

---

---

## DIGITIZATION OF PRINT BOOKS AND FUTURE OF THE LIBRARIANSHIP

*Sanjib K Deka*

### Abstract

*With the advancement of Information Technology (IT) the role of libraries have changed and led to the transformation of library services from traditional services to new services and delivery modes. The book which is the major ingredient in a library has seen a number of transitions over the centuries. This paper discusses the digitization movement of library books in major institutes and libraries. Also enumerates key projects taken over in India and western countries. It also highlights the impact of digitization on library services.*

**Keywords:** Digitization; Digital Archiving; Librarianship; e-publishing

### 1. Introduction

Throughout history the role of the library was to serve as a storehouse, an archive of manuscripts, art, and important documents. The library was the center of information revered by most because each contained the foundational building blocks of information for all humanity. In medieval times, books were valuable possessions far too expensive for most people to own. As a result, libraries often turned into a collection of lecterns with books chained to them. Leading up to today libraries have consisted of large collections of books and other materials. Collections are often used by people who choose not to, or can not afford to, purchase books for themselves. But that definition is changing. We have transitioned from a time where information was scarce and precious to today where information is vast and readily available, and in many cases, free. People who in the past visited libraries to find specific pieces of information are now able to find that information online.

Enabling technologies have led to the transformation of library services from traditional services incorporating card catalogs, printed books and periodicals, bibliographic instruction, in-person reference, to new services and delivery modes incorporating electronic collections, such as e-books, e-journals and databases, virtual reference services, and other online services. Innovation of new services that are peculiar to the online environment is the trend in modern electronic libraries. During the last decade, many libraries, particularly those serving academic communities, have witnessed the emergence of new service paradigms in areas of information access and delivery, reference, instruction, technology facility and support to patrons. Libraries continue to harness new technologies to offer services in innovative ways to meet the changing needs of their patrons. Technology can now support the creation, storage and transmission of large volumes of information in various electronic/digital formats. This has in turn led to changes in information needs of users as they become aware of the capabilities of technology in facilitating access to these various forms of content.

---

---

## 2. Documents Digitization

In recent years there has been much debate about the future of the book in a digital world. E-books have already become a reality. This is not surprising, as the technology of the book has seen a number of transitions. Over the centuries there has been a series of changes in the way that words are presented. Clay, wax, papyrus, vellum, cloth and paper have all been used and stored as tablets, scrolls or folios or books. The digitization of the world's enormous store of library books, an effort dating to the early 1990s in the U K and the US, come into reality, when the search engine giant Google announced the 'Google Book Search' service in December 2004. Google has already signed agreement with Oxford, Harvard, Stanford, University of Michigan, and New York Public Library, to convert the full text of millions of library books into searchable web pages.

There are many organizations working on initiatives to scan public domain books, and libraries are digitizing their own collections. However, according to an OCLC study, these public domain books represent only 20 percent of the world's books. Estimates put the percentage of books that are in print and available in bookstores at only 5 percent. This leaves 75 percent or more in "Twilight Zone" – that is, they may no longer be in print, but since they were published after 1923, they may not yet be in the public domain. These books are an important focus of digitization efforts at various digital library projects.

## 3. Digital Library Projects

Many digital library projects are going on across the world. Some of them are listed below:

### 3.1 Google Book Search

Google's Book Search service is the product of its Books Library Project, which is digitizing books from major libraries around the world in order to make them searchable online. Its partners include the University of Michigan, Harvard University, Stanford University, Oxford University, the University of California and the New York Public Library. Google is also conducting a pilot project with the Library of Congress. According to Google, they just want to encourage people to find books and authors to write more books. Though they plan to digitize all books, Google display at most three snippets – a few sentences surrounding a user's search term, so that a user can find the book and has a reason to go buy it or find it in a library, and to make the Book Search truly comprehensive and useful.

Now, Google plans to begin letting consumers download and print free of charge classic novels and many other, more obscure books that are in the public domain. Using Google's Book Search service, Web surfers hunting titles like Dante's "Inferno" and Aesop's "Fables" will be able to download PDF files of the books for later reading, to run keyword searches or to print them on paper. Up to now, the service only allowed people to read the out-of-copyright books online. The download initiative does not include any books under copyright. For those titles, Book Search only displays basic bibliographic information and, in many cases, small snippets of text surrounding a search term, unless it has permission from the publisher to show more. The company's display of snippets has riled some publishers, but Google has argued the small bites of text constitute fair use.

---

---

### **3.2 Universal Library Project (Million Book Project)**

The mission of the Universal Library is to foster creativity and free access to all human knowledge. The first major project of Universal Library is the Million Book Digital Library project. As a first step in realizing this mission, it is proposed to create the Universal Library with a free-to-read, searchable collection of one million books, primarily in the English language, available to everyone over the Internet. Currently, the Universal Library project uses a Minolta PS7000 overhead scanner, which is potentially able to scan 10000 pages operating on a 3-shift basis. The Million Book Project will digitize a large body of published literature and offer it free-to-read on the surface web - providing students, faculty, and lifelong learners with rapid, convenient access to quality resources.

The Universal Library aim to provide support for full text indexing and searching based on OCR (optical character recognition) technologies where available. Creating one digital copy and mirroring it in different locations will suffice, and will support the multiple uses at any time. Palm leaves, journals and manuscripts are also digitized. Microfiche, Microfilms also can be digitized and processed under this project. Materials published before 1940 are in the public domain and may be scanned for this project. The project will seek publisher permission to scan books from Books for College Libraries, one source for core academic books in English. A previous study done at Carnegie Mellon University Libraries indicates that 22% of publishers granted permission for scanning and mounting on the web. Again, the National Academy Press has given permission to digitize all of their books published prior to 1995.

The Million Book Project (MBP) is a multi-agency, multi-national effort that will require the database to be globally distributed. For location independent access, this globally distributed database will appear to be a virtual central database from any place around the world. At present over 600,000 books have been scanned; 170,000 in India, 420,000 in China, and 20,000 in Egypt. Roughly 135,000 of the books are in English; the others are in Indian, Chinese, Arabic, French, or other languages. Most of the books are in the public domain, but permission has been acquired to include over 60,000 copyrighted books. Twenty-two scanning centers are operating in India, including four mega-centers. Eighteen centers are running in China, including a mega-center in a free-trade zone to avoid customs delays with shipments of books from the U.S. The National Agriculture Library and the United Nations' Food and Agriculture Organization (FAO) have joined the project, along with academic libraries in the United States that have large agriculture collections. Agriculture has become a collection focus for the project, and plans are being developed to create a knowledge network aimed at improving rural community access to critical agricultural information. An important byproduct of the collection will be the existence of a test bed that stimulates and supports much-needed research in information storage and management, search engines, imaging processing, and machine translation.

### **3.3 Digital Library of India**

The Digital Library of India, a portal launched in 2003 to provide access to newspapers, journals, book collection and manuscripts in Indian languages, is part of the Universal Digital Library project of Carnegie Mellon University in Pennsylvania and is also hosted by the Indian Institute of Science in Bangalore, the Ministry of Communications and Information Technology, and the Education and Research Network of India, and 21 participating centres.

---

The mission is to create a portal for the Digital Library of India which will foster creativity and free access to all human knowledge. As a first step in realizing this mission, it is proposed to create the Digital Library with a free-to-read, searchable collection of one million books, predominantly in Indian languages, available to everyone over the Internet. This portal will also become an aggregator of all the knowledge and digital contents created by other digital library initiatives in India. This portal would provide a gateway to Indian Digital Libraries in science, arts, culture, music, movies, traditional medicine, palm leaves and many more. The result will be a unique resource accessible to anyone in the world, without regard to socioeconomic background or nationality. The Digital Library of India will also partner with other country specific Digital Libraries initiatives as part of the Universal Library Project spearheaded by Prof. Raj Reddy and Carnegie Mellon University.

Under this project all the books that are currently available online have either been checked against their copyright re-registration directory in which books prior to 1963 which are not re-registered are deemed to be out of copyright, or have been agreed by the authors. However, in case of a possible error in copyright checking, if the author or publisher sends a written request for removal to the address below, such a request will be validated and complied with.

### **3.4 World Digital Library Project by Library of Congress**

The United States Library of Congress has announced on November 2005 the plan to build a World Digital Library for use by other libraries and public research institutions. Its first funding pledge \$3 million, come from Google Incorporation. They have already worked together to digitized 5000 books in the public domain, whose copyrights have lapsed. The plan behind the World Digital Library is to use public and private money to create a virtual library giving anyone access to the world's rare and unique cultural materials.

The plan will expand upon work already done for the American Memory Project, a Library of Congress effort that has been digitized American documents, images, sound recordings and more. The collection now contains more than 10 million items that can be searched or browsed as a whole or via often powerful specialized search interfaces available via each collection.

The World Digital Library is intended to go well beyond American material. It aims to be a repository of one-of-a kind collections from nations around the world, a much expanded version of the Global Gateway digital library site that the Library of Congress maintains in partnership with other national libraries such as those of Russia, Spain, Brazil, France and the Netherlands. As for the content included in the World Digital Library, the Library of Congress plans to draw from its own resources as part of the effort. More than half of the book the library holds is in languages other than English. Only public domain materials or materials with special permission granted will be placed in the World Digital Library. It is also planning to increase the content from other sources, by offering up a plan allowing others to contribute.

A number of universities in abroad are creating bookless digital libraries that reflect a growing understanding of how today's tech-savvy users access information. Besides the above mentioned projects, a number of projects are underway in the move to digitized works. To name only a few are:

- 
- The Internet Archive, Yahoo & MSN- backed Open Content Alliance
  - The European Union French led i2010 Digital Library Project
  - Amazon's Search Inside The Book & Planned Ebook Programs
  - Publisher Macmillan's BookStore Project
  - The long-standing Project Gutenberg

#### **4. Conclusion**

With the introduction of computer and internet in the libraries, the role of library within a community and the way people interact with the library is changing. We are in an awkward transition period. Every five years we change how we store information, going from an 8" drive to a 5.25" drive to a 3.5" drive to CDs and DVDs to Flash drives and thumb drives. What will future information look like? The role of the library 20 years from now is still a mystery.

Libraries have played a vital role in the advancement of human society. Societal advance depends on young people having access to books via libraries and other means. Many believe that information is now doubling every two years. Machine summarization, intelligent indexing, and information mining are tools that will be needed for individuals to keep up in their discipline work, in their businesses, and in their personal interests. These large digitization projects will enable extensive research in these areas. We can expect that making this unique web resource available free to everyone around the world will enhance the learning process. More than one individual will be able to use the same book at the same time. The availability of online search allows users to locate relevant information quickly and reliably thus enhancing student's success in their research endeavors. Research reveals that students and faculty look online first when they need information because of the speed and convenience of online access. They prefer remote access to electronic resources rather than having to go to a physical library facility. Most of the undergraduate students tend to start with popular Internet search engines like Google because these search engines are more convenient and easier to use than library databases.

In our rapidly changing world, lifelong learning and access to books have become essential to employment, health, peace, and prosperity. The expectation is that greater access to information will enhance respect for diversity and pluralism, alter the ways in which people work and deliberate together, and better equip people to understand and challenge the world around them. We stand on the brink of a new age. Familiar certainties and old ways of doing things are disappearing. Jobs are changing and with them the skills needed for the world of tomorrow. We have no choice but to prepare for this new age, in which the key to success will be the education, knowledge and skills of our profession. The dissemination of knowledge, especially the scanning in of the university collections and making them available to all in the world is truly a noble job. This will revolutionize the research and to make the libraries of the world into the world's library, at the same time there are problems of cataloging and preserving digital holding. Without the proper metadata attached that once appeared in libraries physical card catalogs, a digital book is as good as lost. The web couldn't make the library and its services irrelevant. Digitization of documents won't degrade the library collections, although it could bring new users to the library and increase the interactions. We could summarize that the digitization and growing power of the internet are making the librarian's job more complex and challenging.

---

---

**References**

1. Arms, William Y., "Automated Digital Libraries", D-Lib Magazine, V. 6 (7/8), 2000.
2. Bridges, Karl, "Expectations of Librarians in the 21st Century", Greenwood Press.London. 2003.
3. Roush, Wade, "The Infinite Library", SPAN, January 2006, pp.8-14.
4. The Future of Cataloging. < <http://www.guild2910.org/marcum.htm>>
5. <http://hosted.ap.org>
6. <http://www.google.com>
7. <http://books.google.com>
8. <http://news.com.com>
9. <http://www.computer.org>
10. <http://www.searchenginewatch.com>