NOW YOU SEE IT, NOW YOU DON'T: MAINTAINING DIGITAL LEARNING OBJECTS FOR THE FUTURE

Professor Ross Harvey
School of Information Studies
Charles Sturt University
Locked Bag 675, Wagga Wagga, NSW 2678
AUSTRALIA
Telephone: +61 2 6933 2369
Mobile: 0411 898 292
Fax: +61 2 6933 2733
http://athene.riv.csu.edu.au/~dharvey/

Abstract

Failure to address the issues surrounding the preservation of digital information will result in the loss of digital learning objects in telelearning environments. Factors affecting the longevity of digital information need to be monitored to minimise frustrating and expensive inefficiencies that result from such loss. Strategies based on emerging international best practice (currently refreshing and transfer of data, migration of data, development of emulation software, and encapsulation) must be adopted in educational institutions to ensure that accessibility to and usability of digital learning objects is maintained.

Introduction

Plagiarism is only one of the issues which have become more rampant as teaching and learning take place increasingly in online environments. A less discussed issue is the need to maintain information in digital form for long periods of time. Borgman (2000, p.66) has noted that "digital preservation looms as one of the greatest challenges of information management technology and policy". The major issue is that it is much more difficult to keep information in digital form than it is in most other formats. While a number of solutions are being proposed (such as migration and emulation), they all have significant drawbacks. The increasing reliance on digital information will continue because of its many advantages, but this will be at a cost which has not been fully appreciated yet by most who work in the telelearning environment. This paper notes the implications, especially the long-term consequences, of failing to pay sufficient attention to the preservation of digital information, an essential component and product of telelearning environments, and assesses the best strategies for maintaining digital information in the future.

Why there is a problem

This paper is concerned with some implications – and they are significant implications – of the move from classroom teaching and print-based distance education to the increasing use of digital learning objects and the increasing reliance on digital information in learning environments. Recent interest in the use of *learning objects* – "reusable units or components in a learning environment" (Ip & Morrison 2001, p.289) – is founded upon the continuing availability of these learning objects so that they can be selected and used when required. *Digital learning objects* include "such forms as Web pages, pdf documents, database applications, animations, Java applets, PowerPoint presentations and QuickTime movies" (Oliver 2001, p.454). *Telelearning environments* (by which is meant distance education, based predominantly on the use of the World Wide Web - although print-based or other media may complement such use) rely, of course, on digital information.

Information in digital form is fragile and vulnerable. Unlike information on paper (books, manuscripts, photographic prints and so on), digital information will not last without human intervention and will not remain accessible for any period of time beyond a few months, or, at most, a short number of years. For example, while the archival attributes of CD formats are commonly thought to ensure that digital information on a CD will last for many years – decades or perhaps even as long as a century – the hardware and software required to access that information almost certainly won't be available in the future, unless we make deliberate efforts to ensure that they survive in workable order. Unlike information printed or written on paper, which has been proven to survive and remain accessible to users for centuries, even without any human intervention, digital information requires planning and deliberate actions to remain accessible. Telelearning environments are constructed from digital learning objects. These digital learning objects will not remain available to us and usable in the future unless we do the planning and implement the deliberate actions at, or very close to, the point of their creation. Active intervention is required; the "benign neglect" which has in the past sufficed to preserve information will not do.

So, why is there a problem with the preservation of digital information? Why can't we expect to keep access to this information for as long as we need it, as we can with information printed or written on paper or some other media? These questions have been explored in detail elsewhere (for example, Gilbert 2000, Ross 2000). The main factors affecting the longevity of digital information, those which "can render resources non-interpretable", have been summarised as:

- degradation of the media
- loss of functionality of access devices
- loss of manipulation capabilities
- loss of presentation capabilities
- weak links in the documentation chain
- loss of contextual information. (Ross 2000, p.12)

And why is it necessary, or merely even useful, to pay attention to the keeping of digital information? At the larger societal level, the answers to this question lie in the effective functioning of society which, "of course, has a vital interest in preserving materials that document issues, concerns, ideas, discourse and events … The ability of a culture to survive into the future depends on the richness and acuity of its members' sense of history" (Task Force on Archiving of Digital Information 1996, p.1). More immediately, there are economic concerns; as Ross states, "the preservation and re-use of digital data and information forms both the cornerstone of future economic growth and development, and the foundation for the future of memory. We are increasingly aware of the economic value information and the variety of ways it can be repackaged, marketed and re-used" (Ross 2000, p.2). Universities, as Tony Dean points out in a paper in this issue, are increasingly influenced by e-commerce drivers, and this increases the importance of keeping digital information, for there are economic consequences if we do not.

What kind of digital information?

To keep digital learning objects for the future, we need to be able to preserve a wide range of digital information – text, data, graphics, video, sound, and also (and significantly) combinations of these. In earlier thinking about solutions to digital preservation, making paper or microfilm printouts was often suggested, but of course "digital storage is not simply an alternative means for storing print formats" (Hedstrom 1998, p.193). We deal with "a range of composite documents of varying complexity:

- Static documents composed of text, tables and images;
- Multimedia or data-rich documents such as the kind of documents that we encounter in the networked environment (on the world wide web or on www-based corporate intranets); and,
- Dynamic documents dependent upon data that might have variable instantiations and be held in databases and spreadsheets." (Ross 2000, p.9)

For a simple illustration of this point, consider the software used to create the learning material for a subject and the software and ICT requirements to study this subject. The subject is INF430 *Audiovisual Archiving*, offered by Charles Sturt University. Students receive a printed *Subject Outline* and a CD-ROM. The subject material is also available from the CSU Web site. It uses two textbooks available on the Web, and requires students to access many Web sites.

Type of information	**Software used to create or edit**	**Software required to access**
Text	Word 2000	Word (recent version)
Videoclips	Video editing software	Quicktime
Soundclips	Microsoft Sound Recorder	WAV file player or MP3

		player
Images	Picture editing software	Web browser (e.g. Explorer 5)
Web sites	HTML editor	Web browser (e.g. Explorer 5)
Web forum	Various, e.g. database software, text editors, etc.	Web browser (e.g. Explorer 5)

Even for retaining for future use, a relatively unsophisticated digital learning resource such as this Charles Sturt University subject, we need to know about and apply some complex strategies and practices. The simple solutions of the past will not suffice to preserve the complexities of digital learning objects and their combinations in this subject.

What to keep, and for how long?

Two of the changes in thinking which has allowed us to address this seemingly intractable problem are the realisations: first, that not all information needs to be retained for long periods; and, secondly, that we can categorise information in a way that allows us to decide to which categories we wish to devote our limited preserving resources. These principles have been adopted from records management and archives management practice and applied to the digital environment. It is not sufficient simply to keep the final product. Education is an incremental process, which means that our ability to learn from the historical development of digital learning objects and telelearning environments is hampered if the current versions only are stored. (An analogous example is that of institutions that collect archives, such as the National Library of Australia, which put considerable effort into acquiring and preserving the working drafts of Australian authors, as well as the final published versions.)

For digital learning objects, we may need to keep and ensure the accessibility of parts of them, or early versions of them, for:

- the short term (e.g. working notes, early drafts);
- for the medium term (e.g. to update learning resources from year to year); or for
- the longer term (e.g. to meet legal requirements about retaining material for the lifetime of a course, or to provide a historical record of an individual's or institution's output).

So we need to do some serious thinking about how long we want to keep digital learning objects. For some, it may not matter if they become unusable, by accident or by design (i.e. through not paying attention to matters affecting longevity) in a handful of months or a couple of years.

A first attempt at a typology of how long we need to retain access to digital learning objects used in telelearning environments could look like this.

Product	**Accessibility required (months/years)**
Working notes	Until draft is complete (perhaps 6-12 months)
Draft versions	Until next version is completed (perhaps 1-3 years); a selection retained permanently
Correspondence about copyright and other legal issues	As long as legally required (potentially >50 years for copyright reasons)
Final product	For the lifetime of a course (up to a decade?); a selection retained permanently
Transactions with students	12 months after subject was delivered

This typology, even though it is primitive and incomplete, illustrates that we need to devote serious effort to the question of how to preserve digital information.

Strategies to address the problem

The influential report of a Task Force on Archiving of Digital Information, *Preserving Digital Information* (Washington, D.C.; Commission on Preservation and Access and Research Libraries Group, 1996) set the scene for serious consideration of the issues of preserving digital information. It acted as a stimulus for extensive research, especially in the United States, Canada and several European countries, which have committed major funding for research to develop solutions. We are now, in 2002, starting to see some of the fruits of the commitment of resources initiated by the 1996 Task Force report. Another influential publication, Hedstrom's "Digital Preservation: A Time Bomb for Digital Libraries" (1998) noted that, although some strategies, methods, and technologies for preservation of digital information do exist, they are not yet feasible on a large scale, and they are not affordable". Our ability to create, amass, and store digital materials far exceeds our current capacity to preserve even that small amount with continuing value" (Hedstrom 1998, p.192). In particular, the limitations are:

- An inability to ensure the authenticity and integrity of a source (e.g. through metadata, formal document structures)
- An inability to handle a wide range of digital information
- Affordability
- An inability to handle large quantities – we require, but do not have, *mass* storage systems.

We have not, in 2002, addressed these limitations. We have, however, since Hedstrom noted the conditions in 1998, developed a better understanding of optimum storage conditions for digital media, and we know better what additional data (metadata) needs to be stored with the digital information in order to make it accessible in the future. We have also gone some way towards developing small-scale mass storage systems.

Current preservation practice for digital information

Ross notes that current preservation practice for digital information consists of "preservation of obsolete technologies, migration of digital records to new environments, emulation of obsolete systems (e.g. applications, software, and hardware), bundling, persistent object preservation, and binary targetable code". He also notes that none of these can be considered as "tried and tested". However, "work in developing modelling tools, models, and standards have begun to provide a more secure foundation for ensuring the longevity of digital objects" (Ross 2000, p.17).

We can categorise current preservation practice for digital information in three groups:

- Museum approaches
- Improving digital storage media
- Active methods.

Museum approaches, such as developing and maintaining museums of working computing equipment, software and documentation, are considered to be unviable because of the difficulties of maintaining old equipment and the costs required to operate such collections (CEDARS Report, 2001). The second set of practices relates to improving the capabilities of *storage media* to retain data, and improving storage and handling practices. These are being translated into sets of guidelines capable of immediate application. The National Library of Australia has produced, for example, *Practical Advice for Preserving Publications on Disk* (Woodyard 1999). Gilbert's *Digital Media Life Expectancy and Care* is only one of several publications which provide guidelines for handling to improve the lifespan of digital media (Gilbert 2000).

The third set of practices and strategies provides the solutions most likely to be successful. These *active methods* are based on the principle that digital information can be maintained and kept accessible regardless of the hardware and software platforms on which it was developed and currently resides. These are:

- refreshing – copying the data to a newer carrier of the same type
- transfer – copying the data to a more stable carrier
- migration – porting or modifying the data into a more recent or widely accepted format

- emulation – using software that can emulate or pretend to be a different software or operating system (Woodyard 1999)
- encapsulation – grouping together a digital object and anything else necessary to provide access to that object.

It is now also commonly agreed that there is a crucial need to preserve the metadata about digital objects, that is, the data which describes the content (e.g. name of creator, title, date of creation) and the means by which it was created (e.g. software used, file formats). Documentation about the software and hardware may also need to be retained.

It is not the intention here to provide a detailed summary of the current state of research into these strategies, but a brief indication, together with some references to recent publications, may be of interest. *Refreshing* and *transfer* are probably the oldest of the digital preservation practices. *Migration* – porting or modifying the data into a more recent or widely accepted format – is, similarly, a tried and tested, but not fully reliable, practice (Whatley 2000). These three practices, which together account for most of the digital preservation activities currently taking place, are expensive because of the cost of new media, of appropriate climate-controlled storage facilities for these media, and of the human resources to carry out the tasks. *Emulation* (using software that can emulate or pretend to be a different software or operating system) had its beginning in the games industry, in attempts to ensure that games which were popular and successful on one platform were also available on modern platforms. After initial scepticism about its potential, emulation is now being explored energetically as a possible strategy (Granger 2000, Holdsworth 2001). *Encapsulation* – grouping together a digital object and anything else necessary to provide access to that object – is not a new strategy, but rather the combining of other strategies (http://www.nla.gov.au/padi/topics/20.html).

Much research activity is being directed towards preservation of the Web (Smithsonian 2001). The National Library of Australia is a partner, with the U.S.-based Council of Library and Information Resources, in the *Safekeeping Project* (http://www.nla.gov.au/padi/safekeeping/safekeeping.html); and its *PANDORA Project* is an attempt to capture and preserve Web sites of Australian long-term significance. A recent summary of Australian activity in digital preservation can be found in the proceedings of a conference held in 2001 (Kerry 2001).

Conclusion

The increasingly widespread use of telelearning, reliant as it is on the use and continuing accessibility of digital information in the form of digital learning objects, offers us opportunities for innovation. This can only happen though, if we are prepared to understand the implications, especially the long-term consequences, of failing to pay sufficient attention to the preservation of digital information. As educators, we need to become better informed about

the issues that surround preserving information in digital form, of which only an overview has been presented in this paper. We must lobby for resources to pursue the research agenda which will result in reliable procedures and hardware-software applications to help us manage the long-term retention and continuing accessibility of digital learning objects. If we do not, we are in effect dooming ourselves to high levels of frustration when we cannot access digital learning objects, and to recreating digital learning objects unnecessarily.

There is now little doubt that 'tried and tested' strategies and practices for preserving digital information will be developed and will become standard practice. These are likely to be based on a holistic approach, using a combination of strategies already in common use (such as migration) and newer approaches (such as emulation and encapsulation). This will not be, however, for some years or perhaps decades. What can we do until then, to maintain for use in the future the digital learning objects required to build our telelearning endeavours? Two avenues suggest themselves. The first is to keep informed about research outcomes and bring them to the attention of systems administrators and IT personnel in our universities and TAFEs. Keeping informed is easy to do, thanks to the PADI Web site hosted and maintained by the National Library of Australia: http://www.nla.gov.au/padi/. The second is to incorporate some simple practices into developing digital learning objects material:

- create and bundle metadata
- keep documentation
- seek the most durable media at the time and use them, but keep monitoring what's best
- set up a checking mechanism so that you know when media start to fail
- set up a migration regime.

References

Borgman, Christine L. (2000) *From Gutenberg to the Global Information Infrastructure: Access to Information in the Networked World*. Cambridge, Mass: MIT Press.

Gilbert, Michael W. (2000) *Digital Media Life Expectancy and Care*. Amherst, University of Massachusetts, Office of Information Technologies. http:/www.oit.umass.edu/publications/at_oit/Archive/fall98/media.html

Granger, Stewart (2000). 'Emulation as a Digital Preservation Strategy', *D-Lib Magazine* 6(10).

Hedstrom, Margaret (1998) 'Digital Preservation: A Time Bomb for Digital Libraries', *Computers and the Humanities* 31: 189-202.

Holdsworth, David (2001) 'Emulation, Preservation, and Abstraction', *RLG DigiNews* 5(4).

Ip, Albert & Iain Morrison (2001) 'Learning Objects in Different Pedagogical Paradigms' in G. Kennedy, M. Keppell, C. McNaught & T. Petrovic (eds), *Meeting at the Crossroads: Proceedings of the 18th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education (ASCILITE).* Melbourne: University of Melbourne: 289-298.

Kerry, Angela (2001) *Digital Preservation: The Research and Development Agenda in Australia*. Melbourne: Australian Foresight Institute and Information Resources Group, Swinburne University of Technology.

Oliver, Ron (2001) 'Learning Objects: Supporting Flexible Delivery of Online Learning' in G. Kennedy, M. Keppell, C. McNaught & T. Petrovic (eds), *Meeting at the Crossroads: Proceedings of the 18th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education (ASCILITE).* Melbourne: University of Melbourne: 453-460.

Ross, Seamus (2000) *Changing Trains at Wigan: Digital Preservation and the Future of Scholarship.* London: National Preservation Office, British Library. http://www.bl.uk/services/preservation/occpaper.pdf

Smithsonian Institution Archives (2001) *Archival Preservation of Smithsonian Web Resources*. http://www.si.edu/archives/archives/dollar%20report.html

Task Force on Archiving of Digital Information (1996) *Preserving Digital Information.* Washington, D.C.; Commission on Preservation and Access and Research Libraries Group.

Whatley, Paul (2000) 'Migration: a CAMiLEON Discussion Paper' *Ariadne* 29.

Woodyard, Deborah (1999) *Practical Advice for Preserving Publications on Disk*. Canberra, National Library of Australia. http://www.nla.gov.au/nla/staffpaper/woodyard2.html