

Workshop on
Information Resource Management
13th-15th March, 2002
DRTC, Bangalore

Paper: EB

Archiving the Information Resources in the Digital Era

S.Shanmuga Prabha

Documentation Research & Training Centre

Indian Statistical Institute

Bangalore- 560 059

email:ssprabha_78@hotmail.com

Abstract

As we know, the growth of Information Resources in this electronic age is enormous and the demand for the access of such resources also has been increased by the public. So we, the library professionals have the responsibility to manage the available information in an ever-accessible way by the user community. Archiving is the term called for managing the availability of information resources. This article gives an idea about digital archiving, its benefits, barriers in archiving, etc.

1 Introduction

The principles and tenets of archival practice have evolved over the centuries and have gained international acceptance and application, enabling archivists and archival institutions to acquire, process, describe, preserve and make archives accessible to the public and to researchers. A preserve and make archives accessible institutions also exist to service the wide diversity of records creating, entities, such as central and local governments, parastatals and private sector organizations as well as individuals, non-governmental and non-profit-making organizations.

The preoccupation of archival institutions over the last few centuries has been with records and archives in paper form. The post Second World War period, however, has dramatically changed the scenario, not only enormously increasing the quantity of information generated, but also introducing a multiplicity of mediums in which information is recorded and stored. Major developments in this regard have been the introduction of information in digital form, the advent of the multi-media environment providing a linkage between data, still images, video and sound, and the rapid advances made in communications technology (1).

2 What is ‘Archiving’?

According to the UK’s CEDAR project, the definition of Digital Archiving is as follows:

‘Storage, maintenance, and access to a digital object over the long term, usually as a consequence of applying one or more preservation strategies’, whereby a ‘digital object’ in this sense, is a ‘any resource that can be stored or manipulated by a computer and can be applied to digitised and born-digital material” (2).

3 Who does the Archiving?

There are at least three possibilities:

3.1 ELECTRONIC JOURNAL PUBLISHERS

Many publishers do archive the electronic journals they publish. The January 1994 ARL (Association of Research Libraries) survey indicates that 28% of the libraries rely on the publisher for archiving. The problem here is that many observers feel the publisher cannot be dependent upon for permanent archiving.

3.2 LOCAL LIBRARIES

Libraries could do the archiving themselves. Traditionally archiving of print journals is done by libraries when they bind them so the archiving of electronic journals by libraries would be consistent with this precedent. Thirty-eight percent of ARL libraries reported archiving electronic journals themselves.

3.3 ON A COOPERATIVE REGIONAL OR NATIONAL BASIS

During the last decade there has been more talk than action about cooperative collection development on the part of libraries. As a variation on this theme there may be significant potential for cooperative archiving electronic journals. Twenty-six percent of ARL libraries reported depending on consortiums for archiving. One should mention the CICNet project which seeks to archive all the electronic journals that are available for free on the internet (3).

4 Which electronic journals should be archived?

Not every print journal has enough enduring contributions to scholarship to warrant preservation through binding. Likewise an electronic publication might not have enough scholarly merit to justify permanent archiving. Constantly, part of the collection management responsibility in regard to electronic journals involves deciding which journals should be archived and which should not be (3).

5 In which form it should be archived?

As we are dealing with a variety of born-digital or (to be) digitised documents or a special form of these documents or resources called publications, it makes sense to distinct the

following categories of documents (2):

- off-line printed (paper, microfilm)
- off-line electronic or discrete physical digital media, such as diskettes, tapes, CD-ROM, DVD
- e-books (online distributed, but used offline)
- hybrid electronic documents (discrete documents which contains links to online material) and
- online documents available on the Internet or via proprietary networks.
 - Online documents can be further characterized by their 'fixity'
 - static documents (form and content is recognised as substantially fixed throughout at the point of publication and throughout its lifecycle)
 - cumulative resources, whereby 'fixed' content is added to throughout its lifecycle (broadly the equivalent of a serials publication) and
 - dynamic resources, those whose form and/or content change continuously or 'dynamically' throughout its lifecycle.

6 What is the strategy of 'Archiving'?

The first step in setting an archive strategy is to define the requirements. There are really two factors to consider: the nature of the data; and the frequency and mode of access (4).

Taking data first, a distinction can be made between complete independent entities such as images or text documents, and database records that may be relatively meaningless without links to associated data. Then on the access front, the most obvious distinction is by frequency of access. At one extreme is information that is virtually never accessed, where the motive for going digital is to save space. In this case, a storage medium that is slow to access can be chosen, with the emphasis being on immutability and durability. On the other hand some archived data, particularly when relatively young, may be accessed quite a lot and so needs to be retained on a faster storage medium, possibly disc and certainly some form of immediately online storage.

The mode of access should also be considered, such as whether users want the ability to search the data and download extracts on demand, or print off copies of documents.

There is also the size of the archive to consider. For smaller records, optical-based storage systems such as CD-Rom or Write Once Read Many (Worm) drives are ideal because of their high reliability and durability. For very high data volumes, tape systems are more convenient and cost effective, because of their greater capacity and lower unit storage price. But tape is not suitable when high-speed online access is required.

One archive that has to satisfy the worst of all cases is the US Government's National Satellite Land Remote Sensing Data Archive, comprising data supplied largely by NASA. It currently holds 120 Tbytes, but this is expected to increase to 2,400 Tbytes by 2005. It is one of the world's largest archives of calibrated data, and certainly the largest continuously available online. With terabyte drives expected to be available by 2005, it will still need 2,400 units to serve its users.

7 How to maintain the physical medium of Digital Archives?

While digital records are fragile and delicate and their preservation demands more stringent, in essence, the principles for their preservation are similar to those of the traditional paper records (1).

Digital records need to be stored in a controlled and regulated environment that minimizes fluctuations in temperature and humidity. They should be protected from sunlight and heat, from dust and dirt, and from attack and infestation by insects and rodents. Because of their composition, they should be stored away from magnetic fields and from water and damp, and the disks will need to be rotated and cleaned frequently. These are fairly standard requirements, although in the case of digital records their absence is more damaging.

8 What are the advantages of Archiving?

The electronic documents' archival, storage and delivery is beneficial to the libraries as follows (5):

1. The productivity is immediately increased, notably the bibliographic processes (the document analysis), and there is a better dissemination of the primary documents.

2. The services is highly improved through a quicker access. Every library patron is able to consult the same information at the same time (information sharing). The documents cannot be unavailable any more because they are already in use.
3. The preservation conditions are optimal. A fragile document (press clips) may thus be made available. The space needed for the documentation center or the library decreases dramatically.

9 What are the disadvantages of Archiving?

There are also other problems relating to the archiving of digital information (1).

9.1 NATURE OF THE DIGITAL INFORMATION

The nature of digital information, however, also makes it a complex and difficult media to archive. Because digital information, by its very nature is open to manipulation, alteration, reformation and erasure, and because those who create and use digital information usually have only consideration for the current use and value of the information, it is essential that the archivists become involved at the point at which digital information being created. In a typical digital environment, records only survive immediate usage if the creators feel they have still further use, or if procedures have been introduced and are in place for the retention and transfer of the records to semi-current and non-current status.

9.2 IDENTIFYING AND DEFINING THE RECORDS BEING CREATED

Archiving of information is heavily dependent on being able to identify and define the records being created. The “Records series”, for instance, is a fundamental and underpinning concept in archival management. In an electronic environment dominated by digital information, however, there are difficulties in identifying and defining the records series. This is so because information technology effectively distributes and decentralizes the information creation process by enabling various units and individuals within an organization to generate their own offices. The immemorial enforcement of traditional registry practices consequently collapses, and it becomes difficult to get all the records creators to identify and reference information as per the file classification systems. Admittedly, information

technology provides mechanisms, which index and access. This, however, does not negate the need to identify the records series as this has consequences in terms of the selection and appraisal of the records.

9.3 DIFFICULTY IN IDENTIFYING THE ORIGINAL FORM OF THE INFORMATION

Among the other problems faced with archiving digital information is the difficulty of identifying, authenticating and preserving the integrity of the records to ensure that users find them in “original form”. The contents of a particular disk may be the result of the manipulation of previous generations of data and depending on the availability of the system documentation, the archivist can have immense difficulties in making an accurate determination. Absence or the inadequacy of supporting documentation may render the digital information inaccessible or virtually worthless. Without documentation, users will not know what software package is required, how the data was configured and tagged, and what is meant by the coding at various levels within the database.

9.4 DIFFICULTY IN PRESERVATION OF DIGITAL INFORMATION

The preservation of digital information also creates difficulties. While advances in technology are resolving issues to redundancy and inaccessibility of information transfixed into a technologically obsolescent format, there still remain serious burdens on archival institutions, in terms of the preservation and servicing requirements for the digital information and the equipment and infrastructure required both for maintenance and user access.

9.5 INADEQUACY OF STANDARDIZATION

In archiving digital information, it is essential to standardize formats and practices and ensure compatibility of accessioned data with the operating environment of the archival institution. Reformatting and conversion may have to be undertaken where necessary to service the holdings should be reduced. The archival institution must also keep abreast of changing technologies in the records creating agencies.

9.6 HIGH COST OF ELECTRONIC MEDIUM

The cost of electronic medium compared to paper, when the original is in paper form and has to be digitised. At a time when archival institutional budgets are shrinking, it is difficult to sustain programmes for the preservation of digital information because of the high costs involved (5).

9.7 PERMANENCE OF THE MEDIUM

Preservation of digital information also raises the question of the permanence of the media. While it is acknowledged that magnetic media is not as fragile and impermanent as previously feared, nevertheless, hard disks are not a final archiving media in the same way that paper is, and it is for this reason that information has to be backed up for safety and assurance.

9.8 COPY RIGHT PROCEDURES

One of the most difficult access issues for digital archiving involves Copy Rights management. What rights does the archive have? What rights do various user groups have? What rights has the owner retained? How will the access mechanism interact with the archive's metadata to ensure that these rights are managed properly? Rights management includes providing or restricting access as appropriate, and changing the access rights as the material's copyright and security level changes (6).

9.9 DIFFICULTY IN ACCESS

Access, however is a bit trickier. If the information is stored on a floppy disk, optical disk or tape, one must first have the hardware and software to accommodate the medium such as an optical reader or system with appropriate disk-to-tape drives, and the correct version of the software programme used to store the data. If the information is stored in a remote location, but is accessible through software, one must have the necessary software to locate and retrieve or download it. Additionally, one must have the necessary software to access or read

the data. This introduces many problems and questions. Electronic journals, just like print journals, are produced with a variety of software programmes that run on several different hardware platforms. As electronic journals become more sophisticated to include photos or graphics or even become multimedia, the number of formats will likely increase. Many programmes or formats for these journals will come and go, just like word processing products have done and continue to do (7).

10 What is Metadata?

All archives use some form of metadata for description, reuse, administration, and preservation of the archived object. There are issues related to how the metadata is created, the metadata standards and content rules that are used, the level at which metadata is applied and where the metadata is stored.

The majority of the projects created metadata in whole or part at the cataloging stage. However, there is increasing interest in automatic generation of metadata, since the manual creation of metadata is considered to be a major impediment to digital archiving.

11 Conclusion

The international community has a role to play in the archiving of digital information. The most urgent priority is the developments of standards for the archiving of digital information. These standards should embrace the whole spectrum or life cycle of information from the point of creation, through storage and usage to eventual disposition or archiving. The International Council on Archives could embark on this in the same way that it is developing standards for records description. Training programmers should also be developed for archivists and records managers, especially in developing countries (1).

A variety of metadata formats, content rules and identification schemes are currently in use, with an emphasis on crosswalks to support interoperability, while standardizing as much as possible. Issues of storage and preservation (maintaining the look and feel of the content) are closely linked to the continuous development of new technologies. Current practice is to migrate from one storage medium, hardware configuration and software format to the next. In addition, there are concerns about rights management, security and version control at the access and re-use stage of the life cycle (7).

12 References

1. **Mazikana (Peter)**. The challenges of Archiving Digital Information. *The International Information & Library Review*, 29, 1997, pp. 307-317.

2. **Long-Time** Archiving: Preserving the access to (Licensed) Electronic Documents.
<http://www.stk.cz/elag2001/Workshop/ws6.doc>
3. **Nisonger (Thomas E.)**. Collection management issues for electronic journals. *IFLA Journal*, 22(3), 1996, pp. 235-236.
4. **Archiving** for the digital age.
http://www.findarticles.com/cf_0/m0COW/2000_Nov_30/67641905/p1/article.jhtml?term=archiving
5. **Deschamps (M. Christine)**. The Electronic Library. *IN Bielefeld Conference, Libri*, 44(4), 1994, pp. 306.
6. **Best** practices for Digital Archiving.
<http://mirrored.ukoln.ac.uk/lis-journals/dlib/dlib/dlib/january00/01hodge.html>
7. **Luijendijk (Wim)**. Archiving Electronic Journals: The Serial Information Provider's Perspective, *IFLA Journal*, 22(3), 1996, pp. 209-210.