

## Invited Paper

# Digitisation and Digital Preservation

Jagdish Arora

Director, INFLIBNET, Ahmedabad-380 009

## 1. INTRODUCTION

It is a privilege for me to be here amidst you today for delivering Valedictory Address for the National Conference on Digitisation and Digital Preservation, the two topics that I am most interested in and I have written extensively on both the topics. May be that is why organisers of the Conference found it befitting to call me to say last words on it. They do not want any more writings from me on these topics.

It is indeed befitting to tackle the two topics together since digitisation and digital preservation are related to each other in more than one dimension. One of the main benefits of digitisation is to preserve rare and fragile objects by enhancing their access to multiple numbers of users simultaneously. The librarians have been concerned about the digital preservation ever since the first computer was introduced and its products and services found its way into the libraries. The libraries have been migrating and refreshing their OP AC records ever since automation in libraries started. Since mid 1980s, the libraries in India also started building their in-house databases and began subscribing electronic resources such as Current Contents on Disc (CCOD) as well as other computer-based services that were delivered on 5 1/4 inch floppy discs. Several books in 1980s and 1990s had accompanied floppy discs. These floppies are already obsolete and floppy drives that were used for reading these have completely disappeared. CD ROMs, once respected for its longevity, are known to dysfunction much faster than expected.

Moreover, in time to come, the CD ROM may completely be phased out in favour of its more evolved avatar, i.e., DVD ROM with greater storage capacity. Institutions such as national archives, data archives, and other cultural institutions with preservation as one of their main mandate, have established digital preservation programmes way back in late 1960s. These programmes addressed the issues of preservation of technology and digital contents that existed at that time (paper tapes, punch cards, etc.).

Libraries acquire digital materials through different channels that include buying digital contents from publishers or aggregators and licensing access to online databases and journals.

Moreover, libraries and institutions around the world are taking projects to convert their analogue collections into digital form with an aim to increase their access thus far confined to the four-walls of their libraries, many a times, without ensuring their long-term accessibility. The crucial issue of moving a digitisation pilot to a fully operational system with elements of preservation and sustainability built-in, is generally not given serious consideration that it deserve. The fact that the risk of loss of data in digital form is much greater than any other physical form is required to be addressed in greater details.

There are enough evidences to suggest that many potentially valuable digital materials have already been lost and it incurs substantial costs to recover these digital contents as observed in the following examples:

- ✧ The Census Bureau saved the 1960 census on Univac paper tapes that could be read only with a UNIVAC type II-A tape drive. By the mid-seventies, these paper tape drives were obsolete. When it was decided to archive the information on computer tapes containing the raw data from the 1960 federal census, there were only two machines in the world capable of reading those tapes: one in Japan and the other already deposited in the Smithsonian as a relic.
- ✧ NASA/NSFINOAA rescued valuable 20-year-long TOVS/AVHRR satellite data documenting global warming.
- ✧ In the late 1960s, the New York State Department of Commerce and Cornell University undertook the Land Use and Natural Resources Inventory Project (LUNR). The LUNR project produced a computerized map of New York State depicting patterns of land usage and identifying natural resources. It created a

primitive geographic information system by superimposing a matrix over aerial photographs of the entire state and coding each cell according to its predominant features. The data were used for several comprehensive studies of land use patterns that informed urban planning, economic development, and environmental policy. In the mid-1980s, the New York State Archives obtained copies of the tapes containing the data from the LUNR inventory along with the original aerial photographs and several thousand transparencies. Staff at the State Archives attempted to preserve the LUNR tapes, but the problems proved insurmountable. The LUNR project had depended on customised software programs to represent and analyse the data, and these programs were not saved with the data. Even if the software had been retained, the hardware and operating system needed to run the software were no longer available.

## 2. CHALLENGES FOR PRESERVING DIGITAL CONTENTS

Although, the digital technology offers several advantages over their print counter part, it along with other associated Internet and web technologies are in a continuous flux of change. New standards and protocols are being defined on a regular basis for file formats, compression techniques, hardware components, network interfaces, storage media and devices, etc. The digital contents face the constant threat of “techno-obsolence” and transitory standards.

Magnetic and optical discs as a physical media are being re-engineered continuously to store more and more data. There is a constant threat of backward compatibility for products, including software, hardware and associated standards and protocols that were used in the past. The challenges in maintaining access to digital resources overtime are related to notable differences between digital and paper-based material. Some of the important challenges for preserving digital contents are:

- ✘ Dynamic nature of digital contents.
- ✘ Machine dependency.
- ✘ Fragility of the media.
- ✘ Technological obsolescence.
- ✘ Shorter life span of digital media.
- ✘ Format and styles.
- ✘ Copyright and intellectual property rights (IPR) issues.

## 3. LONGEVITY

Density of media to record information has increased exponentially over time while its longevity to store the information has decreased proportionately. Density of media to store information is inversely proportional to its longevity.

Take for example, clay tablets, you can write only a few sentences on a clay tablets, but it can survive for centuries. Papyrus, paper with printed papers, has greater density of storing information, but has lesser longevity in comparison to clay tablets. The magnetic media and optical media have highest storage capacity, but are highly fragile and therefore longevity for these media is much lower.

## 4. DIGITAL PRESERVATION STRATEGIES

Digital preservation activities can broadly be divided into two components: (i) activities that promote the long-term maintenance of digital image, and (ii) activities that provide continued accessibility of its contents. Several strategies have been proposed but it is unlikely to find a single solution that is appropriate for all data types, situations, or institutions (Tristram, 2002). A set of digital preservation strategies can be applied depending on the lifespan of a digital object as mentioned below:

- ✘ *Long-term preservation:* Continued access to digital materials, or at least to the information contained in them, indefinitely.
- ✘ *Medium-term preservation:* Continued access to digital materials beyond changes III technology for a defined period of time but not indefinitely.
- ✘ *Short-term preservation:* Access to digital materials either for a defined period of time while use is predicted but which does not extend beyond the foreseeable future and/or until it becomes inaccessible because of changes in technology.

Preservation in the digital world is a challenging task for librarians and archivists. However, protocols, strategies and technologies involved in digital preservation have now been well defined and understood. Digital preservation is a cost-intensive activity of continuing nature. Library, archives, or museum cannot make a decision to adopt digitisation with long-term preservation and storage of research collections without deep and continuing commitment to preservation by the parent institution. The preservation in digital world is no more a prerogative of the libraries, but has become the mandate of the parent institution. The necessary financial and technological commitments to maintain digital contents and to migrate it

to future generations must be an organisational commitment. Failure to address to the well-defined digital preservation problems and strategies may result in loss of valuable digital data and may contribute to cultural and intellectual loss resulting in exorbitant costs for recovery, if at all possible. Librarians are compelled to meet the research challenge to resolve the conflict between the creation context and the use context to facilitate digital information preservation.

Digital resources, undoubtedly have several advantages over its analogue counterpart, however, preservation is definitively not one of them. The fact that the risk of loss of data in digital form is much greater than any other physical form is well understood and addressed

to. Long-term preservation of digital information is plagued by short media life, obsolete hardware and software, slow reading time of old media, and defunct websites (Chen, 2001). Dr S.S. Murthy and his team have completed the recommendations methodically.

The lesson learnt from the Conference, therefore, should be that any digitisation project or digital resource development project that we take up should have components for digital preservation built into it with institutional support built into it. The digital projects sanctioned by various funding agencies are given for a specific period of time. It is essential that the institutional support is sought to ensure continuing availability of digital contents.

#### About the Author

**Dr Jagdish Arora** is presently working as Director of Information and Library Network (INFLIBNET), Ahmedabad. Dr Arora is recipient of the Fulbright Professional Fellowship in Library and Information Science (1997-98), SIS Fellow (1999), Young Librarian of the Year (2001, SATKAL), Librarian of the Year (IASLIC, 1999), and ILA-Kuala Best Librarian Award, 2004. He was the Principal Investigator for several projects sponsored by agencies like AICTE, Department of Biotechnology (DBT), Ministry of Information Technology (MIT), Ministry of Human Resource Development (MHRD), the National Highway Authority of India and University Grants Commission. Dr Arora has more than 60 research articles to his credit in learned research journals and as chapters in books and conference proceedings. He has organised several workshops and training programmes at the IIT Delhi. He was resource person for several training programmes, refresher's courses and workshops, and has also presented research papers in several national and international conferences and symposia in India and abroad.