

Marek Sroka*
University Library, University of Illinois
Urbana Champaign, USA



THE GOOGLE LIBRARY PROJECT, INTERNET ARCHIVE AND OPEN CONTENT ALLIANCE: STRUGGLE FOR UNIVERSAL ACCESS TO DIGITAL COLLECTIONS FROM THE AMERICAN PERSPECTIVE

[GOOGLE LIBRARY PROJECT, INTERNET ARCHIVE I OPEN CONTENT ALLIANCE:
WALKA O POWSZECHNY DOSTĘP DO KOLEKCJI CYFROWYCH Z PERSPEKTYWY
AMERYKAŃSKIEJ]

Abstrakt: The article examines the future of free and fee-based access to digital collections in the United States, following the recent settlement of the litigation concerning the Google Library Project. The settlement made many librarians uneasy about potential limitations on access to and use of books for academic and public libraries. Librarians need to ensure that the principles of borrowing and sharing are expanded in a digital world, not circumscribed by contract law or limited by corporate interests. Alternatives to fee-based services are discussed, including the Open Content Alliance and the Internet Archive that strive to offer permanent and universal access to researchers, historians, and scholars to historical collections that exist in digital format. Both organizations continue to build a permanent archive of multilingual digitized text and multimedia content. The article offers insights into American digital world that may be useful for Polish and other European designers and policy makers involved in building digital collections in their countries.

DIGITIZATION – GOOGLE LIBRARY PROJECT – INFORMATION ON THE INTERNET – INTERNET ARCHIVE – USA

Abstract: Przedstawiono perspektywy rozwoju form płatnego i darmowego dostępu do kolekcji cyfrowych w Stanach Zjednoczonych w kontekście niedawnych rozstrzygnięć sądowych zapadłych w sporze wokół Google Library Project. Wielu bibliotekarzy obawia się, że następstwem tych rozstrzygnięć mogą być ograniczenia w dostępie do książek i w prawach do ich wykorzystywania przez biblioteki akademickie i publiczne. Bibliotekarze chcą mieć pewność, że zasada wypożyczenia i dzielenia się materiałami obowiązuje także w świecie cyfrowym, i nie podlega restrykcjom wynikającym z prawa handlowego ani z ochrony interesów wielkich korporacji. Omówiono rozwiązania alternatywne wobec form dostępu płatnego, m.in. inicjatywy organizacji Open Content Alliance i Internet Archive, starających się o zapewnienie społeczności badaczy, historyków i uczonych powszechnego, stałego

* Professor MAREK SROKA, Associate Professor of Library Administration and Coordinator of Area Studies Division, University of Illinois Library. Master of English Philology at the Jagiellonian University; Master of Library Science, University of Illinois Graduate School of Library and Information Science. Two the most important publications: (2007) *The Music Collection of the Former Prussian State Library at the Jagiellonian Library in Kraków, Poland: Past, Present, and Future Developments. Library Trends* No. 55(3), p. 651–664; (2005) *Commercial and Noncommercial Internet-based Information Services: A Threat or a Chance for Libraries*. [In:] M. Kocójowa red. *Profesjonalna informacja w Internecie*. Kraków: Wydaw. UJ, p. 62–66. E-mail: msroka@illinois.edu

Prof. MAREK SROKA, Associate Professor of Library Administration and Coordinator of Area Studies Division, University of Illinois Library. Magister filologii angielskiej w Uniwersytecie Jagiellońskim oraz MLS (*Master of Library Science*), University of Illinois Graduate School of Library and Information. Dwie najważniejsze publikacje: (2007) *The Music Collection of the Former Prussian State Library at the Jagiellonian Library in Kraków, Poland: Past, Present, and Future Developments. Library Trends* No. 55(3), p. 651–664; (2005) *Komercyjne i bezpłatne serwisy informacyjne w Internecie*. [W:] M. Kocójowa red. *Profesjonalna informacja w Internecie*. Kraków: Wydaw. UJ, s. 62–66. Adres elektr.: msroka@illinois.edu

dostępu do cyfrowych kolekcji dokumentów archiwalnych. Obie te organizacje wciąż rozbudowują swoje kolekcje zdigitalizowanych wielojęzycznych dokumentów tekstowych i materiałów multimedialnych. Zaprezentowany opis rozwiązań amerykańskich może okazać się przydatny dla decydentów i innych podmiotów uczestniczących w tworzeniu kolekcji cyfrowych w Polsce i w innych krajach europejskich.

GOOGLE LIBRARY PROJECT – DIGITALIZACJA – INFORMACJA W INTERNECIE – INTERNET ARCHIVE – USA

*

* *

Libraries and the Google Library Project Settlement

“Google’s mission is to organize the world’s information and make it universally accessible and useful” (from Google’s Web site).

Four years ago, the Authors Guild, the Association of American Publishers and a handful of authors and publishers filed a class action lawsuit against Google Book Search. Under the Google Library Project, Google has been scanning into its search database, Google Book Search, millions of books provided by major research libraries and other sources. For those books not in the public domain, the publishers and authors claimed that Google’s scanning infringed their copyrights.

On October 28, 2008, Google, the Authors Guild, and the Association of American Publishers announced the settlement of the litigation concerning the Google Library Project [Google Book Settlement, on-line doc.]. It will take some time for this agreement to be approved and finalized by the Court. The 200-page agreement is extremely complex and includes numerous attachments. The settlement presents significant challenges and opportunities to libraries [cf. Band, on-line doc.].

Opportunities

The settlement still requires the approval of the US district court in New York. If approved, the settlement will provide libraries with opportunities as well as challenges.

The Google deal offers a great opportunity to increase exponentially the scanning of books and use of them. For example, for the in-copyright, not commercially available books that have been scanned through the Google Library Project, Google will display up to 20% of the book’s text. Currently, Google displays only a few snippets of the book’s text. The settlement should allow “a significant expansion of the amount of text users could read of an estimated 70% of published books” [cf. Band, on-line doc., p. 4]. The agreement does not affect how Google displays out-of-copyright (public domain) books. Google will continue to allow Book Search users to read, download and print these titles, just as they do today. All users in the United States will have the ability to search Google’s entire database for digitized books.

Under the agreement, every out-of-print book that Google digitizes will become available online for preview and purchase unless its author or publisher chooses to “turn off” that title [Google Book Settlement, on-line doc.]. Google believes “it will be a tremendous boon to the publishing industry to enable authors and publishers to earn money from volumes that are hard to obtain.”

For institutions that choose to purchase institutional subscriptions, their users will be able to access all of the books in the subscription at no cost to the individual.

Challenges and Concerns

The 200-page settlement has been described by some as “a raw deal for libraries.” Considering the fact that libraries have made huge investments in the books that Google is digitizing, the settlement seems to have ignored libraries’ demands for unrestricted and universal access to their collections.

For example, libraries have always been strong advocates of inter-library loan as an important tool of resource sharing. Under the settlement, the library of a higher education institution may permit faculty and research staff to read, print, download, or otherwise use five pages of any book in its LDC (library digital copy set) that is not commercially available for personal scholarly use and classroom use that is limited to students in the class for the term in which the class is offered. The library, however, cannot use the LDC for interlibrary loan, e-reserves, and course management systems.

Google will make available institutional subscriptions that will allow users within an institution to view the full text of all the books within the Institutional Subscription Database (ISD). Access will continue only for the duration of the subscription and it will not be perpetual. Through agreements with the subscribing institution, Google will limit access to ISD books to “appropriate individuals” within the institution. These will include faculty, students, researchers, staff members, librarians, personnel, business invitees, and walk-in users from the public. Each authorized user will be able to copy, paste, and print the entire book. Once freely available material will now be subject to a contractual subscription with no guarantee that the subscription will remain affordable for the near future.

Google will provide free Public Access Service (PAS) to each public library and not-for-profit higher education institution that requests PAS. Google can provide free PAS to one terminal in each library building in a public library system. A user sitting at a PAS terminal will be able to view the full text of all books in the Institutional Subscription Database and print pages of materials for a “reasonable” per-page fee. No copying and pasting will be allowed. A limited number of terminals allowed under the settlement will restrict the access. For instance, for higher education institutions that do not qualify as Associate Colleges under the Carnegie Classification of Institutions of Higher Education, Google can provide free PAS to one terminal for every 10,000 full-time equivalent students. For those institutions that qualify as Associate Colleges under the Carnegie Classification of Institutions of Higher Education, Google can provide free PAS to one computer terminal per 4,000 full-time equivalent students.

These are just some examples of the restrictions to access included in the Google settlement. According to Francine Fialkoff, editor-in-chief of *Library Journal*, “librarians must do better than acquiesce in an arrangement that relinquishes ownership of books online in favor of contractual provisions and for-pay schemes that subvert the ideals of the public library and academic inquiry” [Fialkoff, on-line doc.]. The question that remains unanswered is what will happen when Google decides to no longer support its book database by making it too expensive for libraries to purchase. Will Google become the sole owner of the world’s information and knowledge? Fortunately, there are still many who believe in fair use and the principles of borrowing and sharing expanded in

a digital world. Any monopoly and concentration of power is dangerous and restricting access to human knowledge is not acceptable.

Internet Archive

“I’m not interested in building an empire. Our idea is to build the future” (Internet Archive founder Brewster Kahle, in the *Economist*, March 7, 2009).

Founded in 1996, the Internet Archive is a non-profit whose main goal is to build a digital library of Internet sites and other cultural artifacts in digital form, with “the purpose of offering permanent access for researchers, historians, and scholars to historical collections that exist in digital format” [About the Internet Archive, on-line doc.]. The archive contains various forms of materials, including texts (1,298,559 texts), audio (321,848 recordings [MP3 files]), live music archive (59,422 concerts), and moving images (166,795 film and videos – figures quoted as of early April 2009; for the most recent numbers see <http://www.archive.org/index.php>). The most famous part of the archive is the Wayback Machine, which stores copies of Internet sites. According to Paul Courant, the dean of libraries at the University of Michigan, “the Internet has become the medium of choice for a great deal of cultural production,” and the Wayback Machine “gives us access to what people were producing at different points in time” [The Internet’s librarian, p. 34–35]. The Internet Archive is also involved in a large-scale digitization project, scanning more than 1,000 books every day. Libraries have to pay a reasonable price for each book, and both parties can then make the digital copy available.

However, all these things are steps towards a wider goal: to create an electronic version of the ancient library of Alexandria referred to by the Internet Archive founder, Brewster Kahle, as “Alexandria 2.0.” [The Internet’s librarian, p. 34–35]. To that end, the Internet Archive embarked on “the Open Library” project, which is funded in part by a grant from the California State Library [The Open Library, on-line doc.]. The project has the lofty goal of creating “one web page for every book ever published,” with links to its full text where available [The Open Library, on-line doc.]. The site currently features approximately 22,900,000 book records, including more than one million books with full-text availability that are searchable and downloadable.

The Internet Archive remains a major player in the Open Content Alliance (OCA), which is “a collaborative effort of a group of cultural, technology, nonprofit, and governmental organizations from around the world that helps build a permanent archive of multilingual digitized text and multimedia material” [What is the Open Content Alliance? on-line doc.]. Members of the Open Content Alliance include various library consortia such as CARLI (the Consortium of Academic and Research Libraries in Illinois) with its 152 member libraries, Boston Library Consortium and other American libraries. International members include Australian, British, and Canadian libraries, including the National Library of Australia, the British Library, and the University of Toronto. The Internet Archive administers the Open Content Alliance.

Future Access to Digital Collections

Online access to human knowledge and information is being shaped by many approaches, including commercial, practical, and idealistic factors. Google represents commercial interests and is now the only big company to be building a significant digital collection of copyrighted books. The recent settlement would allow

Google to scan and sell digital access to almost all books published in the last 85 years and make money providing access to them online.

Libraries are under tremendous pressure from scholars, faculty, students, and college and university administration to provide easy access to their vast holdings through digitized collections. Since Google covers the cost of scanning libraries' books (at least part of it), and libraries are feeling the current financial squeeze, many library administrators have little choice but to embrace Google as a practical alternative to doing their own scanning.

Finally, there is an idealistic approach represented by Internet Archives, and the like, who favor free and universal access to digitized collections and the doctrines of fair use and public domain as well as the expansion of the universe of copyright-free works.

Libraries need to ensure that the principles of borrowing and sharing of their rich resources are expanded in a digital world. Potential limitations on access to and use of books for academic and public libraries should be strongly opposed and the development of national and regional digital libraries and depositories should be encouraged to counterbalance Google's monopolistic tendencies. The ownership of information and knowledge is too critical to be left to the marketplace and leased or licensed to future generations. Libraries have always supported public access to their print collections and the same principal should apply to a digital environment.

References

- A Guide for the Perplexed: Libraries and the Google Library Project Settlement, doc. electr. <http://www.arl.org/pp/ppcopyright/google/> [visited: March 2009].
- About the Internet Archive, doc. electr. <http://www.archive.org/about/about.php> [visited: April 2009].
- Fialkoff, F. doc. electr. (2008). Editorial: Google Deal or Rip-Off? Librarians need to protect the public interest. *Library Journal* [blog], 12/15/2008. <http://www.libraryjournal.com/index.asp?layout=articlePrint&articleID=CA6618842> [visited: March 2009].
- Google Book Settlement, doc. electr. <http://books.google.com/booksrightsholders/agreement-contents.html>. Also available in Polish: Uгода dotycząca programu Google Book Search [visited: March 2009].
- The Internet's librarian (2009). *The Economist* No. 7, p. 34–35.
- The Open Library, doc. electr. <http://openlibrary.org/about> [visited: April 2009].
- What is the Open Content Alliance?, doc. electr. <http://www.opencontentalliance.org/about/> [visited: April 2009].