Brief Review of Books and Products
Along Came Google: A History of Library Digitization

John Adkins

BRIEF REVIEW OF BOOKS AND PRODUCTS

Along Came Google: A History of Library Digitization

By Deanna Marcum and Roger C. Schonfeld

Beginning with the founding of the Library of Alexandria under the Ptolemies of Egypt, the goal of a universal library collecting all of the world's knowledge and making it available to as many people as possible has driven numerous ambitious projects. The scope of these projects has evolved with the technology available at the time. In Egypt, the Ptolemies would have ships searched and any books or scrolls found would be copied for the library (Casson, 2001). Centuries later, the development of the printing press made it easier to make multiple, accurate copies of books and helped to spread literacy across first Europe and then the world. Still later, worries arose about the potential for aging printed materials to deteriorate and become lost, as well as worries regarding the ever-growing demands of expanding collections for more and more room within finite library spaces. The new technology of microfilming was seen as a method to preserve and make copies more widely available (Baker, 2001). The dawn of the computer age saw a few limited attempts to create digital collections that would work toward this goal, and then, as the title states, "along came Google."

There are few individuals better suited to produce an account of these efforts than Deanna Marcum and Roger C. Schonfeld. Not only are both astute observers of and frequent commenters on library and digitization issues, but they also have extensive personal and professional ties to many of the principal actors in the Google Books project and other digitization efforts. Marcum has held leadership roles with the Council on Library Resources, the Council on Library and Information Resources, and at the Library of Congress, and is currently with Ithaka S+R. Schonfeld has worked with the Andrew W. Mellon Foundation, served on the board of the Center for Research Libraries, and has written a well-regarded history of the JSTOR project, which has done much to preserve the journal literature. It should be noted that, at times, these close ties do seem to influence the writers to be less critical of individuals than what might be true otherwise.

The authors' intentions in this work are, as they say, "not to offer a final judgement of Google but rather to explore deeply one example of its efforts to target … the important legacy of published materials held by libraries…" (p. 3). They pursue this through first an examination of previous efforts in the library arena to preserve and share knowledge. The development of the Interlibrary Loan system and the production of the National Union Catalog to facilitate borrowing requests was perhaps the most significant pre-digital age program (p. 15). They also summarize the increasing professionalization of the library field and the development of various cooperative organizations, including the development of the Ohio College Library Center (now OCLC) in the 1960s (p. 17). Libraries continue to take important steps in collaboration with each other on projects such as the Digital Public Library of America.

Next, they detail the influence of technologists from outside of the library field on efforts to create a universal library. Visionaries such as Brewster Kahle of the Internet Archive are interviewed, and their influence on digitization is explained. The discussion here centers around the inherent conflict between the visions of librarians for a universal library and the technologists' belief that technology makes this preorganization of materials unnecessary. When librarians first imagined the universal library, they tended to see it as organized along the existing principles of library science, involving manually generated metadata following established cataloging rules. The technologists believe that the application of enough computing power applied to the problem will allow algorithmically generated metadata and natural language, full-text searching to provide adequate access (p. 39-40). For a full discussion of this perspective, see *Everything is Miscellaneous: The Power of the New Digital Disorder* (Weinberger, 2007).

The narrative goes on next to describe the entry of Google and their Google Print and Google library (soon renamed Google Books) projects in 2004. These efforts were very much in line with Google's stated vision of "organiz[ing] the world's information and mak[ing] it universally accessible and useful" (Google, 2022). At first, these projects were met with approval from both publishers and librarians. Soon, however, publishers and librarians alike became concerned with Google's interpretation of copyright law, and librarians were also concerned with the quality of the scans. The authors conducted numerous interviews with the library leaders who were directly involved with the projects at the University of Michigan and at Stanford University, detailing the negotiations, scanning, and accessing procedures involved.

The entry of Google into the arena brought with it technical expertise in organizing vast quantities of information, near limitless funding, and an optimism and even audacity that would quickly produce over ten million scanned books within a few short years. This accomplishment was accompanied by challenges and frustrations from multiple quarters. Publishers objected to the scanning of non-public domain works and filed suit in federal court. Librarians pointed out the poor quality of some of the scans and the lack of proper cataloging and

organization of the scans. Personally, I am forever frustrated by the difficulty in tracking down volumes in multivolume works, such as the venerable *Letters and Papers of Henry VIII*. A Tumblr site showcasing some of the most egregious examples of blurred images, worker's hands captured in the scans, and other oddities was even created (https://theartofgooglebooks.tumblr.com/).

Google and the publishers were able to arrive at a settlement that promised to provide benefits to Google, the publishers, and to public libraries. This proposed settlement was rejected by the courts, leading Google to lose interest in continuing the project. The benefits of the project live on in the public domain: in the snippet scans that are still available to view on Google Books, the more complete scans stored at the HathiTrust Digital Library, and in the spirit of the venture that lives on in many projects that continue to this day by other hands. These projects benefited from the technological advances of the project and the lessons, good and bad, that were learned.

The story of the founding of HathiTrust is told in detail in chapter seven. The authors discuss the various rivalries and conflicts between the founding universities but are overwhelmingly positive about the impact of the Trust on digitization and library collaboration. A perspective that is not considered is that of libraries of limited resources. For smaller colleges and universities who regularly read such praise in the literature written by those with access, it seems a bit of a cruel joke. We read about all that is available through the Trust but are unable to take part. Although membership has become more affordable in recent years and has expanded beyond the original elite institutions, it would be gratifying to see more accessibility of HathiTrust resources to smaller libraries.

Nonetheless, this volume is an essential text to anyone interested in the history of digitization in libraries. It is a fascinating read that shares the stories of the many well-intentioned people who worked toward this latest attempt at a universal library. Finally, it shows how, despite all the best intentions, sometimes it is impossible to be completely successful.

REFERENCES

Baker, N. (2001). *Double fold: Libraries and the assault on paper*. Random House.

Casson, L. (2001). *Libraries in the ancient world*. Yale University Press.

Google. (2022, February 2). About. https://about.google

Schonfeld, R. (2003). *JSTOR: A history*. Princeton University Press.

Weinberger, D. (2007). *Everything is miscellaneous: The power of the new digital disorder*. Henry Holt.

AUTHOR BIOGRAPHY

John Adkins is the library director at the Schoenbaum Library at the University of Charleston in Charleston, West Virginia. He holds an MA in Humanities (Historical Studies) from the West Virginia Graduate College (now Marshall University) and an MSLS degree from the University of Kentucky.