Yellow Light Digitization: Proceed With Caution

Puanani Mateaki

July 31, 2007

LI827 Preservation Strategies

Randy Silverman

Digitization is our dynamic information paradigm. Techniques to maintain digital information must be identified to manage the mass of material created and discarded each second. Curtailing digital projects is not the answer, but neither is proceeding without care. Software is outdated sooner and sooner. Formats keep changing and obsolescence is a genuine concern to historians and scholars alike. We need to find a way to preserve items longer than the decade most digital information has remained in the recent past.

The Problem

Rejecting digitization is not an option. Digitization allows materials to be manipulated effortlessly. Scholars used to spend tedious hours pouring over manuscripts, documents, and books. Now conclusions are formed without taking copious notes or wading through superfluous text. Finding the significant in the mass of the inconsequential is not as difficult anymore. Material is searchable and that is colossal.

Equally significant is the access provided with digitization and the World Wide Web. Today we can digitally examine objects with a few clicks of the mouse. It is a precious gift to see seminal material personally at a museum or library but the internet is letting more enjoy that privilege.

Technology has increased global communication. It's imperative for the United States to remain an integral part of the global information infrastructure. Globalization is a direct outcome of our digital age. Preservation efforts are not perfected yet but this doesn't mean we should halt production or limit our own technological progression. Financially we can't afford to lag behind as a country.

The time to say no to digitization is past. Technology will continue to speed ahead at a dizzying rate. More material will get lost in the shuffle. We need to make a concerted

effort to find solutions to preserve digital sources. It is not a simple matter of not digitizing, material is being created digitally. The Library of Congress is appropriately taking the lead in the United States to explore digitization. Its website has a digital preservation link with articles, video spots and additional links.

Concern is expressed in the site about preserving email, presidential candidate websites, social networking sites, and other data produced in cyberspace. Audio and video digital formats are also an issue since these degrade faster than traditional audio and video media. These together help form our material culture. Material culture is the psychological influence digital and other media has on people and forms our cultural identity as a people. Material culture scholars study manufactured items because these uncover and reveal the relationships people have with objects and the meaning they assign to these objects.

Finding Solutions

I like how Hedstrom defines digital preservation in her article, "Digital Preservation: A time bomb for digital libraries." She writes it is the "planning, resource allocation, and application of preservation materials and technologies necessary to ensure that digital information of continuing value remains accessible and usable." Libraries and archives traditionally worked on preserving manuscripts, documents, and books. The media difference necessarily calls for new techniques to maintain digital information.

For a small number of materials, digital preservation actually allows access. Some items are so fragile that digital copies are the only way to allow scrutiny. Digital images can keep items protected and free from destructive handling. It's after these images are captured that the preservation of these images becomes problematic.

Like traditional media digital media needs a preservation budget. This is just one instance of when caution is needed. The temptation may be to put all money into digital preservation and ignore time-tested techniques used to preserve other media formats. Since digital preservation is relatively new, a hybrid approach is wiser. Traditional media needs to be preserved too. Taking preservation steps for all types of material stored will lessen possible loss.

Migration should be expected and planned for to avoid costly attempts to change once a format or rather the way to decipher it becomes obsolete. Digital information can be stored in software independent forms like ASCII. Standardization limits obsolescence and enhances interoperability. Globalization makes this ideal for many areas in addition to preservation.

Preservation is dynamic with digital information. Migration occurs again and again. Before each migration the organization must assess whether the cost of changing formats is warranted. Conway states a digital collection requires, "a deep and longstanding institutional commitment to preservation, the full integration of the technology." Information professionals need to be prepared to explain again and again why a document is preservation-worthy.

Leaders committed to digital preservation are essential. A digital collection needs constant care. Changing formats need to be addressed. Decisions about the same collection have to be made over and over again. Money has to be spent to migrate data properly yet still provide access.

Organizations exist specifically to address digital preservation. The Library of Congress website is a fantastic starting point for studying developments in the field and

for identifying additional resources. It's encouraging to learn steps are being taken to make digital preservation a priority. Any preservation measures are filled with uncertainty because the only way to truly learn how long something will last is to wait and see. Despite this measures are being taken based on what we know.

What we know is that standards make preservation easier. The National Information Standards Office (NISO) is establishing these to increase inter-operability throughout the globe. Standards in turn help preservation efforts. NISO provides standards free of charge and is able to do so because of the support of its organization members.

In December 2000 Congress under Public Law 106-554, gave the Library of Congress $100 million to lead the country's digital preservation effort. This created the National Digital Information Infrastructure and Preservation Program (NDIIPP). NDIIPP works with public and private organizations to identify digital information that should be preserved and test how to best accomplish this. NDIIPP's goal is to collect, archive and preserve digital content, especially items only created in digital form.

NDIIPP recognizes digital preservation demands a collaborative approach. Its website states, "No single institution can do the job of collecting, preserving and making available all the information in digital form that the students, teachers, researchers and lifelong learners have come to expect will be available at the touch of a mouse." Partners include universities, businesses and local governments both nationally and internationally and the network keeps growing.

NDIIPP identifies five specific things to meet its goal of collecting, archiving and preserving digital content. The five areas are:

1. Identify and collect at-risk born-digital content.

2. Build and support a national network of partners working together to preserve digital content.

3. Develop and use technical tools and services for preservation.

4. Encourage public policy to support digital preservation.

5. Show why digital preservation is important for everyone.

Since NDIIPP was established in 2000, it has made remarkable progress in all of these areas.

The Library of Congress is archiving thematic websites based on major events like the national elections, tragedies like 9/11, Hurricane Katrina and the Iraq war. The libraries web capture team continued and replaced the work of (MINERVA) Mapping the INternet Electronic Resources Virtual Archive team web archiving project. The library is also a member of the International Internet Preservation Consortium that collects and preserves Internet content from all over the world. The library funded two other projects to accomplish its first aim to identify and collect at-risk born-digital content. The first is Web-at-Risk from the University of California's digital library. It maintains United States political information. The other project is the ECHO DEPository project through the University of Illinois. It is creating resources to assess websites.

In 2005 the Library extended its collecting scope past thematic websites and initiated the project Selecting and Managing Content Captured from the Web (SMCCW). Twenty-five staff members helped to select additional websites for archiving. The staff was trained by the Web Capture team. It helped more Library employees understand the

importance of digital preservation even though these staff members were not part of the digital preservation team.

Teamwork is exactly what NDIIPP wanted so it actively recruited organizations. Today 67 partners collaborate to preserve relevant digital data for future generations. One such partner is Portico, an organization preserving digital journals. Portico began with an archive of JSTOR articles. Portico gets data from publishers then transforms the articles to a standard. Libraries link with Portico for the articles. It migrates material for member libraries when formats change. This cuts cost. Instead of multiple libraries each migrating their own digital journal articles, Portico does it for all participating libraries. Another partner is Satellite Communications for Learning (SCOLA). (SCOLA) preserves popular foreign news broadcasts from Al-Jezeera, Pakistan, Russia and the Phillipines.

LOCKSS is also worth mentioning. Lots of Copies Keep Stuff Safe (LOCKSS) was formed over 10 years ago. LOCKSS is an open-source system of shared e-journals. A LOCKSS box is easily created from an off-the-shelf computer dedicated exclusively to LOCKSS. The box is always, "collecting new content, auditing its content against other LOCKSS boxes and repairing any damage, [and] monitoring reader's accesses to preserved content and transparently stepping in to supply it if the publisher can't or won't supply it." LOCKSS was given additional tasks by the Library of Congress because of their innovative work. They were asked to determine how the library might use LOCKSS technology. They were also charged with creating a checklist to indicate threat models to their digital information as a form of risk management. LOCKSS demonstrates the second mission to collaborate with partners and it also fills the third to develop and use technical tools and services for preservation.

The Library of Congress encourages public policy to support digital preservation. This is the fourth task identified to be one of the NDIIP's duties. The website makes this very evident. It is extremely user-friendly and filled with bright colors, lots of links, and diverse medium forms including video clips and preservation pioneer profiles. There is actually a web page entitled "What you can do."

The last self-defined task of the NDIIPP is to show why digital preservation is important for everyone. Involving other staff in their SMCCW project that captured websites outside thematic events was a great way to do this because staff members learned first-hand just how important it is to have a digital preservation plan. It is crucial for everyone to know how fragile digital media is now. The Library of Congress' website is extremely informative. It is a fantastic way to introduce digital media to the public and it is filled with enough information to quell most questions.

Conclusion

Before an organization develops a digital collection it needs to consider the implications. It requires an on-going commitment since digital preservation is especially filled with uncertainty. Migration and obsolescence should be expected. Digital preservation problems however are being confronted. While we don't have all the solutions now, we are moving in the right direction. Several organizations are searching for ways to best preserve digital information. I am certain material will be lost as we continue to look for optimal digital preservation conditions and procedures. It has always been this way and I expect it always will be that way. Our challenge is to decrease this loss.

References

Archaeological Culture. (2007). Material Culture. Retrieved July 31, 2007 from

http://en.wikipedia.org/wiki/Archaeological_culture

Conway, P. (1996). Preservation in the digital world. Retrieved July 31, 2007 from

http://www.clir.org/pubs/reports/conway2/#gen8

Digital preservation pioneers: The LOCKSS team. Retrieved July 31, 2007 from

http://www.digitalpreservation.gov/importance/pioneers/detail_lockss.html

Importance of digital preservation. *Digital Preservation.* The Library of Congress.

Retrieved July 31, 2007 from http://www.digitalpreservation.gov/importance/

About NISO. Retrieved July 31, 2007 from http://www.niso.org/about/index.html

Hedstrom, M. A time bomb for digital libraries. (1997). *Computers and the humanities.*

Retrieved July 31, 2007 from http://www.uky.edu/~kiernan/DL/hedstrom.html

Rauber, A. and Aschenbrenner. A. Part of our culture is born digital: On efforts to

preserve it for future generations. In: TRANS. Internet-Zeitschrift für

Kulturwissenschaften. No. 10/2001. WWW:

http://www.inst.at/trans/10Nr/rauber10.htm.

What you can do. Retrieved July 31, 2007 from http://www.digitalpreservation.gov/you/