Managerial issues in building digital collections

Deepali Talagala
Director / Information
Unites States Information Centre
44, Galle Road
Colombo 03

e-mail: deepali@gmail.com

INTRODUCTION

For many, the term digital collections is like a magic word which brings up an image of extensive universe of information available through the Internet and accessible from home or office workstations, fulfilling the promise to provide information at any time, to any place, and for any user. Many consider the World Wide Web to be "the" digital collections. Accumulated vast collection of this kind is too unstable to be called a storehouse, however; but thinking of digital libraries merely as vast repositories of materials is also too limited. These digitized collections may be available at many places such public and private sector organizations. Digital collections can be considered as one way of preserving the objects/records created by the society and making them available for the future generations.

To enhance the value of the collection the parent organization can develop digital libraries. The Library of Congress's National Digital Library Program's "American Memory" is an online archive of over 100 collections of rare and unique items important to America's heritage. The collections contain more than 7 million primary source documents, photographs, films, and recordings that reflect the collective American memory. They are a treasure trove of unique personal items from another period in time – perhaps old records, letters with exquisite penmanship and arcane language, clothing, keepsakes, or faded photographs. These collections are 'snapshots' providing a glimpse into America's past. Digital libraries can offer a wealth of information to a large community of users and will become an increasingly valuable part of the Internet as they are developed in a coherent and imaginative way. Any organization which intends to develop a useful digital collection should understand that it would be of no use unless it is organized properly so that these can be searched and information could be obtained easily and in time saving manner. For his purpose any digital collection should have the following characteristics:

Organizing policy: There should be a coherent, organizing principle, sorting materials by topic, by type of material or by subject. Subject experts in each area can assemble the collections, which will then be combined and maintained by institutions for their clientele and the global Internet community. Ideally, different

institutions will take responsibility for building digital collections of various types and then share those collections with others, thus optimizing the advantages of the distributed nature of the Internet.

Accessibility: Although currently, many Internet sites are open, free of charge, an increasing number of sites are limiting access to subscribers or to those who pay a fee. It may also be possible to seek for continued funding and are consider licensing arrangements or user fees.

Searchability: The ability to find and use materials on the Internet is a major challenge for most users. Developing more sophisticated search engines is one approach. In addition, having a standard format for embedding metadata built into the document, will allow the search engines to retrieve information more efficiently.

Preservation: There should be a systematic effort in place to ensure continued access to the digital collections on the Internet. Information on the Internet is, in many cases, ephemeral and presents considerable challenges for archiving. Preservation must also ensure that information resources remain intact; unauthorized tampering with the content of electronic resources could have grave implications for its continued value.

Service: Many internet users may not be competent in searching on these digital collections which are usually available as databases. Few digital libraries have built services into their offerings of collections. There can be built in services where users can ask questions, and can develop FAQ (frequently asked questions) files to assist users.

WHY DIGITAL COLLECTIONS?

Members of the society create records describing its activities, and these records become resources for study and learning. Any country will need a well established infrastructure to ensure that these products of our societies are recorded, archived and stored so as to guarantee accessibility over long periods of time. Cultural institutions, libraries and museums play a vital role in organizing, preserving and providing access to these cultural, intellectual and historical resources of society. They have established formal preservation programs for traditional materials and have understood their role in safeguarding these resources. Though digitization is sometimes referred to as preservation, it is the best way to facilitate access to information. One could argue that digitization is a type of reformatting, like microfilming, but a superior form. Digitization cannot be considered as a reliable medium for preservation.

Digital collections have proved to be an excellent medium for access to information. Remote access to digital records has allowed users to study

research materials, compare the findings etc. limitlessly. Technology is enabling the governments to make available the materials of the publicly funded institutions and also the means to locate them. In addition the students can gain access to teaching materials not available physically. As Donald Waters of the Digital Library Federation have expressed it, the promise of digital technology is for libraries to extend the reach of research and education, improve the quality of learning, and reshape scholarly communication.

However presently the cultural, educational, scientific, public and administrative resources are produced in different formats such as print, audiovisual and digital. These have in turn resulted in the creation of different information products and services used today. ICT has changed forever the way information is created, managed, archived and accessed, and that digital information is now an integral part of our cultural and intellectual heritage. Due to these advances in technology the institutions that have traditionally been responsible for preserving information now face major technical, organizational, resource, and legal challenges in creating, maintaining and preserving the digital collections.

Challenges and issues

When developing digital collections we may have to address new trends that can come in the future. As such we may have to redefine the records/information management. Because in this process the **data processing becomes document management**. In many organizations, public/private computers were used either to produce financial reports using data processing. With the advent of document digitization the most of the organizations now have to improve electronic records management.

In the traditional organizational environment official records were static documents. These traditional documents when transformed into electronic records they have to be managed as dynamic documents. These electronic documents become multi dimensional, because they can be linked to other documents and multi media files.

In traditional records management, the focus was on the media the information was stored such as file folders, microfilms etc. In the case of electronic record keeping systems there may be several media formats which will run on several platforms. Therefore the electronic records management will need to link all these and provide access. Hence the main focus will be on accessibility of the information. As such the transformation of the records management from traditional to the electronic era has created many issues and challenges to the information managers.

Technological issues: Due to new technologies it has been easier to create digital collections which are beneficial to users. At the same time the rapid

changes in the development of ICT make digital media short-lived and inherently fragile. Changing hardware and software has made the task of preserving digital records much more difficult and has created a new set of challenges and issues. Traditionally preserving things meant keeping them unchanged. Digital collections are highly dependent on technologies which make them highly vulnerable to technical obsolescence. This leads to the requirement of a policy of continuous conversion, and/or migration to new formats. As such the publicly funded institutions such as archives, libraries and museums may be the organizations that can provide a long-term commitment for preservation of digital collections.

In libraries the technological revolution has created a hybrid environment including paper, microfilm, video, magnetic and optical media. This environment has redefined the expectations of the users in getting information. In the digital world users can now access information stored anywhere, but they often have no idea about the information provider, whether the data is valid, or whether the content actually appears as the original. At the same time the users want digital information to remain intact and to appear as originally created and at the same time they want to dynamically access and interact with digital content using the most advanced tools available to them. In any case today's society lacks proven methods for ensuring that digital content will continue to exist, that user will be able to continue to access that data using the most up to date tools available, and that any information accessed is authentic and reliable.

In this context the digital preservation is a major challenge. It is not only a problem for those custodians such as librarians/ information managers, who assure the maintenance, preservation and accessibility of digital collections, but it also affects vital information for long-term research and other legal purposes. Hence it is important that such collections must be preserved in a way that ensures cost-effective access over long periods of hundreds of years.

Records management: In addition digital preservation is also a consumer problem. Individuals are storing an increasing amount of personal information on digital media thinking that this will ensure the availability at any time for their future use. This is a mistaken belief since no one is able to provide such guarantees.

In any organization managing electronic records depends not only on Electronic Record Management technology, but also on employees managing their content properly before it is declared a record, or destroyed. A digital collection building involves different phases such as **Records Creation; Records maintenance; and Records Disposition.**

There should be a consistent way of creating; saving and tagging the documents which is not an easy task where there are no proper established systems. To formulate a proper system there should be standard formats, tools and metadata.

To adhere to these standards, and to develop proper document management practices it is necessary to have ongoing awareness programs, training and enforcement. For any organization embarking ERM projects time is a critical factor. The best approach to address this issue may be to break it down into different stages. At the end of each stage, and at the end of each stage to review it and find out the mistakes/weaknesses and address them so that the same mistake will not be repeated in the next stages. This will also help you to understand how long it will take to realize the benefits you are hoping to achieve from ERM. Getting adequate support for ERM within the organization is another issue.

Funding: One of the main problems with digitization and with any significant, is the high cost. It would be extremely expensive for many public organizations unlike for private autonomous institutions. The Library of Congress's "American Memory" project has cost more than 100 million Dollars. In Sri Lanka how many of our public and private organizations are capable of handling such projects unless they are funded by an international funding agency. Even with some kind of international funding many institutions miscalculate the huge amount of funding which is required to set up the infrastructure to supply networks to deliver documents electronically.

Document Selection: Apart from the enormous costs involved in any digitization project the next problem is the selection of material for digitization. In the case of libraries some would wish to digitize their unique collections such as ola leaves collection while the others would prefer to digitize the publications of their parent organization. This will help to clear some storage space which is a constant problem in many libraries. It would be useful if all the libraries could have a Register of Digitization which can be combined to form a National Register of Digitization. This will help anyone to search for digitized publications. In turn this will also eliminate possible duplication among libraries. However for this to work properly it is necessary for the libraries to have a good network with continued co-operation among them. This register also should be updated continuously to provide current information.

Authenticity: Another concern in digitization is the authenticity of the records. With the very sophisticated computer software which is used to create all sorts of images and animations it is difficult to guarantee the originality of the digitized documents. The printed original documents can maintain its authenticity by using many methods such as watermarks but with an electronic copy this will not be possible. With the advent of the new technologies so many authors are writing on a personal computer on the same subject. This makes it difficult to find which version is the first version or first edition. Some uniform guidance and general standardization at an international level would be very useful here.

Conservation: Most libraries and Museums have used the microfilm as the media for preservation of documents. There is no guarantee that the microfilms

will last long. There is evidence that microfilms start deteriorating even after about 30 to 40 years. In such cases we are faced with the problem about whether to preserve the original as it is or is it advisable to depend on the microfilm format. If we cannot rely on the microfilm version of the original can we destroy the original or do we keep the original which takes our precious space. This will lead to another issue, who decides on this and on what criteria. The criteria can be based on many factors; rarity, accessibility, fragility of the original and the usage. Even under these factors one can decide to digitize the rarely used materials to save space. The next issue is in what format are we going to conserve these, CDROMS, magnetic tapes or any other format. The life time of CDROMS is also not really known. There is evidence that early deterioration occurs even in CDROMS. Presently many publications are produced digitally and this will keep on increasing due to the advances in new technologies in which case the cost of digitizing will also reduce.

Standardization: Since there are many different software being produced, the compatibility among these has become a problem. Hence it is necessary to have some Standards in ERM software. In libraries the use of standard thesaurus to create indexing terms has helped them to maintain standards in library catalogues. Similarly there should be standards in creating Metadata. It would be helpful if Dublin Core Metadata is used to maintain standards.

There is one common factor in every digital book, i.e. each e-book has its own device (e-book reader) with a specific license.

Copyright: Each country has its own copy right laws according to which rights of the publishers and the authors are protected. Digitizing the copyrighted material has many issues because the publishers and authors might not be able to control reusing their materials. Even if these digitized materials are only accessible through costly databases there is no guarantee that they are not copied by others. Once these are accessible globally there is very little or no control over the usage.

We all are aware of the fact that there are so many pirated versions of software and other digitized products available in the open market. Concern about digitization of copyrighted information is not surprising.

Staff: The lack of competent staff and the negative attitude of the staff can be a major barrier. The advent of new technologies also created a feeling of threat among the professional staff who are not competent in handling technology. On the other hand using the expertise and knowledge of the staff within the library is a real money-saver. Staff plays a key role in ensuring the success of any digitizing project.

Digital environment also created different job categories also created interesting new work to the staff. Librarians also felt the change in their role from traditional custodian to information manager. Continuing staff development training

programs for staff has to be organized. They have also brought pressure on library schools to enhance the skills of their graduates in these areas.

Concluding remarks

The success of managing a digital collection depends on several things:

- Selection of a proper ERM system and learn more about how to do digital preservation;
- Educating the library and archive professionals responsible on collection management and preservation to be able to cope with digital material alongside the print material;
- Coordinate and cooperate with others who are involved in similar projects

These changes have also brought pressure on library schools to enhance the skills of their students in these areas. In addition, the advent of the Web made easy access to information over the Internet. With the advanced search engine technology, came a revolution in information literacy and information use. Information collections are no longer geographically bound. New forms of digital libraries and information collections are providing more information to more users more easily and on demand. Information professionals throughout the world have realized that these changes have forced them to change their roles. The information has become more valuable than ever. Due to the increase in complex products it has become very important to train users and library personnel. Using Web access, it is possible to search the OPACs of many of the world's libraries and online resources. Digitization has provided the potential for access and preservation for huge collections.

At present, however, the cost of digitization and of creating and maintaining a migration path for preserving the files is very expensive. The benefits of making an underused collection more accessible should be viewed in conjunction with other factors such as compatibility with other digital resources and the collection's intrinsic intellectual value.

As Donald Waters of the Digital Library Federation have expressed it, the ***promise of digital technology is for libraries to extend the reach of research and education, improve the quality of learning, and reshape scholarly communication.***

References:

Fletcher, Patricia T. **Electronic records management in state government: Planning for the information age.** *Records Management Quarterly*, Oct 1990, Vol. 24 Issue 4

Forde, Helen. **Access and preservation in the 21st century: What has changed?** *Journal of the Society of Archivists*, 2005, Vol. 26 Issue 2,

Menkus, Belden. **Defining electronic records management.** *Records Management Quarterly*, Jan 1996, Vol. 30 Issue 1

Sprehe, J. Timothy; McClure, Charles R. **Lifting the Burden.** *Information Management Journal*, Jul/Aug 2005, Vol. 39 Issue 4

Young, Jeanne **Electronic Records Management on a Shoestring: Three Case Studies.** *Information Management Journal*, Jan/Feb 2005, Vol. 39 Issue 1