



DIGITAL PRESERVATION

Why should institutions care about it
and what can they do today?

A practical approach and call to action

THE CHALLENGE	3
Digital Preservation: The Problem that Keeps Growing	3
The Cost of Lost Information	5
Time is <i>Not</i> on Our Side	6
THE SOLUTION	8
What We Have in Place May Not be Enough	8
Requirements of a Digital Preservation System.....	9
CALL TO ACTION.....	13
A Practical Approach to the Digital Preservation Challenge	13
APPENDIX	
From Theory to Action: The National Library of New Zealand.....	15

All of the information and material inclusive of text, images, logos, product names is either the property of, or used with permission by Ex Libris Ltd. The information may not be distributed, modified, displayed, reproduced, in whole or in part, without the prior written permission of Ex Libris Ltd.

TRADEMARKS

Ex Libris, the Ex Libris logo, and other Ex Libris products and services referenced herein are trademarks of Ex Libris, and may be registered in certain jurisdictions. All other product names, company names, marks and logos referenced may be trademarks of their respective owners.

DISCLAIMER

The information contained in this document is compiled from various sources and provided on an "AS IS" basis for general information purposes only without any representations, conditions or warranties whether express or implied, including any implied warranties of satisfactory quality, completeness, accuracy or fitness for a particular purpose.

Ex Libris, its subsidiaries and related corporations ("Ex Libris Group") disclaim any and all liability for all use of this information, including losses, damages, claims or expenses any person may incur as a result of the use of this information, even if advised of the possibility of such loss or damage.

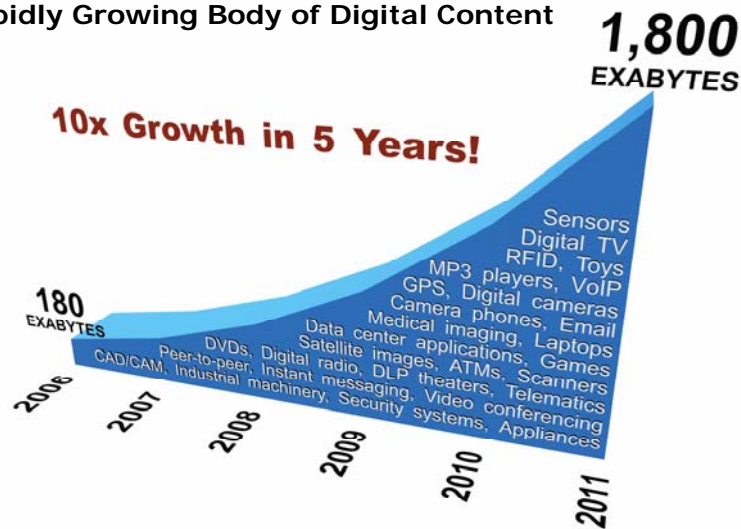
THE CHALLENGE

Digital Preservation: The Problem that Keeps Growing

“The challenges we have today cannot be solved with the same mindset with which they were created.” —Albert Einstein

The need for preserving digital assets is only a few decades old, but is one that becomes more pressing and keeps growing by the day. According to the International Data Corporation (IDC), by 2011 information that is either created, captured, or replicated in digital form will increase 10 fold over that produced in 2006 (see below). The compound annual growth rate between now and 2011 is expected to be almost 60%¹.

Memory Institutions are Facing the Challenge of Preserving a Rapidly Growing Body of Digital Content



Source: IDC White Paper, "The Diverse and Exploding Digital Universe," sponsored by EMC, March 2008.

In the last two decades, digital technology has enabled us to create, use, and be enriched by information in ways that were unthinkable a generation ago. With the technology available today, any internet user can become a global publisher in a matter of minutes by posting video to YouTube. Researchers can use the Web for posting scholarly communication or for documenting experiments and research. This

¹ IDC White Paper "The Diverse and Exploding Digital Universe" sponsored by EMC, March 2008

data may in turn be deposited into an institutional repository—searchable through Google—or published in a blog for the use of colleagues worldwide.

The very same technological advances which make obtaining and sharing information so easy also pose some modern challenges. A 4,000-year-old stone tablet displayed at the Metropolitan Museum of Art is still readable today. Without proper preservation, the digital image showing the tablet on the museum's Web site² may not be viewable in as little as ten years from now.

Possibly of even greater concern is the preservation of the metadata related to the tablets. If the records describing the tablet—which are all stored digitally today—become unreadable, this knowledge will be lost for future generations.

The problem of preserving digital material is most acute when it comes to items that have been produced in digital format that is, born digitally. Representing over 93% of the world's information,³ the vast majority of this material exists exclusively in digital format, a fact that makes the preservation of digital information critical to the perpetuation of our cultural heritage and collective memory.

The preservation of digital information is dependent on a number of components including storage media, hardware, operating system, software application, and file format. If any of these is missing or does not function properly, the digital object may become unusable:

- The media on which digital information is stored have a very limited usable lifespan.⁴
- The preservation of digital material is also dependent on the applicability of the hardware, operating system, or software application required to operate them may no longer be readily available.

2

http://www.metmuseum.org/works_of_art/collection_database/ancient_near_eastern_art/Cuneiform_tablet_and_case_with_a_record_of_court_testimony/ViewObject_enlarge.aspx?depNm=ancient_near_eastern_art&Title=Cuneiform_tablet_and_case_with_a_record_of_court_testimony&PID=0&kWd=tablet&OID=30008409&vW=0&Pg=1&St=5&StOd=1&vT=2&RID=3

3 <http://www.metaarchive.org/about.html>

4 <http://www.dpconline.org/graphics/medfor/media.html/#media1>

- If all of the other components are in place to correctly read and display digital files, these files may yet become corrupted as a result of malicious virus attacks or hardware or software failures.

As Seamus Ross eloquently noted:

Digital objects break. Digital materials occur in a rich array of types and representations. They are bound to varying degrees to the specific application packages (or hardware) that were used to create or manage them. They are prone to corruption. They are easily misidentified. They are generally poorly described or annotated; they often have insufficient metadata attached to them to avoid their gradual susceptibility to syntactical and semantic glaucoma. Where they do have sufficient ancillary data, these data are frequently time constrained.⁵

The Cost of Lost Information

Digital preservation is an expensive undertaking.⁶ A recent study by the Academy of Motion Picture Arts and Sciences provides a vivid example of the magnitude of the costs involved:

- yearly cost of saving a film master: US\$1,059
- yearly cost of saving a digital master converted from film: US\$12,514
- yearly cost of saving a digital master born digitally: US\$208,569

While the cost of preservation may be high, the cost of recreating a digital resource may be much higher. For example, a NASA study has estimated the cost of preservation for electronic engineering records at about \$5-7 per megabyte per year;

5 S. Ross (2007), Digital Preservation, Archival Science and Methodological Foundations for Digital Libraries. Keynote Address at the 11th European Conference on Digital Libraries (ECDL), Budapest (17 September 2007). ©Seamus Ross, HATII at the University of Glasgow

6 Additional information on the cost of digital preservation can be found at <http://www.dpconline.org/graphics/inststrat/costs.html>

according to the same study, the cost of recreating the information is estimated at about \$1,250 per megabyte.⁷

According to the 2005 Heritage Health Index Report published by Heritage Preservation and The Institute of Museum and Library Services, over half of the digital materials in U.S. collections—including DVDs, CDs, and MP3 files—are in probable jeopardy.⁸

Given that many organisations do not know the extent of the problem they face, it is not surprising that the loss of digital data is commonplace, and in some circumstances seems to be accepted as an inevitable hazard. Only 29% of respondents in the 2005 DPC survey could say they have not lost access to some digital information, as a result of it being impossible or too expensive to recover. Even when referring to their most important type of data, this proportion only rose to 43%.⁹

Time is *Not* on Our Side

The window of opportunity for the preservation of digital information is much shorter than what pre-digital era preservationists were accustomed to. A substantial volume of digitally born information has already been lost, and the pace of this loss increases daily. For example, as many as 44% of all the Web sites that existed in 1998 vanished just one year later.¹⁰

According to Adam Jansen, digital archivist for the State of Washington, approximately 50% of the cultural and historical information of the State has been lost due to the lack of foresight to move quickly towards digital preservation.¹¹

7 <http://ssdoo.gsfc.nasa.gov/nost/isoas/us01/minutes.html>

8 <http://www.heritagepreservation.org/HHI/HHISummary.pdf>

9 Mind the gap, Assessing digital preservation needs in the UK, report published by the Digital Preservation Coalition (www.dpconline.org).

10 <http://www.digitalpreservation.gov/importance>

11 <http://newsinfo.inquirer.net/breakingnews/infotech/view/20080127-115141/Expert-says-digital-preservation-inevitable-in-21st-century>

Similar evidence is frequently heard in different parts of the world:

In an e-mail, Svanhildur Bogadottir, city archivist of the Reykjavik Municipal Archives, Reykjavik, Iceland, writes:

At my archives, we are now receiving information in digital form almost every week. In endless sizes and formats, up to 15 to 20 years old. Most of it is impossible to read for us today and some of it has permanent value, but it is not available on paper, so it is really lost. It is a great and GRAVE problem how we can be sure to preserve our information today for the future. This is not a problem for the future but for today.¹²

Noted historian Roy Rosenzweig states:

Over centuries, a complex (and imperfect) system for preserving the past has emerged. Digitization has unsettled that system of responsibility for preservation, and an alternative system has not yet emerged. In the meantime, cultural and historical objects are being permanently lost. . . . Historians have time to think about changing their methods to meet the challenge of a cornucopia of historical sources. But they need to act more immediately on preserving the digital present or that reconsideration will be moot.¹³

12 Digital preservation: A global information management problem, Information Management Journal, Jul 2000 by Stephens, David O, www.findarticles.com/p/articles/mi_qa3937/is_200007/ai_n8917888

13 Roy Rosenzweig, Scarcity or Abundance? Preserving the Past in a Digital Era (2003), <http://www.historycooperative.org/journals/ahr/108.3/rosenzweig.html>

THE SOLUTION

What We Have in Place May Not be Enough

While most organizations have systems in place for storing and managing digital objects, these systems were not designed with preservation in mind.

Digital preservation is about **risk management**—guaranteeing future usability of and accessibility to digital content. Digital Asset Management (DAM) systems and Digital Repositories (DR) are focused on **access management**—facilitating the day-to-day use of digital content by the user community.

These distinctions in goals are manifested in the functional requirements of these two classes of systems. DAM/DR functionality focuses on discovery and access, while digital preservation functionality and workflows are centered on large-scale ingestion of material, on ongoing risk analysis, and on ensuring the integrity of stored objects.

To illustrate the differences, let us examine how the two types of systems would be used to handle similar objects, such as the digital scans of a special collection of rare books owned by a library.

To make the collection available to patrons online, the library would store the digital images in a Digital Repository system. As the emphasis in this case is on providing users with access to easily view the images, these would likely be converted to JPEG or JPEG 2000 formats.

For long-term preservation of these digital images, the library would require a preservation system. In order to maintain the provenance of the image files and track the history of any changes made over time, they would be stored in the original TIFF format in addition to JPEG. Metadata describing the content of each image as well as its technical properties would be stored as a self-contained object. In addition, the preservation system enables the library to monitor the compatibility of the files with current image presentation software versions and to support workflows for safe migration of the files to a relevant format when necessary.

It is also worth noting that although preservation focuses on risk management, it would be a mistake to equate preservation with disaster recovery and backup, as indicated in the Digital Preservation briefing paper published by the Joint Information Systems Committee (JISC):

Disaster recovery strategies and backup systems are not sufficient to ensure survival and access to authentic digital resources over time. A backup is a short-term data recovery solution following loss or corruption and is fundamentally different to an electronic preservation archive.¹⁴

Requirements of a Digital Preservation System

The Open Archival Information System (OAIS) reference model¹⁵ describes the characteristics of a digital preservation system. Written by the Consultative Committee on Space Data Systems, it is now an ISO standard.¹⁶ The model has become widely accepted among preservation bodies and experts worldwide and has been used as a guideline to evaluate current implementations of preservation and archiving initiatives.¹⁷

The OAIS model describes six high level functions that must be present in a preservation system:

1. Ingest
2. Storage
3. Data management
4. Administration
5. Preservation planning
6. Access

¹⁴ JISC. Digital Preservation: Continued Access to Authentic Digital Assets (November 2006)

¹⁵ <http://public.ccsds.org/publications/archive/650x0b1.pdf>

¹⁶ ISO 14721:2003

¹⁷ See an example at http://www.jisc.ac.uk/uploaded_documents/oaismets.pdf - Assessment of UKDA and TNA compliance with OAIS and METS Standards

The following table provides a high level overview of the requirements associated with digital preservation and the limitations of existing systems in addressing these requirements. This is not an attempt to cover all the requirements of the OAIS model but rather to highlight some of the functions that are more relevant to preservation.

Digital Preservation Challenges	Preservation System Requirements	Limitations of Digital Repositories
<i>Ingest</i>		
How do we manage the flow of incoming material (drinking from the fire hose)?	A digital preservation system should be capable of ingesting large volumes of material automatically.	Very limited: "...the handicraft approach will not scale to support the longevity of digital content in the diverse and large digital libraries..." ¹⁸
<i>Storage</i>		
How do we manage storage capacity to handle the constant inflow of large quantities of material?	Advance storage space planning is an important function that ensures adequate capacity at all times.	Very limited: No special tools for storage capacity are planned as storage requirements are more predictable.
How do we store objects to ensure the long term viability of the information?	Objects and their metadata are self-contained and comply with the Preservation Metadata: Implementation Strategies (PREMIS) data model.	Very limited: Objects and metadata are assumed to be rendered by external software applications and are not self-contained.
How do we protect stored information?	In addition to standard measures such as backup, virus protection, and redundant storage, objects are stored in a permanent repository, which is separate from the deposit and staging storage areas.	Very limited: Digital repositories do not have a separate permanent storage area. New material is entered directly into the "permanent" storage area.
<i>Data Management</i>		
How do we manage metadata associated with the objects we store?	Preservation requires the ongoing management of objects and their associated metadata.	Adequate: Digital repositories are generally required and equipped to perform similar data

18 Keynote Address at the 11th European Conference on Digital Libraries (ECDL), Budapest (17 September 2007).
©Seamus Ross, HATII at the University of Glasgow

Digital Preservation Challenges	Preservation System Requirements	Limitations of Digital Repositories
		management functions.
Administration		
How do we deploy a system that is flexible enough to handle future and yet unknown requirements?	To ensure long-term viability, the administration function must allow the embedding of additional tools to accommodate the ever-changing world of technology and formats.	Limited: Digital repositories are built with a shorter time horizon in mind and are not generally highly adaptable in nature.
Preservation Planning		
How do we ensure authenticity and integrity of the material over time?	The system must store the object's original representation as well as a detailed audit trail recording the history of its change.	Very limited: Not part of the functionality of a digital repository.
How do we know when an object is at risk of being inaccessible?	The system must support a risk analysis process with integration to technical registries such as PRONOM.	Very limited: Not part of the functionality of a digital repository
Access		
How do we manage ongoing accessibility to stored content while maintaining its compliance with ownership rights and user access privileges?	To protect stored material, a preservation system creates publishable copies of the material that can be used for accessing the stored content.	Adequate: A Digital Repository is focused on access management to stored objects.

CALL TO ACTION

A Practical Approach to the Digital Preservation Challenge

The stakes are high. Libraries, museums, and academic institutions that have long taken a role in preserving our collective knowledge and cultural heritage can no longer ignore the digital preservation challenge.

Given the high cost of preservation and the even higher potential cost of a lack thereof, a practical approach to digital preservation is required. The following steps provide a direction and recommendations for organizations that are looking to take action.

1. Recognize the need for preservation and create executive accountability for implementing the process. One of the common issues with preservation is the absence of accountability since the results of preservation (or lack thereof) can only be seen in the long run.
2. Translate awareness into action by establishing a team to lead the preservation effort.
3. Develop detailed plans to conduct an information audit, a preservation assessment, and a risk analysis in order to determine the organization's preservation needs.
4. When evaluating a solution to address your preservation needs, look for a system that supports **preservation as a process**, not just the storage of digital objects.
5. Given the extent of the investment and effort required, look at the implementation of a preservation system as a multi-stage effort.

6. The cost of developing a preservation system makes it prohibitive for any single organization to do alone. A collaborative approach is usually the only practical way to develop and maintain a preservation system in the long run.¹⁹

7. Take the first step now. The window of opportunity for preservation is quickly closing on many important digital information assets, and the magnitude of the problem is growing daily. As Roy Rosenzweig said in *Scarcity or Abundance? Preserving the Past in a Digital Era* (2003), "What is often said of military strategy seems to apply to digital preservation: **'The greatest enemy of a good plan is the dream of a perfect plan.' We have never preserved everything; we need to start preserving something.**"²⁰

19 "All organisations need to encourage an international 'market' for digital preservation tools by linking up with other projects around the world and engaging with software vendors. This would deliver economies of scale and reduce risk for individual institutions." *Mind the gap, Assessing digital preservation needs in the UK* report published by the Digital Preservation Coalition (www.dpconline.org).

20 <http://www.historycooperative.org/journals/ahr/108.3/rosenzweig.html>

APPENDIX

From Theory to Action: The National Library of New Zealand

While many organisations are still debating the preservation issue, the National Library of New Zealand has been a leader in charting a course of action toward digital preservation.

The National Library of New Zealand Act 2003 required the National Library to collect and preserve digital content in ways that ensure current and future access to New Zealand's documentary heritage and provide the legislative foundation for the library's digital preservation programme.

*"The purpose of the National Library is to enrich the cultural and economic life of New Zealand ... by, as appropriate, collecting, preserving, and protecting documents, particularly those relating to New Zealand, and making them accessible for all the people of New Zealand, in a manner consistent with their status as documentary heritage and taonga;"*²¹

Following funding approval in 2004, the library launched the National Digital Heritage Archive program (NDHA)²² to address the digital preservation challenge and to put in place the organisation, strategy, planning, and technology required to make the preservation of the digital cultural heritage of the nation of New Zealand a reality.

In addition the NDHA is closely aligned with the New Zealand government's initiative *Creating a Digital New Zealand: New Zealand's Digital Content Strategy*²³ which stresses the need to "unlock our store of valuable content by putting it in digital form so that its value can be rediscovered and renewed. As New Zealanders and end-users, we need to see ourselves on air and online, because this is the opportunity to

²¹ Section 7 of the National Library of New Zealand Act 2003

²² <http://www.natlib.govt.nz/about-us/current-initiatives/ndha>

²³ <http://www.digitalstrategy.govt.nz/Parts-of-the-Digital-Strategy/Content/New-Zealand-Digital-Content-Strategy/>

truly promote our unique heritage, cultures, and achievements and find our place in the digital world.”

The NDHA will serve as the National Library’s storehouse for digital material such as Web sites, CDs, DVDs, images, and digitised copies of print and audiovisual assets that make up its digital heritage collections. The NDHA will preserve these items in their original form and ensure that they can be viewed, listened to, and explored in the future, even if the original technology has become obsolete.²⁴

Addressing the mission critical nature of the system, the NDHA will use a standards-based, commercial digital archive and preservation system developed in partnership with Ex Libris Group™ and operated on Sun Microsystems hardware. The ensuing commercial preservation system will be replicable in other organisations wishing to preserve and provide access to information.

Major program milestones achieved to date include development of the following:

- an object management system to test elements of the digital preservation process and to provide interim storage facilities for the National Library’s digital heritage collections
- a Web Curator Tool for the harvesting of Web sites in partnership with the British Library and under the auspices of the International Internet Preservation Consortium

Recognizing the magnitude of the task and the pioneering nature of this effort, the program has incorporated a strong international collaboration component. The program serves as an international model for the implementation of digital preservation and the ongoing management of digital materials.

An international peer review group consisting of recognised thought leaders and innovators from the international library and academic communities with expertise in

²⁴ <http://www.natlib.govt.nz/catalogues/library-documents/ndha-factsheet-2007>

the areas of digital preservation and permanent access serves as an independent resource for the partnership.

The NDHA will be operational for the ingest, storage, and access of digital materials by the end of 2008. The balance of functionality, including support for digital preservation management, will be in place by the end of 2009.²⁵

"The NDHA is an important component of New Zealand's knowledge infrastructure. Not only does this project deliver the electronic legal deposit requirements under the National Library Act but it is also a strategy for the preservation of New Zealand's digital memory. To build the archive, we knew that we needed to engage both software and hardware partners who shared our vision for the importance of preserving a nation's digital memory in perpetuity."

Penny Carnaby

National Librarian and CEO of the National Library of New Zealand

²⁵ <http://www.natlib.govt.nz/catalogues/library-documents/ndha-factsheet-2007>