WHO WILL PRESERVE ELECTRONIC PUBLICATIONS?

David Haynes and David Streatfield

A paper based on a study funded by JISC through the National Preservation Office, Digital Archiving Working Group, the lead body in the UK for all aspects of preservation digitisation and digital archiving

Although some organisations have their own policies for the preservation of digital material, there is no national policy. The issues involved in creating such a policy are complex but a national agency could co-ordinate the development of the guidelines and standards necessary for the maintenance of a distributed archive, funded from the public purse and supported by changes in legal deposit legislation.

David Haynes and David Streatfield worked together on a consultation exercise to establish who ought to be responsible for long-term preservation of digital material. They are joint authors with Monica Blake and Tanya Jowett of a report to the Digital Archiving Working Group[1]. E-mail: d.haynes1@compuserve.com

Introduction

Why archive digital materials?

There is a growing concern about the lack of provision for the archiving of electronic publications. Currently, the legal deposit libraries have no specific remit to acquire and preserve electronic publications, although they have interpreted their overall responsibility to include preservation of electronic publications. In order to fulfil their roles as guardians of the intellectual heritage of the UK and Ireland, the legal deposit libraries need to address the issue of long-term preservation of digital materials. Recent initiatives, such as the proposed revisions to the legal deposit legislation, could provide a mechanism for this to happen. There is considerable concern that no additional provision has been made to fund the acquisition and preservation of digital materials. This is despite the fact that the deposit libraries are having to cope with an increasing burden of paper-based publications. Inevitably the deposit library have to put in place a selection process to ensure that the most significant works are kept in perpetuity.

Background

The JISC-funded study on 'Responsibility for long term preservation and access to digital materials' extended beyond rights holders to include all stakeholders in the production, exploitation, distribution and preservation of digital materials (see Figure 1).

Much of the current debate on digital archiving was started by the Task Force on Digital Archiving in the USA, which recommended the development of a national system of digital archives to act as repositories for digital information. The Task Force felt that:

Without the operation of a formal certification program and a fail-safe mechanism, preservation of the nation's cultural heritage in digital form will likely be overly dependent on marketplace forces, which may value information for too short a period and without applying broader, public interest criteria.
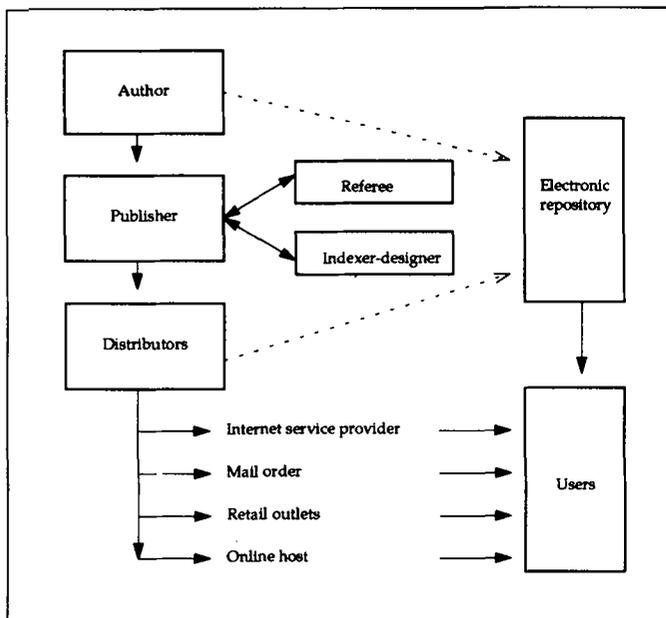
Figure 1 Production of electronic publications

The RLG/CPA Report[2] was subsequently analysed by a team at Loughborough University[3] which came up with eight prioritised actions, including:

□ appointing a National Digital Preservation Officer;

□ establishing a National Digital Preservation Body;

□ devising guidelines on practice and a digital preservation policy.

Policy issues

Is there a need for a policy?

During our study we wanted to establish whether organisations had developed their own policies on digital archiving. Of the 32 organisations that responded 16 had a policy and 8 of the remaining organisations were working towards a policy or felt that a policy was necessary. Some publishers and depositories had comprehensive policies:

> As a publisher, our concern is to preserve our own material. For this, there is a strict archiving policy. Material that is licensed from a third party is archived when it comes into the building and immediately prior to publication. It is kept in raw and built forms. All versions of software are archived including ones that are never published. Archiving includes type definitions and SGML data.

Responsibility for preserving it. Processing, validation, supplementing, scanning, creating metadata in digital form. Take all these together with data, strip off software dependencies. Checks for corruption, write to three different media. Then there is a ;continuing programme of checking for corruption, deciding when to move to new media. Archive held at different sites.

Who should have access to digital archives?

The issue of who should have access to archived electronic publications is a sensitive one for publishers. Many maintain that so long as someone wishes to see a publication it has commercial value and that the publisher must therefore obtain some benefit. This needs to be off-set against the costs of preservation. It is clear that the issue is how to provide users with access to material while protecting the interests of copyright holders.

There is some scope for a consensus on access based on current access policy for printed legal deposit material. However, digital material made available via a network is impossible to control once downloaded to a user. There is a very real danger that publishers could lose potential sales, if their material were freely available from a public archive:

> At the moment with legal deposit, any individual can look at archives and photocopy under fair dealing. With electronic publications, the publishers say it is impossible to fair deal and they are threatening to sue. They are much more concerned about depositing e-publications. Some publishers ... invest an enormous amount in a CD-ROM which will only sell about 10 copies at an enormous price, don't like the idea of it being on a network. They have a point. Similarly with high value financial information.

One approach might be to restrict access to a single stand-alone PC. This could be limited to one PC or terminal per deposit library, or one terminal in one designated library.

Access and copyright are tied in together. The concept of 'fair use' could be applied to electronic publications in the way in which it applies to printed publications. New legislation on legal deposit, requested by the British Library and other key institutions, could provide the basis for licensing use of electronic material.

Different arrangements could be put in place for different types of user. For instance, researchers and academics might have free access to certain materials, whereas commercial exploitation of material would attract a charge or royalty fee. The legal deposit 'copy' could provide a resource of last resort, as suggested by some respondents during our survey:

> Anyone who can show they can't get the information elsewhere. Condition: When it is no longer available commercially, or out of copyright.

> In a legal deposit context any bona fide user who can demonstrate need. Serious research interest. Not available frivolously.

Who will look after the archives?

One of the recommendations arising from the Loughborough report was the idea of a National Digital Preservation Body with a National Digital Preservation Officer. The role of the preservation body would be to co-ordinate digital archiving in the United Kingdom and Ireland. This was reflected in some of the responses in the survey:

> Most secure way of preserving digital material is for responsibility to be assigned to national repositories.

> Need to rely on national repositories. Can not be held locally. Economies of scale. Staffing levels and technology needed nationally. Makes most sense to split the archive up by disciplines as long as they talk to each other...It needs a national body to pull it together and ensure there are no gaps and nothing duplicated.

> ...establishment of a National Digital Preservation body in which all the stakeholders are represented. It is important to have a national forum for all the different interests to get together and find out what

each other is doing and what research is going on etc;

although there was one clear warning that:

> A central depository is unlikely to fulfil the needs of different depositors.

As to who should take on this role, suggestions included the national libraries such as the British Library. Another popular suggestion was for a system of distributed depositories or a body to co-ordinate different depository agencies:

> In the case of the refereed learned journal literature, I think there should be an international consortium of Universities and learned societies, with governmental support, overseeing the archiving and preservation.

> There could be a network of data archives, each funded separately and in different ways.

> We would need some kind of co-ordinating body to decide where the individual material went. This body could also decide on issues like saving snapshots of things every month etc.

> Network of organisations. Central policy, but carried out by lots. One organisation could not cope. Has to be a distinction between public and private information.

However, several publishers suggested that they or the rights holders should archive their own material, although there were cautionary notes as well:

> Publishers keep an archive in case it is needed for further business etc. But they cannot be relied on as they are not doing it for the same reasons as an archive would.

> Feel that the publishers have a responsibility for maintaining the databases of electronic products. Each publisher should maintain own in-house archive. On a national scale it would have to be sophisticated operation to accommodate range of systems and operating systems etc.

Technology concerns

One of the first concerns in developing a strategy is to decide whether to store the material in a standard format (for instance storing documents in *.PDF or *.HTM formats)

or to try and preserve it in the original format. Some materials may be kept in both a standard format, with the associated changes to the look and feel of the product, and in the original format, so that the product is intact and maintains its integrity.

A product which is conserved in standard format by migrating the content to currently-supported software environments and operating systems, is likely to be readily accessible. However there is an associated loss of information by changing the appearance and form of the document. With increasingly sophisticated software, electronic documents are much more than a linear sequence of text, and the environment in which the information is presented can be seen as an integral part of the product. One of the difficulties of preserving the original format is that the technology to read it must also be preserved.

How long should digital materials be kept?

Most respondents felt that selected material should be kept for ever. This raised the issue of selection criteria and the high management cost of selective preservation. There is already a precedent for this in the Public Record Acts in the United Kingdom. Material is selected and reviewed according to criteria applied by the Public Record Office. This makes the selection process a routine administrative function for the majority of material thus reducing the cost.

Who pays?

There is strong support for the view that a digital archive should be funded by the government. It ought to be a national responsibility:

> Payment should be a national responsibility. On the one hand, our organisation [a repository] is partly government funded, and thus has a certain responsibility for its data. However, it would not want to spend additional money for digital archiving. There should be central funding for a national archive.

> Should be national expense. Top-sliced from national organisations or institutions responsible for archiving.

If national repository, then funding should be part of general funding, i.e. state funded. If it is an added responsibility, this should be recognised in the setting of funding. Libraries must make a case for funding. If funds are not forthcoming, then priorities will have to be adjusted.

Government responsibility was often linked to the British Library and other deposit libraries. Many people thought the system for print publications was suitable for digital archives. Although publishers already incur costs by providing 'free' copies of their publications, some respondents suggested that publishers have an additional responsibility to contribute financially to the preservation of their publications. However the counter view is that this 'tax' would act as a considerable disincentive to publishers in the UK and could encourage them to base their activities elsewhere.

The British Library has considered alternatives such as funding for electronic publications under the Private Finance Initiative (PFI).

Several respondents have mentioned in passing the possibility of charging for the use of an archive. One was more specific:

> The beneficiaries become the obvious source of funding. Preservation of scholarly knowledge, commercial value, scientific value, defence against threats. We have devised a matrix, which will be published, of who benefits against what services offered, in order to prioritise allocation of resources.

There was also a strong view that access should be free at the point of use. Institutions might pay a global licence fee.

The level of commitment on the part of funding agencies is a key limiting factor for long-term preservation and needs to be taken into account in putting forward any recommendations. One possibility is that the level of resources, that are devoted to digital archiving, now will not be available in the future, and any strategy for long-term preservation must take this into account.

Funding will depend on the comparative costs of transfer and of emulation. Such estimates would provide a basis for the NPO, or

British Library, to develop an integrated consistent strategy for archiving digital materials.

In the past, academic institutions have paid for journals by subscribing to individual titles. Now, as journals become available electronically (held nationally), national funders are negotiating block deals with publishers for the whole of academe in the UK to have access to electronic titles. National block funding for groups of users could be used as a model for archiving.

Public funding of digital archives may be the best way of ensuring continuity. Keeping costs down would depend on publishers providing one free copy of each electronically published title or issue that they produce, under the proposed provisions for new legal deposit legislation.

A model for future development

The result of the consultation exercise was the development of a model for digital archiving. The intention is that this model should be used to focus the discussion on the development of a national policy. It is a way of testing the ground and has not been adopted as a formal policy at this stage.

National Office for Digital Archiving (NODA)

A national body would be established with the remit for archiving digital materials. Its main role would be to co-ordinate the activities of specialist repositories and to put in place funding to ensure the continuity of the archives. Another key role would be to develop standards for archiving and to conduct research in support of its preservation role. An existing body such as the National Preservation Office could take on this role and in effect become NODA. However, it would need additional resources and funding to be able to take on an extended role.

NODA would develop or co-ordinate the development of guidelines for retention and preservation of digital materials. Specific guidelines, which apply to the creation of electronic documents or digital data, would be developed so that individual items could be easily identified and managed in a digital archive.

The choice of document format will have a significant impact on the ease of maintenance of digital archives. The use of standard formats, or proprietary standards which are widespread ,will enable the data repositories to concentrate their resources on migration of materials from a few well-supported formats.

In some instances there may be good reason for preserving the original formats and in these cases a parallel version could also be kept in a standard format, to allow for migration to other supported formats in the future.

The co-ordinating role would be separated from the archiving role which could be sub-contracted or delegated to specialist agencies. The national co-ordinating body would be responsible for development of appropriate standards of service (in consultation with the stakeholders) and for arbitrating where there are disputes about who should be responsible for which materials. The possibility of adopting a policing role, as suggested for similar bodies in other countries, should be considered as a way of ensuring that the archiving bodies are properly vetted.

Different approaches for different materials - a distributed archive

The decisions on which organisations are contracted to keep digital archives of national importance will depend on a number of factors, such as regional interest, type of data held, and ownership of material. For instance, there may be a strong case for the National Library of Wales and the National Library of Scotland to keep digital material relating to Wales and Scotland. However, certain specialist material, such as sound recordings or experimental data, may need to be kept by agencies with particular knowledge and understanding of data formats involved.

Arising directly from this would be the establishment of a national register of archived digital material. The starting point for this would be a national audit of existing digital archives. Detailed catalogue records would be the responsibility of the individual archiving agencies because of the widely varying nature of the digital materials that need to be conserved. The feasibility of using a common

descriptor, such as a Digital Object Identifier (DOI), should also be investigated.

Publishers will have different priorities from those of NODA and many of them are too small and short-lived to sustain an archive for an appreciable length of time. However, it is possible that some larger publishers with the expertise could bid to provide preservation services on behalf of NODA, on a commercial basis.

A strong point has been made that there should be a separation between service provision and agency responsibility for conservation:

> ...all responsibility need not lie with one body. A split between legal custodians (who must of necessity be long-lived bodies such as national libraries or archives) and service providers, who actually hold and distribute material, can and does work.

International dimension

Many publications and some experimental data fall across national boundaries. It is sometimes difficult to attribute a nationality to some of the larger publishing groups, especially when it comes to electronic publications, which may be released in several different locations simultaneously. NODA could provide a focus for liaison with other national bodies, as well as transnational corporations and inter-governmental organisations.

It is also important to keep track of what is going on in other countries[2,4,5,6].

Legal deposit

There is a widely-held view that the most efficient way to ensure that digital material is archived is to extend the legal deposit legislation, so that it includes electronic publications. The British Library has already submitted a proposal to the Department of National Heritage (now the Department of Media, Culture and Sport) suggesting[7]:

▫ the national published archive be a distributed archive (i.e. it recommends that more than one repository be given the right to receive publications through legal deposit);

▫ new legislation should enable the comprehensive deposit of non-print publications (designated repositories to apply selection criteria, if appropriate);

▫ new legislation should apply not only to publications traditionally associated with print on paper and now appearing in electronic format, but also to all current and future UK non-print publications, including films and sound recordings, microform publications and digital publications (both off-line and on-line)

The need for changes in legislation was endorsed by the library and information professional bodies[8].

Selection and permanent retention

Material that is selected for preservation should be kept for ever. The basis for selection should be the permanent value of the data or product. It is not possible to lay down rules that apply to all categories of digital material that may be subject to preservation. However, general selection criteria should be established and should form the basis for the development of an archiving policy.

Databases present a particular problem, however, because of their dynamic nature. One approach is to take a snapshot of the database at particular point in time. For some databases it may be necessary to keep an audit trail of all the changes made to a database over a period of time. This will give a comprehensive view of the data held. It is not possible to establish a universal archiving policy for databases and decisions will probably have to be made on a case-by-case basis.

Although some of the people consulted put forward the idea of periodic reviews of archived material, as part of a disposal policy, this approach is incompatible with the philosophy of selecting material for permanent retention. It adds a significant and increasing overhead to the administration of the system and could lead to problems of continuity.

Conclusion

There is still some way to go before the ground has been prepared for a national policy for

archiving electronic publications. The authors believe that the way forward is to extend the copyright legislation in order to ensure that at least two copies of every electronic publication is deposited with a digital archiving authority.

Consulting users

It is difficult to predict who the future users of digital archives might be. There is the problem of identifying and locating potential users. However, there are already some digital archives with established user groups that could provide a starting point for a user survey.

An economic model

Cimtech's study on the *Preservation of Digital Material*[9] considered the main options for the long-term preservation of digital publications and estimated the costs involved with each one. The costings are broad estimates and could be refined and developed.

Sharing experience

Finally we were encouraged by the degree of interest and level of response to the initial models that were made available on the Internet. We believe that there is some demand for a meeting or seminar, possibly international in scope, to share experiences and to discuss how the issues have progressed since the last conference in 1995[10] and a proposal for such a seminar has been put forward to the National Preservation Office.

References

1. Haynes D, Streatfield D, Jowett T, and Blake M. *Responsibility for Digital Archiving and long Term Access to Digital Data.* (JISC NPO Studies on the Preservation of Electronic Materials. JISC/NPO, 1997 (in press)

2. Water, D and Garrett, J. *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 Inc 1996. ISBN 1-887334-50-5 (http://www.rlg.org/ArchTF/)

3. Matthews, G, Poulter, A and Blagg, E. *Preservation of Digital Materials: Policy and*

    Strategy for the UK. JISC/NPO Studies on the Preservation of Electronic Materials. British Library Research and Innovation Centre, 1997. ISBN: 0-7123-3313-4, ISSN: 1366-8218. British Library Research and Innovation Report 41. (A version is available at http://lboro.ac.uk/departments/ls/staff/apoulter/digpres.html)

4. National Library of Australia, National Preservation Office. *Statement of Principles: Preservation of and Long-Term Access to Australian Digital Objects.* 1997. (http://www.nla.gov.au/3/npo/natco/princ.html)

5. National Library of Australia. *Legal Deposit in Australia,* Fourth edition, 1997. Last updated 12 May 1997. (http://www.nla.gov.au/1/services/ldeposit.html)

6. National Library of Canada Electronic Publications Pilot Project. *Summary of the Final Report,* 1996. (http://www.collection.nlc-bnc.ca/e-coll-e/ereport.htm)

7. Department of National Heritage, Scottish Office, Welsh Office, Department of Education Northern Ireland. *Legal Deposit of Publications: a consultation paper.* Department of National Heritage, February 1997.

8. *Copyright and the digital environment.* Managing Information 3 (1) Jan 96, p.25-6. ISSN: 13520229. (Statement prepared by the Library Association, UK /JCC working party on copyright which includes representatives of: Aslib, the Association for Information Management, the Institute of Information Scientists, the Standing Conference on National University Libraries, and the Society of Archivists.)

9. Hendley, A. *The Preservation of Digital Material.* London, British Library, 1996 (BL R&D Report 6242)

10. *Long Term Preservation of Electronic Materials. A JISC/British Library Workshop as part of the Electronic Libraries Programme (eLib).* Organised by UKOLN 27th and 28th November 1995 at the University of Warwick. Report prepared by the Mark Fresko Consultancy. The British Library, 1996. BL R&D Report 6328. (http://ukoln.bath.ac.uk/fresko/warwick/intro.html)