# Digital Preservation at the British Library

*Deborah Woodyard*

**The author**
Deborah Woodyard is the Digital Preservation Coordinator at The British
Library in London. She began working on the preservation of digital
materials in 1996 at the National Library of Australia and joined the
British Library in January 2001.

Email: Deborah.Woodyard@bl.uk

**Abstract**
Ensuring the comprehensive coverage, recording and preservation of the UK national
published output is one of the British Library's key responsibilities. A significant
proportion of the Library's cultural heritage collections will be in digital formats and
require long term access, so the Library has to be at the forefront of national and
international efforts to develop digital preservation. The Library is addressing digital
preservation by establishing policy, negotiating and legally obtaining the rights to
preserve and provide access digital publications. It is participating in research and
test-bed projects, and working collaboratively in research and distributed archiving
with other national and international organisations with similar interests.

**Digital developments**
Digital information is not going to go away. Or is it? We receive and create an
increasing amount each day. Innovations mean we are able to provide greater access
to collection material because of digitisation, and use digital publications in new and
more efficient ways. However, as technology races ahead we must be careful not to
leave important digital collections behind in outdated systems.

The preservation of digital material is a high priority for the British Library. *New
Strategic Directions* (The British Library, 2001), the Library's vision for the next five
years, released last year, lists our key responsibilities. They include 'ensuring the
comprehensive coverage, recording and preservation of the UK national published
archive'.

Our collection strategies have been developed in this context and all have a direct or
implied digital component. We aim to ensure improved coverage of the UK's
National Published Archive, to increase our collecting of digital materials, to focus on
user needs and develop greater collaboration with other libraries in collecting,
preserving and providing access to research material.

As a result, a significant proportion of our cultural heritage collections will be in
digital formats and require long term access, so the Library has to be at the forefront
of national and international efforts to develop digital preservation.

**Digital diversity**
The wide variety of materials we already acquire include:

- Digital publications via Voluntary Deposit agreements, such as journals and interactive multimedia on CD-ROM
- Disks accompanying print publications
- Purchased networked/online publications sent by email or FTP
- Special collections material such as manuscript collections containing email, disks, tapes of all types and even computers
- Digitisation and in-house projects producing images and structured products, for example 'Turning the Pages'
- Web sites
- Audio
- Geographic information system data (GIS) from the Ordnance Survey

The diversity of material presents various challenges from understanding and addressing our legal rights and responsibilities for preservation, to the range of technology issues for providing long-term preservation and access.

**Paths for progress**
The Library has undertaken to address the broad concerns around digital preservation by establishing policy, negotiating and legally obtaining the rights to preserve and provide access. It is also participating in research and test-bed projects, and working collaboratively in research and distributed archiving with other national and international organisations with similar interests.

**Digital preservation policy**
A specific digitally oriented policy is required in addition to a traditional preservation statement, due to the fundamental differences in the nature of the material. Digital material is different from analogue material in that it requires active and continued intervention to ensure its survival. Digital material also requires an intermediate computer environment to provide access, which introduces unique complications due to frequent changes in technology. Therefore preservation decisions need to be made right at the beginning of the life cycle management of the digital object, before technology has moved on.

The British Library's digital preservation policy states that the library will address both the physical media and intellectual content aspects of the material. However the digital storage media is unlikely to affect the appearance of the intellectual content, and so the content and media can be fundamentally separated. Therefore it is the content and not the media which is the primary concern of digital preservation.

The aim is to preserve permanently an original version of the digital content, even when modified versions have been created for future use. To aid this, the Library intends to record information specifically for preservation purposes (preservation metadata) at the point when the material is deposited, acquired or created by the Library, and make preservation decisions at that point.

The policy does not dictate which preservation strategies will be applied to digital materials. The Library anticipates 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 development, whether managerial or technological.

**Preservation functions**
To accompany the chosen digital preservation strategies the Library aims to provide support for preservation functions such as maintaining preservation metadata and negotiating with publishers to obtain a suitable digital version of material for preservation.

The Library seeks to build on the work of other comparable organisations engaged in the care of national electronic written and documentary heritage, for example the National Library of Australia (National Library of Australia, 2002). It will work collaboratively, both nationally and internationally, and with different library and archive sectors, in the further development and implementation of its preservation strategy for digital material.

**Rights and responsibilities**
The need to support the Library's collections strategy compels us to continue to press for new legal deposit legislation. The Library is also urging the government to include enabling powers in new primary legislation, so that new categories of publication may be included when it is possible to specify what form their deposit should take.

Until new legislation is achieved we want to improve coverage of digital publications under the voluntary deposit agreement signed and supported by the Department of Culture, Media and Sport, publishers and the legal deposit libraries. These are essential activities to clarify the rights we have to collect digital material but also our responsibility and rights to take action to preserve it.

Collaboration with other interested organisations in the collection and preservation of digital materials is an essential area to be explored to ensure comprehensive coverage, minimal redundancy and the best use of resources.

**Research and development**
There are number of projects the Library is involved in which are looking at different aspects of digital preservation.

*Electronic manuscript materials*
A project being conducted in collaboration with the National Library of Wales is looking into archiving the email of literary figures. Correspondence has always been a rich source of information for researchers and in a digital environment we need to re-examine many selection and collection issues. For example, how do we define which email correspondence is worth keeping, how and when do we collect it from the authors, and how should it be stored, described and accessed?

The Library is also examining the issues surrounding the archiving of digital components of specific scientific collections, which might include a wide variety of computer storage media from punch cards to whole computers, with a multitude of data types. Challenges in this project include reading and understanding the out-dated

data formats to work out what the data represents and finding a way to take the data into the future in a useful form.

## Web archiving
Part of the Library's Collections Strategy is to develop selection criteria for archiving websites which will be significant primary sources to researchers of the future. This will be done in partnership with other libraries and the research community. The major issue with web sites is that these often have an unpredictable life span, possibly just days. Access to many valuable resources will be lost if they are not preserved by institutions such as the Library. The Library has conducted a modest selective web site archiving project called Domain UK. This project selected and collected 100 web sites, with the agreement from the web site owners, to sample the UK internet.

The Library aims to develop this selective approach into a larger archive of material as well as explore collecting an annual comprehensive snapshot of the UK domain. This will require research in conjunction with other international institutions and current proposals under examination  include a suggested project to create a European Web Archive (European Web Archive, 2002) and evaluating collaboration such as that between the Library of Congress and the Wayback Machine (Wayback Machine, 2002).

## Cedars
The recently completed Cedars (CURL Exemplars in Digital Archives) project (Cedars, 2002) included The British Library as a test site in its studies. The broad objective was to explore digital preservation issues ranging through acquiring digital objects, their long-term retention, sufficient description including preservation metadata, and future access.

## LOCKSS
The Library is also a test site for the LOCKSS (Lots of Copies Keeps Stuff Safe) project (LOCKSS, 2002) based at Stanford University in the USA. This is a distribution model for digital material, specifically journal articles, which ensures that items in the system do not suffer from data loss or corruption. The intent is to make it feasible and affordable even for smaller libraries to preserve access to the e-journals to which they subscribe.

## National and international collaboration
The British Library has developed many connections with other national and international bodies in relevant areas to increase the effectiveness and progress in work on digital preservation.

A significant national partnership has been formed in the Digital Preservation Coalition (DPC) (Digital Preservation Coalition, 2002) - of which the Library is a founding member - launched in February this year. The Coalition is co-ordinated under the auspices of the JISC Digital Preservation Focus.

The DPC will foster joint action to address the urgent challenges of securing the preservation of digital resources in the UK and international collaboration to secure our global digital memory and knowledge base.

The Library has also been represented on international working groups such as two directed by OCLC and RLG on preservation metadata and digital archives.

The working group on Preservation Metadata (OCLC/RLG Preservation Metadata Working Group, 2002) has produced a synthesis of international work on the information required to facilitate preservation management and long term access to digital materials. The group also constructed a framework for use in producing and storing this information.

The Digital Archive Attributes Working Group (OCLC/RLG Digital Archive Attributes Working Group, 2002) defined the characteristics of reliable archiving services for heterogeneous research collections. The report from the working group "Trusted Digital Repositories: Attributes and Responsibilities" identifies tools that support research institutions as they seek either to build their own archiving capacity or contract with third-party services for archiving functions.

Strong collaborative links have been forged with other institutions such as the National Library of the Netherlands (Koninklijke Bibliotheek) (Koninklijke Bibliotheek, 2002) which is developing a digital storage facility for long-term preservation.

Apart from looking to similar institutions for collaborative work, we also want to encourage cross-sectoral work in digital preservation reaching into the publishing, scientific and computing communities as well as other cultural and archiving institutions. The aim is to raise awareness and invite participation to find the best possible solutions. For example, we are working in partnership with publishers and libraries, to develop guidance for publishers on the metadata (which describes the content, structure and licensing rights) of digital works that libraries need in order to record and preserve them.

**One step at a time**
These activities are making significant progress in digital preservation, but there is still a great deal of work to be done. The development of the digital preservation policy accompanied by the research and collaborative exploration will significantly improve the Library's ability to provide access to digital materials into the future. Coordination of the Library's work with other national and international efforts through the Digital Preservation Coalition will ensure maximum benefit is obtained through our work.

**References**

The British Library (2001) *New Strategic Directions, 2001*. URL: http://www.bl.uk/about/strategic/planfor.html [6.11.02]

Cedars (2002) *Home page*. URL: http://www.leeds.ac.uk/cedars/ [6.11.02]

Digital Preservation Coalition (2002) *Home page*. URL: http://www.jisc.ac.uk/dner/preservation/prescoalition.html [6.11.02]

European Web Archive (2002) *Expression of Interest*. URL: http://eoi.cordis.lu/dsp_details.cfm?ID=36291 [6.11.02]

Koninklijke Bibliotheek, National Library of the Netherlands (2002) *Home page*. URL: http://www.kb.nl/ [6.11.02]

LOCKSS (2002) *Home page*. URL: http://lockss.stanford.edu/ [6.11.02]

National Library of Australia (2002) *Home page*. URL: http://www.nla.gov.au/ [6.11.02]

OCLC/RLG Digital Archive Attributes Working Group (2002) *Home page*. URL: http://www.rlg.org/longterm/attribswg.html [6.11.02]

OCLC/RLG Preservation Metadata Working Group (2002) *Home page*. URL: http://www.oclc.org/research/pmwg/ [6.11.02]

Wayback Machine (2002) *Home page*. URL: http://www.archive.org/ [6.11.02]