The British Experience in Preserving Digital Library Collections

by NEIL BEAGRIE

1. INTRODUCTION

The increasing amount of information being produced digitally is impacting on and changing forever the expectations of users in all spheres of life. Digital access has many advantages over paper-based access in terms of convenience and functionality. However there are challenges, not yet wholly resolved, associated with ensuring that the many and varied digital resources being created will continue to be accessible. The long-term preservation of research collections in digital form is therefore an issue of increasing concern to research libraries around the world. Preservation and access to this information is dependent on impermanent media and technologies and adapting or re-engineering existing organisational and legal frameworks.

The British experience in preserving digital library collections is accumulating rapidly but it should be stressed is still relatively recent. Historically digital library developments in both the UK and elsewhere have to date primarily focussed on improving access to collections. Relatively few have engaged in preservation of digital materials per se or investigated the long-term preservation implications of digitisation (although a number of them have investigated the use of digitisation to produce surrogates for use in the preservation of analogue originals. However this is a different theme to the issues covered in this paper).

Although the preservation of digital data has a relatively long pedigree in the social sciences and scientific data communities in the UK and elsewhere[1], awareness of the issue of preserving digital materials in libraries in the UK really only emerged in the mid-1990s. What has followed has been a series of important research studies and pilot projects which are detailed in this article. It should be noted that this represents a personal selection and is by no means comprehensive in coverage or exhaustive in its descripion of activity within individual projects or institutions.

2. THE WARWICK I WORKSHOP AND THE JISC/NPO RESEARCH STUDIES

In 1996, a specially commissioned Taskforce on Digital Archiving in the USA published the final report of its work[2]. For the first time the complex inter-relationships and dependencies which impact on the ability to provide the same assurance of long-term access to important digital resources that we currently provide for other materials held in libraries and archives were stated and recommendations were made for the way forward. The impact of the work of the Taskforce has been felt world-wide.

In the U.K, the interim draft of the report was a key influence in a workshop sponsored by the Joint Information Systems Committee of the Higher Education Funding Councils (JISC) and the British Library in 1995[3]. As a recommendation of this workshop, a Digital Archiving Working Group was formed by the National Preservation Office (NPO) and other interested orga-nisations which, among other things, commissioned seven research studies into various aspects of digital preservation. These studies covered a wide range of investigations into policy and requirements, digital collection ma-nagement, and preservation techniques, which have been highly influential[4]. The studies were as follows:

- A Strategic Policy Framework for Creating and Preserving Digital Col-lections

- Digital Archaeology: the recovery of digital materials at risk

- Responsibility for Digital Archiving and Long Term Access to Digital Data

- Preservation of digital materials; policy and strategy issues for the UK

- An Investigation into the Digital Preservation needs of Universities and Research Funders

- A framework of Data Types and Formats, and Issues affecting the long term Preservation of Digital Material

- Comparison of methods and costs of digital preservation

These studies are now also available in a synthesised form which makes them accessible to a wider audience[5].

In addition to this programme of research there have also been projects and test-beds since the mid-1990s with a major element devoted to digital preservation.

3. The Arts and Humanities Data Service

The first of these, the Arts and Humanities Data Service (AHDS), emerged from a fruitful conjunction of circumstances. Complementary activity within the British Library, British Academy, and Joint Funding Council's Library Review Group, lead to a JISC commissioned feasibility study to address the need for a national strategy to preserve and provide access to electronic resources in the arts and humanities. This scoping work in 1995 lead to the establishment of an Arts and Humanities Data Service in 1996[6]. The AHDS was required to respond to several significant challenges:

- The need to preserve the rapidly growing unpublished primary digital research materials being generated within the higher education arts and humanities community and beyond;

- The identification and promotion of standards to ensure the quality and longevity of created resources;

- The need to develop new partnerships with stakeholders outside the higher education scholarly community in order to pursue common aims of data preservation and access.

The AHDS' response to these challenges involved reshaping and developing existing data service provision; at the Oxford Text Archive (University of Oxford) and History Data Service (University of Essex). It also developed three new services from scratch: the Archaeology Data Service (University of York), the Performing Arts Data Service (University of Glasgow) and the Visual Arts Data Service (Surrey Institute of Art & Design University College). In doing so it embarked on developing a comprehensive policy framework for the development and management of the service[7] outreach and publication of standards and best practice[8], and collaboration and partnership with a range of other organisations[9].

4. CEDARS

Another important and more recent test-bed for digital preservation within research libraries in the Higher Education sector has been CEDARS – the CURL Exemplars in Digital Archives[10]. The Cedars Project is a Higher Education initiative funded by the Joint Information Systems Committee which officially began on the 1st April 1998.

The funding was awarded to the Consortium of University Research Libraries (CURL) and the work is carried out on behalf of CURL by three CURL institutions – Leeds, Oxford and Cambridge.

369

The main objective of the project is to address strategic, methodological and practical issues and provide guidance in best practice for digital preservation. It will do this by work on two levels – through a practical demonstrator project which will provide concrete practical experience in preserving digital resources, and through strategic working groups based on broad concepts or concerns which will articulate preferences and make recommendations of benefit to the wider community. The main deliverables of the project will be recommendations and guidelines as well as practical robust and scaleable models for establishing distributed digital archives. One of the key deliverables for CEDARS is the development of a robust metadata schema to support digital preservation and a consultation draft is currently available[11]. The demonstrator project is being developed to following principles outlined in the Open Archival Information Systems model and a paper on its implementation has also bee prepared[12].

A further recent development out of CEDARS has been CAMILEON a JISC and National Science Foundation (NSF) funded project to examine emulation as a preservation strategy for digital materials[13].

5. THE BRITISH LIBRARY AND LEGAL DEPOSIT

Within the British Library, the National Preservation Office has developed a digital remit and has been an active partner with the JISC and others in moving this forward for example through the JISC/NPO studies[14]. There has also been significant progress elsewhere within the BL particularly recently, both in developing policy and proposals for legal deposit of digital materials and tendering for a digital storage system.

Under existing UK legislation, all printed books published in the UK must deposited, free of charge, in the British Library. Five other deposit libraries (Cambridge, Oxford, National libraries of Scotland, & Wales, Trinity College Dublin) are also entitled to receive, on request, free copies of any book published. In 1997, the Secretary of State for Culture, Media and Sport convened a Working Party to discuss the extension of legal deposit to „non-print" materials[15]. Following the publication of the report there has been substantial discussion between the BL and publishers and a voluntary scheme of legal deposit for selected categories of electronic publications was introduced in January 2000[16]. A voluntary code of practice has provided an opportunity for the deposit libraries and publishers in the UK to work through the issues which arise in the deposit of electronic publications[17]. It is anticipated that

legislation will follow within the next 18 –24 months for mandatory deposit of selected electronic materials.

The BL is also currently acquiring a Digital Library System through an EU procurement process[18]. Like CEDARS and several other library initiatives, the project will take the OAIS reference model as its standard for design of the system. The library currently houses digital collections that include thousands of CD-ROMS and other electronic publications. These holdings will increase substantially in the next few years through purchase, the voluntary deposit scheme and any future mandatory deposit legislation for electronic publications. It is a crucial period for the library. It is liaising closely with similar projects in the UK and abroad in building the system and developing the policies and working practices which will be essential to underpin its operation.

Building on and extending the BL existing collaboration with other libraries and organisations in the UK is also a significant theme in its strategic plan for coming years. The BL Partnership and Collaboration Programme announced this year has also included digital archiving in its scope and a project to examine the implementation in libraries of a workbook on the preservation of digital materials (see below) is being supported in the first round of funding.

6. PRESERVATION OF DIGITAL MATERIALS WORKBOOK

The British Library has been a partner and „case study" for a workbook on the preservation management of digital materials[19] by the author and Maggie Jones, funded by the former Library and Information Commission (now Re:source: The Council for Museums, Archives and Libraries). This has provided a valuable partnership and insight into current developments in the national libraries in the UK. The research study to produce the digital preservation workbook has been in progress since July 1999 and will be completed in September 2000[20].

The workbook aims to build on the AHDS study „A Strategic Framework for the Creation and Preservation of Digital Collections" by providing a more detailed description of existing knowledge, strategies and activities, and developing tools such as decision trees and checklists to add implementation.

7. COLLABORATION

Collaboration has been at the centre of many digital preservation activities in the UK. Although it is not always easy to adjust organisations to greater partnership or ensure that collaboration is implemented successfully, it has been widely recognised that the challenges of digital preservation and the solutions require multi-institutional effort and collaboration. A second workshop at Warwick in 1998 („Warwick II") to re-visit digital preservation in the UK with representation from a wide range of organisations, made a number of recommendations for action over coming years. It particularly emphasised the need to further increase co-ordination and collaboration both within the UK and with international partners[21]. The proposal for a National Digital Preservation Coalition made at Warwick II is now under active discussion and the JISC has funded the JISC Preservation Focus post to help begin this process[22].

8. CONCLUSIONS

From the above I hope it is evident that substantial lessons have been learnt in the UK from pilot and research projects, and other sectors in last five years. Much remains to be done and arguably we are now moving into a new phase of development. Significant investment in digital content is now beginning to drive forward the agenda to address the long-term preservation of digital publications and resources. The next few years will see the UK moving from pilot projects to digital preservation services across a range of sectors and institutions. I hope collaborative activity to identify solutions and share responsibilities will be increasingly prevalent and will build on what has already been achieved.

The introduction of voluntary & statutory legal deposit in the UK will be very important in securing a framework for the preservation of many of our electronic publications but will clearly not be a panacea for securing all areas of our electronic heritage. Over the next decade it is likely that a range of responsibilities for the preservation of different collections and classes of material will emerge for electronic materials (as can be seen in traditional print collections today). With increasing globalisation of electronic information, some of these arrangements are likely to be more international and may well lead to new forms of archiving institutions and organisational partnership.

REFERENCES

1   The Essex Data Archive <http://www.data-archive.ac.uk> for example has
    preserved digital files from social science research for more than two decades
    and has developed procedures and standards for documentation and preserv-
    ation in collaboration with similar data centres worldwide.

2   Waters, D. & Garrett, J. (1996). *Preserving Digital Information: Report of the
    Task Force on Archiving of Digital Information commissioned by the
    Commission on Preservation and Access and the Research Libraries Group.*
    Washington, D.C: Commission on Preservation and Access. <http://www.rlg.
    org/ArchTF/>.

3   *Long Term Preservation of Electronic Materials; a JISC/British Library
    Workshop as part of the Electronic Libraries Programme (eLib). 27-28
    November 1995 at the University of Warwick.* <http://www.ukoln.ac.uk/
    services/papers/bl/rdr6238/>.

4   All seven studies are available online <http://www.ukoln.ac.uk/services/elib/
    papers/supporting/#blric>. The studies are also published in hard copy by
    LITC. Each study is priced £20.00 plus postage and packaging and sales are
    handled by TBC Distribution email: tomlinsons@easynet.co.uk.

5   Feeney, M. (ed) (1999). *Digital Culture: maximising the nation's investment.*
    London: National Preservation Office. Free copies can be obtained from: Julia
    Foster of the National Preservation Office at the British Library Freepost
    LON8974, London WC1B 3BR, telephone +44 (0)171 412 7612 or email
    julia.foster@mail.bl.uk. The document can also be downloaded as a pdf file
    from   <http://www.ukoln.ac.uk/services/elib/papers/other/jisc-npo-dig/intro.
    html>.

6   Further information on the AHDS is available from its website at <http://
    ahds.ac.uk>.

7   See *A Strategic Policy Framework for Creating and Preserving Digital Col-
    lections* <http://ahds.ac.uk/manage/framework.htm> and *Managing Digital
    collections: AHDS Policies, Standards and Practices* <http://ahds.ac.uk/
    public/srg.html>.

8   Its publication series „AHDS Guides to Good Practice" <http://ahds.ac.uk/
    public/guides.html>and „Managing Digital Collections" <http://ahds.ac.uk.
    manage/manintro.htm> contain particularly useful information on preserv-
    ation practice and standards.

9 See for example the joint C&IT policy with the Arts and Humanities Research Board <http://ahrb.ac.uk/citpol.html>.

10 Further information on CEDARS is available on their website at <http://www.leeds.ac.uk/cedars/>.

11 *Metadata For Digital Preservation: The Cedars Outline Specification* at <http://www.leeds.ac.uk/cedars/OutlineSpec.htm>.

12 *Demonstrator system – A blueprint for Representation Information in the OAIS model* <http://gps0.leeds.ac.uk/~ecldh/cedars/nasa2000/nasa2000.html>.

13 For further details on the project see <http:www.si.umich.edu/CAMILEON/>.

14 Information on the digital remit for the NPO is available at <http://www.bl.uk/services/preservation/digital.html>.

15 *Report of the Working Party on Legal Deposit* <http://www.culture.gov.uk/LDWGRPT.HTM>.

16 The voluntary code is currently unpublished but has been widely circulated in draft form.

17 A valuable discussion of the issues involved can be found in *Digital Preservation: an introduction to the standards issues surrounding the deposit of non-print publications,* M. Bide et al (1999) <http://www.bic.org.uk/digpres.doc>.

18 For further information see <http://ww.bl.uk/diglib/dlp/dls.html>.

19 Further details at <http://ahds.ac.uk/manage/licstudy.html>.

20 For further information see <http://ahds.ac.uk/manage/licstudy.html>. A consultation draft of the workbook will be available in August 2000 and linked from this url.

21 The recommendations of the workshop are available online <http://www.leeds.ac.uk/cedars/OTHER/warwick2.htm>.

22 Details of the proposed coalition and work of the JISC Preservation Focus will be available through email list announcements and dedicated webpages.

Neil Beagrie
JISC Digital Preservation Focus
Tel/Fax/Voicemail: +44 709 2048179
preservation@jisc.ac.uk
<http://www.jisc.ac.uk>