

Marissa Antosh
ILS655
Digital Library Aspect Paper
August 1, 2010

Topic 8a: Digital Libraries and Preservation

The move of libraries to the digital arena has uncovered a number of unique challenges related to preservation. Digital libraries themselves have been used to preserve materials on paper, but now they themselves need to be preserved. This is especially important because so much information is originating in digital form as opposed to moving from paper to digital. It only exists in bits and bytes.

Why should we preserve digital information?

We live in an information society where virtually anyone can access information at any time. Furthermore, as Witten, Bainbridge, and Nichols (2010) noted, digital libraries are helping rural organizations, creating a repository for scholarly research, and collecting traditions of a dwindling tribe (p. 3-5). Knowledge is no longer reserved for the few. With digital technology, it is the right of many.

Doyle, Viktor, and Paquet (2009) identify a handful of organizations that digital preservation is most important to: e-government, e-commerce, e-health, digital libraries, and digital culture heritage are at the top of their list (p. 33). However, it can be argued that digital preservation is important to more than just organizations, it is important to individuals. When is the last time you sent a letter through the Postal Service or went to get film developed? Instead, you probably send emails or take digital photos. If there are not strategies in place to preserve digital information, all of this information may be lost. This is bigger than keeping records for corporations or the government. It comes down to collective memories of people.

In addition, how will future generations find out about who came before them if digital information is lost? We no longer carve words and images into stone or make goods by hand that will last thousands of years. If there was any question about digital preservation being important, it was quelled by the announcement this past April that the social networking site Twitter donated its entire archive of tweets to the Library of Congress. Librarian of Congress James Billington said it was important to have this archive in order to record our contemporary way of life and how we use social media in discussing societal issues (Raymond & Garrett, 2010). As Ross and Hedstrom (2005) said, "...[W]e use, trust and create e-content, and expect that this content will remain accessible to allow us to validate claims, trace what we have done, or pass a record to future generations (p. 317).

What needs to be preserved?

There is more to preserving digital information than the information itself. The means for viewing the information must be preserved as well. This is a challenge because digital materials rely on software to make them viewable, and the software relies on hardware to run it. If the hardware or the software becomes obsolete, the digital data may be as well as lost (Doyle, Viktor & Paquet, 2009, p. 34).

In addition, the means for discovering digital information needs to be preserved. Doyle, Viktor, and Paquet (2009) called this "preservation metadata" and said, "...[Preservation metadata] plays a vital role in increasing the usability of a preserved digital object and its preservation environment as it provides future users with all the necessary information they need to open, render and interact with the preserved digital object, in addition to providing its preservation history" (p. 34).

How do we preserve digital information?

The goal in digital preservation is to have a preservation plan in place before the information needs to be archived. To this end, Doyle, Viktor, and Paquet (2009) listed how the information should be preserved in order to fulfill potential requirements of end users, including accessibility, executable and understandable content without errors, authenticity, accompanying metadata, and durability (p. 34-35). Thinking about what users may need in the future will help formulate a plan.

Breeding (2002) reminded that there is no permanent digital media and all media has its limits, especially in varying environmental conditions. While optical media may last longer than its magnetic counterpart, the main problem is having the equipment to view the information (p. 48). Also, although digital media makes it easier for anyone to copy and store information, it also makes it that much worse when hardware and software become obsolete because so much more may be lost.

With Doyle et. al.'s requirements of end users and the potential unreliability of magnetic and optical media, how does one go about preserving digital information? Witten, Bainbridge, and Nichols (2010) suggested five strategies: printing to paper or microfilm, preserving hardware and software in museums, duplicating content, migrating to new systems, or emulating hardware and software (p. 416).

Printing to paper or microfilm provides a hard copy of the information and may be useful if the item is printed to archival material (acid-free paper or film) and stored in a controlled environment. However, considering the sheer amount of digital information, this would be costly and time consuming. Also, how would the material be stored or cataloged?

Preserving hardware and software in a museum would definitely be a window to the past for future generations, but this method also strips away all functionality. It is interesting to see the objects, but in archives, they are not being used to access the digital information. Similarly, I could go to a museum and look at an antique sewing machine, but that does not help me sew a pair of pants.

Witten, Bainbridge, and Nichols (2010) used the slogan “lots of copies keep stuff safe” for the content duplication strategy (p. 416). Having multiple copies of information ensures that if one becomes lost or damaged, there is another to replace it. The slogan translates to the acronym LOCKSS, which is open-source software that helps libraries collect and preserve digital documents through a special Web crawler that verifies permissions and repairs damaged documents by matching it against others in the network and finding inconsistencies (p. 416). There is no indication that LOCKSS is hardware-specific, but steps should be taken to make sure that it does not fall prey to obsolescence.

Emulation is essentially keeping documents in their current digital form and emulating the obsolete software on which it was originally run. However, this requires copying the information to new media and checking for errors to make sure the copy is a replica of the original (Witten, Bainbridge & Nichols, 2010, p. 416-417). It is very important that any software used for the emulation strategy is open for others to build upon in order to run the obsolete software (Witten, Bainbridge & Nichols, 2010, p. 417).

According to Doyle, Viktor, and Paquet (2009), migration was the most widely used preservation strategy, but also the most criticized. It involves copying files into a new format after the previous format can no longer be viewed on hardware or software. In addition to being costly, time-consuming, and error-prone, it also compromises the integrity of the original digital

object because it is no longer in its original format (p. 35). This raises questions of validity—is the item still the same item once it’s been written into a new format?

Emulation seems to be the best option of these preservation strategies because the behavior of the document and its content can be preserved and run on an emulator that mimics the environment or the original (Doyle, Viktor & Paquet, 2009, p. 35).

One final method for preservation of digital information is the Open Archival Information System (OAIS), whose framework was created by the National Aeronautics and Space Administration (NASA). Breeding (2002) said it is the most definitive model for long-lasting digital archives because it has strategies for dealing with changing technology and it’s not wedded to any specific hardware or software. Furthermore, “...it provides a comprehensive and well-accepted methodology for creating digital collections in a way that ensures permanence” and it’s a good blueprint for future, permanent digital collections.

Conclusion

As keepers of information in the digital age, we must preserve digital information because it benefits society, it enables collaboration, and it ensures accountability in regulated fields such as healthcare and finance (Ross & Hedstrom, 2005, p. 318). In addition to preserving the information itself, the means to access and interact with that information needs to be retained. Hardware, software, and storage devices are all subject to obsolescence and decay, so other strategies must be implemented, like migration to new formats, emulation of old software to view old data, or OAIS. Each digital library should have a preservation plan at its outset, and what preservation method it uses will depend on factors such as cost, timeliness, and staff

needed. However, every effort should be made to preserve digital information so it is accessible and usable to end users in the future.

References

- Breeding, M. (2002). Preserving digital information. *Information Today*, 19(5), 48. Retrieved July 30, 2010 from Academic Search Premier database.
- Doyle, J., Viktor, H., & Paquet, E. (2009). Long-term digital preservation: preserving authenticity and usability of 3-D data. *International Journal on Digital Libraries*, 10(1), 33-47. Retrieved July 30, 2010 from Academic Search Premier database.
- Raymond, M., & Garrett, S. (2010). Twitter donates entire tweet archive to Library of Congress. Retrieved July 30, 2010, from <http://www.loc.gov/today/pr/2010/10-081.html>
- Ross, S., & Hedstrom, M. (2005). Preservation research and sustainable digital libraries. *International Journal on Digital Libraries*, 5(4), 317-324. Retrieved August 1, 2010, from Academic Search Premier database.
- Witten, I., Bainbridge, D., & Nichols, D. (2010). *How to build a digital library*. Boston: Morgan Kaufmann.