Canadian Conservation Institute

Institut canadien de conservation

Preservation of Electronic Records: New Knowledge and Decision-making

La préservation des documents électroniques : Information récente et prise de décisions

Canadian Heritage | Patrimoine canadien

Canada

Table of Contents/Table des matières

Foreword

The preservation of electronic records is one of the latest challenges facing the conservation and heritage communities. It is a complex and multifaceted task that includes content, media, hardware, and software.

The program for *Symposium 2003 - Preservation of Electronic Records: New Knowledge and Decision-making* was developed to deal systematically and logically with the various issues. The organizing partners — the Canadian Conservation Institute (CCI), Library and Archives Canada (LAC), and the Canadian Heritage Information Network (CHIN) — all have extensive experience in the field of electronic record preservation, and each brought a unique perspective on specific aspects of the topic.

But the challenge of preserving electronic records extends well beyond the traditional heritage community. It really includes all custodians of electronic information — from corporations to government agencies to individuals. To meet this need, the symposium also included a separate half-day event for the general public.

The organizing partners were delighted to welcome more than 350 delegates to the symposium; of these, 85% were from Canada, 10% from the United States, and the rest from a variety of countries including Australia, Bermuda, the Cayman Islands, Cuba, France, Italy, Malaysia, Mexico, the Netherlands, New Zealand, Taiwan, and the United Kingdom. Everyone participated actively in the discussions, and returned to their institutions with not only a better understanding of the challenges but also with viable and practical solutions that can be implemented immediately.

Charles Costain
Acting Director General
Canadian Conservation Institute (CCI)

Ian E. Wilson
National Archivist
Library and Archives Canada (LAC)

Roch Carrier
National Librarian
Library and Archives Canada (LAC)

Jean-Marc Blais
Acting Director General
Canadian Heritage Information Network (CHIN)

Avant-propos

La préservation des documents électroniques compte parmi les plus récents défis que doivent relever les collectivités de la conservation et du patrimoine. Cette tâche complexe concerne de nombreux éléments tels que le contenu, les médias, le matériel et les logiciels.

L'objectif du programme du *Symposium 2003 – La préservation des documents électroniques : Information récente et prise de décisions* est de se pencher systématiquement et logiquement sur les divers enjeux. Les organisateurs partenaires – l'Institut canadien de conservation (ICC), Bibliothèque et Archives Canada (BAC) et le Réseau canadien d'information sur le patrimoine (RCIP) – possèdent tous une vaste expérience de la préservation des documents électroniques, et chacun d'entre eux apporte un éclairage unique à des aspects précis de la question.

Mais le défi que représente la préservation des documents électroniques dépasse de loin la collectivité du patrimoine. Il inclut en fait tous les détenteurs de renseignements électroniques – des sociétés aux organismes gouvernementaux, et jusqu'aux particuliers. En vue de satisfaire ce besoin, une activité d'une demi-journée à l'intention du grand public fut présentée dans le cadre du symposium.

Les organisateurs partenaires ont été ravis d'accueillir plus de 350 participants, dont 85 % venaient du Canada, 10 % des États-Unis et 5 % de divers pays y compris l'Australie, les Bermudes, les îles Caïmans, Cuba, la France, l'Italie, la Malaisie, le Mexique, les Pays-Bas, la Nouvelle-Zélande, Taïwan et le Royaume-Uni. Tous ont pris part activement aux discussions, et tous sont retournés dans leurs établissements avec non seulement une meilleure compréhension des défis à relever, mais également avec des solutions pratiques qu'ils seront en mesure de mettre en œuvre sur-le-champ.

Charles Costain
Directeur général par intérim
Institut canadien de conservation (ICC)

Ian E. Wilson
Archiviste national du Canada
Bibliothèque et Archives Canada (BAC)

Roch Carrier
Administrateur général/ Bibliothèque nationale du Canada
Bibliothèque et Archives Canada (BAC)

Jean-Marc Blais
Directeur général par intérim
Réseau canadien d'information sur le patrimoine (RCIP)

Organizing Committee

Co-Chairs: Jane Down (CCI), Mary Murphy (LAC), Richard Green (LAC), Pat Young (CHIN)

Program: Joe Iraci (CCI) (*Chair*), Tom Strang (CCI), René Paquet (LAC), Bruce Walton (LAC), Gilles St-Laurent (LAC), Fay Hjartarson (LAC), Sheila Carey (CHIN)

Posters: Wanda McWilliams (LAC) (*Chair*), Alison Bullock (LAC), Janet Mason (CCI), Greg Hill (LAC), Bonnie Szirtes (CHIN)

Publications/Signs: Barbara Patterson (CCI) (*Chair*), Linda Leclerc (CCI), Sophie Georgiev (CCI)

Tours: Mary-Lou Simac (CCI) (*Chair*), Brenda Muir (LAC), Mike Thompson (LAC), Denis Gagnon (LAC)

Preservation Quest: Linda Street (CCI) (*Chair*), Joe Iraci (CCI), Lise Perron-Croteau (CCI), Carol MacIvor (CCI), Shanna Ramsay (CCI), Yvette Hackett (LAC), Louisa Coates (LAC), Randall Ware (LAC), Chris Dicks (LAC), Paul Lima (CHIN), Jane Down (CCI), Greg Hill (LAC)

Local Arrangements: Greg Hill (LAC) (*Chair*), Katherine Phillips (LAC), Natalie Labonté (LAC), Judith Falardeau (LAC), Colette Landry (CCI)

Trade Show/Marketing: Shanna Ramsay (CCI) (*Chair*), Chris Dicks (LAC)

Registration/Correspondence: Christine Bradley (CCI) (*Chair*)

Finance: Colette Landry (CCI) (*Chair*)

Media Relations: Carol MacIvor (CCI) and Louisa Coates (LAC) (*Co-Chairs*), Pauline Portelance (LAC), Paul Lima (CHIN), Shanna Ramsay (CCI)

Comité organisateur

Coprésidents : Jane Down (ICC), Mary Murphy (BAC), Richard Green (BAC), Pat Young (RCIP)

Programme : Joe Iraci (ICC) (*président*), Tom Strang (ICC), René Paquet (BAC), Bruce Walton (BAC), Gilles St-Laurent (BAC), Fay Hjartarson (BAC), Sheila Carey (RCIP)

Affiches : Wanda McWilliams (BAC) (*présidente*), Alison Bullock (BAC), Janet Mason (ICC), Greg Hill (BAC), Bonnie Szirtes (RCIP)

Publications/Panneaux : Barbara Patterson (ICC) (*présidente*), Linda Leclerc (ICC), Sophie Georgiev (ICC)

Visites : Mary-Lou Simac (ICC) (*présidente*), Brenda Muir (BAC), Mike Thompson (BAC), Denis Gagnon (BAC)

À la sauvegarde de votre patrimoine : Linda Street (ICC) (*présidente*), Joe Iraci (ICC), Lise Perron-Croteau (ICC), Carol MacIvor (ICC), Shanna Ramsay (ICC), Yvette Hackett (BAC), Louisa Coates (BAC), Randall Ware (BAC), Chris Dicks (BAC), Paul Lima (RCIP), Jane Down (ICC), Greg Hill (BAC)

Organisation locale : Greg Hill (BAC) (*président*), Katherine Phillips (BAC), Natalie Labonté (BAC), Judith Falardeau (BAC), Colette Landry (ICC)

Salon commercial/Promotion : Shanna Ramsay (ICC) (*présidente*), Chris Dicks (BAC)

Inscription/Correspondance : Christine Bradley (ICC) (*présidente*)

Finances : Colette Landry (ICC) (*présidente*)

Relations avec les médias : Carol MacIvor (ICC) et Louisa Coates (BAC) (*coprésidentes*), Pauline Portelance (BAC), Paul Lima (RCIP), Shanna Ramsay (ICC)

Acknowledgements

The hosting partners would like to extend their thanks to the Organizing Committee and all volunteers, speakers, poster presenters, session chairs, exhibitors, and sponsors, including the following organizations:

Audio Conservation Studio of Library and Archives Canada
Audio Visual Preservation Trust
Gatineau Preservation Centre
Canada Museum of Science and Technology
Canadian Conservation Institute
Canadian Heritage Information Network
Canadian Museum of Civilization
Department of Canadian Heritage
Library and Archives Canada
The Willow Group
Translation Bureau, Public Works and Government Services Canada
Tunstall & Tunstall Data Recovery Services

Remerciements

Les partenaires hôtes désirent remercier le Comité organisateur et tous les bénévoles, les conférenciers, les personnes qui ont présenté des affiches, les présidents de séances, les exposants et les commanditaires, y compris les organismes suivants :

Bibliothèque et Archives Canada
L'Institut canadien de conservation
Le Bureau de la traduction, Travaux publics et Services gouvernementaux Canada
Le Centre de préservation de Gatineau
Le ministère du Patrimoine canadien
Le Musée canadien des civilisations
Le Musée des sciences et de la technologie du Canada
Le Réseau canadien d'information sur le patrimoine
Le studio des enregistrements sonores de la Division de la musique (Bibliothèque et Archives Canada)
Le Trust pour la préservation de l'AV
Tunstall & Tunstall Data Recovery Services
The Willow Group

Preface

Symposium 2003 - Preservation of Electronic Records: New Knowledge and Decision-making took place in Ottawa on September 15–18, 2003, with a program that took the form of a decision tree. Following this format, the sessions could focus on areas that related to each other in a structured manner and all the key issues relating to the preservation of electronic records could be introduced in a logical sequence from the opening to closing speakers. The topics discussed included not only storage media but also the issues that must be addressed before considering how long an electronic record will last. [For the purpose of this symposium, the term "electronic record" was used to describe a large variety of records, e.g. audio and video recordings in analog or digital form and data files such as text, spreadsheets, e-mails, etc.]

The decision tree was initially formulated by the Program Committee, but refined by Tom Strang from CCI and Bruce Walton from LAC. Addressing the key decisions/choices that an institution needs to make when considering the acquisition and preservation of electronic records led to five main session categories: Appraisal of Electronic Records; Authenticity of Electronic Records; Developing a Preservation Strategy for Electronic Records; Preservation Strategies for Electronic Records; and Media Knowledge. The call for papers clearly outlined this approach, and resulted in the submission of more than 45 abstracts. Because of the decision tree program structure, the process to select the papers was very specific. The final program consisted of 29 papers from 6 countries: Canada (14); United States (8); Australia (3); United Kingdom (2); France (1); and Germany (1). In addition to the main program, 4 posters that fit into the program structure were also presented (1 from Canada, 2 from the United States, and 1 from France).[1]

Another objective of the program was to feature a wide variety of small- to medium-sized institutions that included not only archives and libraries, but also cultural institutions such as art galleries and museums that are faced with preserving electronic records. For example, art galleries often include video art in their collections, but the needs of these electronic records are quite different than the needs

1. One of the posters that was presented at the symposium was the decision tree that appears on p. 2 of the "Introduction" of this book of postprints. Hence, only three abstracts are included in the "Posters" section.

Préface

Symposium 2003 – La préservation des documents électroniques : Information récente et prise de décisions a eu lieu à Ottawa, du 15 au 18 septembre 2003, avec un programme qui a pris la forme d'un arbre de décision. Selon ce format, les séances pouvaient porter sur des sujets liés entre eux de façon structurée et tous les points importants concernant la préservation des documents électroniques pouvaient être présentés logiquement, du premier au dernier conférencier. Les sujets discutés incluaient non seulement les supports de préservation, mais aussi les questions à traiter avant de se demander quelle est la durée de vie d'un document électronique. [*Nota* : Pour les fins de ce symposium, le terme « document électronique » a été utilisé pour désigner un large éventail de documents, comme les enregistrements audio et vidéo sous forme analogique ou numérique et les fichiers de données tels que les textes, les tableurs, les courriels, etc.]

L'arbre de décision a d'abord été formulé par le comité du programme et raffiné ensuite par Tom Strang, de l'ICC, et Bruce Walton, de BAC. Les choix les plus importants que doit faire une institution et les décisions clés qu'elle doit prendre au moment de penser à l'acquisition et à la préservation des documents électroniques ont mené à l'établissement de cinq catégories principales de séances : Évaluation des documents électroniques; Authenticité des documents électroniques; Élaboration d'une stratégie de préservation des documents électroniques et Connaissance des supports. Cette approche a été soulignée dans la demande de communications, et plus de 45 résumés ont été soumis. Grâce à la structure du programme en arbre de décision, le processus de sélection a été très précis. Le programme final a consisté en 29 communications représentant six pays : le Canada (14); les États-Unis (8); l'Australie (3); le Royaume-Uni (2); la France (1) et l'Allemagne (1). En plus du programme principal, quatre affiches qui cadraient avec la structure du programme furent également présentées (une du Canada, deux des États-Unis et une de la France).[1]

Un autre objectif du programme consistait à présenter une grande variété de petites à moyennes institutions possédant non seulement des archives et des

1. Une affiche présentée au symposium portait sur l'arbre de décision qui apparaît à la page 2 de l'introduction du présent ouvrage. Par conséquent, seules trois affiches font partie de la section sur les affiches.

of the electronic records typically found in archives and libraries. Thus, the speakers list included several individuals from the museum and gallery communities.

Finally, it was important that the program include some case histories that highlighted what various institutions have actually done or are doing to preserve electronic records. While discussing preservation strategies is important, seeing strategies actually being implemented and working in the real world provides useful information and models to follow. Among the case histories presented were the preservation of audio language recordings from Aboriginal elders in the Northwest Territories in Canada, the preservation of American poet Robert Creeley's computer files, and the implementation of the Victorian Electronic Records Strategy (VERS) within the Government of the State of Victoria, Australia. The case histories were arguably the most useful aspect of a program that included something for everyone.

As a whole, the program provided delegates with a sense of the broader issues involved in collecting and preserving electronic records, as well as knowledge about the challenges that other institutions are facing and how they are dealing with them.

We are pleased to present the papers from the symposium program in this book of postprints. However, because speakers were given the opportunity to revise their papers after the symposium and the final submissions were lightly edited, the text herein may differ slightly from the original presentations.

Jane Down, CCI
Richard Green, LAC
Joe Iraci, CCI
Mary Murphy, LAC
Patricia Young, CHIN

Note: The papers in this book are published in the language in which they were presented, but all include an abstract in both English and French.

bibliothèques, mais aussi des institutions culturelles telles que des musées qui doivent conserver des documents électroniques. Les musées, par exemple, incluent souvent de l'art vidéo dans leurs collections, mais les besoins de consultation de ces documents électroniques sont bien différents de ceux qu'on retrouve habituellement dans les archives et les bibliothèques. Ainsi, la liste des conférenciers comprenait plusieurs personnes du monde muséal.

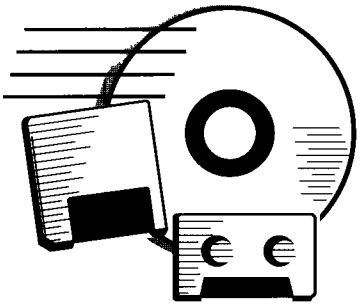
Enfin, il était important de retrouver dans le programme des histoires de cas afin d'illustrer ce que diverses institutions ont fait ou font concrètement pour préserver leurs documents électroniques. La discussion des stratégies de préservation est certes importante, mais le fait de connaître des stratégies qui sont mises en œuvre et qui fonctionnent dans une situation réelle permet d'obtenir de l'information utile et d'avoir des modèles à suivre. Parmi les cas décrits, il y a eu la préservation des enregistrements sonores des aînés autochtones des Territoires du Nord-Ouest, au Canada, la préservation des fichiers informatiques du poète américain Robert Creeley et la mise en application de la stratégie VERS pour les documents électroniques de l'État de Victoria, en Australie. On peut soutenir que les cas présentés formaient l'aspect le plus utile d'un programme répondant aux attentes d'un grand nombre de participants.

En général, les délégués sont repartis avec une vue d'ensemble des grandes questions liées à la collection et à la préservation des documents électroniques de même qu'une connaissance des problèmes auxquels se heurtent d'autres institutions ainsi que de leur manière de les résoudre.

Nous sommes fiers de vous présenter les communications du symposium dans cet ouvrage. Cependant, étant donné que les conférenciers ont eu l'occasion de modifier leurs textes après le symposium et qu'ils ont été légèrement révisés, il est possible que les textes soient quelque peu différents de ceux présentés durant le symposium.

Jane Down, ICC
Richard Green, BAC
Joe Iraci, ICC
Mary Murphy, BAC
Patricia Young, RCIP

Remarque : les communications sont publiées dans la langue utilisée lors de la présentation, mais toutes sont accompagnées d'un résumé en français et en anglais.



Authenticity and Accessibility, Objects and Technologies: Digital Preservation Basics

Bruce Walton

Library and Archives Canada
395 Wellington Street
Ottawa ON K1A 0N3
Canada
Tel.: (613) 996-1579
E-mail: bwalton@archives.ca

Abstract

Digital preservation is one of the principal challenges facing modern preservation institutions, whether they are archives, libraries, museums, or galleries. A considerable part of the discussion on this subject has concerned the relative merits of different preservation methods, such as emulation and migration. Preservers faced with the choice of a preservation method, while correctly concerned about how best to act, may not in fact appreciate the fundamental nature of the preservation that those methods purport to achieve. This paper therefore discusses the nature of digital materials and what their preservation actually consists of (based mainly on the work of the InterPARES project), and reviews briefly the nature and characteristics of several potential methods as they relate to the twin objectives of digital preservation — accessibility and authenticity.

Introduction

Although action in a given domain is usually best preceded by an understanding of the fundamental nature of the objects and processes involved, this is not always the case. The pressing urgency for action to preserve digital materials in archives, libraries, and museums has led to considerable activity in recent years, one part of which has been a spirited debate regarding the relative merits of specific preservation methods — of which emulation and migration have probably received the most attention.

This is all to the good, but in many cases this activity has proceeded without reference to the fundamental nature of digital objects and digital preservation, presumably because those actively involved in the research and debate have an intuitive understanding of what is involved. Most preservers,[1] however, lack the theoretical and/or practical experience with the subject required for such understanding. The research undertaken in the first phase of the InterPARES[2] project, and particularly that of the project's Preservation Task Force, provides a solid conceptual basis on which preservers may base their work in digital preservation.[3]

Using the InterPARES work as a starting point, this paper considers three questions:[4]

1. What is a digital record?[5] It does not "exist" in the same way as a traditional record — it is not information inscribed on a relatively durable medium and readable without mechanical intervention, so what is it?
2. What exactly is meant by preservation of digital records? For traditional records, it is possible to come up with a basic statement about preservation. Something like "preservation consists of proper containerization, correct storage environments, proper handling when in use, plus copying or stabilization/treatment as required" would probably be accepted by most who are active in the field. But what is the analogous statement for digital records?
3. What methods are available to carry out digital preservation, and how will their use affect the authenticity and accessibility of the records being preserved?

What is a digital record?

To understand digital preservation, it is first necessary to see digital records as existing simultaneously on three linked levels — as conceptual, logical, and physical objects:[6]

- The conceptual object is the record as *human artifact*, able to convey meaning from its creator to its reader, e.g. a textual document, a photo, a map, or a sound recording. It is what we experience with our senses in the everyday world. The conceptual object is what is visible on the computer monitor as a document is created or read/heard/seen; conceptual objects are what get preserved for clients.

- The logical object associated with a given conceptual object is a *computing artifact*, and may be thought of as the conceptual object at the software level; the logical object requires some appropriate software for its use. For the computer, the logical object is what is being worked with as a document is created/edited, saved, or opened.
- The physical object is also a computing artifact, and is as close as one gets to a digital equivalent of a traditional record in storage. The physical object is what the computer is working with when a file manager such as Windows Explorer is used to move or delete a file; the operating system does not need to know whether the file in question is a textual document, a spreadsheet, a photo, etc. — it is simply dealing with a bit string, and no application software is required for the physical object to be acted upon.[7]

There is a fundamental distinction between conceptual and logical/physical objects. The conceptual object is what *humans* recognize and understand; the logical and physical objects associated with a given conceptual object are what *computers and information networks* process, store, and send and receive. In the digital realm (as in the analog), what preservers set out to preserve for long-term accessibility are authentic conceptual objects such as books and records; computers, however, deal only in logical and physical objects. For the computer, the conceptual object doesn't exist — what is seen on the monitor is simply a temporary output or product resulting from the application of some method of reproduction to logical and physical objects.[8] In other words, computers do not and cannot deal with records as records, because records are conceptual in nature, not logical or physical.

What is digital preservation?

The logical and physical object(s) associated with a given conceptual object are what the InterPARES project calls its "digital components."[9] These digital components are all that a preserver has to work with — they are all that there is to "preserve," in the traditional sense of storage and care. The physical maintenance of the digital components over time, while essential, is the lesser part of the challenge of digital preservation,[10] and by itself is not a sufficient action to enable the preserver to present the authentic conceptual object to clients who want it. In order to carry out digital preservation, more than the preserved logical and physical objects are needed — an appropriate method of reproduction is required to enable the conceptual object to be reproduced from its digital components. So the basic statement of what digital preservation consists of, analogous to the basic statement for traditional preservation suggested above, is: *digital preservation consists of maintaining digital components and maintaining the ability to reproduce the record from them.*[11] Put another way, digital preservation requires some combination of digital objects with technology that yields the authentic conceptual object being preserved. Regardless of the method employed to arrive at that objective, the same two functions are being carried out — maintaining digital objects and maintaining the ability to reproduce the authentic record from them.

This definition can be illustrated with an example with which every reader will be familiar — a document attached to an e-mail. Such attachments are opened with one or two clicks of the mouse. The very ease with which the attachment can be opened and displayed might suggest that the attachment *is* the document, but, looked at in the light of the preceding discussion of conceptual objects and their digital components, it be can seen that the attachment is actually the logical object associated with the record, and that the mouse-clicks have called into operation the reproduction method required for that record.

Authenticity and accessibility in digital preservation

The reason it is so easy to read the attachment in the preceding example is that the person using the attached document and its creator inhabit the same technological space (specifically, the user's computer has either the software used by the creator, or another software that can recognize it and reproduce the document). But a technological operation has nonetheless taken place in order to allow the user to see the conceptual object that is the document. The challenge of digital preservation is a dual one. First, how is that reproduction going to work when there are years and decades of technological change between the creation and use of the document, i.e will it remain *accessible*? And second, assuming that there is a working method of reproduction, will the reproduced document be the same in all essential respects to the documents as first consigned to the preserver's custody, i.e. will it remain *authentic*? Any method of digital preservation must keep those things that it is preserving both accessible and authentic.

Accessibility and authenticity are the necessary twin objectives of digital preservation. There is no point in having a completely authentic digital document if it can only be accessed using methods that are technically or financially beyond the means of most

preservers, just as ready accessibility is of little value if the authenticity of the thing being accessed is doubtful. Achieving both objectives therefore requires an appropriate balance between the two. As will be seen below, there are methods that, because they involve no changes to the digital components of the document, are very good at maintaining authenticity but, for the same reason, are suspect in terms of maintaining accessibility. Similarly, other methods, because they take care to periodically move the digital components out of obsolete formats into current ones, are good at maintaining accessibility but can threaten authenticity precisely because they change the digital components. So, if digital preservation is about working with digital objects and related technologies in order to produce accessibility and authenticity, how might a preserver actually do this?

What methods are available?

The graphic in Figure 1 shows five different methods that might be used in digital preservation, arranged on a spectrum according to how they address the problem of technology obsolescence.

Imagine a vertical line between "Emulate Old Technology" and "Migrate Formats." The two methods to the left of the line accept technology obsolescence and incorporate it, either wholly or in part, into the "technology" side of the preservation equation. The three methods to the right of the line are focused on avoiding technology obsolescence, or eliminating it altogether. Looking at each method, from left to right:

"Preserve Old Technology" — In this case the digital components are preserved as is, meaning that the only possible method of reproduction is therefore the combination of the application software used to create the record, plus the operating system on which that creating software was designed to run, and the hardware that could run that operating system.

"Emulate Old Technology" — This method is similar to preserving the original technology in that the digital components remain logically as created,[12] but instead of old hardware and software being preserved, they are emulated on new technology, i.e. new software is written to enable a current computer or operating system to behave like the original technology.

"Migrate Formats" — With this method, a line has been crossed in that the digital components are no longer left as is (with the focus of preservation therefore being on maintaining the ability to reproduce the record from the components in their original format); instead, they are changed as required, and the focus of preservation shifts to ensuring that those changes are consistent with ongoing accessibility and authenticity. There are several variants of this approach,[13] but the simplest is *version migration*, in which documents are regularly re-saved in the current version of the software used to create them. Migration is at the mid-point on the spectrum in that, at any given point in time, reproducing the record from its digital components is still reliant on a specific software (as with the two methods already discussed) but, over time, the technology obsolescence problem is avoided (as with the two methods discussed next).

"Standardize Formats" — In this method, the digital components are also changed by being converted to standard non-proprietary formats such as ASCII for databases, tiff for images, ArcInfo Export format for geographic information systems, or pdf or some Web markup language for textual documents.[14] Once in these formats, the records remain accessible because their reproduction is no longer dependent on a specific software, i.e. because the logical object is encoded in a standard form, there are multiple softwares that can be used to reproduce the document.

"Persistent Object Preservation"[15] — This method is still being developed and has been the subject of intensive research in recent years by the United States government [the National Archives and Records Administration (NARA) and other agencies]; initial development of a working system is about to go to contract. This approach to digital preservation, which will be the basis of NARA's

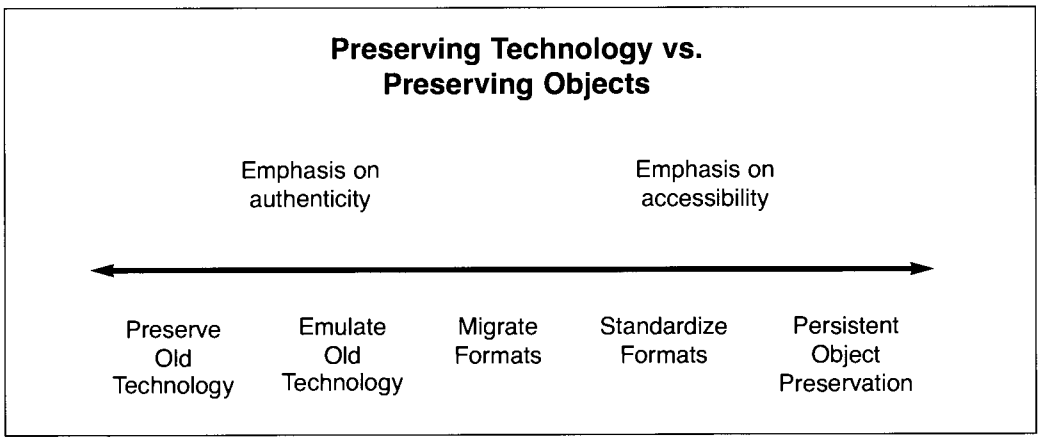


Figure 1. Methods for digital preservation.

Electronic Records Archives,[16] represents a bold attempt to meld record-keeping and archival requirements with advanced computing techniques in order to transcend technology obsolescence. Persistent object preservation is the polar opposite of preserving old technology because it focuses exclusively on the thing to be preserved rather than the technology that will be used to reproduce it. The focus is on all three levels — the conceptual, logical, and physical, the essential characteristics of which are captured in tags. In other words, all aspects of the record are abstracted, or represented by other data, to a very high degree in order to render them as independent of specific technology as possible.[17]

Assessing the options

Looking at the preceding discussion of the five different methods of digital preservation, it can be seen that the two methods that accept old technology (in whole or in part) are focused on the question "how are we going to reproduce an authentic conceptual object from these digital components which we have left in their original software?"; in other words, they are about preserving *technologies*. The other three methods are trying to find the right answer to "how are we going to keep the digital components so that we can use current technology to reproduce the record in an authentic form?"; in other words, they are about preserving *objects*. The methods on the left of the spectrum, by not changing the document, do nothing that might potentially damage authenticity; the methods on the right, by removing the document from its original technological context, avoid the problem of technology obsolescence, and are therefore more likely to preserve accessibility through time.

All these approaches have their pros and cons:

- Preserving old technology offers initial ease of application and no apparent threat to authenticity, but it is simply not practical through time. With ample resources, ingenuity, perseverance, and bit of luck, one or a only a few original technologies might be kept in working order for a decade or two; absent those success factors, and/or with multiple technologies to preserve (which will be the reality in large institutions and, over time, in small ones too), this method is guaranteed to fail — the preserver will be left with unusable digital components. Even if this method could be made to work, it would still not be advisable because records preserved with old technology would require clients to use old technology with which they would be unfamiliar. Nor would it be suited for access over the Internet, where the preserver supplies the client with the digital components and any information necessary to reproduce the document, but the client supplies the technology.
- Emulating old technology, unlike the "preserve" method discussed above, avoids the trap of attempting the impossible task of keeping a museum of functioning technology into the indefinite future; in this sense it is a feasible method. Emulators are sophisticated programs, however, and their development and maintenance is therefore labour- and cost-intensive. This in itself probably means that this method is not suited to most preserving institutions. And even if that were not the case, this method (like the previous one) still requires users to work with old technology and is not suited for access over the Internet.
- Migrating formats overcomes the problem of technology obsolescence by regularly updating the digital components so that they remain accessible with current technologies. This method also has the advantage of being something that most institutions can do now without undue expense. Given the rate of change in the software industry, however, and the fact the backwards compatibility between different versions of the same product is not a high priority for vendors, migration would require ongoing effort to stay on the "software treadmill," which means considerable effort and expense over time. Moreover, the cumulative effect of multiple migrations over time will be to introduce changes to the digital components that could damage the authenticity of the document as reproduced. Finally, quite apart from these considerations, migration's dependence on proprietary softwares puts the entire preservation enterprise at risk — if the product line being used to store the digital components disappears from the marketplace, the preservation effort comes to a dead end.
- Standardizing formats also overcomes technology obsolescence, in this case by converting digital components to widely supported standard formats. It is feasible now (e.g. it is the approach used for archival records at Library and Archives Canada), and allows clients to access records using current technologies — which allows for access over the Internet. On the other hand, conversion to a standard format may be inadequate for some authenticity requirements (e.g. formatted text preserved as an ASCII .txt file would be in a standard form, and therefore accessible, but it would lose many special formatting features of the original). Finally, because standards themselves are subject to change, over time some form of migration — with all that entails — would be required.

- Persistent object preservation, although it has not yet been deployed in an operational system, promises to overcome technological obsolescence and maintain authenticity. One factor to consider is that it has been developed for initial use in an environment (NARA) that must deal with vast collections of records and is sized accordingly in terms of the computing technologies used. Feasibility of use in smaller institutions will depend on this approach being scalable to and affordable in those settings.

As yet there is no consensus in the digital preservation community as to the best method of digital preservation. However, it is worth noting that the four major national archives that have existing digital preservation programs [the National Archives of Canada (NA) and the National Archives and Records Administration (NARA)] or are beginning one [the United Kingdom National Archive (formerly the Public Document Office) (PDO) and the National Archives of Australia (NAA)], all have chosen to use a strategy that emphasizes preserving objects over preserving technologies [conversion to standard format (NA, NARA, NAA) or format migration (PDO)], and that the major research initiative underway among this group (NARA's Electronic Records Archives) also focuses on objects rather than technologies.

Conclusion

This paper has provided a basic overview of the nature of digital records and what is involved in their preservation. It is *far* from being either comprehensive or expert in its coverage of the subject, but may be useful for those faced with digital preservation responsibilities who understand that digital preservation is not the same as traditional preservation but may not be sure how to begin. A little bit of knowledge is a dangerous thing, however, and knowledge uninformed by experience is of less use than knowledge that has been validated and enriched by some real-world practice. Preservers with digital responsibilities should consider the specifics of their situation,[18] identify an appropriate preservation strategy (i.e. select a method), and begin an active digital preservation program, if only on a test basis. In the field of digital preservation (assuming that a copy of the digital components has been set aside just in case), doing something is definitely better than doing nothing.

Endnotes

1. "Preserver" is the term used in the InterPARES research to denote the legal or actual person with responsibility for the maintenance of authenticity and accessibility of digital materials.

2. The InterPARES (International Research on Permanent Authentic Records in Electronic Systems) project was a collaborative, multidisciplinary, and international research endeavour that involved 60 researchers from 11 countries. Its aim was to develop the theoretical and methodological knowledge necessary for the long-term preservation of the authenticity of electronic records.

3. This paper is not formally a report on the findings of the InterPARES project, but its content is very much informed by the results of that project and by the work of Ken Thibodeau (of the United States National Archives and Records Administration) on digital preservation.

4. The physical side of preservation (i.e. using appropriate storage media, storing those media in the correct environment, and periodically copying records forward onto current physical formats) is not considered here. Although physical preservation is a critically important part of digital preservation, it is far more straightforward.

5. Although this paper talks principally of "documents" and "records," the information is generally applicable to electronic publications and other documents in electronic form.

6. See Ken Thibodeau's "Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years," which can be found on the Internet at www.clir.org/pubs/abstract/pub107abst.html.

7. The exception to this (in the Microsoft Windows environment) occurs when you try to delete a .exe file — in which case the file manager will ask you if you are sure you want to do this because it could make a program impossible to use.

8. "Method of reproduction" here simply means "software."

9. The full report of the InterPARES Project can be found on the Internet at www.interpares.org/book/index.htm.

10. The preservation of complete bit strings on readable media must be accompanied by whatever changes to the logical format are required by the preservation method being used.

11. See www.interpares.org/book/index.htm.

12. The physical storage format on which they are recorded will of course change over time.

13. See Ken Thibodeau's "Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years," which can be found on the Internet at www.clir.org/pubs/abstract/pub107abst.html.

14. It should be noted that pdf is not, strictly speaking, a non-proprietary format. An effort is currently underway to define an archival version (pdf(A)) and to have it adopted as an ISO standard.

15. Given that the method specifically addresses the requirement to preserve structured aggregations of records and not just individual records, the term "Persistent *Archives* Preservation" would be more appropriate.

16. See www.archives.gov/electronic_records_archives/index.html.

17. This approach is explained in far more detail in Reagan Moore's "Collection-based Persistent Digital Archives, Part 1" which can be found on the Internet at www.dlib.org/dlib/march00/moore/03moore_pt1.html.

18. That is, preservers should identify the quantity and type(s) of records to be preserved, assess the state of the technology available to them, and consider any specific institutional requirements that apply, such as available budget, in-house information technology and access policies, and relevant legal frameworks.

Résumé

La conservation des données numériques est l'un des principaux problèmes auxquels doivent faire face les établissments modernes voués à la préservation, qu'il s'agisse d'archives, de bibliothèques ou de musées. Une grande part des discussions sur ce sujet ont porté sur les mérites relatifs de diverses méthodes de préservation dont l'émulation et la migration. Les spécialistes devant choisir une méthode de conservation, qui se soucient à juste titre de faire pour le mieux, pourraient ne pas apprécier la nature fondamentale du type de conservation que ces méthodes prétendent offrir. Cette communication examine donc la nature des documents numériques et en quoi consiste véritablement leur conservation (surtout à partir du travail du projet InterPARES), et étudie brièvement la nature et les caractéristiques de quelques méthodes possibles dans la mesure où elles ont un lien avec le double objectif de la conservation de documents numériques : l'accessibilité et l'authenticité.