Digital Preservation: Strategies for Indian Libraries

Ravinder Kumar Chadha

Keywords: Digital Preservation, IPR

1. Introducrtion

Easy Internet access presumes that everyone can capture, access, and use the world's accumulated digital information. People, businesses, institutions, and governments invest time and effort to create and capture digital information for instantaneous access by anyone. Researchers are striving to make this information available to communities worldwide. Unfortunately, nobody can guarantee the continued preservation and accessibility of digital information generated in this era of rapid technological advances.

Traditionally, preserving things meant keeping them unchanged; however, the digital environment has fundamentally changed our concept of preservation requirements. If we hold on to digital information without modifications, accessing the information will become increasingly more difficult, if not impossible. Digital preservation presents its own unique challenges, arising from the basic nature of digital data—it is machine-readable, not eye-readable. Unlike the fairly straightforward process of decoding other machine-dependent media, such as microfilm or LPs, maintaining digital data in a form that is intelligible to humans involves the use of a complex set of tightly interwoven technologies.

Even if we could find a physical medium to contain unaltered digital data permanently, formats for recording the information would change and the hardware and software needed to recover the information would become obsolete. Digital preservation is plagued by the short media life, obsolete hardware and software, and slow read times of old media. Rapid technological advances do not solve the problem; instead, we need to migrate digital materials from one technology generation to another every few years. For digital records, the preservation issues extend beyond media life considerations. Devices for reading these media rapidly become obsolete; the various formats for digital documents and images introduce additional complications. Using research to develop policies, procedures, standards, and protocols based on solid frameworks provides accurate concepts and essential attributes of preservation in the digital information life cycle.

2. Advantages of Digital Access

Digital technology offers distinctive advantages to institutions with impressive collections of scholarly resources. Information content can be delivered directly to the reader without human intervention. Readers can retrieve information content in digital form remotely, although such delivery may tax the capabilities of even the most sophisticated projection equipment and networks. Digital image quality is extraordinary and is improving constantly. It is now possible to represent almost any type of traditional research material with such visual quality that reference to the original materials is unnecessary for most, if not all,

purposes. The power of full-text searching and sophisticated, cross-collection indexing affords readers the opportunity to make new uses of traditional research resources. Newly developed system interfaces (the look and feel of the computer screen) combined with new ways to deliver manageable portions of large image data files promise to revolutionize the ways in which research materials are used for teaching and learning. It is no wonder that there is a nearly overwhelming rush to jump on the digital bandwagon.

3. Digital Preservation Strategies

UNESCO's Guidelines for the Preservation of Digital Heritage (2003) group these strategies under the following four categories:

3.1 Short-term Strategies

Bit-stream Copying

Refreshing

Replication

Technology Preservation or Computer Museum

Backwards Compatibility and Version Migration

3.2 Medium- to Long-term Strategies

Migration

Viewers and Migration at the Point of Access

Emulation

Canonicalization

Emulation

3.3 Investment Strategies

Restricting Range of Formats and Standards

Reliance on Standards

Data Abstraction and Structuring

Encapsulation

Software Re-engineering

Universal Virtual Computer

3.4 Alternative strategies

Analogue Backups

Digital Archaeology or Data Recovery

4. Combinations

These strategies have demonstrated to work in certain circumstances over limited periods of time. None of them have proven themselves against unknown threats over centuries of change. Most of these strategies are, however, being used in the management of data, and it is likely that combinations of these strategies will continue to be researched and proposed for large-scale, long-term preservation. It is, therefore, reasonable for preservation programmes to look for multiple strategies, especially if they are responsible for a range of materials over extended periods.

4.1 Short-term Strategies

Short-term digital preservation strategies are likely to work for a short period of time only. These strategies include bit-stream copying, refreshing, replication, technology preservation or computer museum, backwards compatibility and version migration.

4.2 Bit-stream Copying

Bit-stream copying, commonly known as "backing up data" refers to the process of making an exact duplicate of a digital object. It deals only with the question of data loss due to hardware and media failure, whether resulting from normal malfunction and decay, malicious destruction or natural disaster.It should be considered the minimum maintenance strategy for even the most lightly valued, ephemeral data.

4.3 Refreshing

Refreshing essentially means copying digital information from one long-term storage medium to another of the same type, with no change whatsoever in the bit-stream (e.g. from an older CD-

RW to a new CD-RW). "Modified refreshing" is the copying to another medium of a similar type with no change in the bit-pattern that is of concern to the application and operating system using the data, e.g. from a QIC tape to a 4mm tape; or from a 100 MB Zip disk to a 750 MB Zip disk. Refreshing is a necessary component of any successful digital preservation project. It potentially addresses both decay and obsolescence issues related to the storage media.

4.4 Replication

Replication is used to represent multiple digital preservation strategies. Bit-stream copying is a form of replication. LOCKSS (Lots of Copies Keeps Stuff Safe) is a consortia form of replication, while peer-to-peer data trading is an open, free-market form of replication. LOCKSS uses low-cost tools to crawl the Web to cache "redundant, distributed, decentralized" e-journal presentation files for which a library has a subscription or license. LOCKSS supports the traditional model whereby individual libraries build and maintain local collections of journals, and work is underway to develop a user interface for local collection management of e-journals cached using the LOCKSS system. A LOCKSS Alliance of participating libraries has been formed and the system is currently in beta test mode. The intention of replication is to enhance the longevity of digital documents while maintaining their authenticity and integrity through copying and the use of multiple storage locations.

4.5 Technology Preservation

Technological preservation is based on keeping and maintaining the technical environment that is used for creation of contents including operating systems, original application software, media drives, etc. It is sometimes called the "computer museum" solution. In other words, technological preservation becomes applicable to digital materials that are left on obsolete storage media and hardware and software required to access them are discarded. Technology preservation is more of a disaster recovery strategy for use on digital objects that have not been subjected to a proper digital preservation strategy.

4.6 Backwards Compatibility and Version Migration

This strategy relies on the ability of current versions of software to interpret and present digital material created with previous versions of the same software and to save them in current format. In the case of backwards compatibility, the presentation may be limited to temporary viewing, whereas version migration permanently converts documents into a format that can be presented by the current version of the software.

5. Medium to Long-term Preservation Strategies

Strategies proposed for medium and long-term preservation are likely to work for a long period of time. Such strategies should be used for digital materials that are likely to be of value for a long period of time. Medium and long-term preservation strategies include:

5.1 Migration

Migration is a broader and richer concept of digital preservation than "refreshing". Migration is a set of organized tasks designed to achieve the periodic transfer of digital materials from one hardware / software configuration to another, or from one generation of computer technology to a subsequent generation. The purpose of migration is to preserve

the integrity of digital objects and to retain the ability for clients to retrieve, display, and otherwise use them in the face of constantly changing technology. Migration includes refreshing as a means of digital preservation but differs from it in the sense that it is not always possible to make an exact digital copy or replica of a database or other information object as hardware and software change and still maintain the compatibility of the object with the new generation of technology.

5. 2 Viewers and Migration at the Point of Access

Migration or providing viewing facility at the point of access has been proposed as an alternative to recurring and incremental migration. The process involves use of appropriate viewers, software tools or transformation methods that provide accessibility at the time of access, using the original data stream. For example: The VERS strategy converts documents to a PDF format on the basis that third-party viewers for PDF may be constructed from the format specification.

Limitations of this approach includes i) viewers may not be available for all formats such as executable files; ii) Viewers may be able to represent some, but not all, elements of digital materials; iii) the gap between the original format and the prevailing technologies at the time of access may be too great for the tools or methods to cope with; and iv) Viewers, tools or methods, and corresponding metadata must also be maintained or adjusted as technologies change.

5.3 Cannibalisation

Cannibalisation is a technique designed to allow determination of whether the essential characteristics of a document have remained intact through a conversion from one format to another. Cannibalisation relies on the creation of a representation of a type of digital object that conveys all its key aspects in a highly deterministic manner. Once created, this form could be used to algorithmically verify that a converted file has not lost any of its essence. Cannibalisation has been postulated as an aid to integrity testing of file migration, but it has not been implemented.

5.4 Emulation

Emulation uses a special type of software, called an emulator, to translate instructions from original software to execute on new platforms. The old software is said to run "in emulation" on newer platforms. This method attempts to simplify digital preservation by eliminating the need to keep old hardware working. Emulation combines software and hardware to reproduce in all essential characteristics the performance of another computer of a different design, allowing programs or media designed for a particular environment to operate in a different, usually newer environment. Emulation requires the creation of emulator programs that translate code and instructions from one computing environment so it can be properly executed in another.

6. Investment Strategies

Investment preservation strategies involve investment of efforts at the time of archiving digital materials. Such strategies include: Restricting Formats and Standards, Reliance on Standards, Data Abstraction and Structuring, Encapsulation, Software Re-engineering and Universal Virtual Computer.

6.1 Restricting Formats and Standards

Preservation programmes may decide to only store data in a limited range of formats and standards. This can be achieved either by only accepting

material in specified formats or by converting material from other formats before storage. All digital objects within an archival repository of a particular type (e.g., colour images, structured text) can be converted into a single chosen file format that is thought to embody the best overall compromise amongst characteristics such as functionality, longevity, and preservability. For, example most of the textual and graphical information can be converted into PDF format. The UK Archaeology Data Service (ADS), for example, specifies a preferred (but not exclusive) range of formats for deposit and provides guidelines for depositors on creating or preparing materials for submission.

The strategy does not necessarily solve the access problem unless the obsolescence of formats and standards used are handled effective through some other strategy. This strategy imposes serious restrictions on the range of materials that a preservation programme can accept. Moreover, the process of conversion from original format may cause some loss of essential elements.

6.2 Reliance on Standards

This preservation strategy involves the use of open, widely available and supported standards and file formats that are likely to stable for a longer period of time discarding proprietary or less-supported standards. Such standards or formats may either be formally agreed or may be de facto standard formats that have been widely adopted by industry. For example, majority of digitisation programmes choose TIFF (Tagged Image File Format) as an open, stable and widely supported standard for creation of preservation master images. Similarly, most publisher use PDF as de facto standard for electronic distribution of their research articles, due to the availability of PDF readers for all platforms. Reliance on standards may lessen the immediate threat to a digital document from obsolescence, but it is not a permanent preservation solution.

6.3 Data Abstraction and Structuring

Data abstraction, sometimes also called normalization, involves analyzing and tagging data so that the functions, relationships and structure of specific elements can be described. Using data abstraction, the representation of content can be liberated from specific software applications, the digital contents can, however, be read using different applications as technology changes. Data abstraction makes a document application-independence and simplifies the transport of data between platforms and over generations of technology. The technique, however, has its limitation, it requires extensive development of tools and methods for analysis and processing in order to correctly represent and tag each type of data. Moreover, technology eventually used for presentation may still limit what functions can be represented.

6.4 Encapsulation

Encapsulation may be seen as a technique of grouping together digital objects and metadata necessary to describe and provide access to that object. The grouping process lessens the likelihood that any critical component necessary to decode and render a digital object will be lost. Encapsulation is considered a key element of emulation.

Encapsulation may also bundle metadata that describe or provide link to the software applications and platform used for original contents considering the fact that it is impractical and unnecessary to

encapsulate the software. Open Archival Information System (OAIS) Reference Model, for example, describes incorporating data objects and their associated metadata into Archival Information Packages (AIPs).

6.5 Software Re-engineering

Digital materials are mostly tied to the application software used for creating them. The application software, in turn, are dependent on a specific system or platform in order to function. Application software get most affected by changes in technology. Moreover, they are also usually unsuited

for preservation strategies, including regular migration. Software reengineering may offer a number of strategies for transforming software as technologies change, similar to transformation of data formats. Some possibilities include:

Adjustment and re-compiling of source code for a new platform;

- Reverse-engineering of compiled code into higher level code and porting that to the new platform;
- Re-coding of the software from scratch, or re-coding in another programming language; and
- Translation of compiled binary instructions for one platform directly into binary instructions for another platform.

Reengineering application would require source code, which may not be available except for open source programmes and software that are developed in-house. Even when source code is available, porting to other platforms is not a trivial job, it requires considerable time and effort per object. Moreover, compilers or interpreters are required for the new platform for the code language.

6.6 Universal Virtual Computer

Universal Virtual Computer is a form of emulation. It requires development of a computer program independent of any existing hardware or software that could simulate the basic architecture of every computer since the beginning, including memory, a sequence of registers, and rules for how to move information among them. Users could create and save digital files using the application software of their choice, but all files would also be backed up in a way that could be read by the universal computer. To read the file in the future would require only a single emulation layer—between the universal virtual computer and the computer of that time.

This approach requires substantial investments both at the time of archiving while developing encoding methods or UVC-native interpretive programmes for each data type as well as at the time of restoration in developing a UVC emulator and restore programmes. Moreover, if original data objects are abstracted or transformed for encoding purposes, such transformation may discard essential characteristics.

The proof-of-concept prototype for the UVC approach (Lorie, 2002) has been used to produce a logical schema, decoder programme and representation mechanism for PDF documents, such that the document content can be represented using a UVC interpreter and restore programme.

7. Alternative Strategies

Alternative strategies to digital preservation include taking analogue backup of document (print or microfilm) or recovering data from obsolete digital media.

7.1 Analogue Backups

Analogue backups combine the conversion of digital objects into analogue form with the use of durable analogue media, e.g., taking high-quality printouts or the creation of silver halide microfilm from digital images. An analogue copy of a digital object can, in some respects, preserve its content and protect it from obsolescence, without sacrificing any digital qualities, including sharability and lossless transferability. Text and monochromatic still images are the most amenable to this kind of transfer. Given the cost and limitations of analogue backups, and their relevance to only certain classes of documents, the technique only makes sense for documents whose contents merit the highest level of redundancy and protection from loss.

Limitation of this strategy includes i) advantages offered by digital technology such as convenience of use, storage efficiency, search and navigation possibility is lost; ii) the strategy does not completely remove the threat of technological obsolescence; and iii) long-term stability of analogue material may depend on expensive storage environments that prove to be less reliable than well-managed computer systems based on high levels of redundancy.

7.2 Digital Archaeology

Digital archaeology includes methods and procedures to rescue content from damaged media or from obsolete or damaged hardware and software environments. Digital archaeology is explicitly an emergency recovery strategy and usually involves specialized techniques to recover bit-streams from media that has been rendered unreadable, either due to physical damage or hardware failure such as head crashes or magnetic tape crinkling. Digital archaeology is generally carried out by commercial data recovery companies by maintaining a variety of storage hardware (including obsolete types) plus special facilities such as clean rooms for dismantling hard disk drives. Given enough resources, readable bit-streams can often be recovered even from heavily damaged media (especially magnetic media), but if the content is old enough, it may not be possible to make it renderable and /or understandable.

7.3 Combination Strategies

As mentioned before, no single strategy is appropriate for all data types, situations, or institutions. A number of strategies may, therefore, be necessary to cover the range of objects and characteristics to be preserved. Preservation programmes should also consider the potential benefits of redundancy in pursuing more than one strategy. It may be noted that even with good planning, a single strategy may fail leaving the programme with no means of access. Several digital preservation projects use more than one approach, for example:

- Standards such as TIFF for image collections are often chosen in preparation for eventual migration to other standard formats over the long-term;
- The VERS strategy couples the use of standards (PDF, XML) to the future use of viewers and the likely migration of XML encoded metadata in the future;
- Persistent archives use data abstraction with the view to eventual migration – migration of the data, the mark up system and the supporting software, and upgrading of hardware;
- The Universal Virtual Computer (UVC) approach combines data abstraction with rules for migration of data objects at the point of

access, and an emulation approach for software objects. The "durable encoding" approach ads the use of fundamental standards for encoding data, including encoding that could be understood by the UVC.

8. IPR Issues

IPR issues are not simple in the digital preservation world, where migration copies, archival copies, derivative versions, and other states of an object exist, over a period of time. Meeting legal requirements for preserving digital objects requires careful, comprehensive, ongoing approaches that avoid risk to the organization

Each institution must determine if the material it is seeking to preserve is in the public domain, or if someone other than the institution owns the copyright. It may want to try to locate the copyright owner and license (perhaps at a cost) those rights needed to preserve a work. Alternatively, it may conclude that preservation activities are authorized by "fair use". As per the Copyright laws the author/ Copyright holder has exclusive right to copy an item of work. In addition he has right to:

- prepare derivative works based upon the copyrighted work
- distribute copies of the copyrighted work to the public
- perform some copyrighted works publicly
- display some copyrighted works publicly
- in the case of sound recordings, to perform the copyrighted work publicly by means of a digital audio transmission in the United States
- control access to a work protected by the use of a technological measure

Digital preservation strategies may impinge on these rights. Migration, for example, may be a violation of the copyright owner's right to prepare a derivative work. Making a digital work broadly available may impinge on the copyright owner's distribution, performance, and display rights. Preserving a password-protected or encrypted file may require violating the copyright owner's exclusive right to control access. Given that almost everything is copyrighted and the copyright owner has extensive exclusive rights, how can digital preservation occur without risking copyright infringement?

8.1 US Copyright Law section 108 (Limitations on exclusive rights: Reproduction by libraries and archives): The Section 108 of US Copyright Law as modified by Digital Millennium Copyright Act of 1998 has a provision that allows libraries and archives to copy, digitise and make accessible published documents in their collections. The amended, section 108 also permits up to 3 digital copies of unpublished and damaged works for preservation of copyrighted material provided that digital copies are not made available to the public outside the library premises or put on the Internet. This further permits a library to copy a work into a new format if the original format becomes obsolete—that is, the machine or device used to render the work perceptible is no longer manufactured or is no longer reasonably available in the commercial marketplace.

In order to be able to take advantage of the exemptions, certain ground rules must be met. The library or archives must be open to the public; the copying cannot be for "direct or indirect commercial advantage;" and any copies made must carry a notice of copyright.

Assuming that those conditions are met, libraries and archives can engage in limited copying for preservation purposes without fear of infringement. However, certain other requirements apply:

- You must own a copy of the original.
- The copying must be solely for preservation or security.
- The original must be "damaged, deteriorating, lost, or stolen," or the existing format in which the work is stored is obsolete.
- A reasonable investigation reveals that an unused copy cannot be obtained at a fair price.

8.2 US Copyright Law section 107 (Limitations on exclusive rights: Fair use): Another exemption libraries and archives can use for their digital preservation programs is Section 107, Fair Use. Fair use is a judicially interpreted doctrine decided on a case-by-case basis. You have no assurance that any specific use is fair until a judge tells you it is fair. And while fair use is supposed to favour reproduction done for the purpose of teaching, scholarship, or research, not all copying done for such purposes is automatically fair. In determining whether a use is fair, a court must consider no less than four factors. These are:

- the **purpose** of the use (including whether the use transforms the original into
- something new or merely replicates the original)
- the **nature** of the original material (whether it is primarily creative or factual)
- the **amount** of the original duplicated
- the effect on potential **market** or value of the original

Given the social benefit of preservation, it seems likely that the courts would tolerate a preservation program that sought only to preserve digital information but did not seek to distribute it to others.

Any digital preservation program is likely to exist in a grey area of legality. It is important for those charged with digital preservation responsibilities to understand that, while many actions may be acceptable, the area in which they can work with legal certainty (primarily the exceptions afforded by Section 107 and 108) is extremely limited. It is imperative, therefore, that digital preservation programs remain in close contact with their institution's legal advisors to ensure that they do not place their institution at an unacceptable level of risk.

9. Strategies for Indian Librarianship

Indian IT industry is progressing at a very fast pace. India is considered as superpower in the IT field. We have been generating huge amount of digital information since last 25 years. We, however, have not fully realised the importance for preservation of digital data being generated by government agencies, research organisations, academic institutions, cultural and commercial organisation, etc. There is an immediate need to address the issue of digital preservation at national level and formulate a National Digital Preservation Policy. This Conference can recommend to Ministry of Culture and Ministry of Information Technology for the same.

In India digital preservation will need to be a distributed responsibility. This is partly because of enormous amount of digital material being produced by a large number of organisations and partly because of the problems related to digital preservation mentioned. However, decisions regarding preservation of digital information need to be taken at an early stage so that those creating digital data are logically the ones best able to undertake that initial activity. It is also a factor that solutions are not going to be of the nature of "one size fits all". Different approaches have to be

adopted for different types of digital resources and, while duplication of effort is to be avoided, a certain amount of judicious overlap can be beneficial, particularly in these early stages of developing digital repositories.

The role of organisations creating digital materials is both crucial and difficult to integrate into a coherent infrastructure for preserving digital materials. This is mainly due to the reason that they may be reluctant to hand over their materials elsewhere. Libraries and archives have established their credentials for preserving print materials over a very long timeframe. In these very early stages of developing digital repositories, it may be difficult for creators to assign the same level of trust to the librarians for preserving digital materials. So the library Professionals have to take lead and undertake the responsibility of preserving the important data of their organisations.

Some creators of digital materials may be best placed to undertake preservation responsibility because of their in-depth knowledge of the subject matter, but they may well lack the necessary archiving skills. The optimum solution in these cases might be an alliance between an organisation skilled in managing digital data, and the creators, so that those with the greatest knowledge of the material maintain control over decisions on what content needs to be preserved and at what intervals.

The need for the development of reliable tools and services has been recognised throughout the world and related developments are taking place everywhere. This is yet another example of the global nature of digital preservation and the tendency for the same issues to emerge in different parts of the globe at much the same time. While focussing primarily on developing the Indian digital preservation agenda, we have to recognise that digital preservation is very much a global issue and it is critically important to establish good lines of communication with all those engaged in digital preservation efforts.

The Indian libraries should also work with respect to the cultural record is being creating in digital forms. The digital environment is still relatively uncultivated at this stage, but the need is urgent, the time is opportune and the conditions are fertile for a strong, far-sighted set of cultivating actions to help ensure that the digital record ultimately matures and flourishes. By analysing the emerging digital environment, and by setting up digital archives so as to identify the most demanding preservation. The following issues need to be addressed urgently.

i) Primary responsibility of preserving digital data rests with the creators, providers and owners of digital information.

ii) Long-term preservation of digital information on a scale adequate for the demands of future research and scholarship will require a deep infrastructure capable of supporting a distributed system of digital archives.

iii) A critical component of the digital archiving infrastructure is the existence of a sufficient number of trusted organizations capable of storing, migrating and providing access to digital collections.

iv) A process of certification for digital archives is needed to create an overall climate of trust about the prospects of preserving digital information.

v) Certified digital archives must have the right and duty to exercise an aggressive rescue

function as a fail-safe mechanism for preserving valuable digital information that is in jeopardy of destruction, neglect or abandonment by its current custodian.

In view of the above, it is proposed that in India we must set up a National Centre for Digital Preservation. The NCDP will not be a repository for digital data but will work towards a more effective and efficient infrastructure for digital preservation within the country. It will set tone for initiating digital preservation activities in a coordinated manner.

The NCDP may undertake activities like developing pilot projects, propose support structures, and the development of best practice. It may undertake the following activities:

9.1 To coordinate with existing and potential digital archives around the country and provide coordinating services for better preservation of digital data.

Action is urgently needed to ensure that documents, software products and other digital information objects are preserved before they slip irrevocably away. The proposed NCDP may undertake a project designed with this particular focus as a cooperative venture so as to develop strategies for preserving precious data in distributed digital archives. Because the objects in this focal area are at such risk of loss, the project would also provide a useful means of exploring the operations of certification and fail-safe mechanisms for digital archives.

9.2 To initiate national debate on setting up of advance digital archives, particularly with respect to removing legal and economic barriers to preservation.

A national debate may be sponsored by the NCDP to generate an enormous amount of creative thinking about the commitment to the development of digital archives. It might be focused on fostering creative alliance, especially with publishers, and practical, joint efforts designed to lower the legal and economic barriers to the effective operation of digital archives.

9.3 To recommend archival application of technologies and services, such as hardware and software emulation algorithms, transaction systems for property rights and authentication mechanisms, which promise to facilitate the preservation of digital data.

Only through early and active use will digital archives be able to influence the development of key new technologies and services and help to ensure that they support information longevity. Moreover, there is growing need for evidence that digital archives can practically and effectively incorporate in their daily operations automated systems for emulating obsolete hardware and software, transacting intellectual property and using cryptographic and other mechanisms for creating trusted distribution channels for digital information.

9.4 To develop national information infrastructure to ensure that longevity of digital information is an explicit goal.

NCDP may work for developing distributed network of linked digital information archives in which digital information will flourish over the long-term. Communication and information network policy decisions regarding pricing, security and network extension will greatly affect the viability of these archives and their efforts to preserve digital information. These policy decisions need to be informed with an understanding of the importance and complexity of digital preservation.

9.5 To prepare of a white paper on the legal and institutional foundations needed for the development National Depository of digital data.

To work out a proposal for amending the Delivery of Books Act 1956 so as to have a legally mandated system of deposit for published works, in which publishers are required to place with a certified digital archives a copy of a work in a standard archival format in addition to the printed copies.

9.6 To create subject digital repositories.

ICSSR, NISCAIR, DESIDOC, NML, INSA and other such apex bodies in specialised subject areas should be encouraged to create subject digital repositories in their subject area.

9.7 To examine test and implement emerging standards and tools regarding formats, hardware/software, security, access rights management, etc.

NCDP may work out standards for digital archives and administer the process of digital archival certification. The appropriate individuals and organizations need now to begin systematically to identify and describe the standards, criteria and mechanisms for archival certification and thereby launch the process that would lead ultimately to a formal certification program.

9.8 To act as national agency for coordinating digital preservation initiatives in the country and also coordinate with other countries.

There is considerable evidence of worldwide interest in the means of preserving digital information. A number of agencies are working at national level viz. the European Union, the Consortium of University Research Libraries in Great Britain and a national Working Party in Australia on the management of material in electronic format. They have generated working papers on the topic of digital preservation and invited international collaboration. India need a nodal agency to identify and facilitate international collaborative efforts in the field of digital preservation.

9.9 To identify current best practices and to benchmark such practices for being used in the country.

a. The design of systems that facilitate archiving at the creation stage.

There is a need to provide long-term access to government and scholarly data being produced in the country. There will be need to study how are publishers redesigning the creation process to support their electronic publishing programs? What software are they using and how have they influenced software producers to modify their development of their products and suggest solutions for various organisations in the country.

b. Storage of massive quantities of culturally valuable digital information.

There is an immediate need to develop large digital archives for social and culturally important information. Examples include the archives of census data, remote sensing satellite imagery, weather data land records, scholarly output, research data, etc.

c. Requirements and standards for describing and managing digital information.

Descriptive information about the content of digital objects, their origins and provenance and their management over time is critical for both long-term preservation and future use of digital information. Standards and best practices for describing and managing digital information are needed to track changes in ownership or control over digital objects throughout their life cycle, to administer intellectual property rights, and to document any changes in the format and structure of digital objects that may ensue from migration.

A responsible digital archive must provide to its users what it knows about the provenance and context of its objects so that users can make informed decisions about the reliability and quality of the evidence before them. Ministry of IT, Ministry of Culture, Professional bodies, Library and Information Science professionals, need to collaborate in an evaluation and expansion of descriptive standards and practices so that they satisfy the special requirements of digital preservation and access.

d. Migration paths for digital preservation of culturally valuable digital information

Data migration is a common practice as organizations preserve their essential data records through successive changes in hardware and software. Cultural archives that have been collecting digital objects have also had to begin migrating them as the hardware and software on which they were created has become obsolete. What is the range of experience of different organizations with archiving different types of content? What can be learned and generalized from these experiences? How do strategies compare among different organizations for archiving similar materials. Are there economies of scale that could be achieved by combining efforts across digital archives? What are the costs of the different strategies employed? What strategies have failed? In what ways have practices improved over time? On the basis of analysis of the above, the NCDP may formulate recommendation for national digital programme and initiate dialogue at international level with similar agencies

10. Conclusion

Decisions about preserving information should consider the costs. We can use current technology to determine the costs of retaining information; however, both expenditures and technology will evolve. Whereas we can project the costs for basic elements of technology—such as digital media per unit volume of information and unit processing by computers—there are no proven techniques for estimating he costs of long-term digital information preservation.

We can now make information easily available to communities worldwide via the Internet. We, however face the challenge of preserving digital information with its paradox of short media life, obsolete hardware and software, slow read times of old media, and defunct Web sites. Despite the wealth of accumulated, technology- generated information, we currently lack proven methods for preserving this information or for using optimal technology tools to access it and determine its authenticity. Failure to address these digital preservation problems is analogous to squandering potential professional, personal, and economic gains, contributing to cultural and intellectual poverty, and resulting in exorbitant costs for recovery. We are compelled to meet the research challenge to resolve the conflict between the creation context and the use context to facilitate digital information preservation.

There are numerous challenges before us, but also enormous opportunities to contribute to the development of a national infrastructure that positively supports the long-term preservation of digital information. Such an infrastructure is a desirable outcome that will benefit us only if we conceive and structure it to benefit posterity.

About Author

Dr. Ravinder Kumar Chadha, Joint Secretary, Parliament of India, Lok Sabha Secretariat, New Delhi 110001.

E-mail: ravichad@gmail.com