



Preservation by digitization

by Mike Sharun

Technology used to preserve information and our history has become outdated. Now digitization is taking centre stage as the tool of choice.

EVERY DAY, IN CANADA and throughout the world, we lose a little of our history.

Some of the losses are high profile: a devastating fire at the Anna Amalia Library in Weimar, Germany, destroyed more than 50,000 irreplaceable volumes. South Korea has conducted a highly publicized campaign to locate the missing Volume I of the two-volume *Jikji*, a collection of Buddhist teachings acknowledged as the first book ever produced using movable metal type. In the US, NASA announced it had lost track of the original magnetic-tape recordings of man's first walk on the moon – one of humanity's most spectacular achievements.

Around the world, events such as these have raised awareness and alarm and given rise to a growing movement to preserve humanity's collective memory. According to the United Nations Educational, Scientific and Cultural Organization (UNESCO), 80 percent of anthropological audiovisual documents – dialect recordings, musical ceremonies, etc. – are not properly archived. They are playable, but increasingly chemically unstable. Sooner or later, all magnetic

tapes deteriorate to an extent that they are no longer playable. And old replay machines are disappearing. So even if we have the best-preserved collections, we will not have the machinery to play them.

What can be done? Many of the world's leading libraries and museums have embarked on projects to digitize their most valuable collections. A far-ranging conversation is underway within and among the disciplines most concerned with preserving, classifying and accessing information. Professionals in library sciences, information technology and knowledge management are discussing emerging standards and best practices for archiving priceless assets, making them universally available and seamlessly integrating them into the vast sea of information that is "born digital."

Challenges of preservation

Institutions that are undertaking preservation projects tend to encounter similar challenges. They range from the profound (what to preserve and why), to the pragmatic (how many labour hours should be budgeted for preparing and scanning documents), to the technical (what the metadata industry standard will be five years from now). Based on our conversations with industry experts, curators and technologists, here are the nine most frequently encountered challenges in preserving our heritage.

Challenge 1: Setting priorities in the context of tight budgets

Digitization is expensive, particularly when unique materials are involved. Although some people might think digitization is much like photocopying, that's not the case. You need to prepare the material for digitizing. The material may be fragile and need to be conserved so you won't damage it during scanning. With the labour costs that go into organizing and assembling the materials, it is an expensive

Protecting the history of indigenous peoples

One organization that's striving to protect and share critical cultural information is the U'mista Cultural Society in remote Alert Bay, British Columbia.

The U'mista Cultural Society is collaborating with other organizations on a project that uses digital media to thoroughly document the largest and oldest collection of Kwakwaka'wakw (Kwakiutl) cultural material in Europe. <http://www.umista.ca>

The Kwakwaka'wakw First Nation is an indigenous nation residing in British Columbia on northern Vancouver Island and the mainland. The collection of cultural material was assembled by collector Johan Adrian Jacobsen around 1881 and is held by the Ethnological Museum in Berlin, Germany. <http://www.umista.ca>

"The collection contains pieces many of our people have never seen before," said Andrea Sanborn, executive director of the U'mista Cultural Centre. "By digitizing the collection and making it available to our community, we can help our people learn more about their culture. We can also share information with other researchers who are working on projects involving the Kwakwaka'wakw."

The U'mista Cultural Society's efforts were recently recognized by EMC, which gave the organization one of seven grants distributed to groups around the world under the first annual EMC Heritage Trust project. The project provides grants for protecting and improving access to information.

The EMC Heritage Trust project began accepting nominations for next year's awards on July 1, 2008. Nomination forms and complete program details are available at http://canada.emc.com/heritage_trust.

process. These costs, combined with chronic budget constraints, force institutions to make choices about what to preserve first.

Challenge 2: Preserving context

Even though the original recording of Neil Armstrong's 1969 moonwalk is gone, documents available on the Kennedy Library's website capture some of the historical, political, and scientific currents surrounding the event. For example, part of the collection includes the impetus behind the entire initiative: a one-page memo Kennedy wrote in April 1961, instructing Vice President Lyndon Johnson to evaluate the space program.



ANNE PHILLIPS

A living culture. Old and new totems in Alert Bay, BC



To effectively preserve an historic event or item, we also have to protect and store information that helps put that piece of history in its proper context. Only then can we truly get a clear picture of why a piece of history is significant. Of course adding contextual information can greatly increase the scope of a preservation project, which contributes to Challenge 1 – funding.

Challenge 3: Describing collections

One of the most intellectually challenging and time-consuming steps in digitizing large collections involves arranging and describing the materials beforehand, both to preserve their original context and to enhance their searchability. The arrangement and description of materials and the addition of metadata require significant resources.

While some institutions describe their collections to the level of individual items, it's not always practical or necessary. Many times the best approach is to mimic what a traditional archive does, creating a collection-level description and a series-level description – then adding a folder-level description that summarizes the contents of individual physical file folders.

Individual pages within the folders then are scanned and indexed to preserve that taxonomy. Optical character recognition will enable search engines to identify the contents of individual pages, eliminating the need to describe text-based items individually.

Challenge 4: Adding metadata – information about information

After teams describe and organize a collection, they need to enter descriptive

information, or metadata, for each digital item or asset. The metadata provide the “electronic hooks” that connect a keyword search to a specific item or items.

Metadata also can supply technical information that identifies what device and device settings were used to capture the digital asset, the original file type, and whether the file type was converted into other file types. This information is important in preserving the integrity of the digital assets and in reconstructing them in the future as standards and technologies evolve. It is important to include clearly written narrative descriptions in the metadata that provide the details required by scholars and researchers. Folksonomic tagging – allowing online visitors to define keywords and descriptive information about a digital asset – is one approach some museums and libraries are trying.

Challenge 5: Automating metadata capture

A new generation of intelligent indexing and active archiving tools will enable organizations to capture metadata and index documents in a more efficient, rigorous way. The challenge is that the task of extracting metadata and indexing files is still largely a manual process. To do it right with the current tools is very labour-intensive and time-consuming. But a new generation of intelligent indexing and active archiving tools will enable organizations to capture metadata and index documents in a more efficient and rigorous way.

Challenge 6: Handling documents

For newer documents, photos and recordings that are in fairly good condition, the task of preparing an item to be reproduced tends to be labour-intensive but uneventful. Staples need to be removed. Pages must be carefully unfolded. Sheets may need to be hand-fed into the scanner if using an automated feeder is too risky. Some books can only be opened to 90 degrees, after which they must be supported. Some books can't be digitized at all. Taking these necessary precautions adds to the time and expense of a project.

Challenge 7: Respecting copyrights

Current laws can be unclear on whether it is allowable to digitize rights-protected

materials for preservation purposes, even if there is no intention to publish the content. Consequently, many institutions struggle with the question of whether it is permissible to digitize and index protected materials for preservation purposes only, without making them public. Look for these laws to evolve as more institutions seek to digitize most, if not all, of their collections.

Challenge 8: Keeping up with changing technologies

As data capture has become easier, some forms of data preservation have become less reliable because of shorter media life. Paper has been around for thousands of years, but how many video formats have we seen in the last 20 years? One of the ironies of the information age is that as data capture has become easier, data preservation has become less reliable because of shorter media life.

More efficient, precise data migration technologies would provide a short-term fix to the problem. Standards bodies are working on long-term solutions. The

→ pg 20

Storage Networking Industry Association has established a 100-year archive task force to develop standards and best practices for long-term archiving.

Challenge 9: Creating a sound information infrastructure

The success of a preservation project depends in part on the underlying information infrastructure. The right infrastructure can greatly streamline the process of digitizing, indexing, and archiving collections, managing archived data, and ensuring a “future proof” solution compliant with current and emerging standards.

Digital curators: Where IT and IS meet

Meeting and overcoming these challenges to preserving our information heritage digitally will require a new generation of “digital curators,” information professionals whose role is to manage a trusted body of information in digital form for current and future use. This new class of information expert will draw insight and knowledge from the fields of information sciences (IS) and information technology (IT) and will help bridge the gap that has existed between the two.

On the IS side of the equation, digital curators will be responsible for evaluating and selecting digital collections, adding value to archived information in imaginative ways, and fostering collaboration with like-minded institutions. On the IT side, they will be equipped to manage vast repositories of archived data, make it searchable, protect it from accidental or malicious destruction, and keep it accessible over time as technology standards change.

For a relatively young country, Canada’s natural, cultural and, now, digital history is varied, rich and compelling. For the enrichment of future Canadians, I believe as a country we should make every effort possible to utilize information technology to protect our heritage. ♪

Mike Sharun is the managing director for EMC Canada, a leading developer and provider of information infrastructure technology and solutions that enable organizations of all sizes to transform the way they compete and create value from their information. For more information, visit www.emc2.ca.