CHALLENGES OF DIGITAL PRESERVATION IN DIGITAL LIBRARIES

Satish M V **Umesh M.**

Abstract

Today Digital Technology (computers and communication technology) is enabling information to be created, manipulated, disseminated, located and stored with increasing ease over the traditional one. On the other side Technological Obsolescence is the result of evolution of technology: as newer technology appears, older ones ceased to be used. Consequently, information, which relies on obsolete technology, becomes inaccessible. And preserving access to this (digital) information poses a significant challenge to the Library and Information Community. The present study highlights the digital preservation techniques and issues involved with it.

Keywords : Digital Library, Digital Preservations

1. INTRODUCTION

The purpose of preservation is to ensure protection of information of enduring values for access by present and future generations. Today, we are in the transition period from traditional library to digital library environment. In this concern digital technology is revolutionizing the traditional library concepts of preservation, access and archival of information. With this digital preservation is getting more importance.

2. WHAT DOES DIGITAL PRESERVATION MEAN ?

“Digital preservation” or “digital archiving” means taking steps to ensure the longevity of electronic documents. It applies to documents that are either “born digital” and stored on-line (or on CD-ROM, diskettes or other physical carriers) or to the products of analog-to-digital conversion, if long-term access is intended.

3. WHY IS DIGITAL PRESERVATION NECESSARY ?

More and more is information is being created in digital form, either through converting existing materials to digital form or, increasingly, “born digital”, where there is no other format but the digital original. There are increasing expectations in all spheres of life that the information we all need will be available on the Internet or at least in all offline digital format, such as CD-ROM. Digital access has many advantages over paper-based or microform access in terms of convenience and functionality. The increasing proliferation of digital information, combined with the considerable challenges associated with ensuring continued access to digital information, means that it is imperative that there be concerted action to overcome these challenges.

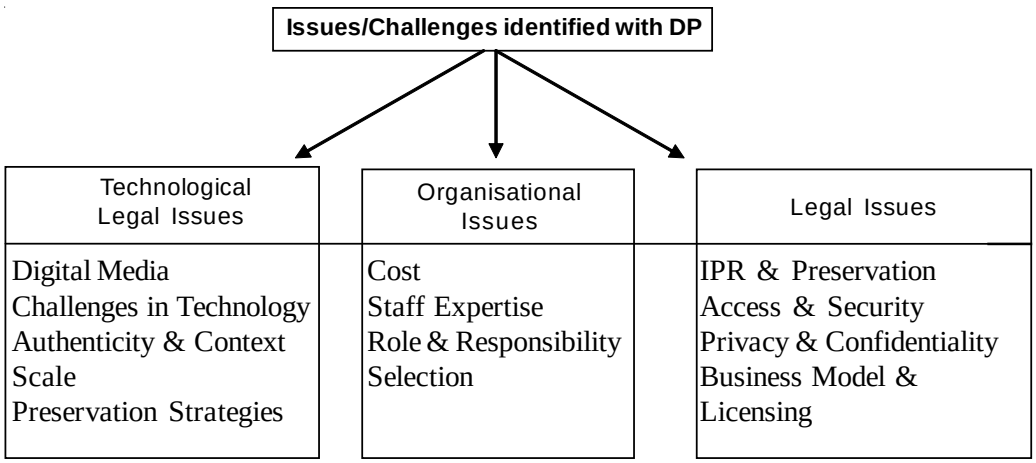
4. HOW ARE DIGITAL MATERIALS DIFFERENT?

As a Research Libraries Group (RLG) survey noted: “Digital materials, regardless of whether they are created initially in digital form or converted to digital form, are threatened by technology obsolescence and physical resources over time are related to notable differences between digital and paper-based material:

- Machine Dependency. Digital materials all require specific hardware and software in order to access them.
- The speed of changes in technology means that the timeframe during which action must be taken is very much shorter than for paper-timeframes during which action needs to be taken is measured in a few years, perhaps only 2-5, as opposed to decades or even centuries we associate with the preservation of traditional materials. Technology obsolescence is generally regarded as the greatest technical threat to ensuring continued access to digital material.
- Fragility of the media. The medial digital materials are stored on is inherently unstable and without suitable storage conditions and management can deteriorate very quickly even though it may not appear to be damaged externally.
- The ease with which changes can be made and the need to make some changes in order to manage the material means that there are challenges associated with ensuring the continued integrity, authenticity, and history of digital materials.
- The implications of allocating priorities are much more severe than for paper. A digital resource, which is not selected for active preservation treatment at an early stage will very likely be lost or unusable in the near future.
- The nature of the technology requires a life-cycle management approach to be taken to its maintenance. A continual programme of active management is needed from the design and creation stage if preservation is to be successful. This in turn leads to much more involvement both within and between institutions and changing roles.

5. ISSUES IDENTIFIED WITH DIGITAL PRESERVATION (DP)

There are number of issues involved in the preservation of digital objects.



5.1 Technological Issues

i. Digital media

"Digital materials are especially vulnerable to loss and destruction because they are stored on fragile magnetic and optical media that deteriorate rapidly and that can fail suddenly from exposure to heat, humidity, airborne contaminants, or faulty reading and writing devices."

Precautions can be taken which will help significantly to reduce the danger of loss and include:

- Storing in a stable, controlled environment.
- Implementing regular refreshment cycles to copy onto newer media.
- Making preservation copies (assuming licensing/copyright permission).
- Implementing appropriate handling procedures.
- Transferring to "standard" storage media.

However, while the media on which the information are stored may or may not fail, what is certain is that technology will change rapidly so that even if the media is retained in pristine condition, it may still not be possible to access the information it contains. No matter how exemplary the care of the media is, it will not remove the requirement to deal with changes in technology, though responsible care should make it easier to manage technology changes.

ii. Changes in technology

"Unlike the situation that applies to books, digital archiving requires relatively frequent investments to overcome rapid obsolescence introduced by galloping technological change."

Because digital material is machine dependent, it is not possible to access the information unless there is appropriate hardware, and associated software which will make it intelligible. Technology advances even in the past decade illustrate this point:

- $5\ ^{1}/_{4}$ inch floppy disks have been superseded by $3\ ^{1}/_{2}$ inch floppy disks;
- There have been several upgrades to Windows software since it was first introduced and it would now be very difficult to convert from earlier versions to the current versions;
- Thousands of software programs common in the early 1990s are now extinct and unavailable.

RLG survey cited technological obsolescence as the greatest threat to successful digital preservation. Precautions can, and should be taken, which will greatly reduce the risk of inadvertently losing access to a resource because of changes in technology. These include:

- Using standard file and media formats, as recommended by reputable sources.
- Providing detailed documentation (covers Metadata) to enable both contexts to be determined and also to facilitate successful management.

iii. Authenticity and context

"At each stage of the cycle, electronic records need to be actively managed according to established procedures, to ensure that they retain qualities of integrity, authenticity and reliability."

For example, scholars will need to feel confident that references they cite will stay the same over time, courts of law will need to be assured that material can withstand legal evidential requirements, government departments may well have legally enforceable requirements regarding authenticity, and so on.

iv. Scale

Although computer storage is increasing in scale and its relative cost is decreasing constantly, the quantity of data and our ability to capture it with relative ease still matches or exceeds it in a number of areas. Some repositories still face significant challenges in developing and maintaining scaleable architectures and procedures to handle huge quantities of data generated from sources such as satellites or the web. The technical and managerial challenges in accessioning, managing and providing access to digital materials on this scale should not be underestimated.

v. Preservation Strategies

Several strategies attempt to address the primary digital preservation problem of technological obsolescence. These include migrating information through successive generations of technology, using software to emulate the behaviour of older machines, preserving original hardware and software to run obsolete programs, and creating hard copies (paper or microform) of digital objects. Each of these strategies meets some, but not all, preservation goals. And choice of strategy will depend upon the nature of the material and what aspects are to be retained.

a) Migration

Migration is the primary strategy articulated by most organizations that plan to preserve digital objects. It covers a range of activities to periodically copy, convert or transfer digital information from one generation of technology to subsequent ones, i.e. it ensures that the digital information is re-encoded in new formats before the old format becomes obsolete. Migration may involve copying digital information from a medium that is becoming obsolete or physically deteriorating to newer one (e.g. floppy to CD-ROM) and converting from one format to another (Microsoft Word to ASCII). Migration certainly preserves the physical presence and the content of digital object. However, it may not preserve preservation, functionality and context. For example, presentation elements such as bolding and italics may disappear, and functionality and context provided by links between databases entries may be lost because the links break. Successive migrations may eventually result in unacceptable data loss.

There are a number of ways to increase the chances of using migration successfully as preservation strategy. These include:

- Capturing the context by documenting hardware and software required to access digital objects (and ensuring the technology is subsequently available).
- Monitoring backward compatibility of software to determine when to migrate.
- Choosing a small number of standard formats and converting non-standard formats to standard ones to minimize the amount of conversion required.

b) Emulation

Emulation refers to creating new software that mimics the operations of older hardware or software in order to reproduce its performance. Thus, not only are physical presence and content preserved, but

digital objects could display original features (e.g. layout) and functionality available with the older software. Emulation has recently attracted attentions as a potential strategy to assist preservation, recognizing that some electronic material that is highly dependent on particular hardware and software will not lend itself to migration.

One of the benefits of the emulation strategy compared with migration is that the original data need not be altered in any way. It is the emulation of the computer environment that will change with time. This should help maintain the integrity and "look and feel" of the material. Another advantage of implementing emulation is its possible efficiency. Once the data is archived with appropriate metadata and software, no other action is required apart from media refreshing until access is desired. One emulator can also be used as a solution for several data objects requiring the same operating environment.

c) Output to permanent paper or microform

Outputting a hard copy of a digital file is a "low tech" solution that can result in a well-standardized product with a life expectancy of several hundred years. Certainly, this strategy could fix the object as a whole and preserve content and to some extent layout. However, a decreasing number of publications (flat files, printable formats) lend themselves to such methods. For example, output to paper will lead to great functional loss for hypertext documents, and cannot capture multimedia. Despite these drawbacks, a "hybrid strategy" of creating both microfilm and digital copies is gaining support as a technique for reformatting paper originals. The digital copy enhances access and functionality, and the microform copy acts as an archival surrogate.

d) Technology preservation

Another method for ensuring ongoing access to digital objects would be to simply keep older technology available for use. Although this would preserve content and enable future generations to view digital objects in their native format with original layout and functionality, creating hardware or software "museums" is prohibitive in cost, space and technical support requirements. At best this method is an interim measure when migration is not possible.

5.2 Organisational Issues

While technological issues are undeniably challenging, there are also numerous challenges, which relate to the ability of organisations to integrate the management of digital materials into their organisational structure. In addition, there is an increasing need to go beyond the confines of individual organisations, or even countries, to maximise the benefits of the technology, address issues such as copyright, and also to overcome the challenges cost-effectively The following issues are being faced, and in many cases, systematically addressed, by organisations world-wide.

i. Costs

"Part of the difficulty in understanding costs has been the lack of working examples from which to learn and the difficulty in extrapolating costs from pilot projects (of which there have been many) to full-scale public services."

As the above quote illustrates, there is a wide and potentially misleading amount of project-related data on costs which may or may not have any bearing on the costs of managing digital materials long-term. It is important to differentiate costs for digitization, for which there are more reliable cost models (though still needing to be interpreted with care), from the costs of managing digital materials, whether those materials are produced as a result of digitizing analogue materials or whether they are "born digital".

Expertise, the ability to employ and develop staff with appropriate skills is made more difficult by the speed of technological change and the range of skills needed. It is also limited by resource constraints on organisations which may well need to retain the same level of ongoing commitment to and management of traditional collections and may need to integrate commitment to digital collections without additional resources.

Because costs for both technical and organisational infrastructure are still not well defined organisations are confronted with the requirement to commit to the principle to safeguard significant digital assets, without a clear idea of the associated costs over time. This makes forward planning a somewhat more hazardous activity and one in which organisations need to begin to take action but may be unwilling to do so without more concrete assurances of costs. This requires faith that cost-effective solutions are more likely to emerge once organisations have sufficient practical experience in managing digital collections. This experience is also likely greatly to improve the prospects for effective collaboration, which is based on a shared understanding of the practicalities involved. An approach which builds incrementally on practice within the institution and collaboration with others who are confronting the same challenges will reduce risk and help develop effective strategies and practices.

ii. Expertise

The dramatic speed of technological change means that few organisations have been able even fully to articulate what their needs are in this area, much less employ or develop staff with appropriate skills. In addition, there is little in the way of appropriate training and "learning by doing" can often be the most practical interim measure. Continuous professional development will be at least as necessary for dealing with digital materials as it is for other developmental needs.

iii. Organisational structures

"In addition to redefining responsibilities of organisations, it may be necessary to redefine roles within organisations to ensure long-term access to digital information." (PADI)

The nature of the technology and dependencies in the preservation of digital materials are such that there are implications for organisational structures. Organisational structures tend to be segregated into discrete elements for the efficient processing of traditional collections, but will need to cross boundaries in order to draw on the full range of skills and expertise required for digital materials

iv. Roles and Responsibilities

"Although there is continuity of purpose and value within cultural institutions, these exist alongside a fundamental examination of roles and practices."

There are some existing repositories which undertake responsibility for specific subject areas or specific formats. In the UK, for example, the Arts and Humanities Data Service and Data Archive are two examples of institutions undertaking responsibility for social science and humanities research data, while the National Sound Archive assumes responsibility for its collection of sound recordings. In addition, there is work going on in other countries to establish national co-operative models for digital preservation. In time, it is expected that these efforts in individual countries will crystallise into clearly defined roles and responsibilities where it is as obvious which institution is likely to be the major preserver of specific digital materials as it is for non digital materials. Despite these encouraging developments, at the present time the question of who should be responsible for ensuring long-term preservation is by no means as established in the digital environment as it is in the analogue environment.

Even when it has been determined which organisations will undertake to act on their long-term digital preservation responsibilities the environment will demand far greater engagement with a much larger group of stakeholders than has previously been the case. Some will inevitably choose to contract out all or part of their digital preservation responsibilities to a third party provider. The lifecycle approach advocated by Beagrie and Greenstein has significant implications for the way organisations responsible for long-term preservation need to interact and collaborate with data producers and publishers and each other.

Roles are also changing within as well as between institutions. Assigning responsibility for preservation of digital materials acquired and/or created by an organisation will inevitably require involvement with personnel from different parts of the organisation working together. This can potentially present difficulties unless underpinned by a strong corporate vision which can be communicated to staff. Similarly, staff working in an increasingly electronic environment are needing to modify their role to reflect the different demands of the technology.

Finally, creators of digital materials need to be able to understand the implications of their actions in terms of the medium to long-term viability of the digital material they create. Whether it be a record created during the day-to-day business of the department, a digital copy of analogue collection material, or a "born digital" resource, guidance and support as well as an appropriate technical and organisational infrastructure will assist in facilitating greatly improved prospects for efficient management and preservation.

iv. Selection

"In the network environment, any individual with access to the Internet can be a publisher and the network publishing process does not always provide the initial screening and selection at the manuscript stage on which libraries have traditionally relied in the print environment." (National Library of Canada 1998)

The enormous quantity of information being produced digitally, its variable quality, and the resource constraints on those taking responsibility to preserve long-term access, makes selectivity inevitable if the objective is to preserve ongoing access. In the digital environment, it is possible to by-pass the traditional distribution channels, as well as filtering and quality control processes. While there are benefits for users in terms of swift access, there are also difficulties in terms of quality control. Selecting quality materials for long-term retention therefore places a burden on organisations in terms of resources and also in terms of the potential impact of selection.

With traditional collections, lack of selection for preservation may not necessarily mean that the item will be lost, allowing for a comfort zone of potential changes in criteria for selection at a later stage. No such comfort zone exists in the digital environment where non-selection for preservation will almost certainly mean loss of the item, even if it is subsequently considered to be worthwhile.

In cases where there may be multiple versions, decisions must be made in selecting which version is the best one for preservation, or whether more than one should be selected. Sampling dynamic resources as opposed to attempting to save each change, may be the only practical option but may have severe repercussions if the sampling is not undertaken within a well-defined framework and with due regard to the anticipated contemporary and future needs of the users.

Some consideration also needs to be given in the selection to the level of redundancy needed to ensure digital preservation. A level of redundancy with multiple copies held in different repositories is inherent in traditional print materials and has contributed to their preservation over centuries. Although in a digital environment a single institution can provide worldwide access and accept preservation responsibility, it remains an issue of concern to many that a level of redundancy should exist in the digital environment.

Such concerns need to be balanced against the potential cost in duplication of effort. Either scenario points to a greater level of overt collaboration in selection between institutions to preserve electronic publications. In any scenario, it will be critical to establish sustainability and unequivocal acceptance of responsibility to avoid the danger of losing access over time. There still needs to be assurance that preservation responsibility will be undertaken, and a clear understanding of who will undertake that responsibility and for what period of time. Otherwise there can be no guarantee that, even if several copies are stored in various repositories, all of those repositories might, for a variety of reasons, cease maintenance of the digital object at some point.

Finally, in all successful preservation strategies it may well be necessary to repeat steps in the selection process, with appropriate documentation, as part of the long-term cycle of actions to maintain access in new technological environments.

5.3 Legal Issues

Compounding the technical challenges of migrating digital information is the problem of managing the process in a legal and organizational environment that is in flux as it moves to accommodate rapidly changing digital technologies." (Waters and Garrett 1996)

i. Intellectual property rights (IPR) and preservation:

Copyright and other intellectual property rights (IPR) such as moral rights have a substantial impact on digital preservation. As outlined in Technological Issues the preservation of digital materials is dependent on a range of strategies, which has implications for IPR in those materials. The IPR issues in digital materials are arguably more complex and significant than for traditional media and if not addressed can impede or even prevent preservation activities. Consideration may need to be given not only to content but to any associated software. Simply copying (refreshing) digital materials onto another medium, encapsulating content and software for emulation, or migrating content to new hardware and software, all involve activities which can infringe IPR unless statutory exemptions exist or specific permissions have been obtained from rights holders.

As both migration and emulation will involve manipulation and changing presentation and functionality to some degree (especially over any period of time) important issues of principle and practice are raised in negotiations. It is important to establish a dialogue with rights holders so that they are fully aware of these issues and the actions and rights required to ensure the preservation of selected items.

What is different about IPR and electronic materials?

Traditional materials are relatively stable and well-established legal and organisational frameworks for preservation are in place. This is not the case for electronic materials. Digital materials need consideration of both content and also hardware and software, and require very different methods of preservation

The duration of IPR in electronic materials will often extend well beyond commercial interests in them and the technology, which was used to generate them. Long-term preservation and access may require migration of the material into new forms or emulation of the original operating environment: all of which may be impossible without appropriate legal permissions from the original rights owners of the content and underlying software.

ii. Access and security

Some of the additional complexity in IPR issues relates to the fact that electronic materials are also easily copied and re-distributed. Rights holders are therefore particularly concerned with controlling access and potential infringements of copyright. Technology developed to address these concerns and provide copyright measures can also inhibit or prevent actions needed for preservation. These concerns over access and infringement and preservation need to be understood by organisations preserving digital materials and addressed by both parties in negotiating rights and procedures for preservation.

iii. Business models and licensing

Consideration of the business models for dissemination of electronic materials and the range of stakeholders also impacts on IPR and preservation. In most cases electronic publications (particularly electronic journals) are not physically owned by the subscribers, who license access from the publisher. Subscribers are therefore concerned that publishers consider the archiving and preservation of these works and include archiving and perpetual access to back issues in licensing of these works.

iv. Stakeholders, contract and grant conditions, and moral rights

Electronic materials are the result of substantial financial investment by public funds (e.g. research councils) and/or publishers and intellectual investment by individual scholars and authors. Each of these stakeholders may have an interest in preservation; the archiving organisation will need to acquire permissions from them to safeguard and maximise the financial investment or the intellectual and cultural value of the work for future generations. Such interests may be manifested through contract, licence, and grant conditions or through statutory provision such as “moral rights” for the authors.

v. Privacy and confidentiality

Information held within the repository may be subject to the Data Protection Act or similar privacy legislation protecting information held on individuals. Information may also be subject to confidentiality agreements. Privacy and confidentiality concerns may impact on how digital materials can be managed within the repository or by third parties, and made accessible for use.

vi. Investment in deposited materials by the repository

Holders of the material over many decades will almost certainly need to invest resources to generate revised documentation and metadata and generate new forms of the material if access is to be maintained. Additional IPR issues in this new investment needs to be anticipated and future re-use of such materials considered.

Where a depositor or licensor retains the right to withdraw materials from the archive and significant investment could be anticipated in these materials over time by the holding institution, withdrawal fees to compensate for any investment may be built into deposit agreements.

6. CONCLUSION

There is a need for further research has been recognized and appropriate strategies are being tested but technology will continue to evolve and will continue to raise new issues. It may well be that there will never be single definite strategy appropriate to different categories of digital materials may need to be employed. In this way a parallel can be drawn with paper environment that also utilizes range of preservation strategies (de-acidification, microfilming, appropriate storage and handling).

7. REFERENCES

1. Rajiv Gupta, Preservation of digital information: some aspects, IASLIC Bulletin, 48 (4), 2003 p206-212

2. Lynch, Clifford A., Integrity of Digital Information: Mechanics of definitional issues. Journal of the American Society for Information Science, 45, 1994

3. Deegan, Marilyn. Digital futures: Strategies for the information age, London: Library Association, 2002.

4. Raitt, David., Libraries for the new millennium: Implications for managers London: Library Association, 1997.

5. http://www.dpconline.org/graphics/digpres/index.html (accessed on 21st Sep 2005)

6. http://www.kulisfd.cjb.net (accessed on 23 sep 2005)

About Author

M V Satish is currently working as a Library Trainee in J R D Tata Memorial Library, Indian Institute of Science, Bangalore

Email : satishmv@library.iisc.ernet.in

M Umesh is currently working as a Library Trainee in J R D Tata Memorial Library, Indian Institute of Science, Bangalore

Email : umesh.iisc@gmail.com