Libraries, Information and Society IS502AO2
Fall 2019, Terry Weech – Assignment 4
December 13, 2019
Amelia Ishmael

Digital Initiatives and the Digital Library Movement

Digital technologies have changed the way that patrons access information. As information centers, libraries have had a tremendous role in these changes—from implementing, to participating, to being affected by them. By the end of the 1960s, computers were beginning to be use for indexing and information retrieval, yet online library catalogues were not widely available until the early 1980s. (Lesk) Since the 1990s libraries have, in a large part, been digital; by this time most documents were produced on computers, technological advances in internet indexing and searching functions made keyword searching practical and nearly instantaneous, and OCR technologies enabled the scanning and processing of legacy materials—including converting old library card catalogues to digital records, and creating retrospective digital databases of journals published before 1990. (Lesk) Soon, it seemed that more information was available online than on paper, and thus information seekers rationally start most research inquiries via the internet—prompting the troubling instinct that "If it's not online, it doesn't exist."

The contemporary availability of information resources on the internet is made possible through many initiatives—including those taken by libraries, individuals, grant agencies, and publishers. Although the somewhat sudden availability of an overwhelming quantity of information has washed most information seekers up in a perception that an information revolution has occurred and that all information is now free and accessible, this couldn't be further from the truth. The digitization of information is actually an extremely expensive process, and is riddled with legal issues involving physical and intellectual copyright, social issues that question what sort of digital initiatives receive funding priority, as well as integral issues regarding quality indexing and metadata, and preservation of digital files. This paper looks at a history of some digitization initiatives and compares them.

A survey of digital libraries

Before the internet was widely available, the potential that its unique architecture had as a digital repository was explored by two significant digitization projects included Michael S. Hart's independent Project Gutenberg (1971) and the Library of Congress's American Memory archive (1990). These were followed by two federally funded digital-library initiative programs: DLI-1 (established in 1993, sponsored by the National Science Foundation, the Defense Advanced Research Projects Agency, and NASA) and DLI-2 (established in 1998, supported by National Library of Medicine, Library of Congress, National Endowment for the Humanities, and the FBI), which had more in common with computer-technology and systems-organization initiatives than public libraries. (Mischo)

In the 1990s and early 2000s academic libraries began implementing digital initiatives that were designed to extend and strengthen their traditional services. Many of these were brought together by the consortium Digital Library Foundation in 1995. Some of the more mature examples include the digital libraries established at University of Virginia and University of Michigan. (Greenstein, 2) These digitization projects were largely funded by digitization grants supplied from businesses and philanthropic organizations (such as those from IBM, Andrew W. Mellon Foundation and Ford Foundation), acquired through competitive application processes. One of the main priorities of these projects has been to digitally reformat physical items with the goal of increasing access to, initially, primary-resource and special-collection materials that are considered high-value resources for academic researchers. Digital libraries also include born-digital resources. The content that these digital-library projects produce is largely academic, fundamentally Western in scope (Jones, 247), and traditionally limited in access to research institutions. The goal of these projects is usually to creating databases that would be accessible to paying members or subscribers, rather than universal open access.

As an alternative to these expensive "boutique" digitization projects, the non-profit digital library Open Library launched in 2006 (emerging out of the Internet Archive project) with the goal to provide wide access to public domain and out-of-print books. Another historical digitization project was the Million Book Project (aka the Universal Library, 2007-2008), led by Carnegie Mellon University, which sought to increase the credibility of valuable information resources available online through a large-scale scanning project. (Lesk)

These large-scale book digitization initiatives were quickly swept under by Google Books (2004), who—armed with incredible technical and financial resources—have attempted to expand the scope of digitization of printed materials comprehensively. To date, Google Books Search has indexed over 40 million books across the world, inhibited—only slightly—by legal restrictions. (Lee) The power of Google Books is directly connected to its searching capabilities, which use OCR technologies to offer keyword search of full texts and external links.

HathiTrust (2008, via HathiTrust Research Center) is another large-scale digital library, with collaboration between 12 libraries at the University of California and over 60 international research libraries. Its repository is based at the University of Michigan and includes a large amount of content digitized by the Google Books project and the Internet Archives initiatives. Though it started with a goal to be open, full access is limited to membership, which consist of largely other research libraries. (Piper)

The Digital Public Library of America (2010)—partnering with the Library of Congress, the Smithsonian Institution, Internet Archives, and more—has expanded beyond books to include library, archive, and museum collections. Collaborations and partnerships across institutions to broaden access to unique resources. As an aggregate library, it has combined collections from over 2000 institutions to create digital exhibitions and digital access to collections. Its scope seems to be expansive, comparable to national-cultural heritage libraries such as Europeana (2008, funded by the European Union), which indeed DPLA plans to provide interoperability with. (Piper) As participates in the linked open data movement, it takes advantage of the particular web architecture of the internet, where most cloud-based ILS library catalogues are now already residing.

Problems and barriers

The concept of a universal library, which many digital libraries purport is their goal, has many logistical barriers.

As with many technology-driven advancements, one of the largest barriers that digital libraries face is financial. Although the initial developments in digital libraries were funded federally, the real progress had been led by commercial entities such as Google—which provides free access to its services but assault users with its marketing and advertisements and its invasive data-collection practices. Yet where access is not provided, digital library paywalls can effectively

dissuade potential members and users—access is typically limited to institutional affiliation credentials, which essentially excludes the public from its services.

Another substantial hurdle that digital libraries face as they struggle to improve their services, is legal. Whereas, in many cases, first-sale doctrine rights protect independent institutions' abilities to digitize of their own collections for onsite access, many of the previously mentioned digital libraries have encountered barriers when publishing materials that they do not own intellectual copyright on the world-wide web. (Title 17, Section 109) Although, at first, many digital libraries worked cautiously inside of legal boundaries by only offering access to resources that were in the public domain, the possibilities to expand their offerings to restricted items, has proved too tempting to bypass—even if the libraries don't own intellectual (or sometimes even physical) copyrights.

Although the digital reproduction, publication, and sale of surrogates of physical items would seem, according to common sense, to be limited to the same restrictions as physical reproduction, publication, and sale, digital library initiatives have taken the conflict to court numerous times to argue for allowance. This situation is complex. As Lesk describes:

> The development of digital libraries does not align well with traditional legal views. Historically, there was a clear separation between the first-use market and the later-use market. The author or publisher was involved in the first-use market: the original purchaser of something had to buy from a legitimate copyright holder. The second-user market, whether involving a library, a used bookseller, or an individual making a gift to a friend, did not involve or further recompense the originator. Digital libraries break both ends of this. On one side, it is easy to copy digital bits, and the copies are perfect and almost instantaneous, so individuals can look for ways to avoid paying the copyright holder even when seeking a brand-new work. On the other side, the possibility of digital rights management means that the copyright holders can impede the sale of second hand books or music, as well as the lending of works by libraries.
>
> For libraries, this is a new kind of problem. Libraries converting their own catalogs didn't generally have to worry about getting anyone else's permission. Many long out of print and commercially worthless books, however, still can't be legally digitized without permission, while at the same time that numerous new best-selling works (particularly music or video) are being illegally distributed online.

Google Book Search was taken to court over infringements, when its keyword "indexing" feature made the entire copyrighted works accessible. (*The Authors Guild Inc., et al. v. Google, Inc.*)

Open Library was publicly criticized for ignoring DMCA takedown notices, and for mass violation of copyrights on an international level, regarding its service of eBook lending resources to materials that are protected by copyright and that they have no licenses for. (Meadows) Whereas digital libraries that serve for commercial profit would seem to have a more difficult time accessing exceptions reserved for Fair Use, due to its transformative use of adding ADA compliance to some resources, the reformatting is sometimes exempted by copyright.

Here is one of the places that large-scale digital initiatives, which historically relied on some perverse exception with first-sale rights, diverge from more localized digital initiatives that are conducted by the libraries themselves over their own archives and collections. Because most special-collection libraries insist on the transfer of intellectual copyrights along with the transfer of physical ownership during acquisition processes, special-collection libraries have a special advantage in digitization projects. Yet, going back to the previously mentioned issue, independent special-collection libraries usually hope to recoup the costs of building and maintaining their digital libraries through expensive subscription sales, which make them exclusive access.

Another problem that universal digital libraries face is social. Many open digital libraries claim that their initial goals were to make all of the world's human knowledge universally accessible to the public through the internet, yet "by ideologically linking LSDIs to public libraries, which remain extremely popular in concept (if not always in terms of tax support), LSDI proponents have attempted to appropriate some of that earlier initiative's legitimacy." (Jones, 253) As Jones also describes, LSDIs largely echo public libraries' founding principle of taking a paternalistic role of providing access to the public to prioritize written materials that are mostly book-based and academic texts that could elevate knowledge according to priorities set in the Western tradition of knowledge being fixed through text and, and those that seem beneficial the public good and even moralistically "good." (Jones, 253) Google Book Search, for example went first to leading academic libraries, and then to monasteries. Prioritizing printed texts over those available on the internet. The critics of GBS's global goals are numerous, including the President of the French National Library' skepticism of Google's cultural equality despite its enormous focus on English-language texts (Lewis), and Jones's skepticism regarding the printed word and Western formats of knowledge as having priority and without sensitivity to the fact that there are over various knowledge resources that are in use throughout the globe. (Jones, 254) Perhaps the real aim is to universalize the Western concept of literacy and through it to colonialize the world with

Western political and cultural sympathies? They assert the quantity and quality of information provided by their services, yet their indexing and accuracy is that of a commercial provider rather than the ethics of the library or the accuracy of the academic.

The reliance on wealthy patrons to fund and administer support digitization projects persists, and thus with it the selection of what gets digitized and who may access it and for what price. Thus the projects that receive funding are not based on the quality of the knowledge that they hold, or the public's need of the information, but rather the interests of the institutions that can fund the project or who own the resource. The public, the users of the services, has none (or extremely limited input) as users. An alternative is digital libraries and LSDIs that are for-profit, whose projects are not fueled by a presumed impetuous that the materials would be for the "overall well-being of individuals and societies," (Jones, 258) yet the selection of materials again are selective in favor of preserving an idealistic culture that us useful for those with wealth to retain.

Additional concerns

Digitization is commonly being aligned today with "digital preservation" in cases where reformatting of analogue resources is conducted as a means of reformatting as part of an attempt to preserve materials that are deteriorating, fragile, and rare. Indeed, special copyright exceptions for books and audiovisual materials exist for the creation of "access copies" and for surrogates of information resources that are either in an irretrievable state or are not otherwise available at a reasonable price on the open market. Yet one of the concerns with digital preservation is that digitization is not enough—it is not enough simply to scan all of one's collection and post it to one's website, or to purchase digital materials or subscription-access to digital materials. As Lavoie describes,

> This suggests that as more and more digital materials come under the stewardship of collecting institutions, preservation will become less like an event occurring at discrete intervals, and more like a process, proceeding relatively continuously over time. As a consequence, it will become more difficult to distinguish preservation activities from the routine, day-to-day management of digital materials.

Preservation never ends, especially when dealing with quickly obsolescent carriers such as digital files, it rather requires a continuous investment in maintenance. (Lavoie 2004)

When the digitization process is not funded by the search engines or the individual publishers, it may be provided by the libraries, archives, or museums that own the resources. These

initiatives have become expected ways for libraries to extend access and promote their collection, by creating digital collections. Additionally, rare book and special collection libraries have participated in digital initiatives to create digital surrogates for materials in their own collection that can be available through remote access. These initiatives are usually sought as part of a preservation plan—to increase access to fragile materials that are at risk of destruction if handled.

For organizations with digitization projects that are under 100,000 items, outsourcing the digitization and maintenance is most practical. Previously this role was taken on by vendors who specialized in microfilm services for libraries. Adam Matthew Digital (1990) was initially a commercial publisher that specialized in microfilms, and has carried on services into digitalization as a digital publisher that assists special-collection libraries in creating digital-subscription packages of their collections. The costs of digitalization are frequently provided by the publisher, which the publisher recoups in part by selling access privileges to subscribers.

Though the internet is an information repository, it is not a library. As Piper affirms, "We cannot expect Google, as a public corporation, to have the same values as libraries, nor should it. To assume that a corporation is acting altruistically, and without the need for profit, is foolish. That's not the kind of beast it is. Selling books online, selling advertising, developing digitizing processes—all these fall into the realm of what a corporation will do. Preservation and access may be peripheral collateral, but they are not the intended target." (Piper, 23) When browser services such as Google implement large-scale digitalization initiatives they do not fundamentally have the same responsibilities or core values as a library does—especially regarding privacy, access, and authoritative-quality assurance.

Preparing for the future

With the growing integration of digital libraries within traditional library frameworks, emerges specific concerns of digital materials—including an increasing need to enhance indexing, metadata description, copyright management, etc.

Within the first step of creating digital libraries was the development of systems and then as an extension of traditional library services into (predominantly) the academic research library. More widely then, digital libraries seek to integrate legacy bibliographic resources into digital catalogues and heighten the user's ability to access and search electronic resources—not only on-site (which is the traditional limitation) but anywhere in the world. Some of the most exciting and

ambitious digital libraries are using the unique architectural qualities of the internet to improve search results and external links. Linked open data uses authoritative standard vocabularies and URI to enhance records. As an outgrowth and expansion of traditional cataloguing duties, metadata description is an integral aspect of the digital library.

One of the ongoing investments related to digital libraries include systems improvement, user services, and technical infrastructure. In order to maintain integrity, digital information needs to be accessible and persistent. To keep up with the programs, the increasing interest in digital libraries has caused a swift rise in the number of digital-librarian positions that are added to library staffs and charged with administering projects and providing stewardship to digital collections. According to Skene's analyzation of recent digital-initiative librarian job descriptions, the top four new skills and responsibilities required from these new positions are: metadata, digital preservation, digital collections, and digitization, (Skene) These new skill-sets were anticipated in 2007 by the Association of College and Research Libraries, who say that that digital libraries were not an adjacent position to librarianship but would soon be essential as libraries recognize the necessity to be more involved in expanding their stewardship to digital collections, digitization projects, and digital storage and retrieval systems. Skene concludes, "It is evident that digital initiatives librarian positions are growing at a considerable rate." (Skene)

Bibliography

ACRL Research Planning and Review Committee (2007), Top Ten Assumptions for The Future of Academic Libraries and Librarians: A Report from the ACRL Research Committee, Retrieved from www.ala.org/acrl/issues/value/tenassumptions

Copyright Law of the United States, Title 17 (2016). Retrieved from https://www.copyright.gov/title17/

Greenstein, D. and S. E. Thorin. (2002). The Digital Library: A Biography (Washington, DC: Digital Library Federation, Council on Library and Information Resources. Retrieved from https://old.diglib.org/pubs/dlf097/dlf097.pdf

Jones, E. (2017). The Public Library Movement, the Digital Library Movement, and the Large-Scale Digitization Initiative: Assumptions, Intentions, and the Role of the Public. *Information & Culture, 52*(2), 229-263. doi:10.7560/IC52205

Jones, Ed. (2016). Linked Data for Cultural Heritage. Association of Library Collections and Technical Services.

Lavoie, B. (2004). 13 Ways of Looking at Digital Preservation. *D-Lib Magazine 10* (7/8). Retrieved from http://www.dlib.org/dlib/july04/lavoie/07lavoie.html

Lavoie, B. (2003). The Incentives to Preserve Digital Materials: Roles, Scenarios, and Economic Decision-Making. OCLC Online Computer Library Center, Inc., Office of Research. Retrieved from http://www.oclc.org/research/projects/digipres/incentives-dp.pdf

Lee, Haimin. (2019). The Keyword: 15 years of Google Books. *Google*. Retrieved from https://www.blog.google/products/search/15-years-google-books/

Lesk, M. (2012). A personal history of digital libraries. *Library Hi Tech*, *30*.4, 592-603. doi:10.1108/07378831211285077

Lewis, B. (2013). Google and the World Brain. BBC. [digital video]

Meadows, Chris. (2017) "The Internet Archive's OpenLibrary project violates copyright, the Authors Guild warns". TeleRead. Retrieved from https://teleread.org/2017/12/19/the-internet-archives-openlibrary-project-violates-copyright-the-authors-guild-warns/

Mischo, William H. (2005) Digital Libraries: Challenges and Influential Work. *D-Lib Magazine 11*(7/8). Retrieved from http://www.dlib.org/dlib/july05/mischo/07mischo.html

Piper, Paul S. (2013). The Library's Future is Digital: HathiTrust and the Digital Public Library of America. *Online Searcher 37*(2), 22-27. Retrieved from http://www.infotoday.com/OnlineSearcher/Articles/Features/HathiTrust-and-Digital-Public-Library-of-America-as-the-future-88089.shtml

Skene, E. (2018). Shooting for the moon: an analysis of digital initiatives librarian job advertisements. *Digital Library Perspectives 34* (2), 84-90. doi:10.1108/DLP-06-2017-0019

Working Group on Defining Digital Preservation. (2007). Definitions of Digital Preservation (Text). ALCTS Preservation and Reformatting Section. Retrieved from http://www.ala.org/alcts/resources/preserv/defdigpres0408