ISSN: 2319-8435 Impact Factor : 2.2030(UIF) Volume - 4 | Issue - 7 | May - 2016

e-Library Science Research Journal

DIGITAL PRESERVATION INITIATIVES IN SELECTED UNIVERSITY LIBRARIES: A CONTENT ANALYSIS

Rajesh K
Library Assistant, National Institute of Advanced Studies,
Indian Institute of Science campus, Bengaluru (India).

ABSTRACT

This study provides an examination of select universities in the world with reference to their libraries and their digital preservation system. Emphasis in laid on the method/process of digital preservation across these library systems. The findings of this study will enable/support future research on the aspect of digital preservation in the domain of library science.

KEYWORDS : Digital library, Preservation, Digitization, DSpace, Metadata.

1.1 INTRODUCTION

Several research libraries are either already involved in large-scale preservation initiatives or are contemplating or planning involvement in such endeavors. Two of the most visible large-scale projects, Google Book Search and Microsoft Live Search Books, have generated lot of interest among the academic community and the audience in general. The digitization and preservation of millions of books under programs such as Google Book Search and Microsoft Live Search Books is dramatically expanding our ability to search and find information. The primary motivation is to make it easier to find and access books. Nonetheless, access and preservation goals are usually interrelated, since access to scholarly materials depends upon their being fit for use over time. The connection between preservation and access in the digital world is complex.

1.2 Need for the study

There is significant uncertainty about the long-term strategies of initiatives such as Google Book Search and Microsoft Live Search Books. These are relatively new programs and there is no

evidence to suggest that the corporate and nonprofit partners have any long-term business plans for maintaining access to digitized collections or for migrating delivery platforms through future technology cycles.Such uncertainties strengthen the case for libraries taking responsibility for preservation, both from archival and access perspectives. Therefore, the present study intend to study the digital preservation initiatives been taken in major libraries around the world.

1.3 Scope of the study

The digital preservation has a short, but successful history. The concept of digital preservation has made deep impact on libraries and users. Over the past, many digital preservation projects have been set up, and presently many are being set-up. For this study, we plan to study the digital preservation projects been taken over in top five universities libraries.

1.4 Objectives of the study

The main objectives of the study are as follows:

- To identify major digital preservation initiatives in academic sector.
- To study the important features of digital preservation.
- To study the services and facilities provided in selected digital preservation projects.
- To study the infrastructure requirements for digital preservation.
- To study the resource formats adopted for digital preservation
- To study the methods adopted by the selected digital libraries to reach out to the library users.

1.5 Statement of the problem

The statement of the problem is “Digital Preservation Initiatives in Selected University Libraries: A Study”

1.6 Methodology

The study used content analysis method as a qualitative approach to study the digital preservation projects. The present study selected the world’s top five universities from the ranking list of 2015-16 available on Times Higher Education website. Further, the library websites of the selected five universities have been analyzed to identify the digital preservation activities.

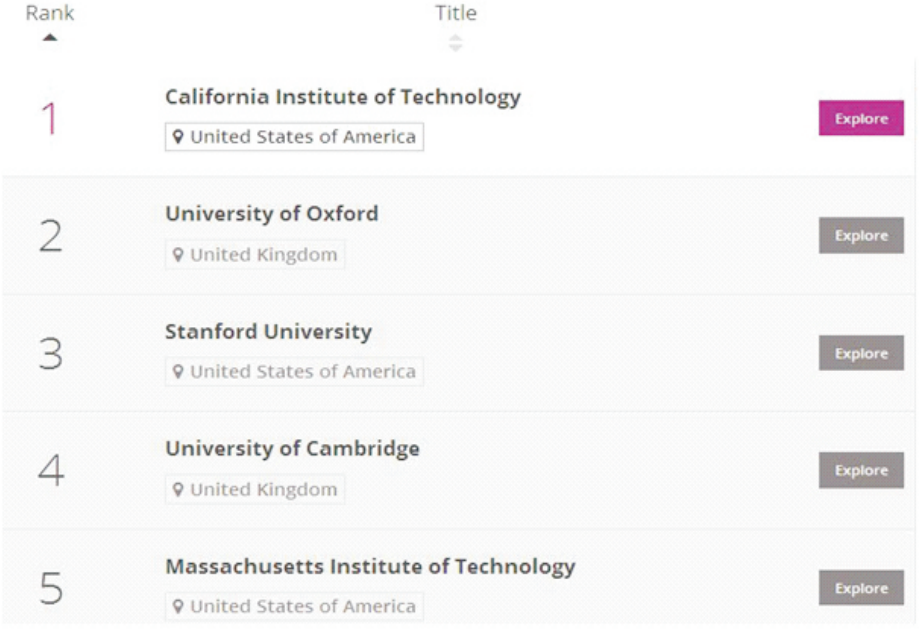
1.7 Review of Literature

The Middle Age Bible is printed on a durable medium for archiving recorded societal information over the centuries. Today we store books, art, and documents in digital form without knowing if they will remain intact over the years (Duranti, 2000). Digital preservation is defined as the managed activities necessary: 1) For the long term maintenance of a byte stream (including metadata) sufficient to reproduce a suitable facsimile of the original document and 2) For the continued accessibility of the document contents through time and changing technology (Research Libraries Group, 2002).Information technology has made it possible to produce more recorded information than any previous decade of human activity and “The fact that the majority of these data is less accessible than ever before is one of the ironies of the modern information age” (Duranti, 2000).Many people around the world are involved in ambitious research projects on long-term digital preservation, with most of them focusing on technical solutions (Ruusalepp, 2005). The lack of expertise in digital preservation appears to be a significant obstacle to developing digital preservation programs (Hedstrom & Montgomery, 1998).

1.8 Libraries under Study

The present study selected the world's top five universities from the ranking list of 2015-16 available on Times Higher Education website. Further, the library websites of the selected five universities have been analyzed to identify the digital preservation activities.

Figure 1: World University Rankings



1.8.1 California Institute of Technology

The California Institute of Technology (Caltech) is a world-renowned science and engineering research and education institution. It is a private coeducational university and research institute located at Pasadena, California, U.S. Caltech was established in 1891 as a school for arts and crafts.

Caltech Archives' Collection

The Caltech Archives' holdings include a wide range of materials arranged i.e. Manuscripts, Photographs and Images, Oral Histories, Digital Media Collection, Online Exhibits, Books and Texts, Audio Visual, Artifacts, Fine Art.

Digital Preservation at Caltech

Caltech has taken many initiatives in building the digital collection with an objective of long-time preservation i.e.

Oral Histories Online: In 2002, Caltech launched its first digital project, titled Oral Histories Online, which brings selected interviews with members of the Caltech community to the public on the web. The project is part of the Caltech Collection of Open Digital Archives (CODA) hosted by the Caltech Library Services.

Lab Notes Online: Lab Notes Online provides access to original laboratory notebooks that record some of the greatest milestones in research at Caltech. The first offering is Robert A. Millikan's celebrated oil-

drop experiment notebooks which recorded the attempts to measure the charge of the electron. Other notebooks and historic data sets are in the process of being added to the collection.

Tech Online: Caltech has taken the digitalization of the Caltech student newspaper, the California Tech, with the intention of moving back in time to the original publication date of 1920.

Caltech Collection of Open Digital Archives (CODA): The Caltech Collection of Open Digital Archives (CODA) is the institute's collections of faculty research publications and other content supporting the mission of the institute. The different collections under CODA are listed in the following sections.

CaltechAUTHORS: The CaltechAUTHORS collection includes research publications by Caltech authors. Mainly articles, but also books, book chapters, technical reports, conference papers and more.

CaltechTHESIS: The CaltechTHESIS includes all Caltech theses written since 2002, as well as digitized copies of earlier theses.

CaltechCampusPubs: The CaltechCampusPubs includes general publications by Caltech faculty, staff, and students that are of interest to the campus community as a whole.

Caltech-hosted Conferences: The Caltech-hosted Conferences contains papers delivered at several conferences hosted at Caltech.

Metadata: The Caltech libraries have adopted the MARC21 standard for its collection.

1.8.2 University of Oxford

Oxford is an independent and self-governing institution, consisting of the central University and the colleges. Thirty-eight colleges form a core element of the University. The University offers undergd postgraduate programs. Oxford meets the needs of its students, academics and the international research community with a wide range of library services provided by more than 100 libraries, raduate anmaking it the largest university library system in the UK. Oxford University is internationally renowned for its scholarly library collections, and in particular for those of the Bodleian Library which has been a library of deposit for almost 400 years. The Bodleian Libraryis the University's main research library and is the second largest in the UK after the British Library. Oxford Libraries Information System (OLIS) contains catalogue records for nearly seven million of the estimated ten million titles held by libraries associated with Oxford University. University libraries have an extensive range of e-resources across all subject areas.

Digital Preservation at Oxford

The University has a long tradition in digital scholarship and there are a number of completed library projects focusing on the digitization of primary resources. The Oxford Digital Library (ODL) is a core service of the Bodleian Libraries. Oxford is actively involved in developing electronic information, including electronic text archives and image databases. Through the Oxford Digital Library, the University is a leader in the digitization of manuscripts and other library material. The Oxford Digital Library started its operations in July 2001 and has been performing a major role in coordinating and stimulating digitization activities in the University. The ODL aims to offer a Digital Library architecture

which will allow centralized access to digital resources, both those it creates itself and those acquired from outside.

Digital Bodleian: The Digital Bodleian is a collection of digitized images from the Bodleian Libraries' special collections.

Archive-It: Archive-It is a subscription web archiving service from the Internet Archive that helps organizations to harvest, build, and preserve collections of digital content. The Bodleian Libraries have partnered with Internet Archive to harvest and preserve its digital collections.

Oxford University Research Archive (ORA): The Oxford University Research Archive (ORA) is an online collection of research publications by scholars at the University of Oxford. It contains freely available copies of journal articles, conference papers and posters, reports, and working and discussion papers. It contains digital copies of doctoral research theses by students of the University. ORA enable researchers to maximize the visibility of their research.

Bodleian Electronic Archives and Manuscripts: The Bodleian Electronic Archives and Manuscripts (BEAM)is being created to provide for the management of born-digital archives and manuscripts acquired by the Library. BEAM will provide the means to gather, describe, manage and preserve the digital components in archive and manuscript collections while maintaining their relationship with more traditional components of the same collection.

Access and Control: Search Oxford Libraries Online (SOLO) is the main search engine for library collections. SOLO offers a one-stop search and delivery solution for quickly accessing Oxford's main library information resources regardless of type, format, or location. Search results are presented with a wide array of additional links to improve resource discovery. The Single-Sign-On will offer easy access to subscription resources, whether on or off campus.

Metadata: The use of established standards for descriptive metadata (i.e. EAD, TEI) is a precondition for this integration process.

1.8.3 Stanford University

The Leland Stanford Junior University commonly referred to as Stanford University or Stanford, is an American privateresearch university. The university was established in 1891 by Leland and Jane Stanford. The Stanford University Libraries include Green Library (the main campus library), Meyer Library (technology services and study spaces), 14 specialized branch and department libraries, 3 auxiliary libraries housing less-used or overflow materials, and 5 coordinate libraries, connected to graduate schools and other organizations.

Stanford University Archives

The Stanford University Archives was created in 1965 to collect, preserve, and make available the historically and legally valuable records of the University and of Stanford community members. The University Archives supports faculty teaching and research, graduate level research, as well as undergraduate thesis and course work. Additionally, the University Archives serves as the major information source for campus offices, such as the President's Office, the News Service, the Office of

Development, the Facilities Planning Office, and the Stanford Alumni Association.

Digital Preservation at Stanford

Electronic Thesis and Dissertations: The Electronic Thesis and Dissertation system was launched in November 2009 and has been a successful project with more volumes being added to the collection. Stanford's PhD and Engineering graduate students have had the option to submit their culminating works either online or on paper. But, students are more interested in submitting their reports online.

Stanford Media Preservation Lab: The Stanford Media Preservation Lab (SMPL) serves to preserve and enhance access to original sound and moving image collection materials held by Stanford University Libraries. The lab's core operations are centered on creating a high-quality copy of the original content in digital formats that are easily accessed by researchers and others, and that enable ongoing, long-term management of the content for future users.

Digital Library Systems and Services: The Digital Library Systems and Services (DLSS) is the information technology production arm of the Stanford Libraries. It serves as the digitization, digital preservation and access systems provider for SULAIR, and it is the research and development unit for new technologies, standards and methodologies related to library systems.

Google Books: Google Books, is an audacious project, begun in 2004, to bring the power of Google's search capabilities to printed books. The project involved a Library Partners program to scan tens of millions of books from major libraries around the world, and apply full-text indexing to them, making them full searchable. To date, Google has scanned and indexed over 12 million volumes of which over 2 million came from Stanford's stacks.

Digitization Services: Stanford University Libraries provides high quality digitization services - digital scanning, reformatting, and capture of born digital materials to support the University's academic and research mission.

Metadata: The Stanford University Libraries have adopted different metadata standards to satisfy the metadata requirements to clearly describe and retrieve the content stored in various formats. The list of metadata standards adopted by Stanford University Libraries are Dublin Core Metadata Initiative, Metadata Object Description Standard (MODS) (for generic descriptive MD across formats), MIX: NISO Metadata for Images in XML Schema, PREMIS for preservation metadata, Metadata Encoding and Transmission Standard (METS)

1.8.4 University of Cambridge

The University of Cambridge is a public research university located in Cambridge, England, United Kingdom. It is the third-oldest surviving university in the world. The institute grew out of an association of scholars that was formed in 1209. Over the course of six centuries Cambridge University Library's collections have grown from a few dozen volumes into one of the world's great libraries, with an extraordinary accumulation of books, maps, manuscripts, and journals. These cover every conceivable aspect of human endeavour, spanning most of the world's cultural traditions.

Digital Preservation at Cambridge

DSpace@Cambridge: DSpace@Cambridge is the institutional repository of the University of Cambridge. This is the place where the Cambridge users' can self-archive their papers, share research data or store digitised image or multimedia collections. The repository was set-up on July, 2006, to facilitate both dissemination and preservation of digital material created by the members of the University of Cambridge.

Cambridge Digital Library

Cambridge Digital Library consists of books, manuscripts and other items from across our collections that are especially significant. They can be accessed at any time, from anywhere in the world and browsed cover to cover. The digital library has a collection of around 13,241 books and manuscripts. The collections under Cambridge Digital Library are: Treasures of the Library, Darwin-Hooker Letters, Newton Papers, Hebrew Manuscripts, Sanskrit Manuscripts.

1.8.5 Massachusetts Institute of Technology

Massachusetts Institute of Technology is a private research university located in Cambridge, Massachusetts, United States. It was established by the Commonwealth of Massachusetts on April 10, 1861.The institute has been producing great students as it is evident by 78 Nobel Laureates it has produced till date. MIT libraries support the Institute's programs of research and study in both innovative and traditional ways. Library locations offer technology-enabled rooms for group collaboration and virtual meetings with peers across the globe, as well as quiet spaces for individual study. The Libraries have over five million items in print and digital formats, including electronic journals and books, images, maps, musical scores, and sound and video recordings.

Digital Preservation at MIT

MIT's Institutional Repository: Through the MIT Libraries' website, MIT students and researchers can search over 3 million printed volumes and tap into over 55,000 databases and electronic journals, and a growing number of digital collections.MIT uses DSpace, the digital library software, developed by the institution in association with Hewlet-Packard which lets the user to save, share, and search research data.The digital library at MIT is designed and developed to meet the multidisciplinary and organizational information needs of the institution. It provides access to the digital work of the whole institute through one interface. The digital library is organized into communities, sub-communities, and collections, each of which retains its identity.The digital library at MIT provides long-term physical storage and management of digital items in a secure, professionally managed repository including standard operating procedures such as backup, mirroring, refreshing media, and disaster recovery. A persistent identifier is assigned to each item to ensure it is retrievable far into the future.

Institute Archives and Special Collections: The Institute Archives and Special Collections, a unit of the Libraries, contains MIT's founding documents and the personal papers of noted faculty. It serves as the "memory" of MIT, collecting and preserving records that document MIT's history and the people who have been a part of that history. The Archives works to build collections of primary sources such as the official records of the Institute and selected personal and professional papers of MIT faculty, staff, and students. The department also collects MIT publications and the record copy of MIT theses, and maintains MIT's rare book collections. Occasionally the Archives accepts collections of records of non-MIT persons and organizations whose activities complement the Archives' holdings. In addition, staff members at both the Institute Archives and the MIT Museum have worked with MIT faculty to create a

list of available oral histories.

DSpace@MIT: DSpace@MIT, an innovation of the Libraries, is a digital repository containing over 60,000 items, including MIT theses and the scholarly works of MIT faculty and researchers, labs, and centers. It contains selected digital theses and dissertations from all MIT departments dating as far back as the mid-1800s.It material also includes conference papers, images, peer-reviewed scholarly articles, preprints, technical reports, theses, working papers, and more.Although much of the digital content is under restricted access, DSpace@MIT'sOpen Access Articles collection provides over 7,500 scholarly articles that MIT faculty have made openly available on the web under their Open Access Policy.Analysis of usage statistics for last fiscal year indicates that DSpace@MIT content was downloaded directly by end-users 24 million times or, on average, at a rate of over 1 million downloads per month.

Digital Content: The digital library at MIT accepts all types of digital formats. Some examples of items that are available in MIT digital library are: Documents, such as articles, preprints, working papers, technical reports, conference papers, Books, Theses, Data sets, Computer programs, Visualizations, simulations, and other models, Multimedia publications, Books, Bibliographic datasets, Images, Audio files, Video files, Learning objects, Web pages.

Search and Access: The DSpace software allows contributors to limit access to items - at the collection and the individual item level. The submission process allows for the description of each item using a qualified version of the Dublin Core metadata schema. These descriptions are entered into a relational database, which is used by the search engine to retrieve items.

1.9 Findings Suggestions and Conclusion

Academic libraries and archival institutions have a unique opportunity in the area of digital preservation. The need for storing the information for future use was released long ago which had led to the development of archives in the past. But, the influence of Information and Communication Technology (ICT) and the increased literature growth have made the archiving task a difficult one.

Major universities around the world have taken up the task of digital preservation to save the data and information for the next generations. Realizing the huge importance of data, all major universities have taken major projects in digital preservation. Although, all the resources cannot be preserved, but it is worth satisfying that we have released the importance of digital preservation and have started up the task of digital preservation some of which have been described in this study.

Based on the observations from the present study, the following factors related to digital preservation should be considered before taking up any digital preservation task.

The selection of documents for digitalization should consider factors such as intellectual value, added value through digitization, preservation needs, and scarcity of holdings.

Subject specialists and curators should take up the responsibility of making selection decisions for digitization according to recommended guidelines. Digital projects should be developed in consultation with faculty and researchers and/or in collaboration with colleagues at other institutions.

Digital capture, processing, and quality control should be performed according to accepted standards and best practices in order to create high quality digital files.

Metadata for the digital resources should be captured and/or created for effective search and access. Standardized metadata formats e.g., MARC, MODS, EAD, etc and controlled vocabulary derived from appropriate thesauri should be considered for greater efficiency. Persistent identifiers should be created for the digital objects.

The digital preservation infrastructure should ensure long-term access to digital materials. It

needs to be ensured that systems are developed in accordance with digital archiving best practices and that the Library's digital assets are controlled using a set of core metadata elements.

1.10 Reference

1. Duranti, Luciana (2000). The InterPARES Project. In: InterPARES Symposium Proceeding, February 2000 - The complete proceedings from the February 2000, Inter PARES Symposium. Available at(http://www.interpares.org/display_file.cfm?doc=ip1_symposium_2000.pdf)
2. Hubbard N, Wisman V. The long-term archiving of European Earth observation satellite data: current status and guiding principles, CEO/282/1998 issue 1.0. European Commission, Joint Research Centre Ispra, 1998. 28pp.
3. Research Libraries Group. (2002). Trusted digital repositories: Attributes and responsibilities. An RLG-OCLC Report. Available at (http://www.rlg.org/longterm/repositories.pdf)
4. Ruusalepp, Raivo (2005). Digital Preservation in Archives: An Overview of Current Research and Practices. Estonian Business Archives.
5. https://www.timeshighereducation.com/world-university-rankings/2016/world ranking#!/page/0/length/25/sort_by/rank_label/sort_order/asc/cols/rank_only(accessed on 16th April 2016)
6. http://libraries.mit.edu/dspace-mit/technology/features.html(accessed on 16th April 2016)
7. www.stanford.edu(accessed on 16th April 2016)
8. http://library.stanford.edu/libraries(accessed on 16th April 2016)
9. http://en.wikipedia.org/wiki/California_University(accessed on 20th April 2016)
10. http://en.wikipedia.org/wiki/University_of_California,_Berkeley_Libraries(accessed on 20th April 2016)
11. https://en.wikipedia.org/wiki/California_Digital_Library(accessed on 20th April 2016)
12. http://www.ask.com/wiki/Cambridge_University_Library(accessed on 20th April 2016)
13. http://www.lib.cam.ac.uk/eresources/index.php(accessed on 20th April 2016)
14. http://www.lib.cam.ac.uk/collections.htm(accessed on 20th April 2016)
15. http://www.lib.cam.ac.uk/preservation/prepare/(accessed on 23th April 2016)
16. http://en.wikipedia.org/wiki/University_of_Oxford(accessed on 23th April 2016)
17. http://www.ox.ac.uk/research/libraries/index.html(accessed on 23th April 2016)