Digital Preservation for Digital Libraries

**Key Concepts**

With the rise of digital libraries, “digital objects have gradually emerged as the primary means in which we create, disseminate, and exchange information” (Xin, Jiang, & Min, 2010). While many objects in digital libraries are digital copies of physical items in a library’s or institution’s collection made through the process of digitization, increasingly items are born-digital. A born-digital item is an item like a photo, text file, or other materials that was created, stored, and disseminated always digitally. Items that are accessible on a digital library are preserved in their original physical copy and on servers depending on the nature of the item. Preservation of all things is impossible, but institutions seek to preserve items because it “is the only means we have for securing long-term access to our cultural heritage” (Zorich, 2007). Digital preservation is the responsibility of not just archivists. Librarians and others can take steps to make digital items available for years to come.

**Changing Preservation**

Ross Harvey’s book *Preserving Digital Materials* attempts to describe how the notion and methodology of preservation has changed over time. The practice of preservation was once about preserving individual items, but is becoming more about preserving knowledge than any one material item. The question arises as to what preservation is. It may be a set of actions, a way of seeing, a way of interpreting information, copying, or reformatting (Harvey, 2010, 14). What must be dealt with is the difference between data, digital object, and database. Data is information provided in a binary digital form, which includes digital objects and databases. A digital object can be simple, like an image file or moving image file, or complex, like a website. Databases are a way of structuring collections of data stored in a computer system. Older
conceptions of preservation included concepts like longevity, choice, quality, integrity, and accessibility. These notions remain vital to the changing, more digital, nature of modern preservation, but they have undergone developments. Harvey asserts that longevity went from a “focus on extending the life of physical media to one on ‘the life expectancy of the access system’” (2010, 15). Choice “is no longer a decision made later in the life cycle of an item, but has become, for digital materials, ‘an ongoing process intimately connected to the active use of the digital files’” (Harvey, 2010, 15). Integrity, or truthfulness, “no longer has maintaining the physical medium as its primary emphasis, but now is about developing procedures that allow us to ensure and be assured that no changes have been made” (Harvey, 2010, 15). Access to artifacts alone is not sufficient; rather Conway suggests that what is needed is access “a high quality, high value, well-protected, and fully integrated digital product” (Harvey, 2010, 15). All forms of digital preservation require alteration because technology changes often enough for digital objects to become inaccessible without changing into a format readable on modern technology.

**Standards of Digital Preservation**

The OCLC’s Digital Archive focuses mostly on born-digital content. It models itself of various standards of digital preservation. “Benchmark for Digital Reproductions of Monographs and Serials,” endorsed by the Digital Library Foundation” states that digital masters are optimally formatted and described digital objects with a focus on their quality persistence, and interoperability (Bellinger, 2002). To accomplish this, the OCLC supports a scanning environment with technology and personal of the Digital and Preservation Resources, DPR. An important feature of digital preservation is incorporation of metadata. The OCLC uses the Metadata Encoding and Transmission Standard, METS, for discovery, administrative, and
structural metadata. Discovery metadata is useful in helping patrons find materials in a digital library, administrative is about management, and structural metadata is related to understanding and use of an item. The Digital Archive complies with the OASIS reference model to ensure interoperability. “OASIS was developed for the sciences but has been embraced by the humanities” (Zorich, 2007). NISO is another important standard about metadata and creating digital libraries. There are also consumer standards for digital objects and collections like the Digital-Surrogate Seal of Approval, DSSOA. It “is a tool for asserting that a digital surrogate accurately and completely replicates the content and presentation of a static, analog original such as a book” that allows institutions to certify that its digital objects are complete and accurate reproductions of original physical items (Jacobs & Jacobs, 2013).

**Process**

Miller and Rhodes note the Library of Congress has an easy to use six step process for digital preservation (2012). The steps include identify, select, store, protect, manage, and provide. An institution must note what materials is has and wants to spend resources on because “digital preservation requires an explicit commitment of resources” (Miller & Rhodes, 2012). To identify efficiently, an institution should have an inventory of its collection in a spreadsheet or other type of document that makes it easy to make resource choices. Selecting is making the choice of what items to preserve and what standards to follow. Storing materials includes creating metadata to help preserve materials on a digital library as accessible files and on servers
as storage versions. Protecting items relates to making sure that materials are not lost and remain accessible as technology changes. The step of protecting is also important to making sure that the content can be accessed across operating systems and internet browsers. Managing requires an institution to develop a digital preservation policy. The final step of providing is about access to end-users.

**Examples from Digital Libraries**

Cornell University performed a “hybrid approach in the use of digital image technology, which coupled the creation of digital surrogates for brittle books to improve access with the production of either COM or paper replacements printed on acid free paper for preservation” in the 1990s (Kenney & Rieger, 2000). Such methods did not preserve the digital objects, but funding from the IMLS, Institute of Museum and Library Services, allowed them in the late 1990s to plan and implement a plan of long-term preservation. The image below shows the cost of managing a single digital collection, but Cornell University has over thirty-one digital collections. Cornell uses TIFF files for images and converts the scanner’s RDO files to an open format named Cornell Digital, CDL, to display the structure of image data in a non-propriety format. A campus wide plan for standardization of future digital collections efforts was underway in 2000.
The University of North Texas, UNT, realizes that “metadata is a key factor for ensuring the long-term access of digital resources” (Alemneh, Hastings, & Hartman, 2002). UNT strives to create metadata at the time of creating a digital object per best practices. To ease creation of metadata, UNT used NoteTab, which reduces the need for data enterers and editors to learn the syntax of a metadata standard. Descriptive metadata is a short-term fix, but “preservation metadata plays a significant role in facilitating preservation decisions, detects preservation threats and provides measures for minimizing risks to long-term access” (Alemneh, Hastings, & Hartman, 2002). By using diverse tools and standards UNT was able to test the capability of migrating, or transferring digital objects from generation to generation.

The most interesting project about archiving is the Gone Dark Project. It is a joint study between the Institute of Social and Cultural Anthropology and the Oxford Internet Institute. The intent is to investigate instances of Web site that are not online and have not been captured by archiving initiatives like the Internet Archive. It is important to preserve online materials because one study determined that a “historically significant social media content decays at an alarming
rate with 11 percent of timely media content lost within one year, rising to nearly 30 percent in two years” (Barone, Zeitlyn, & Mayer-Schönberger, 2015). Important websites like databases or archives can disappear from financial issues, armed conflicts, neglect, technical issues, natural disasters, and competition between big companies who buy up sites and then abandoned or absorbed. Sites like the Wayback Machine at the Internet Archives are vital in preserving glimpses of websites, but are inefficient in keeping websites alive because the site does not contact owners of at-risk websites to help them preserve a website. Sometimes individuals can become archivist and curators of dark web content. Sanjana Hattotuwa runs a personal blog in Sri Lanka that provides zip files of entire sites from NGOs and web based human rights initiatives. Without this website, many webpages in Sri Lanka would disappear without a trace.

https://sitesatrisksl.wordpress.com/
References


