PRESERVATION OF E-RESOURCES: ISSUES, CHALLENGES AND OPPORTUNITIES IN DIGITAL ERA

Satish Kumar,

Information Scientist,

ARIES, Nainital, UK

Email – sklisc@gmail.com

ABSTRACT

Digital preservation is a broad term used to describe the continued accessibility and maintenance of digital resources and safeguarding these for the foreseeable and the distant future. Digital preservation is a vital part of the creation and management of any digital collection. Without the appropriate preservation a digital collection can easily become inaccessible and therefore useless in just a matter of a few years.

Terms such as medium preservation and technology preservation are widely used when discussing issues related to the preservation of electronic records. The advent of digital information age introduces new preservation requirements. Medium preservation has been addressed in discussions on environmental and handling concerns for tapes, magnetic disks etc. Greater attention should instead be directed to the obsolescence of technologies. It is challenge to imagine not only how to technically preserve electronic records indefinitely, but also how to choose what to preserve and how to guarantee the electronic record's reliability and authenticity in the future. The combine problems of immense volume, unstable storage media, and obsolete hardware and software add up to some very tough problems, which have to be dealt with. Digital preservation is becoming a business issue. Not only are historians, librarians and archivists alarmed by the loss of cultural and government records due to a lack in digital preservation, but certain industries have also realized that they need to keep data longer and longer for regulatory or business reasons.

This paper discusses long-term archiving and long-term access to digital documents, with an emphasis on criteria for selection. Selecting materials for digital preservation depends on whether the materials are both valuable and endangered, whether appropriate digitization procedures and standards for these materials exist, and whether copyright allows reasonable access for educational and research purposes.

Keywords: *Digital information, Digital libraries, Digital archiving, Digital preservation, LOCKSS, Long-term preservation, E-resources.*

INTRODUCTION

The world's biggest digital library is internet. We cannot develop a digital library on par with internet. Digital library is very complex system. A digital library can have multi-tier architecture. Different digital libraries follow different architectures and models.

Digitization technology offers the facilities to preserve documents and make them in an easily accessible system. The traditional library problems like conservation, preservation, storage and space can be solved by digitization. The difficulty and expense of preserving digital information is potential impediment to digital library development. Greater attention should be directed to the obsolescence of technologies. It is challenge to imagine not only how to technically preserve electronic records indefinitely, but also how to choose what to preserve and how to guarantee the electronic record's reliability and authenticity in the future. The combine problems of immense volume, unstable storage media, and obsolete hardware and software add up to some very tough problems, which have to be dealt with.

The goal of digital preservation is to maintain the ability to display, retrieve, and use digital collections in the face of rapidly changing technological infrastructures and elements in this modern electronic environment. Digital preservation is a broad term used to describe the continued accessibility and maintenance of digital resources and safeguarding these for the foreseeable and the distant future.

This paper provides a brief overview of the main issues in implementing digital preservation technique in libraries, highlighting current research and standard efforts. It also looks into the aspects involved in the digital preservation and discusses some of the techniques required for preservation of e-resources. It gives a bird's eye view of various projects initiated for archiving internet and digital content of scholarly journals.

DIGITAL PRESERVATION

Preservation for posterity has always been one of the main responsibilities of libraries. Traditional preservation has been an extension of the custodial function, under which importance was given to preservation of the 'physical form' – manuscripts, books, sculptures, gramophone records, etc. In the digital world, this traditional preservation concept is undergoing a transformation, due to the special features of digital resources.

The UK CEDAR project has given the following definition of 'digital preservation': "Storage, maintenance and access to a digital object over the long term, usually as a consequence of applying one or more preservation strategies."

Digital documents are created using computer systems, and depend on them for continued access. Hardware and software systems become outdated very rapidly, leading to technological obsolescence. Digital preservation, therefore, involves taking measures to ensure that the digital documents will remain available, usable and authentic in the future when the applications and systems, which were used to create and interpret them, might no longer are available.

Preservation of digital resources is much more complex than that of non-digital material such as paper. For example, when a book is preserved, all aspects of the book are preserved – its physical presence, format, layout and contents, since all these individual elements are inextricably linked. Digital formats on the other hand can be easily separated into individual elements, hence more efforts are needed to preserve them as a whole. For example, one can retain the content of the digital document while losing the layout (converting HTML to PDF format); or one can maintain its physical presence i.e. the computer file, but fail to preserve its readability. One can think of the following

different levels at which digital resources can be preserved:

- ***Preserve the physical presence***: Also known as 'bit preservation', under which the computer file is preserved.
- ***Preserve the content***: Since actual information content is the most crucial component, one can provide access to it at its lowest level such as ASCII text, without the embellishments such as font variations or layout features.
- ***Preserve the presentation***: The original format or layout may include different font faces and sizes, columns, margins, the use of white space etc. that carry out functions such as emphasizing important parts of the text, improving clarity, etc. In many formats of digital documents (e.g. SGML, XML), the layout specifications are separate from the content. To maintain the original look of the document, these layout specifications must also be preserved.
- ***Preserve the functionality***: Special efforts are required to preserve the functionality of digital documents, which might contain multimedia components, exist in hypertext format and have navigation functions such as toolbars, keyword search or interactive tables of contents.

Digital preservation is the series of actions and interventions required to ensure continued and reliable access to authentic digital objects for as long as they are deemed to be of value.

In order to preserve materials on a scale commensurate with mass storage capabilities and in formats that are accessible and usable, it is necessary to articulate some basic requirements.

- ***User's perspectives***: User's expectation is always changing, yet users, especially research scholars need both traditional documents and electronic documents of old information and current information.
- ***Institutions's responsibility***: Libraries, archieves and other custodians have responsibility for their any properties. So institutions should plan for digital materials including their maintenance, preservation and distribution.
- ***Mission of parent institution***: First object of libraries, archives and other custodians is to satisfy the user's expectation and user's requirements. They should preserve all materials in all formats.
- ***Storage media***: Storage media is having different formats such as text, data graphics, video and sound, different storage capacity like floppy disk, CD-ROM (R/R-W), DVD ROM (R/R-W), Pen drive, External Hard disk drive etc.

PRESERVING INTERNET

The USA based *Internet Archive* launched in 1996 by **Brewster Kahle**, the dot.com multi-millionaire who wanted to back-up the internet, the project very publicly demonstrates what can be done. His internet library quickly became the largest publicly-accessible, privately-funded digital archive in the world, and currently includes books, audio and film collections, cartoons and software. However, the one service of Internet Archive that continues to draw widespread attention is the **Wayback Machine**, which enables users to see archived versions of websites, dating back to 1996. Admittedly, the service has only captured a fraction of the internet's history – but it is still more than 150 billion pages from millions of websites. The Internet Archive made it its mission to broadly harvest the web and do its best to present it back to users.

And it's not just web-harvesting organizations that are scrambling to stay abreast of today's changing digital content. *For example* an article could be published by, say, Springer, and the data that

supports it could be housed at a university or an aggregator or a data-service organization, so how do we preserve the links between the published article and the data, housed in a different places? The not-for-profit organization, **Portico** is investigating research projects to address these very questions. It preserves tens of thousands of scholarly e-journals, digitized historical collections and, more recently, e-books. Instead of providing an outside link now Portico deal with more and more complex content, dynamic forms of content, such as data and multimedia.

NEED FOR DIGITAL PRESERVATION

Librarians currently face many issues and concerns for archiving electronic journals, such as differences between digital and print media, rapid obsolescence of digital technology, shift in the responsibility of archiving to publishers, legal issues such as copyright and intellectual property rights, selection, and many more.

As defined by **Jones and Beagrie** (2002), digital preservation refers to the series of managed activities necessary to ensure continued access to digital materials for as long as necessary. The ease with which digital information can be created, combined with the huge increase in computer power and network bandwidth, has led to the proliferation of a vast amount of '**born-digital'** data, especially in science and engineering, where petabytes (10" 15 bytes) of data are being generated by scientific instruments on a daily basis. This data deluge has forced many to address the issue of long term preservation and curation so as to ensure that data generated today can survive the changes of technology and can be accessed in the future.

Publishers have adopted **licensing models** that require continued subscriptions to ensure continued access to digital content. Often, a library is only ***"renting***" the content and loses access to it once they stop subscribing. In other cases, a library ***"owns***" the content during the subscription period, but must pay an annual fee to the publisher to continue accessing the content online once the subscription is canceled. A library facing difficult budget decisions often cannot afford to pay this fee, and so loses access.

An alternative offered by some publishers is to send the library **CDROMs** with their back content. However, the content is usually not in a format that the library can easily make available to their users. In all these cases, the results are the same: ***users lose access to digital content they rely on***. To guard against this, some libraries also continue their print subscription for critical material as a backup, and warehouse it, an expensive but sometimes necessary option.

The challenge for digital preservation is not just the volume of data. The hardware and software used to store and access digital information are constantly upgraded and superseded. Technology obsolescence is generally regarded as the greatest technical threat to ensuring continued access to digital material.

Projects like **LOCKSS** (***Lots of Copies Keeps Stuff Safe)*** from Stanford University have gone a long way toward making digital preservation a reality (Reich and Rosenthal, 2001).

The **National Library of the Netherlands** or **Koninklijke Bibliotheek (KB)** has become the main library to which several for-profit publishers made an agreement with. After the agreement with Elsevier, who was the first, the KB concluded similar agreements with Kluwer Academic Publishers (2003), BioMed Central (2003), Blackwell (2004), Oxford University Press (2004), Taylor & Francis (2004), Sage (2005), Springer (2005), and Brill Academic Publishers (2005).

All KB agreements dictate that the KB will preserve what the publisher sends to the library. The archived content is exactly the same as the published content. This coverage may change as publications become more complex and include multimedia and dynamic content. For now, however, the KB's policy is to preserve "**as is**." As a part of the agreements, the KB

provides on-site access to the journals on a current basis to all on-site authorized library users. Like the KB initiative, there are many other related services such as-

- PORTICO,
- LOCKSS (Lots of Copies Keep Stuff Safe) and
- CLOCKSS (Controlled LOCKSS).

SELECTION FOR PRESERVATION

"Selection" is another important issue in electronic preservation. The huge quantity of information being produced digitally, its variable quality, and the resource constraints on those taking responsibility to preserve long-term access make selectivity inevitable for archiving.

Traditionally, lack of selection for preservation may not necessarily mean that the item will be lost, but in the digital environment non-selection for preservation will almost certainly mean loss of the item. Although not all resources can or need to be preserved forever, some will not need to be preserved at all, others will need to be preserved only for a defined period of time, and a relatively small sub-set will need to be preserved indefinitely. Selecting materials for digital preservation depends on three criteria:

- Whether the materials are both valuable and endangered;
- Whether appropriate digitization procedures and standards for these materials exist;
- Whether copyright allows reasonable access for educational and research purposes.

However, just as the web allows anyone to publish information on it without having to obtain any formal permission, it also allows its removal. The library must capture relevant free as well as paid web-based resources whenever they are available, in order to safeguard against their possible loss from the web, so that they can provide continued access to these resources to its users. Though libraries will have to cater to remote access to resources (online journals, online full-text databases etc.) as a part of collection development, it would be desirable to download and preserve selected core resources on local servers.

Recently, Portico has signed up an increasing number of smaller, often society-run journals. Many of these journals are not on the kinds of standard platforms that other publishers use, so files are not in a standard format. So preserving content from a smaller publisher is much more expensive per title. *For example*, an organization might write a preservation system for, say, Elsevier's publications, and preserve thousands of titles. Do the same for a smaller publisher and it only preserves a few titles.

DIGITAL PRESERVATION: ISSUES, CHALLENGES AND OPPORTUNITIES

Issues & Challenges

Traditionally, special libraries have retained issues of scholarly journals, either individually or collaboratively, providing their readers access to the content received even after the publisher has ceased or the subscription has been canceled. In the digital age, libraries often subscribe to journals that are only available digitally over the Internet. Although convenient for patron access, the model for digital subscriptions does not allow the libraries to retain a copy of the journal. If the publisher ceases to publish, or the library cancels the subscription, or if the publisher's website is down for the day, the content that has been paid for is no longer available.

The first goal is a fundamental requirement that must be addressed by any digital repository. The second goal means that repositories need to support searching and retrieval to improve access to information. This is an area which has received much attention in the last few years and has been greatly facilitated by standards, such as the **Open Archive**

Initiative Protocol for Metadata Harvesting (OAI-PMH), allowing service providers to create discovery services across repositories by recurrent metadata harvesting. The third and fourth goals are what digital preservation needs to achieve. It is not enough just to keep the original bit-stream that represents the information stored in a digital object. The challenge is to make sure that users can access the content that has been subscribed by the library in the past, despite publisher and vendor changes over time.

Some issues and challenges for implementing digital preservation are -

- Technical issues and challenges related to digital preservation in library include a lack of practical implementations of preservation standards and a lack of technical knowledge, in general, of what information is required to support the digital preservation process within the library.

- Digital preservation is a complex process and there are many unsolved organizational, managerial and technical issues that make digital preservation a challenging task for library subscribing electronic resources. The focus of many library activities to date has been on subscribing resources, giving access and promoting but the challenge is to make sure that users can access the content that has been subscribed by the library in the past, despite publisher and vendor changes over time.

- Digital preservation requires new workflows, new skills and close co-operation across different professions ranging from traditional preservation management skills to computing science. The organizational structure to support this is not yet in place.

- The fact that many institutional libraries rely on short-term funding makes it hard to make long-term subscription. The situation is further complicated because costs for preservation are, in general, difficult to calculate and are poorly understood and it is difficult to segregate costs for preservation from costs for access.

- Another challenge is to maintain the balance between ease of deposit and the need for preservation. The costs and risks associated with digital preservation tend to grow when a digital collection includes a large number of diverse file formats.

- The technology on websites changes so rapidly. There are different tools to create websites and different technologies used to present information and these are changing year after year. If we are trying to capture this information coherently we really have to focus on who we are preserving it for, and why.

Bram van der Werf, executive director of the not-for-profit **Open Planets Foundation (OPF),** a UK-based forum for digital preservation, agrees, and believes rapidly changing and complex content is a major preservation challenge. Until a few years ago, digital objects, such as Word and Excel files, were relatively static, and, over time, libraries and archives have learned to deal with these,' he said. 'But people are now increasingly publishing and disseminating information on websites and blogs. We are no longer dealing with static objects but, instead, highly-interactive digital expressions.'

To make matters worse, today's digital content is also highly dependent on the latest "device", be it a tablet, iPhone or iPad, as well as its flavour of operating system and even the latest app. 'These technologies have a very short life cycle and the speed at which they regenerate is so fast, often less than a year,' observed van der Werf. 'So the real issue is that the objects we want to keep are

becoming increasingly complex and dependent on lots of technologies that we only have very limited control over.'

Opportunities

Stanford University-based, LOCKSS (Lots of Copies Keep Stuff Safe) is, fundamentally, a web-harvesting organisation, set up to preserve libraries' electronics materials across its distributed network. From the start they have been preserving large amounts of audio and video databases, HTML, XML, and PDFs. LOCKSS were set up to preserve academic e-journals exactly as they were presented on publisher websites, so the mime extension or file format really hasn't mattered to the LOCKSS program.
But harvesting and archiving websites is growing in complexity, with preservation becoming increasingly difficult to manage as technology on websites rapidly changes. To preserve the author's words and presentation of his or her material is an increasing technical challenge as the web is evolving from a document model to a programming environment. This is a general web-preservation problem; with new technologies such as Web 2.0 and HTML5, published content is increasingly dynamic.

To tackle this, LOCKSS won a grant from the **Andrew W Mellon Foundation** in April 2012 to develop new ways to gather and preserve some dynamic digital content. According to Victoria Reich, executive director at LOCKSS, a key challenge has been to capture the richness of content presented on the web. However, the organization has made progress and released its first Open Source software addressing this problem in April 2013.

For its part, the OPF has also established several new projects to deal with recent issues. *For example*, **SCAPE**, cofounded by the European Commission and led by the **Austrian Institute of Technology**, is developing open-source, scalable tools for the automated preservation of complex, multi-terabyte-sized data sets. The project is collaborating with global, open-source initiatives to help deal with scalability issues, including access to super-computers as well as existing technologies to identify and characterize files.

International Internet Preservation Consortium – **www.netpreserve.org**, for preserving United Kingdom websites, **UK Web Archive** was developed by British Library in 2004.

LOCKSS (LOTS OF COPIES KEEP STUFF SAFE)

LOCKSS project, under the auspices of Stanford University, is a peer-to-peer network that develops and supports an open source system allowing libraries to collect, preserve and provide their readers with access to material published on the Web. The system attempts to replicate the way libraries do this for material published on paper. It was originally designed for scholarly journals, but is now also used for a range of other materials. *Examples* include the **SOLINET** project to preserve theses and dissertations at eight universities, US government documents, and the **MetaArchive** Cooperative program preserving at-risk digital archival collections, including Electronic Theses and Dissertations (ETDs), newspapers, photograph collections, and audio-visual collections.

INTEGRATING DIGITAL PRESERVATION IN LIBRARY

Long-term preservation is an important mission for a digital preservation system in a library. It ensures that the digital collection the library builds will be preserved for future users, regardless of the vagaries of institutional funding and decisions made by publishers. However, that is only half of the story. Unless users can access preserved content when it is no longer available from the publishers, it is like locking print media in a sealed vault with no key for safekeeping. The content might be safe, but it is certainly not useful.

MISSION

The other mission for a library's digital preservation system is to make the preserved content available whenever it is needed. That may occur at any time for one of many reasons: there could be a temporary interruption of internet access, the publisher may have blocked access to the library because of a mistake or a dispute, or the library may simply have canceled their subscription and permanently lost access. Regardless of the reason, the digital preservation system must be ready immediately to serve its content in place of the publisher's.

STRATEGY/TECHNIQUES

One strategy for making preserved content available is for end-users to interact directly with the digital preservation system, since most systems include a web user interface. There are several problems, however. The first is the uncertainty of users having to decide whether to access the publisher or the preservation system. The second is the need for users to learn a different interface than the one they are accustomed to for accessing the content. The third is the risk of giving users direct access to a digital preservation system that the library relies on for long-term preservation.

Another strategy that does not have these problems is to integrate the digital preservation system with the other library systems, and provide controlled access through the same interfaces that users are already familiar with. For most libraries, the **online public access catalogue (OPAC)** is the main way users access digital content.

Accessing content based on bibliographic data is known as "**context-sensitive linking**". Providing this capability from a library's OPAC requires a database of online targets for every publisher, title, and article held by a library. Creating and maintaining such a database is beyond the capabilities of most libraries, and even most OPAC vendors. In the late 1990s, **Herbert Van de Sompel**, a librarian at the University of Ghent, created a new kind of software system called a "**link resolver**" or "**link server**" that provides this service.

Different strategies are being used to carry out preservation of digital resources. Some of the significant ones are:

- **Migration:** Under this method, there is a periodic transfer of digital materials from one hardware/software configuration to another. For example, file formats are converted into compatible new formats as soon as the original formats face the risk of becoming obsolete (e.g., HTML 3.2 to HTML 4.02 or .doc to .docx)

- **Emulation:** It involves retaining information about how a digital object was created and accessed so that future access can be accurately and authentically reproduced. Emulation retains the functionality, look and feel of the original document.

- **Technology Preservation**: This involves preservation of the technical environment by conserving the software and hardware needed for interpreting digital information.

- **Conversion of data to standardized format:** Data could be transferred to ASCII format. It retaisn information, but not the structure or functionality.

- **Universal Virtual Computer (UVC):** One of the latest techniques to be explored is the UVC-based preservation method, which allows digital objects to be reconstructed in its original appearance anytime in the future, a programme or emulator is written to carry out this interpretation in the machine language of the UVC at the time the record was archived (without requiring any knowledge of the future target machine). The data can be stores in any format and the knowledge required to

decode it is encapsulated in the UVC programme (format decoder). The UVC concept consists of the UVC itself, a logical data scheme with type description, the UVC programme and the logical data viewer.

INTEGRATING LOCKSS WITH A LINK RESOLVER

The LOCKSS (Lots of Copies Keep Stuff Safe) digital preservation system was developed under the auspices of the Stanford University Libraries. The LOCKSS system is open source software that can be installed and run on a local computer by any library. It enables libraries to take local custody of their digital content from participating publishers.

Local custody is critical to long-term preservation. It separates payment from access, and insulates libraries from future price increases for preservation and access to content they already own. It enables libraries to build and preserve real digital collections, with guaranteed perpetual availability, in the same way that they do with print media. This includes subscription and open access materials, e-journals, e-books, and a wide range of other web delivered content.

In addition to on-going development of the LOCKSS software, the LOCKSS Program at Stanford University also operates the Global LOCKSS Network (GLN), which is funded by LOCKSS alliance memberships. A modest annual fee provides member libraries with technical assistance, and pays for ongoing work to make additional publishers, titles, and volumes available for preservation. If a library discontinues their membership, they retain the software and all the content on their local LOCKSS box. Over 500 publishers with over 9,000 journal titles have selected the GLN as their digital preservation and post-cancellation partner.

The LOCKSS software is also used to preserve a number of specialized digital collections. These Private LOCKSS Networks (PLNs) are operated by outside organizations that want to ensure long-term preservation of their digital content.

Integrating the LOCKSS system with a link resolver is an effective way to make content preserved in a library's local LOCKSS box available to users. Most OPACs already use a link resolver to provide access to the library's online content. Adding the library's LOCKSS box as a content target provides users with another way to access content.

THE OPEN ARCHIVAL INFORMATION SYSTEM (OAIS) REFERENCE MODEL

OAIS is a reference model for archiving communities and most of the archiving projects are based on OAIS model For instance, PANDORA, LOCKKS, Chronopolis project, etc. It specifies how digital assets should be preserved for users from the moment digital content is ingested into digital storage area, through various preservation strategies to the creation of dissemination packages for end users. The Open Archival Information System (OAIS) reference model has been widely adopted and used to inform the development of preservation tools and repositories. OAIS was an initiative started by the **Consultative Committee for Space Data Systems** (CCSDS) of the **National Aeronautics and Space Administration (NASA).** The OAIS model was adopted as an ISO standard (**ISO14721:2003 OAIS**). This model comprises four components- ***producers, consumers, management*** and the ***archive***.

The producers generate the information and submit it to the archive for preservation. The consumers use the preserved information. The management is responsible for laying down explicit policies and objectives of the archive. It is not responsible for day to day administration of the archive. This management is accomplished by a functional entity within the archive itself.

The OAIS Reference Model is a conceptual framework for a generic archival system which is

committed to a dual role of preserving and providing access to information.

Central to the reference model is an open archival information system (OAIS) which is "an organization of people and systems that has accepted the responsibility to preserve information and make it available for a Designated Community" (Lavoie, 2004). The reference model includes an OAIS Functional Model that describes the functional components which collectively fulfill the system's preservation and access responsibilities. It also defines the external environment within which the OAIS operates and includes an information model which provides a high-level description of the information objects managed by the archive.

The high-level functional model can be represented diagrammatically in Figure 1 (Consultative Committee for Space Data Systems, 2002).

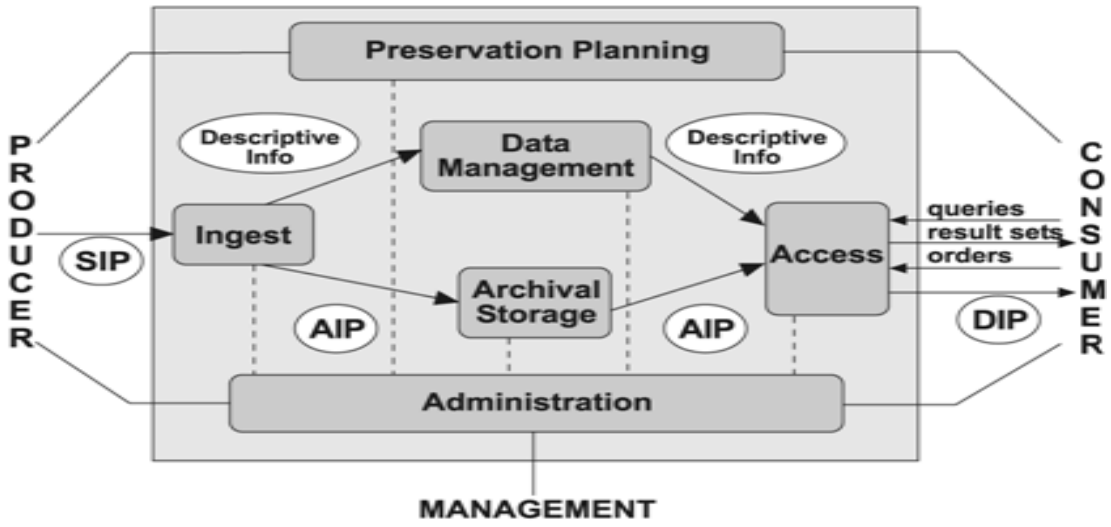

Figure 1: OAIS functional model

This will enable managers of institutional repositories to determine what information needs to be developed and maintained to ensure the usability of the repository content now and into the future. In addition, it also helps to define how to present and enable access to the content and may even determine its actual format. These considerations in turn all have a direct impact on the long-term preservation.

MAJOR DIGITAL ARCHIVING PROJECTS

Portico (http://www.portico.org)

Portico is a digital preservation service provided by ITHAKA, a non-profit organization with a mission to help the academic community use digital technologies to preserve the scholarly record and to advance research and teaching in sustainable ways. Portico was started in 2002 to create a sustainable digital archive; it has collaborated with 151 publishers (on behalf of over 2000societies and associations) and 741 libraries to archive 13690 and 129,890 e-journals and e-books. The participating libraries are supposed to get access to dark archives after 'trigger event' occur. These may include:

- Cessation of a publisher's business
- Discontinuation of a title
- Non availability of back issues
- Vagaries of Nature

- Failure of delivery mechanism of publisher.

LOCKSS /CLOCKSS (http://www.lockss.org / http://www.clockss.org)

The LOCKSS (lots of copies keep stuff safe) program is an open source, library-led digital preservation system built on the principle of 'lots of copies keep stuff safe'. Today, it is named as '**Controlled lots of copies keep stuff safe (CLOCKSS)**'. The CLOCKSS system allows librarians to take custody of and preserve access to the e-content for which libraries have paid for.

Using their computers and network connections, librarians can obtain, preserve, and provide access to purchased copies of e-contents.

The LOCKSS allows preserving all electronic contents locally for which subscriptions are paid by the libraries which allow them to restore their print purchasing models. If the publishers web site contents are unavailable, archived contents are served to the library users through LOCKSS archived program. The LOCKSS delivers a copy of the original publication to authorized users in real time, whenever it is needed. Because LOCKSS preserves the original publisher's copy of each item, it ensures that the most authoritative version persists, remains unchanged, with full credit to the publisher.

At present, the LOCKSS has more than 9000 e-journal titles from 510 publishers and there are 164 libraries and 88 publishers who have entrusted their content to CLOCKSSS for long-term preservation.

PANDORA (http://pandora.nla.gov.au)

PANDORA (Preserving and Accessing Networked Documenary Resources of Australia) is an archive of significant Australian online publications and websites issued on the Internet. The National Library of Australia and its partners are building the archive to ensure lon-term access to significant Australian documentary heritage that is published online. PANDORA has been placed on the Memory of World Australian Register in August 2004. The NLA selects e-journals from the Australian Journal Online database for preservation in PANDORA.

E-journals are one of the six categories of online publications included in PANDORA which lists more than 2000 journals published in Australia.

KOPAL (http://kopal.langzeitarchivierung.de/index_ziel.php.en)

It is a cooperative development of long-term digital archive of German National Library and Universitaets bibliothek, Gottingen. The goal of this project is to develop a technical, organizational solution to ensure long-term availability of electronic resources. It preserves bit streams of digital documents follows three steps of storage, migration and emulation.

PubMed Central (http://www.ncbi.nlm.nih.gov/pmc/)

It is a free archive of biomedical and life sciences journal literature at the US National Institutes of Health's National Library of Medicine (NIH/NLM). In keeping with NLM's literature, mandate to collect and preserve as a digital counterpart to NLM's extensive print journal collection. It was launched in February 2000, and is managed by NLM's National Center for Biotechnology Information (NCBI). As an archive, PMC is designed to provide permanent access to all of its content, even as technology evolves and current digital literature formats potentially become obsolete. NLM believes that the best way to ensure the accessibility and viability of digital material over time is through

consistent and active use of the archive. For this reason, free access to all of its journal literature is a core principle of PMC. It provides access to more than 250 journals from 50+ publishers. It retains all perpetual rights to archive all submitted materials and aims at maintaining the long-term integrity and accuracy of the contents.

National Digital Information Infrastructure and Preservation Programme (NDIIPP) (http://www.digitalpreservation.gov/about/background.html)

The NDIIPP of Library of Congress (LC) aims at collecting, preserving and making available significant digital content for current and future generations. LC through its programme has collaborated with various libraries and organizations for preserving at risk digital content, into over 1400 collections, and built a distributed digital preservation infrastructure.

Ohio LINK (Library and Information Network)

It is a consortium of 88 Ohio colleges and Universities, and the State Library of Ohio. OhioLink's e-services include a multi-publisher E-journals publisher centre which was launched in 1998 provides access to more than 7000 scholarly journal titles from 40+ publishers across different disciplines. Ohio LINK has declared its intention to maintain the EJC content as a permanent archive and has got perpetual archival rights in its license agreement.

Data-PASS (http://www.data-pass.org)

Data-PASS (Data Preservation Alliance for the Social Sciences) is a project supported by NDIIPP, of five major US Social Science data archives. It ensures that all at risk social science data are identified, acquired, archived and preserved for future use.

MetaArchive (http://www.MetaArchive.org)

MetaArchive began in 2004 as a venture led by Emory University with Georgia Technical University, University of Louisville, Virginia Technological University, Aubum University, Florida State University, and Library of Congress. It supports, promotes and extends collaborative approach to distributed digital preservation practices.

Chronopolis Project (http://chronopolis.sdsc.edu)

The Chronopolis project provides long-term archiving and preservation services for digital content provided by the California Digital Library and the Inter-University consortium for Political and Social Science (ICPSR). Under this project, some core archiving and preservation tools and services have been developed under a framework, popularly called ADAPT-approach to digital archiving and preservation technology. This model is based on a layered, digital object architecture which includes a set of modular tools and services, built on open standards and web technologies. The project also borrows from open archival information system (OAIS) reference framework.

e-Depot System (http://www.kb.nl/hrd/dd/index-en.html)

The Koninklijke Bibliotheek (National Library of the Netherlands) (KB) is the first national library in the world to start and own an operational system for the deposit and lon-term preservation of digital publications. In 1994, the KB took the decision to include e-publications in its depository collection (e-Depot). Since then, research and development on long-term digitization and archiving has been an important activity of KB. The e-Depot, an intrinsic part of Dutch National Library aims to ensure long-term preservation of digital objects. The KB has developed a specific workflow for archiving

e-publications. Elements of this workflow are: accept and pre-process; generate and resolve identifiers; search and retrieve publications and identify, authenticate and authorize users. The technical heart of the e-Depot system is IBM's DIAS (Digital Information and Archiving System). The KB ensures, libraries, publishers and end users that the information stored in the archive will outlast the transience of digital documents. In addition, the e-Depot offers publishers a durable check on archived formats as well as guidance on how to create the most durable electronic publications. The library has entered into an agreement with 30+ publishers and libraries to archive the scholarly content. As of November 2007, the Depot has ingested more than 10 million digital objects; most of them are publications by international publishers.

CONCLUSION

The problem of long-term digital preservation is becoming more real in the midst of a digital era. Old assumptions regarding information preservation are no longer valid, and it is clear that aggressive actions are needed to ensure the understandability of data for ages to come. In order to address these challenges, new technologies and systems are being developed. Such systems will be able to better address these vital issues if they are equipped with storage technology that is inherently dedicated to preservation and that supports the different aspects of the preservation environment. An appropriate storage system will make any solution more robust and lower the probability of data corruption or loss.

With the Digital Millennium Act in force, the publishers are not obligated to provide perpetual online access to content to which libraries used to have subscription. Implementing the LOCKSS concept in library can facilitate the perpetual access to such content. This, of course, can be done only through negotiation with the concerned publisher and it is likely to have financial implications.

Despite these agreements and developments, the field of digital archiving is still in its infancy, and much work needs to be accomplished to achieve a secure and permanent archiving of electronic journals. However, the successful agreements between the KB and the leading for-profit publishers could be used as a model for other publishers, especially for not-for-profit publishers and other publishers around the world.

The integration of the LOCKSS digital preservation system and link resolvers has proven to be an effective way for libraries to provide reliable access to the digital content that they own, even if they lose access to it from the publisher. Although the details of integrating a LOCKSS box with a link resolver are specific to the vendor, many of the vendors have adopted common practices that allow work done for one integration to be leveraged to integrations with other vendors' products. LOCKSS has been enhanced to retrieve archived content in terms of its bibliographic data, encoded in industry-standard Open URL format.

Though we could not achieve all the objectives with the LOCKSS system, we believe that the LOCKSS and **CLOCKSS** (Controlled Lots of Copies Keep Stuff Safe) technology provides a cost-effective solution to facilitate long-term access to the scholarly literature. The concept can be used for long-term preservation and access to any digital content that is worth preserving in the Library.

As a result, link resolvers can simply format and redirect a request to the local LOCKSS box. For content that is archived in a local LOCKSS box, the LOCKSS target for a title displayed by a library OPAC is a reliable source, because it automatically "fails over" to the locally archived version if access through the publisher becomes unavailable.

REFERENCES

1. "CLOCKSS", http://www.clockss.org (Accessed on 16 October 2013)
2. "Digital Millennium Act", available at http://www.copyright.gov.in.legislation/dmca.pdf (Accessed on 15 October 2013)
3. "LOCKSS (Lots of Copies Keep Stuff Safe)". Project web site. SourceForge. (Accessed on 11 October 2013).
4. "LOCKSS", Wikipedia, available at http://en.wikipedia.org/wiki/LOCKSS
5. Beagrie, N. and Jones, M., 2002, "Trusted Digital Repositories: Attributes and Responsibilities", Mountain View, CA, Research Libraries Group- OCLC, available at http://www.rlg.org/longterm/repositories.pdf
6. Digital Preservation Coalition. 2002. Preservation Management of Digital Materials: A Handbook, Digital Preservation Coalition, York, available at: www.dpconline.org/graphics/handbook/ (Accessed on 16 October 2013)
7. Giuli, T.J., Maniatis, Petros, Baker, Mary, Rosenthal, David S. H. and Roussopoulos, Mema, 2004, "Attrition Defenses for a Peer-to-Peer Digital Preservation System",
8. Hockx-Yu, Helen, 2006, "Digital preservation in the context of institutional repositories", Electronic library and information systems Vol. 40, No. 3, pp. 232-243.
9. Lavoie, B, 2004, "The Incentives to Preserve Digital Materials: Roles, Scenarios, and Economic Decision-Making", White Paper, OCLC Research, available at: www.oclc.org/research/projects/digipres/incentives-dp.pdf (Accessed on 16 October 2013)
10. Kirchhoff, Amy J., 2008, "Digital preservation: challenges and implementation", Learned Publishing, 21, pp. 285–294.
11. McDonald, J. and Van de Velde, E., 2004, "The lure of linking", Library Journal, available at: www.libraryjournal.com/article/CA405398.html (Accessed on 08 October 2013).
12. Maniatis, Petros, [et al], 2005, "The LOCKSS Peer-to-Peer Digital Preservation System", ACM Transactions on Computer Systems, 23, Issue 1, pp. 2–50.
13. Obaiah, B. and Jayakanth, Francis, 2008, "Long-term preservation and perpetual access to scholarly content: the LOCKSS way", Proceedings of National conference on digitization and digital preservation, pp. 190-194.
14. Pool, Rebecca, 2013, "Mind the gap: catching digital content before it slips away - PRESERVATION- Research Information" available at www.researchinformation.info/features/feature.php?feature_id=397
15. Reich, V. and Rosenthal, D., 2001, "LOCKSS: a permanent Web publishing and access system**."** D-LIB Magazine, Vol. 7, No. 6. Available at: www.dlib.org/dlib/june01/reich/06reich.html (Accessed on 25 October 2013)
16. Rosenthal, David S. H. and Vargas, Daniel L., 2012, "LOCKSS Boxes in the Cloud", (Accessed on 16 October 2013)
17. Rosenthal, David S. H., Lipkis, Thomas Robertson, Thomas S. and Morabito, Seth, 2005, "Transparent Format Migration of Preserved Web Content". D-Lib Magazine (Corporation for National Research Initiatives) 11 (1). doi:10.1045/january2005-rosenthal. (Accessed on 25 October 2013)
18. Stanford University LOCKSS Program, 2013, "LOCKSS: Lots of Copies Keep Stuff Safe, Stanford University, Stanford" available at:

www.lockss.org (Accessed on 10 October 2013).

19. "Transparent Format Migration of Preserved Web Content", D-Lib Magazine (Corporation for National Research Initiatives), pp. 11 (1). doi:10.1045/january2005-rosenthal. (Accessed on 25 October 2013)