PERSPECTIVE OF DIGITAL PRESERVATION: NEED AND STRATEGIES IN THE DIGITAL AGE

Kh Surachand Singh ***Memori Sagolshem*** ***Th Purnima Devi***

Abstract

While digital technologies are enabling information to be created, manipulated, disseminated, located, and stored with increasing ease, preserving access to this information poses a significant challenge. Digital preservation involves the use of well-defined techniques to prevent the original artifact from deteriorating further and to perhaps even improve it to the point where it can be used again. It involves quite different methods, skills, and outcomes and can complement traditional preservation services, while simultaneously providing unique and dynamic new uses of information.

Keywords: Digital Preservation; Digital Attributes; Digital information; Emulation; Digital preservation Strategies

1. Introduction

In a very short time, preservation has developed into a critically important part of managing a library's most precious assets, its collections. — Abby Smith, "Preservation in the Future Tense"

It is widely accepted that information technology is revolutionizing our concepts of documents and records in an upheaval at least as great as the introduction of printing, if not of writing itself. The current generation of digital records therefore has unique historical significance; yet our digital documents are far more fragile than paper. Infact, the record of the entire present period of history is in jeopardy. The content and historical value of many governmental, organizational, legal, financial, and technical records, scientific databases, and personal documents may be irretrievably lost to future generations if we do not take steps to preserve them. Digital preservation is emerging as a trustworthy process, yet there is much ongoing debate, and skepticism abounds, concerning the viability and even the meaning of this process. Given the nature of electronic storage technologies and the ephemeral nature of Web pages, many are doubtful that digital preservation will ever become a reality.

2. Definitions

Digital preservation is defined as a long-term, error-free storage of digital information, with means for retrieval and interpretation, for the entire time span that the information is required for. "Retrieval" means obtaining needed digital files from the long-term, error-free digital storage, without possibility of corrupting the continued error-free storage of the digital files. "Interpretation" means that the retrieved digital files, files that, for example, are of texts, charts, images or sounds, are decoded

and transformed into usable representations. This is often interpreted as "rendering", i.e. making available to a human to access. However, in many cases it will mean able to be processed by computational means.

The Research Libraries Group has proposed the definition of Digital preservations as follows:

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

3. Need for Digital Preservation

The vision of creating digital libraries that will be able to preserve our heritage currently rests on technological quicksand. There is as yet no viable long-term strategy to ensure that digital information will be readable in the future. Not only are digital documents vulnerable to loss via media decay and obsolescence, but they become equally inaccessible and unreadable if the software needed to interpret them—or the hardware on which that software runs—is lost or becomes obsolete. Unlike the preservation of paper or microfilm, the preservation of digital information demands ongoing attention. This constant input of effort, time, and money to handle rapid technological and organisational advance is considered the main stumbling block for preserving digital information. Indeed, while we are still able to read our written heritage from several thousand years ago, the digital information created merely a decade ago is in serious danger of being lost. Preservation is therefore a more immediate issue for digital than for traditional resources.

4. Requirements for Digital Preservations

The preservation of digital documents involves a number of distinctive requirements. In particular, all such documents possess a unique collection of core digital attributes that must be retained. These attributes include their ability-

- to be copied perfectly,
- to be accessed without geographic constraint, to be disseminated at virtually no incremental cost (given the existence of appropriate digital infrastructure), and
- to be machine-readable so that they can be accessed, searched, and processed by automated mechanisms that can modify them, reformat them, and perform arbitrary computations on their contents in all phases of their creation and distribution.

Furthermore, new inherently digital ("born-digital") document forms, such as dynamic, distributed, interactive hypertext and hypermedia, must retain their unique functionality, including their ability to integrate information from disparate traditional sources, such as books, periodicals, newspapers, mail, phone messages, data, imagery, and video.

5. The Digital Longevity Problem

Documents, data, records, informational and cultural artifacts of all kinds are rapidly being converted to digital form, if they were not created digitally to begin with. This rush to digitize is being driven by powerful incentives, including the ability to make perfect copies of digital artifacts, to publish them on a wide range of media, to distribute and disseminate them over networks, to reformat and convert them into alternate forms, to locate them, search their contents, and retrieve them, and to process them with automated and semi-automated tools. Yet, the longevity of digital content is problematic for a number of complex and interrelated reasons.

It is now generally recognized that the physical lifetime of digital storage media is often surprisingly short, requiring information to be "refreshed" by copying it onto new media with disturbing frequency. The technological obsolescence of these media (and of the hardware and software necessary to read them) poses a different and equally urgent threat. Moreover, most digital documents and artifacts exist only in encoded form, requiring specific software to bring their bit streams to life and make them truly usable; as these programs (or the hardware/software environments in which they run) become obsolete, the digital documents that depend on them become unreadable—held hostage to their own encoding. This problem is paradoxical, given the fact that digital documents can be copied perfectly, which is often naively taken to mean that they are eternal. There is currently no demonstrably viable technical solution to this problem; yet, if it is not solved, our increasingly digital heritage is in grave risk of being lost. In addition to the technical aspects of this problem, there are administrative, procedural, organizational, and policy issues surrounding the management of digital material.

6. Strategies for Digital Preservations

There are several strategies, which individuals and organizations may use to combat the loss of digital information.

- **Refreshing:** Refreshing is the copying of data onto newer media or systems. For example, transferring census data from an old tape to a new one or transferring an MP3 from a hard drive to CD. This strategy may need to be combined with migration when the software or hardware required to read the data is no longer available or is unable to understand the format of the data. Refreshing will likely always be necessary due to the deterioration of physical media.
- **Migration:** Migration is the transferring of data to newer system environments (Garrett et al., 1996). This may include conversion of resources from one format to another (e.g., conversion of Microsoft Word to PDF or Open Document), from one operating system to another (e.g., Solaris to Linux) or from one programming language to another (e.g., C to Java) so, the resource remains fully accessible and functional. Resources that are migrated run the risk of losing some type of functionality since newer formats may be incapable of capturing all the functionality of the original format, or the converter itself may be unable to interpret all the nuances of the original format. The latter is often a concern with proprietary data formats.
- **Replication:** Creating duplicate copies of data on one or more systems are called replication. Data that exists as a single copy in only one location is highly vulnerable to software or

hardware failure, intentional or accidental alteration, and environmental catastrophes like fire, flooding, etc. Digital data is more likely to survive if it is replicated in several locations. Replicated data may introduce difficulties in refreshing, migration, versioning, and access control since the data is located in multiple places.

- **Emulation:** Emulation is the replicating of functionality of an obsolete system (Rothenberg, 1998). For example, emulating an Atari 2600 on a Windows system or emulating WordPerfect 1.0 on a Macintosh. Emulators may be built for applications, operating systems, or hardware platforms. Emulation has been a popular strategy for retaining the functionality of old video game systems. The feasibility of emulation as a catch-all solution has been debated in the academic community (Granger, 2000).
- **Universal Virtual Computer:** Raymond A. Lorie has suggested a Universal Virtual Computer (UVC) could be used to run any software in the future on a yet unknown platform (Lorie, 2001). The UVC strategy uses a combination of emulation and migration. The UVC strategy has not yet been widely adopted by the digital preservation community.
- **Trustworthy digital objects:** Digital objects that can speak to their own authenticity are called trustworthy digital objects (TDOs). TDOs were proposed by Henry M. Gladney to enable digital objects to maintain a record of their change history so future users can know with certainty that the contents of the object are authentic (Gladney, 2004). Other preservation strategies like replication and migration are necessary for the long-term preservation of TDOs

7. Preservation at the Digital Age

The goal of any preservation program is to ensure long-term, ready access to the information resources of an institution. —Abby Smith, "Preservation in the Future Tense"

The best way to satisfy the criteria for a solution is to run the original software under emulation on future computers. This is the only reliable way to recreate a digital document's original functionality, look, and feel. Though, it may not be feasible to preserve every conceivable attribute of a digital document in this way, it should be possible to recreate the document's behavior as accurately as desired—and to test this accuracy in advance.

The implementation of this emulation approach involves:

1. developing generalizable techniques for specifying emulators that will run on unknown future computers and that capture all of those attributes required to recreate the behavior of current and future digital documents;
2. developing techniques for saving—in human-readable form—the metadata needed to find, access, and recreate digital documents, so that emulation techniques can be used for preservation; and
3. developing techniques for encapsulating documents, their attendant metadata, software, and emulator specifications in ways that ensure their cohesion and prevent their corruption. The only assumption that this approach makes about future computers is that they will be able to perform any computable function and (optionally) that they will be faster and/or cheaper to use than current computers.

The central idea of the emulation strategy is to emulate obsolete systems on future, unknown systems, so that a digital document's original software can be run in the future despite being obsolete.

There are some difficulties in saving the digital documents. It may be due to factors such as media decay and software and hardware obsolescence, it is sometimes suggested that they be printed and saved in hard-copy form. Printing the traditional documents results in the loss of their unique functionality (such as dynamic interaction, nonlinearity, and integration), and printing any document makes it no longer truly machine-readable, which in turn destroys its core digital attributes (perfect copying, access, distribution, and so forth). Beyond this loss of functionality, printing digital documents sacrifices their original form, which may be of unique historical, contextual, or evidential interest. Proposed alternatives to printing digital documents include translating digital documents into standard forms or extracting their contents without regard to their original nature. Though these approaches have traditional analogues (such as the translation of ancient texts into the vernacular to give them a larger audience), they are fraught with danger. The meaning of a document may be quite fragile, since meaning is in the eye of the beholder: what may be a trivial transformation to a casual reader may be a disastrous loss to a scholar, historian, or lawyer. While it is often useful to create contemporary vernacular transcriptions of historical documents (such as Shakespeare's Sonnets or the Declaration of Independence), society places a high value on retaining the originals so that we may verify that content has not been lost in transcription, as well as for scholarly and aesthetic purposes. For digital documents, retaining an original may not mean retaining the original medium (which rapidly decays and becomes obsolete), but it should mean retaining the functionality, look, and feel of the original document.

8. Conclusion

Digital preservation can, therefore, be seen as the set of processes and activities that ensure the continued access to information and all kinds of records, scientific and cultural heritage existing in digital formats. A serious commitment to preserving digital information requires a legal environment that enables preservation. It also means that specific organizations—libraries, government agencies, corporations—must take responsibility for preservation by enacting new policies and creating the economic means to secure survival of this generation's knowledge into the future.

References

1. Gladney, H. M. (2004). "Trustworthy 100-year digital objects: Evidence after every witness is dead". ACM Transactions on Information Systems 22 (3): 406-436.
2. Granger, Stewart (2000). "Emulation as a Digital Preservation Strategy". D-Lib Magazine 6 (10).
3. Lorie, Raymond A. (2001). "Long Term Preservation of Digital Information". Proceedings of the 1st ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL '01), 346-352.
4. Ross, S. and Gow, A. (1999). Digital archaeology? Rescuing Neglected or Damaged Data Resources. Bristol & London: British Library and Joint Information Systems Committee. ISBN 1900508516.
5. Rothenberg, Jeff (1998). Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation. Washington, DC, USA: Council on Library and Information Resources. ISBN 1-887334-63-7.
6. Rothenberg, Jeff (1999). "Ensuring the Longevity of Digital Information". Expanded version of Ensuring the Longevity of Digital Documents.