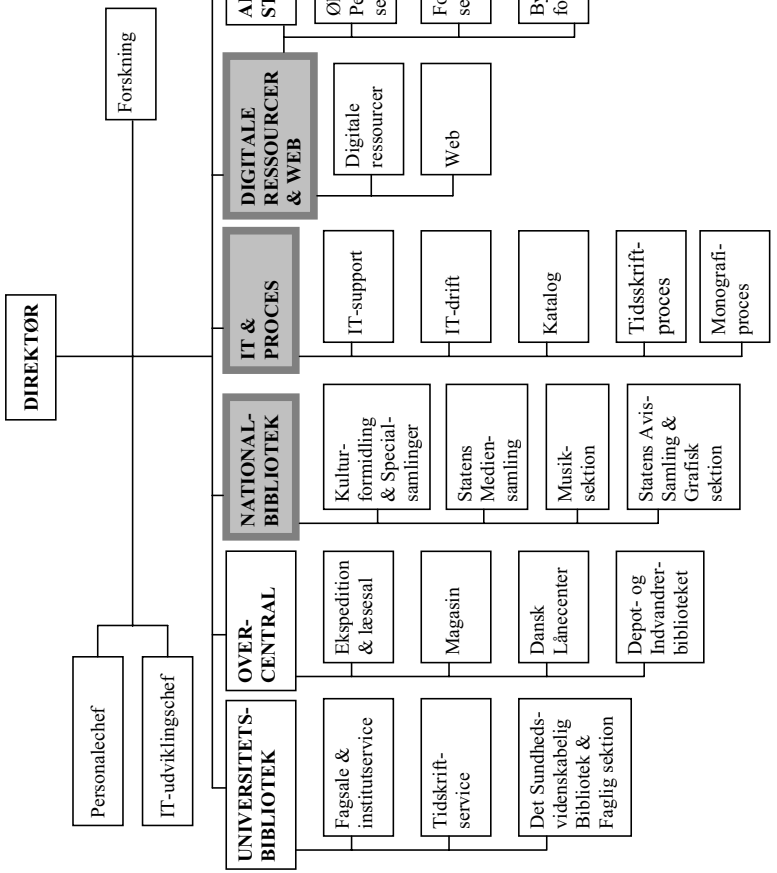
National Library of China –July 2005



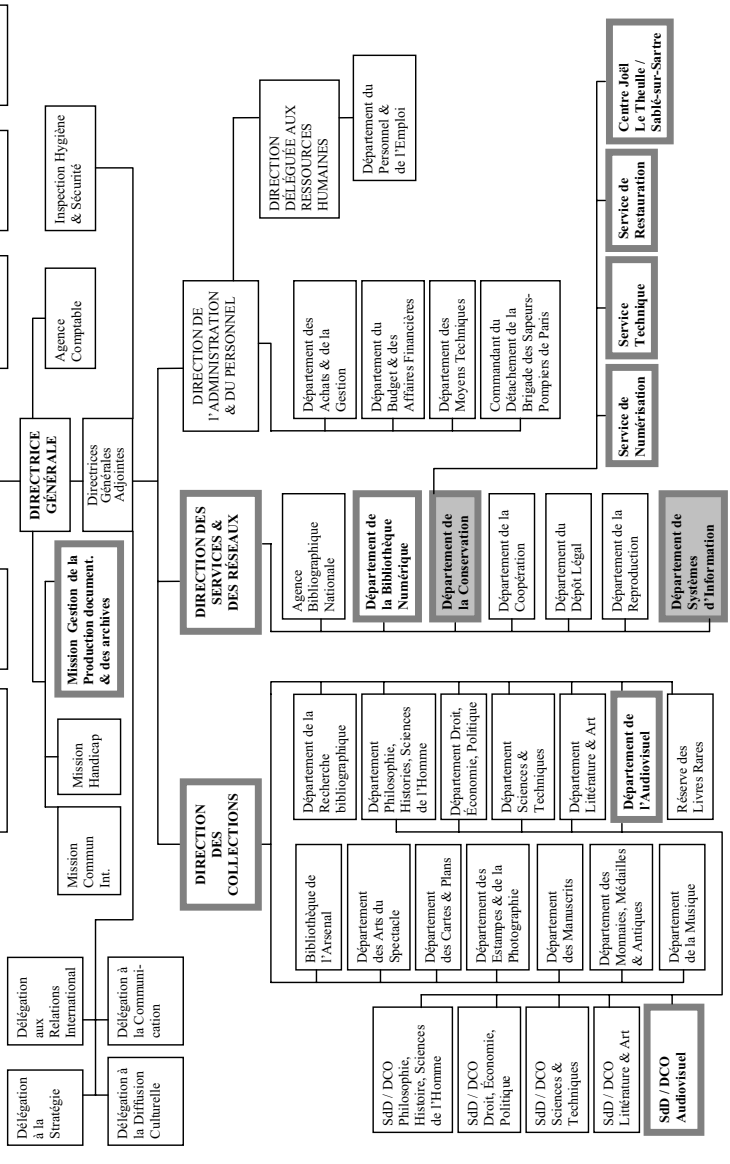
Det Kongelige Bibliotek / National Library of Denmark – July 2005



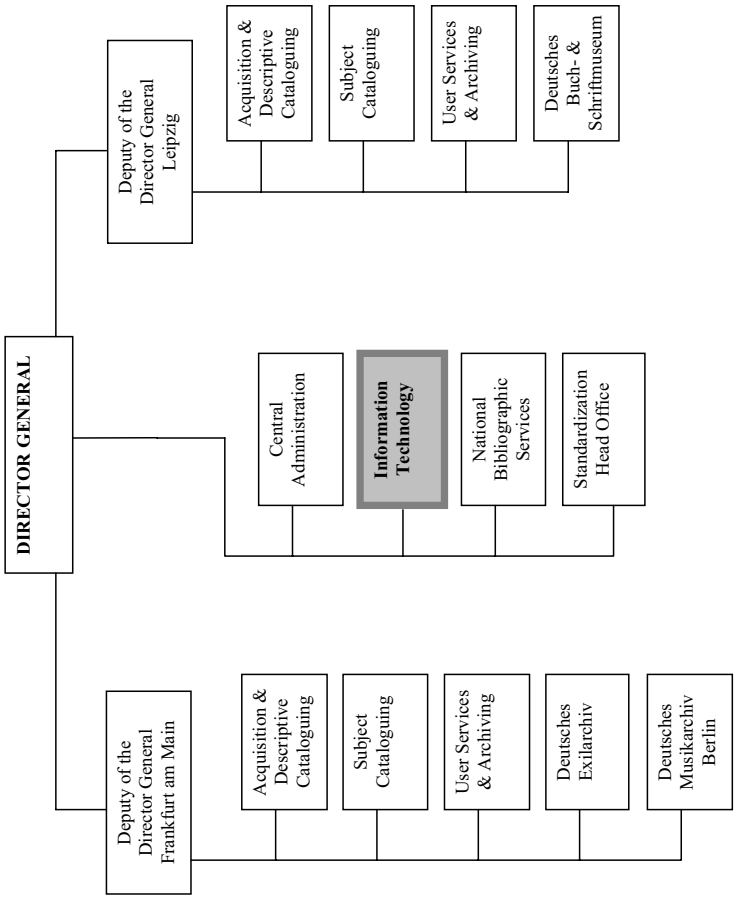
Statsbiblioteket / National Library of Denmark – July 2005



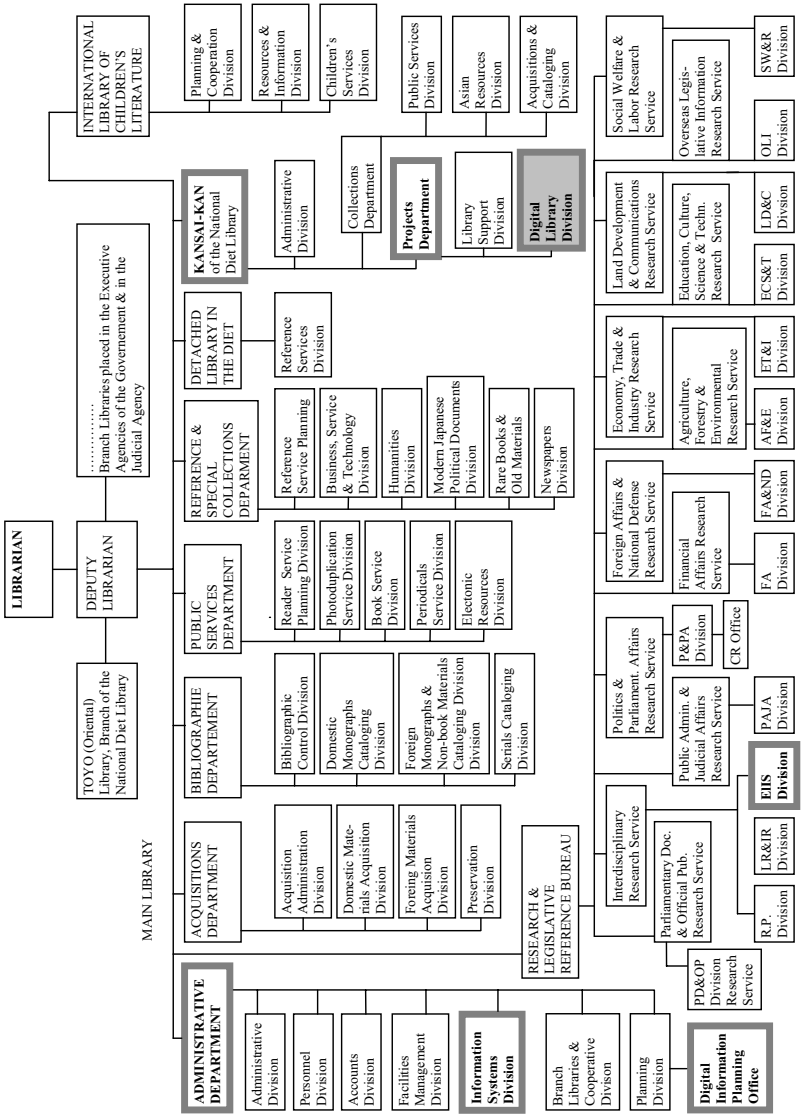
Bibliothèque nationale de France / National Library of France – May 2005

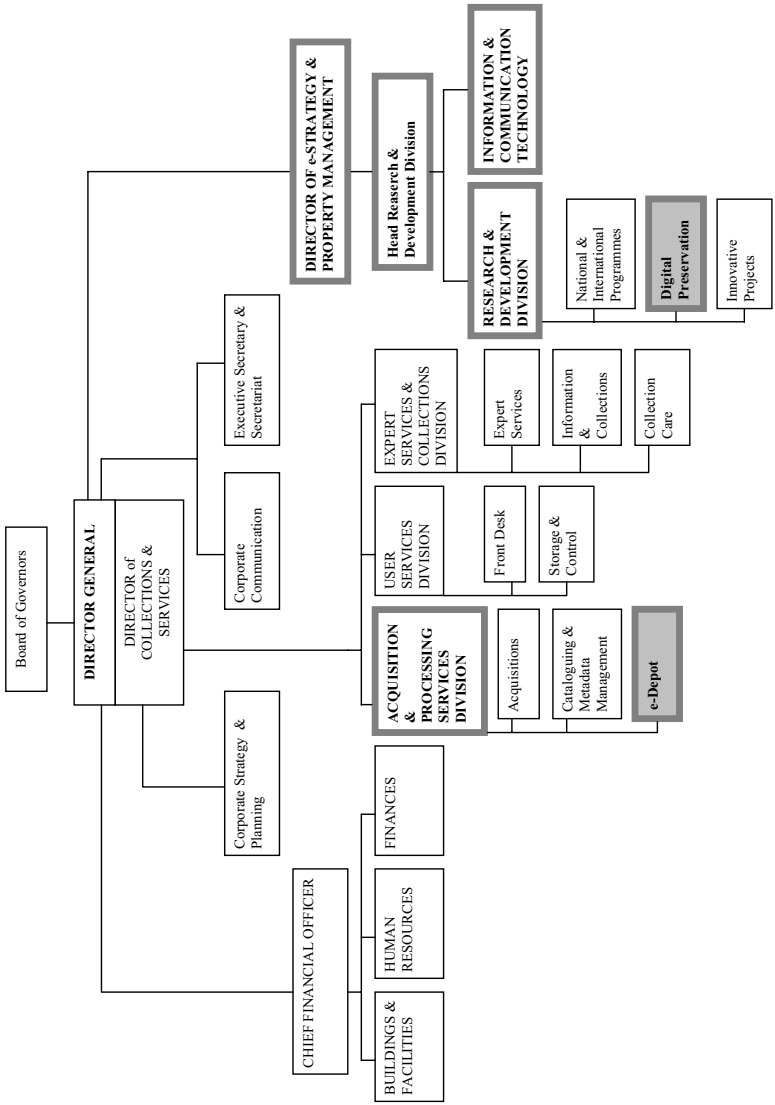


Die Deutsche Bibliothek / National Library of Germany – May 2005

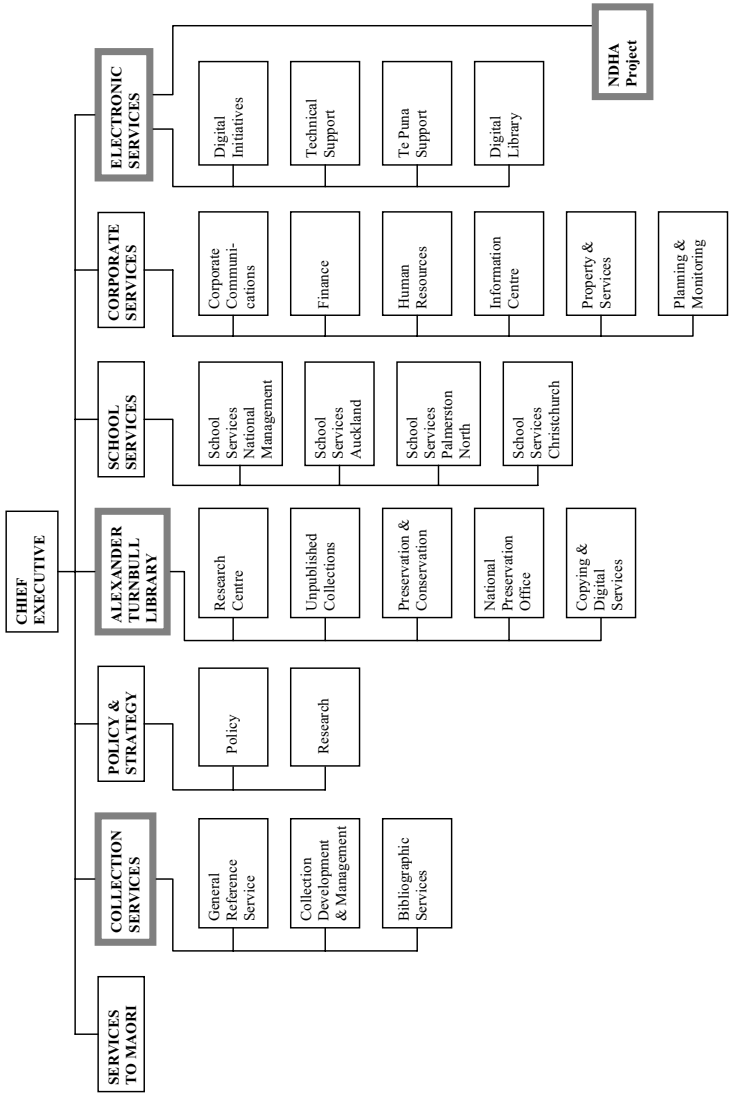


National Diet Library / National Library of Japan – July 2005

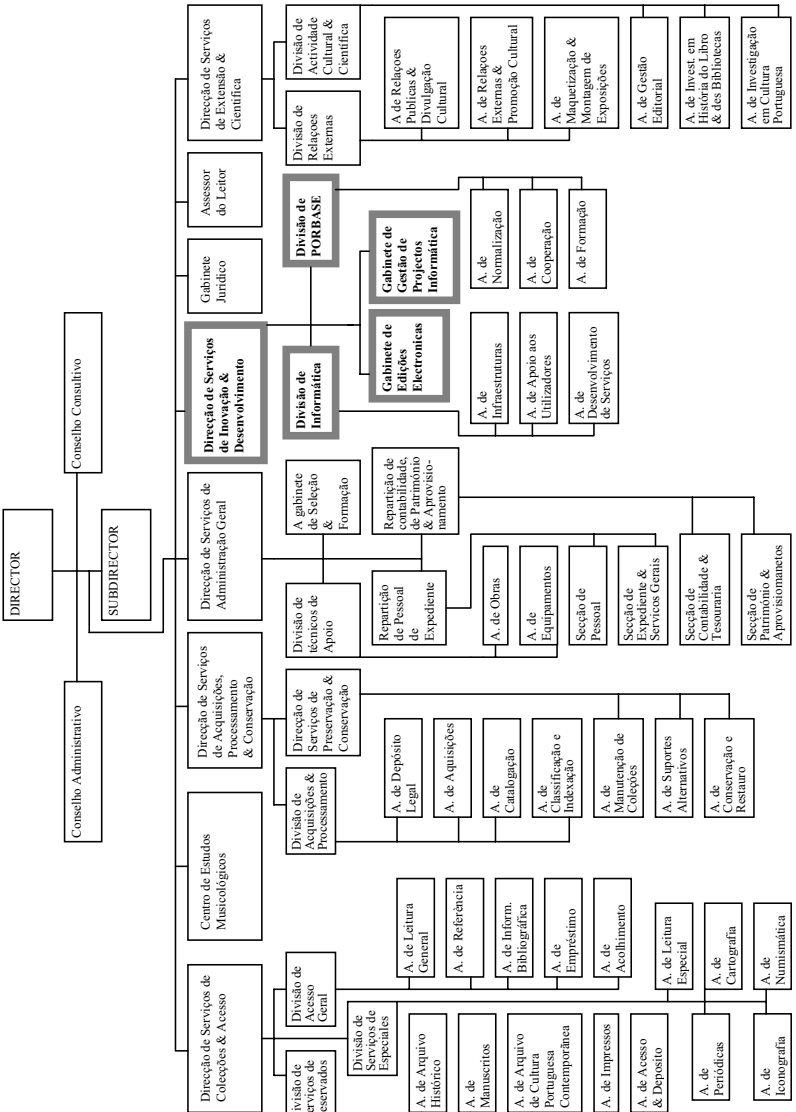




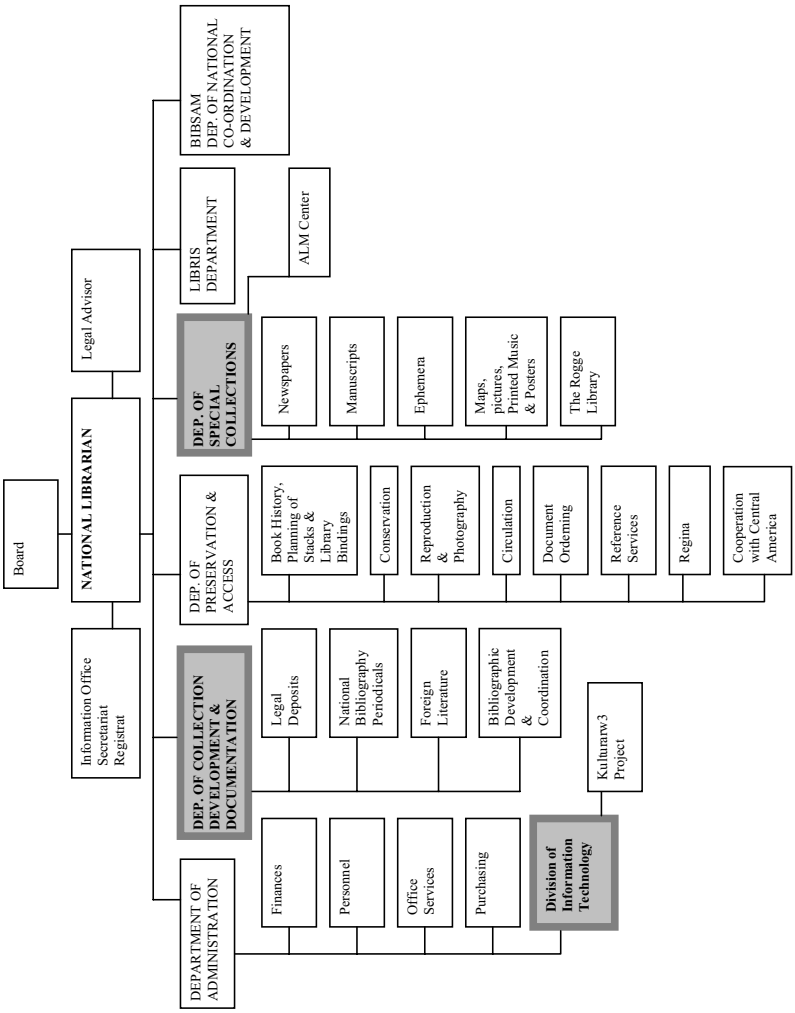
National Library of New Zealand Te Puna Mātauranga a Aotearoa / National Library of New Zealand – July 2005



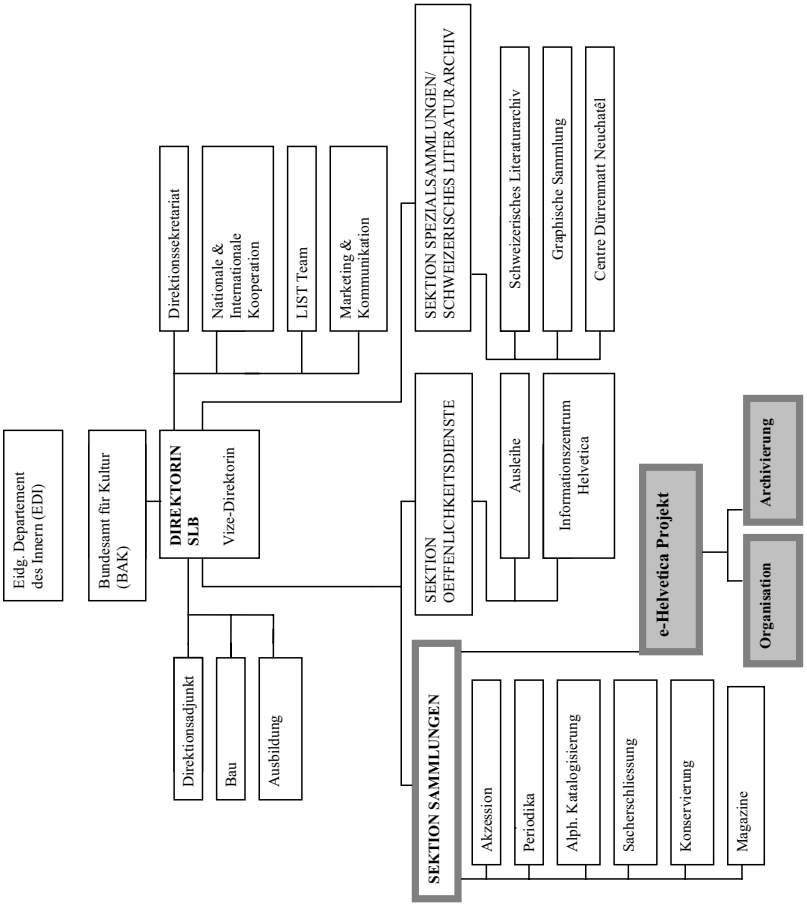
Biblioteca Nacional / National Library of Portugal – May 2005



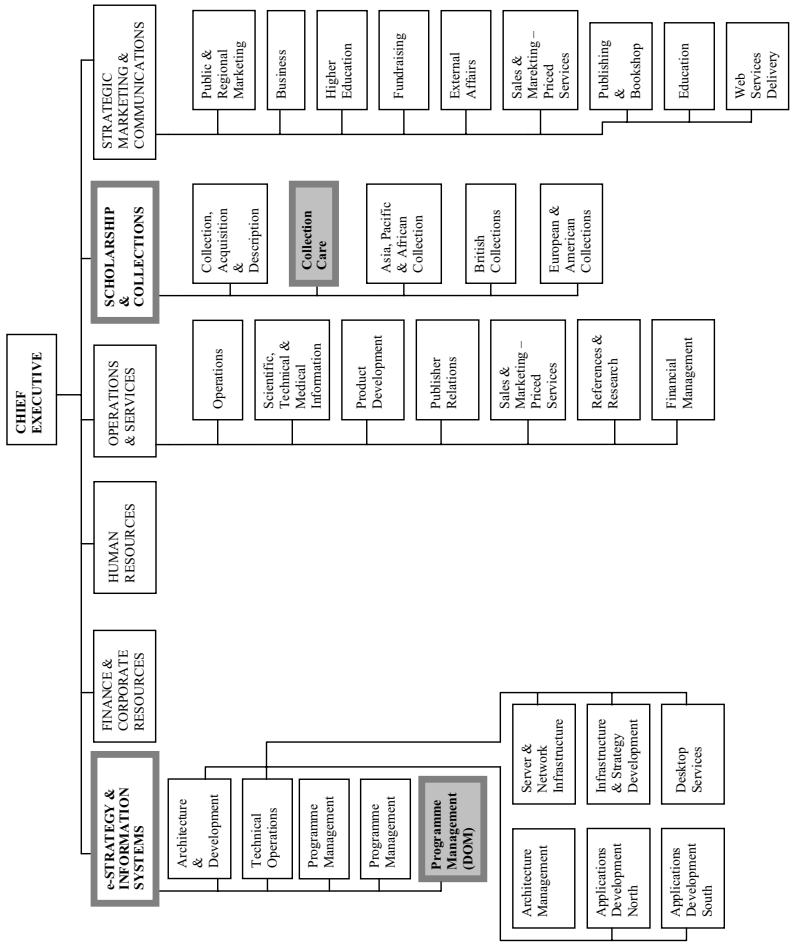
Kungliga Biblioteket / National Library of Sweden – February 2005



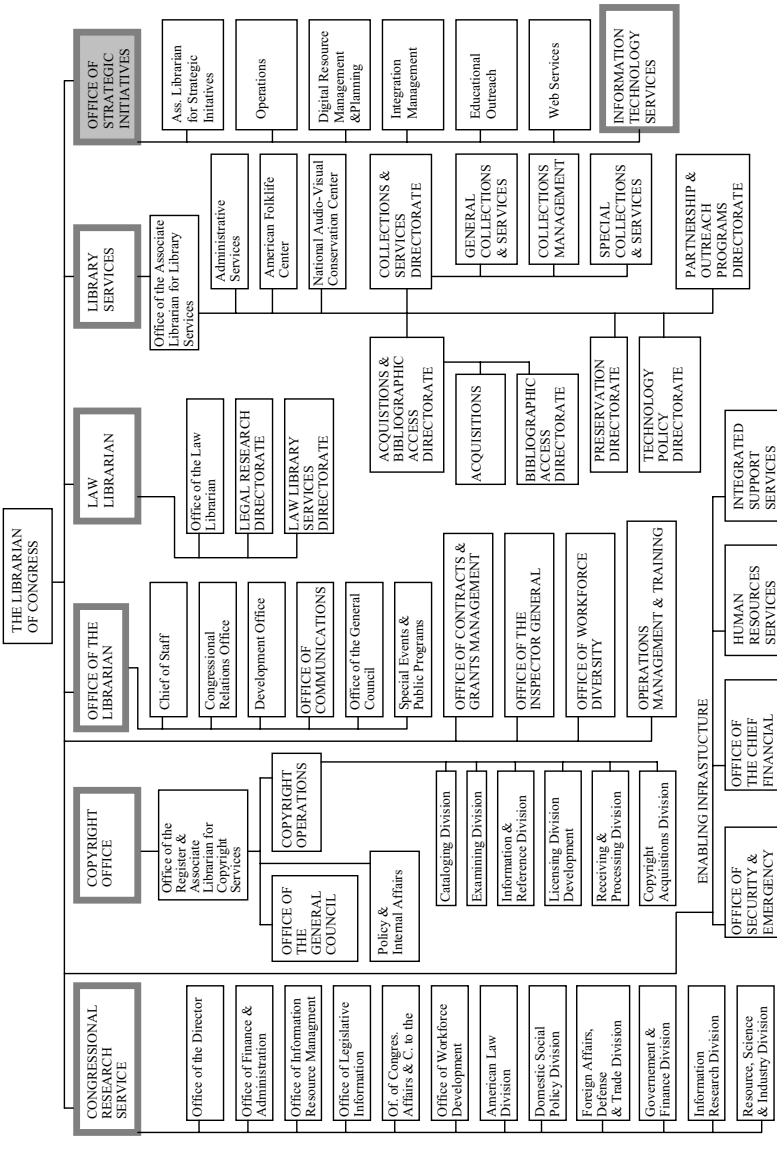
Schweizerische Landesbibliothek / National Library of Switzerland – July 2005



The British Library / National Library of the United Kingdom – June 2005



Library of Congress / National Library of the USA – September 2005



II. Overviews national libraries

3. Cooperation chart

Cooperation chart

	Australia	Austria	Canada	China	Denmark	France	Germany
Australia			IIPC/PADI		IIPC	IIPC	ICABS/ PADI
Austria				informal	PLANETS	Minerva EU	ReUse/ D-A-CH/ Minerva EU
Canada	IIPC/PADI				IIPC	IIPC	PADI
China		informal					
Denmark	IIPC	PLANETS	IIPC			IIPC	
France	IIPC	Minerva EU	IIPC		IIPC		FP7/ Minerva EU
Germany	ICABS/ PADI	ReUse/ D-A-CH/ Minerva EU	PADI			FP7/ Minerva EU	
Japan							
Netherlands	ICABS/ PADI/ PREMIS	PLANETS	PADI		PLANETS	FP7/ RLG-NARA Taskforce TDR	ICABS/FP7/ PADI/ Dias-UG
New Zealand	PREMIS/ IIPC		IIPC		IIPC	IIPC	
Portugal	ICABS						ICABS
Sweden	IIPC/PADI		IIPC/PADI		IIPC	IIPC	PADI
Switzerland	PADI	D-A-CH	PADI				D-A-CH/ PADI
UK	IIPC/ICABS/ PADI/DPC/ PREMIS/ informal	PLANETS	IIPC/PADI	IDP	IIPC/ PLANETS	IIPC/FP7	ICABS/FP7/ PADI
USA	IIPC/DPC/ ICABS/ PREMIS/PADI		IIPC/PADI		IIPC	FP7/RLG-NARA Taskforce TDR	ICABS/PADI

Japan	Netherlands	New Zealand	Portugal	Sweden	Switzerland	UK	USA
	ICABS/ PADI/ PREMIS	PREMIS/ IIPC	ICABS	IIPC/PADI	PADI	IIPC/ICABS/ PADI/DPC/ PREMIS/ informal	IIPC/DPC/ ICABS/ PREMIS/PADI
	PLANETS				D-A-CH	PLANETS	
	PADI	IIPC		IIPC/PADI	PADI	IIPC/PADI	IIPC/PADI
						IDP	
	PLANETS	IIPC		IIPC		IIPC/ PLANETS	IIPC
	FP7/ RLG-NARA Taskforce TDR	IIPC		IIPC		IIPC/FP7	FP7/RLG-NARA Taskforce TDR
	ICABS/FP7/ PADI/ Dias-UG		ICABS	PADI	D-A-CH/ PADI	ICABS/FP7/ PADI	ICABS/PADI
	draft cooperation agreement						informal
draft cooperation agreement		PREMIS	ICABS	PADI	PADI/ informal	ICABS/ FP7/ PLANETS/ PREMIS/PADI	ICABS/RLG-NARA Task-force/PADI/ PREMIS
	PREMIS			IIPC		IIPC/ PREMIS	IIPC/ PREMIS
	ICABS					ICABS	ICABS
	PADI	IIPC			PADI	IIPC/PADI	IIPC/PADI
	PADI/ informal			PADI		PADI	PADI
	ICABS/FP7/ PLANETS/ PREMIS/PADI	IIPC/ PREMIS	ICABS	IIPC/PADI	PADI		IIPC/ICABS/ PREMIS/DCP/ NDIIPP/PADI
informal	ICABS/RLG-NARA Task-force/PADI/ PREMIS	IIPC/ PREMIS	ICABS	IIPC/PADI	PADI	IIPC/ICABS/ PREMIS/DCP/ NDIIPP/PADI	

Appendices

References

Beagrie, Neil. 2003. *National Digital Preservation Initiatives: An overview of the 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 and Library of Congress. CLIR Series on Strategies and Tools for the Digital Library 116. See: http://www.clir.org/pubs/reports/pub116/contents.html

Consultative Committee for Space Data Systems. 2002. *Reference Model for an Open Archival Information System. Blue Book.* http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf

Hodge, Gail and Evelyn Frangakis. 2004. *Digital Preservation and Permanent Access to Scientific Information: The State of the Practice*. Sponsored by The International Council for Scientific and Technical Information (ICSTI) and CENDI, U.S. Federal Information Managers Group. Oak Ridge/Tennessee, Information International Associates, Inc./CENDI Digital Preservation Task Group/National Agricultural Library. CENDI Publication 2004-3. See: http://cendi.dtic.mil/publications/04-dig_preserv.htm

OCLC/RLG PREMIS Working Group. 2004. Implementing Preservation Repositories For Digital Materials: Current Practice And Emerging Trends In The Cultural Heritage Community. Report by the joint OCLC/RLG Working Group Preservation Metadata: Implementation Strategies (PREMIS). Dublin, O.: OCLC Online Computer Library Center Inc. See: http://www.oclc.org/research/projects/pmwg/surveyreport.pdf

Recommended reading

On digital preservation in general:

Literature

Charter on the Preservation of Digital Heritage. Adopted at the 32nd General Conference of UNESCO on 17 October. See: http://portal.unesco.org/en/ev.php-URL_ID=17721&URL_DO=DO_TOPIC&URL_SECTION=201.html

Lavoie, Brian and Lorcan Dempsey. 2004. 'Thirteen way's of Looking at … Digital Preservation', in: *D-Lib Magazine*, vol. 10, nr. 7/8, (july/august 2004). See: http://www.dlib.org/dlib/july04/lavoie/07lavoie.html

Webb, Colin. 2003. *Guidelines for the preservation of digital heritage*. Prepared by the National Library of Australia. (Information Society Division UNESCO, Memory of the world). See:
http://unesdoc.unesco.org/images/0013/001300/130071e.pdf

Webb, Colin. 2005. *Report to ICABS on guidance for digital preservation. Report on a survey of sources.* National Library of Australia. See:
http://www.nla.gov.au/nla/staffpaper/2005/documents/webb1.pdf [110]

Websites

- The Cornell Digital Preservation Management Tutorial: See:
http://www.library.cornell.edu/iris/tutorial/dpm/

- The PADI website: See: http://www.nla.gov.au/padi/

- The Digital Preservation Coalition website: See:
http://www.dpconline.org/graphics/index.html

- The nestor website: See:
http://www.langzeitarchivierung.de/index.php?newlang=eng

Actual information
For current information on the state of the art of digital preservation see also the websites of the national libraries. Web addresses to be found in the Overview Section of this Survey. See also the Annual Country Reports on the CDNL Website: http://consorcio.bn.br/cdnl/2005/HTML/countryreports.htm

For actual information in general see also: *D-Lib Magazine* (see: http://www.dlib.org/) and *RLG DigiNews* (see: http://www.rlg.org/en/page.php?Page_ID=12081), or various international mailing lists on digital preservation (for instance: the DIGLIB Research list the DigiCult List, the Digital-Preservation List, the DPC Discussion List, the EPIC list, the IFLA-L List, JISC mail, PADIforum List, the RLG List).

More digital preservation overviews:

Hedstrom, Margaret. 1998. 'The role of national initiatives in digital preservation.' in: *RLG Digi News* 1998. See:
http://www.rlg.org/legacy/preserv/diginews/diginews2-5.html#feature2

[110] See also the Summary in the Appendices.

The state of digital preservation: an international perspective. Conference proceedings. April 24-25, 2002. Washington: Council on Library and Information Resources. See: http://www.clir.org/pubs/repors/pub107/pub107.pdf

Lunghi, Mario. 2004. *Enabling Persistent and Sustainable Digital Cultural* Heritage in Europe. European Actions for Sustainability and Preservation. 1. The Netherlands questionnaire responses summary. EU Conference 15-16 september 2004. See: http://www.iccu.sbn.it/PDF/AIA-PART_1.pdf

Marin, Elizabeth. 2001. *Management of networked electronic publications. A table of status in various countries*. National Library of Canada. See: http://www.collectionscanada.ca/obj/r7/f2/r7-100-e.pdf

Mackenzie Owen, J.S. and J. van der Walle. 1996. *Deposit collections of electronic publications*. Luxembourg: European Commission, 1996. See: http://cf.hum.uva.nl/bai/home/jmackenzie/pubs/ELDEP.zip

Norris, Michael. 2005. *An international survey of archival, preservation and related clauses in copyright law.* See: http://www.lboro.ac.uk/departments/ls/disresearch/CLDP/DOCUMENTS/Copyright%20survey.doc

List of acronyms

A

AACRII	=	Anglo-American Catalogue Rules
AAI	=	Authentication and Authorisation Infrastructure
ADAM	=	*Archivierung elektronischer digitaler Daten und Akten – data management*
ADIC	=	Advanced Digital Information Cooperation
aDORe	=	a modular, standards-based Digital Object Repository
AHDS	=	Arts and Humanities Data Service
AIHT	=	Archive Ingest and Handling Test
AI	=	Artificial Intelligence
AI	=	Adobe Illustrator (File Format)
AIP	=	Archival Information Package
AIT	=	Advanced Intelligent Tape
ALM	=	Archives, Libraries, Museums
ALPSP	=	Association of Learned and Professional Society Publishers
AMD	=	Audio MetaData
APAC	=	Australian Partnership for Advanced Computing
API	=	Application Programming Interface
APSR	=	Australian Partnership for Sustainable Repositories
ARC	=	ARChiving
ARROW	=	Australian Research Repositories Online to the World
ASCII	=	American Standard Code for Information Interchange
ASM	=	Application Storage Management
ASTOR	=	Archival STOR
ATL	=	Alexander Turnbull Library
AV	=	Audio Visual

B

BBC	=	British Broadcasting Cooperation
BL	=	British Library
BN	=	*Bibliotheca Nacional*
BND	=	*Biblioteca Nacional Digital*
BnF	=	*Bibliothèque nationale de France*
BWF	=	Broadcast Wave Format

C

CAD	=	Computer Aided Design
CALIS	=	China Academic Library and Information System
CC	=	Creative Commons
CCDCS	=	Council of the Consultative Committee for Space Data
CCTV	=	China Central TeleVision

CD	=	Compact Disk
CD-ROM	=	Compact Disk – Read Only Memory
CDC Zantaz	=	*Caisse des Dépôts et Consignations Zantaz*
CDI	=	Committee on Digital Issues
CDLC	=	China Digital Library Corporation
CDNL	=	Conference of Directors of National Libraries
CDSL	=	Computer networked and Distributed Systems Laboratory
CEDARS	=	Curl Exemplars in Digital ARchiveS
CENDI	=	Commerce, Energy, NASA, Defense Information Managers Group
CENL	=	Conference of European National Librarians
CIDL	=	Canadian Initiative on Digital Libraries
CLIR	=	Council on Library and Information Resources
CMS	=	Content Management System
CNC	=	*Centre National de la Cinématographie*
CNES	=	*Centre National d'Etudes Spatiales*
CNKI	=	China National Knowledge Infrastructure
CNRI	=	Corporation for National Research Initiatives
CRS	=	Congressional Research Services
CSTB	=	Computer Science and Telecommunications Board
CURL	=	Consortium of University Research Libraries

D

DAAT	=	Digital Asset Assessment Tool
D-A-CH	=	*Deutschland – Austria – SCHweiz*
DAMS	=	Digital Asset Management Systems
DANS	=	Data Archiving and Networked Services
DARE	=	Digital Academic REpositories
DB2	=	Data Base2
DC	=	Dublin Core
DC	=	District of Columbia (Washington)
DCC	=	Digital Curation Centre
DCM	=	Digital Collections Manager
DDA	=	Department of Documentation and Digitisation
DDB	=	*Die Deutsche Bibliothek*
DEFF	=	*Danmarks Elektroniske Fag- of Forskningsbibliothek*
DEOG	=	Digital Executive Oversight Group
DHC	=	Documentary Heritage Collection
DIAS	=	Digital Information Archival System
DIDL	=	Digital Item Declaration Language
DIGARCH	=	DIGital ARCHiving and Long-Term Preservation
DIMAC	=	*Depósito Digital de Publicações da Administração Centra*
DIMIC	=	*Depósito Digital de Publicações Impressas Comerciais*

DINF = *Divisão de Informática*
DIP = Dissemination Information Package
DiTeD = *Depósito de Teses e Dissertações*
DiVA = *Digitala Vetenskapliga Arkivet*
DKB = *Det Kongelige Bibliotek*
DMA = *Deutsches Musikarchiv*
DOM = Digital Object Management
DOS = Disk Operating System
DOSS = Digital Object Storage System
DPC = Digital Preservation Coalition
DSEP = Deposit System for Electronic Publications
DSID = *Direcção de Serviços de Inovação e Desenvolvimento*
DSTC = Distributed System Technology Centre
DSpace = Digital Space
DSTC = Distributed Systems Technology
DRM = Digital Rights Management
DTD = Document Type Definition
DVD = Digital Versatile Disk
DVD-ROM = Digital Versatile Disk – Read Only Memory

E

EAD = Encoded Archival Description
EBC = Educational Broadcasting Cooperation
ECHO = European Cultural Heritage Online
ECPA = European Commission on Preservation and Access
eIS = E-strategy and Information Systems (Directorate)
ELIS = Electronic Library Information System
ENSSIB = *École Nationale Supérieure des Sciences de lInformation et des Bibliothèques*
EPICUR = Enhancement of Persistent Identifier services – Comprehensive method for Unequivocal Resource identification
EPS = Electronic Publishing Service Ltd (UK)
EPS = Encapsulated Post Script (file format)
ERPANET = Electronic Resource Preservation and Access NETwork
eThos = e-THesis National Service Pilot
EUBAM = *Portal zu EUropäischen Angelegenheiten für Bibliotheken, Archive, Museen und Denkmalpflege*

F

FIFA = *Fédération Internationale de Football Association*
FP6 = 6th Framework Programme
FP7 = 7th Framework Programme

FEDER	=	*Fonds Européen de Développement Régional*
FEDORA	=	Flexible Extensible Digital Object Repository Architecture
FRBR	=	Functional Requirements for Bibliographic Records
FTP	=	File Transfer Protocol

G

GAIE	=	Georgia Association of International Educators
GAP	=	*Gabinete de Gestão de Projectos*
GEDE	=	*Gabinete De Edições Electrónicas*
GIF	=	Graphics Interchange Format
GIS	=	Geographic Information System
GLIN	=	Global Legal Information Network
GUI	=	Graphical User Interface
GWDG	=	*Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen*

H

HATII	=	Humanities Advanced Technologies and Information Institute
HDD	=	Hard Disk Drive
HDTV	=	High Definition TeleVision
HSM	=	Hierarchical Storage Management (System)
HTML	=	Hypertext Markup Language
HTT Track	=	Hyper Text Template Track
HTTP	=	Hyper Text Transfer Protocol

I

IBM	=	International Business Machines
ICABS	=	IFLA CDNL Alliance for Bibliographic Standards
ICP	=	Internet Content Provider
ICN	=	*Instituut Collectie Nederland*
IDP	=	International Dunhuang Project
IFLA	=	International Federation of Library Associations and Instiutions
INA	=	*Institut National de l'Audiovisuel*
INESC-ID	=	*Instituto de Engenharia de Sistemas e Computadores-Inovação Desenvolvimento*
IIPC	=	International Internet Preservation Consortium
IPR	=	Intellectual Property Rights
ISSN	=	International Standard Serial Number
ISO	=	International Organization for Standardization
IT	=	Information Technology
ITS	=	Information Technology Services

J

JCLD	=	Joint Committee on Legal Deposit
JHOVE	=	JSTOR Harvard Object Validation Environment
JISC	=	Joint Information Systems Committee
JPEG	=	Joint Photographic Experts Group

K

KB	=	*Koninklijke Bibliotheek* (National Library of the Netherlands)
KB	=	*Kungliga Biblioteket* (National Library of Sweden)
KNAW	=	*Koninklijke Nederlandse Academie van Wetenschappen*
kopal	=	*Kooperativer Aufbau eines Langzeitarchivs digitaler Informationen*
KUB	=	*Konferenz der Universitätsbibliotheken der Schweiz*

L

LAC	=	Library and Archives of Canada
LAN	=	Local Area Network
LIFE	=	Lifecycle InFormation for E-literature
LINUX	=	LINus Torvald's UniX
LLC	=	Limited Liability Company
LLP	=	Limited Liability Partnership
LMER	=	*LangzeitarchivierungsMetadaten für Elektronische Ressourcen*
LoC	=	Library of Congress
LOCKSS	=	Lots of Copies Keep Stuff Safe
LP	=	Long-Playing phonograph record
LTO	=	Linear Tape Open
Lustre	=	Linux and Clusters

M

MAB2	=	*Maschinelle Austauschformat für Bibliotheken*
MADS	=	Metadata Authority Description Schema
MAENED	=	Multimedia Access across Enterprises, Networks And Domains
MARC	=	MAchine Readable Cataloguing
METS	=	Metadata Encoding and Transmission Standard
MIME	=	Multipurpose Internet Mail Extension
MINERVA	=	MINistrial NetwoRk for Valorising Activities in digitisation (Europe)
MINERVA	=	Mapping the INternet Electronic Resources Virtual Archive (USA)
MIT	=	Massachusetts Institute of Technology (Data Center)
MIX	=	Metadata for Images in XML

MLA	=	Museums, Libraries and Archives
MODIS	=	Moderate Resolution Imaging Spectroradiometer
MODS	=	Metadata Object Description Schema
MPEG	=	Moving Pictures Experts Group

N

NARA	=	National Archives and Records Administration
NAS	=	Network Attached Storage
NBN	=	National Bibliography Number
NDL	=	National Diet Library
NDL	=	National Digital Library (Program)
NDL-OPAC	=	National Diet Library-Online Public Access Catalog
NDHA	=	National Digital Heritage Archive Programme
NDIIPP	=	National Digital Information Infrastructure and Preservation Program
NDNP	=	National Digital Newspaper Program
NEDLIB	=	Networked European Deposit LIBrary
nestor	=	Network of Expertise in long-term STOrage of digital Resources
NISO	=	National Information Standards Organisation
NIWI	=	Netherlands Institute for Scientific Information
NLA	=	National Library of Australia
NLC	=	National Library of China
NLNZ	=	National Library of New Zealand
NLR	=	*Nationaal Lucht- en Ruimtevaart laboratorium*
NSF	=	National Science Foundation
NSTL	=	National Science and Technology Library
NY	=	New York
NYU	=	New York University

O

OAI-PMH	=	Open Archives Initiative Protocol for Metadata Harvesting
OAIS	=	Open Archival Information System
OCLC	=	Online Computer Library Center
OCW	=	*Onderwijs, Cultuur en Wetenschap*
ODBC	=	Open DataBase Connectivity
ÖNB	=	*Österreichische Nationalbibliothek*
OPAC	=	Open Public Access Catalog
OSI	=	Office of Strategic Initatives

P

PADI	=	Preserving Access to Digital Information

PANDORA = Preserving and Accessing Networked DOcumentary Resources of Australia
PANDAS = PANDORA Digital Archiving System
PANIC = Preservation webservices Architecture for Newmedia and Interactive Collections
PBS = Public Broadcasting Services
PDF = Portable Document Format
PDF/A = Portable Document Format/Archive
PI = Persistent Identifiers
PIDDAC = *Programa de Investimentos e Despesas de Desenvolvimento da Administração Central*
PIN = *Perennisation des Information Numeriques*
PLANETS = Preservation and Long-term Access through NETworked Services
PNG = Portable Network Graphics
POC = *Programa Operacional da Cultura*
POP = *Pilotprojekt Übernahme und Archivierung von Online-VerlagsPublikationen*
POSI = *Programa Operacional para a Sociedade da Informação*
PPT = PowerPoinT
PREDICA = *Centro de Excelência em Preservação e Digitalização em Contextos Avançados*
PREMIS = Preservation Metadata: Implementation Strategies
PRESERV = Preservation E-print SERVices
PRESTO
SPACE = Preservation towards storage and access. Standardised Practices for Audiovisual Contents in Europe
PROTEAN = Preservation Over Time by Electonic Archiving and Networking
PS = Post Script

R

R&D = Research & Development
RAD = Record of Archival Description
RFI = Request for information
RLG = Research Libraries Group
RTF = Rich Text Format

S

SAN = Storage Area Network
SB = *Statsbiblioteket*
SC = Scolarship & Collections
SFA = Swiss Federal Archives

SLB	=	*Schweizerische Landesbibliothek*
SHERPA	=	Securing a Hybrid Environment for Research, Preservation and Access
SIP	=	Submission Information Package
SML	=	Standard Markup Language or Simple Markup Language
SMTP	=	Simple Mail Transfer Protocol
SPAR	=	*Système pour Preserver, Archiver et Répartir l'accès aux données*
SQL	=	Structured Query Language
SSG	=	*Sondersammelgebiete*
STM	=	Science, Technology and Medicine
SUB	=	*Staats- und Universitätsbibliothek Göttingen*
SURF	=	*Stichting Universitaire RekenFaciliteiten*
SVEP	=	*Samordning av den Svenska Högskolans Elektroniska Publicering*

T

TAPE	=	Training for Audiovisual Preservation in Europe
TV	=	TeleVision
TIFF	=	Tagged Image File Format
TXT	=	TeXT

U

UBCIM	=	Universal Bibliographic Control and International MARC
UDT	=	Universal Dataflow and Telecommunications
UK	=	United Kingdom
UKOLN	=	United Kingdom Office for Library and Information Networking
ULCC	=	University of London Computer Centre
UMI	=	University Microfilms International
UN	=	United Nations
UNESCO	=	United Nations Educational, Scientific, and Cultural Organization
UNIMARC	=	UNIversal MARC
UNIX	=	UNiplexed Information and Computing System
URL	=	Uniform Resource Locator
URN	=	Uniform Resource Name
US	=	United States
USA	=	United States of America
UVC	=	Universal Virtual Computer

V

VMD	=	Video Metadata

W

WARC	=	Web ARChive file format
WARP	=	Web ARchiving Project
WAV	=	WAVeform Sound
WTO	=	World Trade Organization

X

XML	=	Extensible Markup Language
XWS	=	XML and WebServers

Report of the National Library of Australia on guidance for digital preservation: a summary

In 2004−2005, ICABS initiated a second survey on digital preservation, alongside the survey of KB, which was carried out by the National Library of Australia (NLA). This survey focused on the availability of suitable guidance documents for preserving materials. It also resulted in a report entitled, *Report to ICABS on guidance for digital preservation. Report on a survey of Sources. National Library of Australia* (July 2005). Both reports of KB and NLA had been sent for review to the ICABS Board by the end of July and were jointly presented during the IFLA ICABS Session on digital preservation during the World Library and Information Congress in Oslo in August 2005.

The NLA survey was carried out in 2004 by Jennifer Hodgeman, and was updated and analysed in mid-2005 by Colin Webb. Most of the resources found, were identified through the NLA's PADI subject gateway. As digital preservation is a rapidly evolving field, the conclusion was that the body of guidance literature is by no means statistic. Even during the months between the first and second survey, many new guidance resources appeared. Therefore the report aims to be a reasonably well-informed impression of the state of guidance, based on a subjective analysis of a snapshot of available resources taken in mid 2005. The NLA/ICABS survey will be made available online. A summary of this second ICABS survey is added here.

Although at present few resources offer standardising guidance, the report provides an extensive, though by no means exhaustive, overview of current existing resources with a guidance potential. Six main types of resources are distinguished:

- *Primary sources* − resources written specifically to provide broad guidance (not just describing a specific technique), and with apparent broad community endorsement. Only one standard (for OAIS), a very few guidelines, and some training courses seemed to fit into this category.

- *Secondary sources* − resources written to provide guidance, but considered likely to be dated or without apparent community endorsement.

- *Local sources* − resources written to provide endorsed guidance for specific projects, programmes or communities. In many cases, these situation-specific sources may provide useful guidance for a wider audience, but those intending to use them would need to carefully consider their applicability to their own circumstances.

- *Embedded sources* − resources containing some useful preservation guidance within other information, such as reports on projects or articles debating issues, proposing approaches, or critically analysing experience. The great majority of literature on digital preservation, and most relevant conferences and seminars, may best be approached in this light: as sources of some potentially useful guidance that require a critical effort to extract the guidance and decide what is useful.

- *Standards and tools* − resources describing specific approaches or facilities. Except in a very few cases, most resources of this kind appeared to be marginal as guidance documents. Many relevant standards exist, especially in the area of file formats; there are also many partly-developed tools such as format registries and metadata extraction software, which may play a key role in digital preservation. In terms of guidance, however, these resources are of themselves relatively less useful than guides to where, when and how best to use them in the context of managing collections.

- *Current awareness sources* − resources that provide a convenient way of keeping up to date on recent developments and locating leads that may be worth following. A number of subject gateways, discussion lists, project 'knowledge bases', and conferences seem to fulfil this kind of role.

The NLA Report assumes a definition of digital preservation that covers the processes required to maintain access. As such, digital preservation should not be seen as a single process, but as a cluster of many contributing processes covering the life cycle of information resources and many aspects of their management. The report contains two overviews. The first overview classifies guidance sources in eight digital preservation process areas:
- taking responsibility for archiving and preservation;
- managing digital preservation programmes;
- creating preservable digital content and working with producers;
- deciding what to preserve;
- transferring digital objects to an archive; naming, describing and controlling;
- digital objects;
- managing legal issues that impact on preservation;
- protecting and storing data of digital information resources;
- and maintaining the means of representing archived objects for access.

Since libraries may be interested in a wide range of digital resources, the second overview is subdivided in 13 different kinds of digital materials. These materials range from online web-based materials to digital art, and from scientific data sets to electronic games: general – covering a wide range of materials;

- archival records including e-mail;
- audio-visual materials (audio & moving image);
- digital art and e-literature;
- digital games, virtual reality and software;
- digital 'manuscripts';
- images;
- miscellaneous digital materials;
- online information and web resources;
- physical format digital materials;
- scientific, statistical and research data;
- spatial, geospatial and similar data sets;
- current awareness resources.

Within each category a summary of core issues on which guidance may be needed is given. The criticality of being able to find guidance in each category is considered, along with the availability of guidance sources. Areas of particular strength and weakness are noted. Then an overview of surveyed guidance resources is given, with a short description of the contents, the URL, and some short notes on the source and scope of each resource. The resources chosen for inclusion in the report are considered to be either representative of a wider range of relevant resources, or to be considered to be the most relevant. Most of the resources were identified through the NLA's PADI Subject Gateway (Preserving Access to Digital Information, available at http://www.nla.gov/au/padi) and are therefore limited to guidance resources in English only.

The cautious conclusion of the survey is that since most digital preservation programmes are operating in an environment where much still has to be decided or agreed upon, it is hardly surprising that many people find it difficult to locate guidance that meets their needs. However, there are some 'wellsprings' of guidance that have produced a steady stream of help over the past decade, and to whom it is reasonable to look for further guidance in future. These include the following categories:

- Projects and programmes that investigate existing best practices and standards for their own needs, and then make their findings publicly available (such as literature reviews that occur in the set-up stages of many digital preservation programmes);

- Projects and programmes that undertake original research for their own needs, and then make their results available (such as a number of research projects that have been funded to explore specific processes);

- Collaborative groups that set standards or benchmarks that all partners agree on (such as standards-setting bodies, collaborative digitisation projects, digital preservation alliances);

- Agencies set up to provide training or guidance to a specific industry or sector (such as bodies that organise training programmes, seminars, and reports for the Higher Education sector); and

- Agencies with a mission to enable the preservation of digital heritage, wherever it is found (such as UNESCO, IFLA).

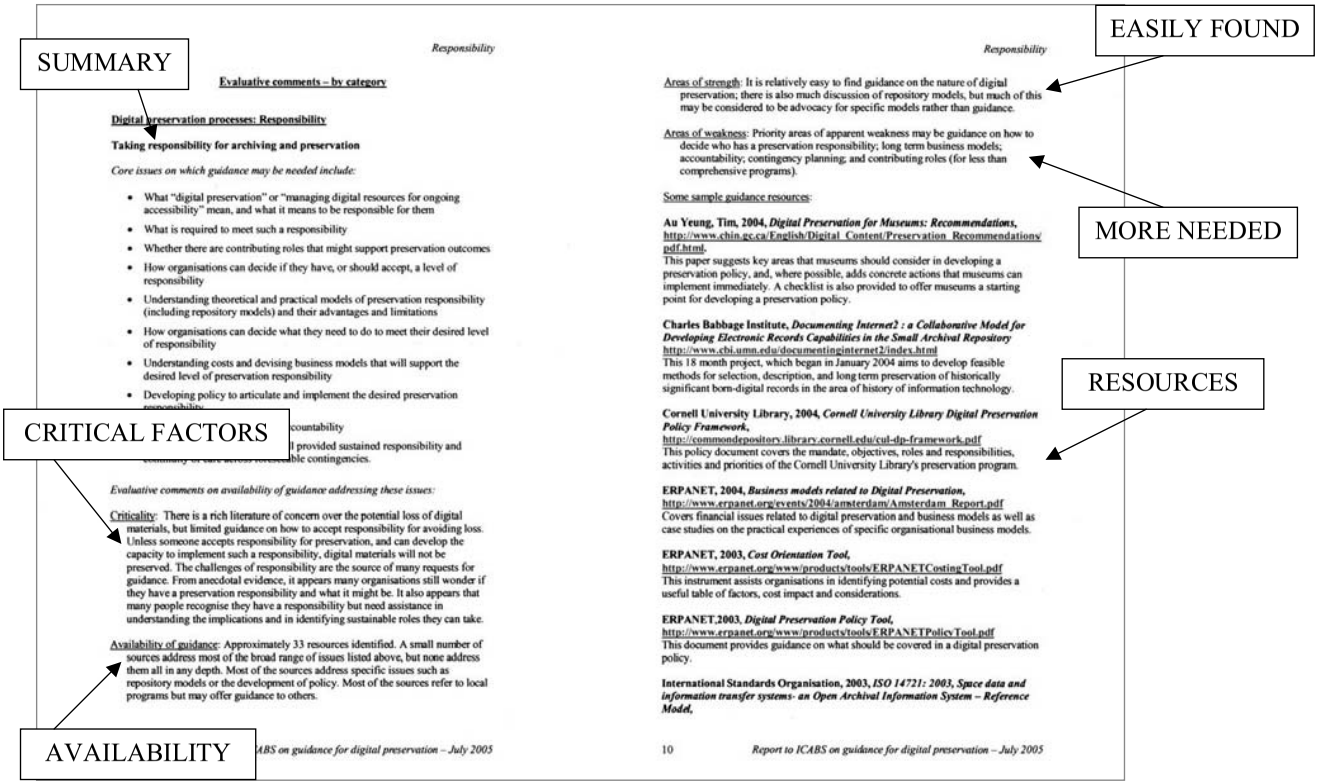
The Report concludes that at present, there is no single best source of guidance available that covers all relevant topics in the field of digital preservation. There is however a growing body of resources that can be considered as core guidance documents in digital preservation. Most of these resources are available in the form of guidelines, rather than as standards or codes of practice. Standards and codes of practice do exist, but are generally restricted to specific subjects and cannot be understood without a context.

Apart from this relatively small group of formal documents, there is a growing amount of literature that can be relevant to digital preservation processes as well. This represents a wealth of guidance, but at the same time leads to a situation in which it can be hard to see the wood for the trees. To tackle this problem, several organisations have already set up training programmes to provide a framework of basic concepts, in order to help digital preservation managers and stakeholders to find the information they need.

For the near future the Report suggests that it would help if more sectoral peak bodies would invest in the development of basic guidelines for managing different kinds of digital materials. This concept has already proven its worth in the fields of records archiving and audio preservation. These guidelines could provide the basis for creating a platform for knowledge sharing and cooperation on new developments. The future development of guidance in digital preservation would ideally consist of a mixture of basic concept guidelines, codes of practice for specific communities, standards covering specific steps and techniques, and strong mentoring and support mechanisms. And in all of this, the value of personal contact for any guidance should not be underestimated.

The full report can be found on the ICABS website and on the NLA website. http://www.nla.gov.au/nla/staffpaper/2005/documents/webb1.pdf

Example of a page of Guidances on digital resources



Libraries all over the world have to deal with fast growing numbers of digital materials that need to be safeguarded. Publications in digital form, online or on CD, digitised images, and born-digital objects need to be preserved and kept accessible. Safeguarding digital heritage is a major issue, especially for national libraries, because of their legal task of preserving the national heritage of a country.

Does day-to-day practice in storing and accessing digital objects illustrate a mutual need for certain standards? Are there currently any standards for the development and building of digital repositories, and how are these being applied? Are there common standards in research on permanent access? Or is it still too early to speak of standards, and is it only possible to distinguish best practices?

Read more on the state of the art of digital repositories, preservation strategies and current projects in the national libraries of Australia, Austria, Canada, China, Denmark, France, Germany, Japan, the Netherlands, New Zealand, Portugal, Sweden, Switzerland, the UK and the USA.

ISBN 3-598-21847-8

THOMSON

K · G · Saur Verlag