Proceedings of the International Multiconference on
Computer Science and Information Technology, pp. 317 – 324

ISSN 1896-7094

Long-Term Preservation of Digital Assets – Some Specific Aspects

Maria Stybliński

Faculty of Informatics and Materials Sciences, Institut of Informatics, University of Silesia,
ul. Będzińska 39, 41-200 Sosnowiec, Poland;
mstyblin@us.edu.pl

Abstract. Increasing amounts cultural, natural and scientific heritage is being created or represented in digital form. The preservation and reuse of these digital assets forms both the cornerstone of future civilization growth and development, and the foundation for the future of memory.

These digital materials represent our heritage and are its future intellectual capital. The fast pace of change in the technological landscape makes ensuring long-term access to this material a challenge. The paper mentions about some specific aspects of preservation of especially assets in digital ages.

Keywords: digital information, assets, preservation of information, technological obsolescence, preservation strategy, migration, emulation, formats, standards, metadata.

Motto: *"As we move into the electronic era of digital objects it is important to know that there are new barbarians at the gate and that we are moving into an era where much of what we know today, much of what is coded and written electronically, will be lost forever. We are, to my mind, living in the midst of digital Dark Ages" (Terry Kuny, 1998)* [1].

Introduction

The growth of digital world of information technology and communication took more then forty years, but the digital world is still relatively young and immature in relation to the larger information universe, parts of which have been under development for centuries.

Therefore nowadays information is being generated in greater quantities and at higher frequency than at any time in human history. Such information is penetrating and transforming nearly every aspect of our culture and life. Increasing amounts cultural, natural and scientific heritage is being created or represented in digital form.

Digital information is inherently more fragile than traditional technologies such as paper or microfilm. It is more easily corrupted or altered, without recognition. Digital storage media have shorter life spans and require access technologies that are changing at an ever-increasing pace. Because of these technological advances, the time frame in which we consider archiving becomes much shorter

Cultural Transformations and Digital Dependence

• The increasing dependence upon digital information is having several dramatic effects. It is changing the way in which our culture is recorded and our culture itself is being transformed. As we think of physical

products of culture as artifacts, so we should also be thinking of digital and electronic products as digital-facts *(d-facts)* or *e-facts*. However d-facts are very fragile, because digital technology is very fragile also[13].

Our ability and commitment as a society to preserve cultural memory, and natural and science heritage are far from secure. As a mean of recording and providing access to our cultural memory, digital technology has numerous advantages and may help relieve the traditional conflict between preservation and access. Digital technology affords multiple, simultaneous uses from a single original in ways that are simply not possible for materials stored in any other form.

The Limits of Digital Technology and the Technological Obsolescence

Digital technology, however, poses new threats and problems as well as new opportunities. Its functionality comes with complexity. Reading and understanding information in digital form requires equipment and software, which is changing constantly and may not be available within a decade of its introduction.

The combination of these factors is both challenging and troublesome. On the one hand, there are considerable opportunities offered by digital technology to provide rapid and efficient access to information. On the other hand, there is a very real threat that the digital materials will be created in such a way that not even their short-term viability can be assured, much less the prospect that future generations will also have access to them.

The need to create and have widespread access to digital materials has raced ahead of the level of general awareness and understanding of what it takes to manage them effectively. Rapid changes in the means of recording information, in the formats for storage, and in the technologies for use threaten to render the life of information in the digital age, but numerous examples illustrate same the danger of losing valuable cultural memories that may appear in digital form.

Early attention to the difficulties in preserving digital information focused on the longevity of the physical media on which the information is stored. Even under the best storage conditions digital media can be fragile and have limited shelf life. Moreover, new devices, processes and software are replacing the products and methods used to record, store, and retrieve digital information on breathtaking cycles of 2- to 5- years. Given such rates of technological change, even the most fragile media may well outlive the continued availability of readers for those media.

Technological obsolescence represents a far greater threat to information in digital form than the inherent physical fragility of many digital media. Digital information today is produced in highly varying degrees of dependence on particular hardware and software.

Precautions which can be taken will help significantly reduced the danger of loss:

- Storing in a stable, controlled environment.
- Implementing regular refreshment cycles to copy onto newer media.
- Making preservation copies (assuming licensing/copyright permission).
- Implementing appropriate handling procedures.
- Transferring to "standard" storage media.

Information Objects in the Digital Landscape

Information objects in digital form move through life cycles. They are created, edited, described and indexed, disseminated, acquired, used, annotated, revised, re-created, modified and retained for future use or destroyed by a complex, interwoven community of creators and other owners, dissemination, value-added services, and institutional and individual users.

The digital world is still too new for us to describe fully the life cycle of the information objects that do now or will in the future reside there, but what surely unites the community of actors in their various information-based activities is their common purpose in support of the pursuit of knowledge.

Knowledge cannot advance without consistent and reliable access to information sources, past and present. It is the archival function in the system of knowledge creation and use that serves to identify and retain important sources of information and to ensure continuing access to them.

The application of computer hardware and software has also generated other new kinds of information objects, including the products of simulation, remote sensing, computer-aided design (CAD) and information systems (IS). These objects come into being and exist as creatures of the digital environment; if nurtured well, digital technologies will certainly beget still other kinds of information objects, which we can now only anticipate.

Digital technologies increasingly serve to integrate information resources. Whatever preservation method is applied, the central goal must be to preserve information integrity; that is, to define and preserve those features of an information object that distinguish it as a whole and singular work. In the digital environment, the features that determine information integrity and deserve special attention for archival purposes include the following: content, fixity, reference, provenance, and context.

Digital information objects acquire the qualities of content, fixity, reference, provenance and context—and thereby their integrity—as they move through a life cycle in a series of relationships with parties, or stakeholders, who have specific interests in their creation, management, dissemination, use or retention. The initial stakeholder in a digital information object is, the creator of its content. Following creation, a digital information object may pass through a series of gateways of increasingly public release and access.

The fundamental problem of preserving electronic documents or "digital objects" comes from the nature of the objects themselves. Ensuring ongoing access, therefore, requires currency with technology changes, and moving digital objects from obsolete to current file formats, storage media, operating systems and so on. This is hard to be done from the following reasons:

- the rapidly increasing number of digital objects and proliferation of document standards and formats;
- the increasing complexity of digital objects (incorporating text, images, audio, video in various formats) and their increasing software dependence (e.g., storage in databases);
- the lack of planning to incorporate preservation needs in systems and lack of availability of off-the-shelf products supporting preservation needs;
- the lack of consideration of long-term access requirements when creating digital produc;
- the absence of widely-accepted standards which will assure access over time;
- copyright/intellectual property rights that may interfere with the ability to preserve digital objects through systematic copying;
- unstable storage media (e.g., diskettes) whose life span is limited;
- a lack of technical expertise in collections managers and preservation experts;
- an emphasis on the creation and/or acquisition of digital material in an era of diminishing resources, rather than ongoing preservation and access to existing electronic holdings [3].

Preservation Requirements

In the digital world, to succeed in the preservation of digital objects, preservation measures must ensure that as many of these aspects as possible persist over time. In preserving a digital object, we aim to:

- *Fix the object as a discrete whole* — t he boundaries of digital objects are less clear, especially if they are compound objects created by assembling different media or by linking to resources from around a network.
- *Preserve the physical presence* — preserving the physical file does not mean that the object will remain accessible.
- *Preserve content* — refers to maintaining the ability to access the content at its lowest level, such as ASCII text, without the embellishments of font variations and layout features.
- *Preserve the presentation* — content is typically rendered in some presentation, format or layout. To retain the original look of a document, these layout specifications must also be preserved, especially when they contribute significantly to the understanding and interpretation of the content.

- *Preserve functionality* — digtal objects can contain multimedia components (i.e., text, graphics, audio, and video), exist in hypertext format, contain dynamic content generated automatically from data stores, or have navigation functions, such as toolbars, keyword search, or interactive tables of contents. Special efforts must be made to preserve the functionality.
- *Preserve authenticity* — activities to guard authenticity include securing digital objects against unauthorized changes and monitoring digital objects through multiple "copying" cycles to ensure that each copy is an acceptable rendition of the original.
- *Locate and refer to the digital object over time* — objects can be readily altered, copied or moved. An individual must be able to match a citation to a digital object, and to distinguish it from other versions or editions.
- *Preserve provenance*—provenance is an archival concept that asserts the origin and chain of custody of an object and contributes to defining it as a whole. Imprint statements and bookplates. Establishing an object's origin and history help confirm that the work is authentic and its content is intact.
- *Preserve context*—digital objects are partly defined by their hardware and software dependencies, their mode of distribution and linkages to other digital objects. Preserving digital objects may mean weaning them from some technical dependencies, changing the mode of distribution, and deactivating links. In these circumstances, preserving context is a particular challenge.

Digital Preservation.

The term "digital preservation" refers to both preservation of materials that are created originally in digital form and never exist in print or analog form (also called "born-digital" and "electronic records") and the use of imaging and recording technologies to create digital surrogates of analog materials for access and preservation purposes [10]. This means taking steps to ensure the longevity of electronic documents. It applies to documents that are either "born digital" and stored on-line (or on CD-ROM, diskettes, DVD, or other physical carriers) or to the products of analog-to-digital conversion, if long-term access is intended [3].

Digital archiving or digital preservation - terms which tend to be used synonymously, refer to the long-term storage, preservation and access to digital information. Digital preservation can be described as using digital technology to preserve the information content. The products are the digital surrogates for the original, as preserving digital objects born digital [16].

Preservation must allow future users to retrieve, access, decipher, view, interpret, understand, appreciate, and experience informational entities (for ex.: documents, maps, data, records, etc.) in meaningful and valid ways [15]. It is required specifically exclude the potential use of digital technology to preserve the original artifacts through digitization.

Digital materials, regardless of whether they are created, initially in digital form or converted to digital form, are threatened by technology obsolescence and physical deterioration. Digital preservation refers to the series of managed activities necessary to ensure continued access to digital materials as long as necessary. It is shared to:

- *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 in 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

Digital Preservation Strategies

The following approaches to digital preservation have been developed:

- Preserve the original software (and possible hardware) that was used to create and access the information. This is known as the technology preservation strategy. It also involves preserving both the

original operating system and hardware on which to run it.

- Program future powerful computer systems to emulate older, obsolete computer platforms and operating systems as required. This is the technology emulation strategy.
- Ensure that the digital information is re-encoded in new formats before the old format becomes obsolete. This is the digital information migration strategy."

Digital preservation strategies are:

I. *Technology preservation* a means of overcoming technological obsolescence by retaining the hardware and software used to access the digital resource. It should be noted that the current definition of this strategy involves individual institutions needing to maintain both hardware and software for all materials they create and/or acquire. If these recommendations were implemented, this variation on the technology preservation strategy could become a much more feasible proposition and provide valuable support for genuinely long-term emulation or migration strategies.

II. *Migration of Digital Information/Assets* refreshing digital information by copying it from medium to another medium and the possibility of maintaining a complex set of emulators describe two distinct points on a continuum of approaches to preserving digital information. Migration is the periodic transferring of digital materials from one media or format, or from one hardware/software configuration to another, because of obsolescence, 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. In migration method can threaten the integrity of the data – over time a greater chance of data becoming lost or corrupt in the process. Data migration includes:

- Medium refreshing (e. g. rewrite a CD);
- Medium conversion (diskette to CD);
- Format conversion (ASCII to .pdf format);
- Version upgrade (Office97 to Office2000),
- Migration of technical environment (W98 to NT);

There are the variety of migration strategies for transferring digital information from obsolete systems to current hardware and software systems, so that the information remains accessible and usable. Nonetheless, there are a number of ways to increase the chances of using migration successfully as a preservation strategy. These include:

- capturing the context by documenting hardware and software required to access digital objects (and ensuring the technology is subsequently available);
- monitoring backward compatibility of software to determine when to migrate;
- hoosing a small number of standard formats and converting non-standard formats to standard ones to minimize the amount of conversion required

There are exist differing degrees of migration, ranging from relatively straightforward conversion to a major paradigm shift. Obviously the latter category will be most relevant to the disadvantages outlined below.

III. *Change Media*—transfer digital materials from less stable to more stable media. Copying from one medium to another has the distinct advantage of being universally available and easy to implement. It is a cost-effective strategy for preserving digital information, yet the simplicity and universality of copying as a migration strategy may come at the expense of great losses in the form or structure of digital information

IV. *Change Format*—the migration strategy for digital archives with large, complex, and diverse collections of digital materials is to migrate digital objects from the great multiplicity of formats used to create digital materials to a smaller, more manageable number of standard formats that can still encode the complexity of structure and form of the original.

V. *Emulation*—refers to the process of mimicking, in software, a piece of hardware or software so that other processes think the original equipment/function is still available in its original form. The essential idea behind emulation is to be able to access or run original data/software on a new/current platform by running software on

the new/current platform that emulates the original platform. If emulation is to be adopted, the first question is what is to be emulated?

Emulation means technological obsolescence of hardware and software by developing techniques for imitating obsolete systems on future generations of computers. Three options are:

- Emulate applications
- Emulate operating systems
- Emulate hardware platforms

The major parts of approach to emulation are:

- Developing generalize able 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;
- 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
- Developing techniques for encapsulating documents, their attendant metadata, software, and emulator specifications in ways that ensure their cohesion and prevent their corruption

Emulation refers to creating new software that mimics the operations of older hardware or software, in order to reproduce its performance. Thus, not only are physical presence and content preserved, but digital objects could display original features (e.g., layout) and functionality available with the older software. Emulation has recently attracted attention as a potential strategy to assist preservation, recognizing that some electronic material that is highly dependent on particular hardware and software will not lend itself to migration. Emulation is used to provide "backward compatibility" for video games, and to model how future systems might run. Emulators exist for some obsolete systems, however, emulation for preserving digital objects over the long term has not been widely tested or priced.

VI. *Adherence to standards*—adhering to stable and widely adopted open standards when creating and archiving digital resources. These are not tied to specific hardware/software platforms and thus can defer inaccessibility of digital resource due to technological obsolescence. Can either be self-imposed by institutions creating digital resources, or imposed by agencies receiving digital resources

VII. *Backwards compatibility*—being able to retain accessibility to a digital resource following upgrade to new software and/or operating systems.

VIII. *Encapsulation*—grouping together a digital resource and whatever is necessary to maintain access to it. This can include metadata, software viewers, and discrete files forming the digital resource**.** This technique aims to overcome the problems of the technological obsolescence of file formats cause the details of how to interpret the digital bits in the object can be part of the encapsulated information [10].

IX. *Converting to stable analogue format*—converting certain valuable digital resources to a stable analogue medium such as permanent paper or preservation microfilm or, more recently, nickel disk readable by electron microscope. This cannot be recommended as more than a pragmatic interim strategy for a small category of digital materials, pending the development of more appropriate digital preservation strategies.

X. *Digital archeology*—rescuing digital resources, which have become inaccessible as a result of technological obsolescence and/or media degradation. Not so much a strategy in itself as a substitute for one when digital materials have fallen outside a systematic preservation programs. Museums of hardware and software where ancient versions of hardware, operating systems, and applications software would be lovingly preserved so that people could read old data. Digital “archeology” would be used to recover lost data; preserving the software and the hardware that is used to access and read the digital data. This idea born some obstacles, but sounding very nostalgic.

Successful digital archiving requires three additional steps:

- *Data Refreshing*—copying data from one set or copy of the digital media to another of the same kind. It is important, but offers no answer to any problems brought about by the technological obsolescence of the chosen media or the infrastructure supporting it. It is a short-term solution.

- *Document data*—record information about the source, content and structure of the digital data such as creator, field names and definitions, standardization rules, relationship to other data, etc.
- *Back-up data*—make duplicate copies of digital data and store the copies in a secure environment. Creating and securing back-up copies, however, does not completely protect digital data from media degradation or from hardware or software obsolescence.

Metadata

Description of a digital object it is "data about data" called metadata. Metadata is an important part of any digital preservation strategy: Within a digital repository, "metadata accompanies and makes reference to each digital object and provides associated descriptive, structural, administrative, rights management, and other kinds of information." [11]. So it is information, which describes significant aspects of a resource. Metadata are required successfully to manage and preserve digital materials over time and assist in ensuring essential contextual, historical, and technical information are preserved along with the digital object.

Metadata structured data that makes the other data useful. Provides contextual information; consistent information adhering to established standards. Such descriptive data should include the contextual information crucial to the long-term management of electronic information. Metadata elements useful in preservation might include:

- identifiers;
- hardware, operating system and software required to access a document;
- physical details of tangible format publications such as CD-ROM, floppy disks;
- encoding standard and version;
- migration history and its success;
- data to assist determining authenticity;
- rights management information;
- versions and dates.

Standards

Digital archives will benefit from the widespread adoption of data and communication standards that facilitate reference to digital information objects and enable their interchange among systems.

Standards can keep digital documents readable: so long as documents conform to standards and so long as those standards remain in common use, means till they evolve or become obsolete or rely on "migration". If it happened it is required to convert (translate) documents into new forms as necessary. Ideally converting into new standard forms. *Standards* allow expanded networks, provides a common understanding and global interoperability [4].

Conclusion.

The information society is charged with the responsibility of preserving information and especially preserving assets and making it available for future and memory from many important points of application. This will need to develop a range of strategies to ensure the preservation of and access to various categories of digital objects. Custodial and non-custodial arrangements will need to be considered both from preservation and an access perspective. Let me finished with citation:

"Our digitally encoded memories are fast becoming obsolete. As we create more and more ingenious sways of encoding and storing them we tend to exacerbate the preservation problem. While we should act positively to address these difficulties we must not underestimate the lengths to which future generations will go to unravel the record we leave behind. We must also not forget how evocative a single record can be to the future.

If we would wish future generations to have a responsibly and collaboratively to ensure that we are leaving a digital record that is durable, processable, and intelligible" [14].

References:

1. *A Digital Dark Ages? Challenges in the Preservation of Electronic Information,* Workshop of the Audiovisual & Multimedia, 63rd Annual IFLA General Meeting, Copenhagen, Denmark. September, 1997. In: International Preservation News 17, May 1998.
2. Beagrie N.:*Digital Preservation: Best Practice and its Dissemination,* ARIADNE, Issue **43**, April 2005.
3. Bullock A.: *Preservation of Digital Information: Issues and Current Status,* Information Technology Services, National Library of Canada, 1999.
4. Day M.: *Metadata for digital preservation: a review of recent developments,* UKOLN, University of Bath; ECDL2001, 5th European Conference on Research and Advanced Technology for Digital Libraries, Darmstadt, Germany, 4-9 September 2001.
5. *Digital Preservation,* `http://www.digitalpreservation.gov/ndiippch.html`
6. *Digital Preservation Coalition,* `http://www.dpconline.org/graphics/orgact/storage.html`
7. Hedstrom M., Montgomery S.: *Digital Preservation Needs and Requirements,* In: RLG Member Institutions, December 1998.
8. Hodge G., Bonnie C.: *Digital Electronic Archiving, The State of The Art and The State of The Practice,* A Report : International Council for Scientific and Technical Information Policy Committee, CENDI, Oak Ridge, TN, 1999.
9. `http://www.collectionscanada.ca`
10. `http://www.nla.gov.au/padi/topics/17.html`
11. *Lynch 1999,* `http;//www.dlib.org/dlib/september99/09lynch.html`
12. *Research Challenges in Digital Archiving and Long-term Preservation,* Report on the NSF Workshop on Research Challenges in Digital Archiving: Towards a National Infrastructure for Long-Term Preservation of Digital Information, 2002. National Science Foundation Digital Government Program, Library of Congress National Digital Information Infrastructure and Preservation Program.
13. Ross S.: *Historians, Machine-Readable Information and the Past Future,* In: S. Ross and E. Higgs (eds): Electronic Information Resources and Historians: European Perspectives, St. Katharinen, 199.
14. Ross S.: *Changing Trains at Wigan: Digital Preservation and the Future of Scholarship,* National Preservation Office, 2003.
15. Rothenberg J.: *Using Emulation to Preserve Digital Documents,* NELiNet Rev: 2000.05.21.
16. Walter Henry: *Stanford University Libraries,* August 2000.