



A GUIDE FOR THE PERPLEXED: LIBRARIES AND THE GOOGLE LIBRARY PROJECT SETTLEMENT

Jonathan Band
policybandwidth
jband@policybandwidth.com

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 (<http://books.google.com/booksrightsholders/agreement-contents.html>). Under the project, Google has been scanning into its search database 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. The settlement still requires the approval of the presiding judge in the US district court in New York because the case was brought as a class action on behalf of all affected rightsholders.

The settlement presents significant challenges and opportunities to libraries. This paper does not explore the policy issues raised by the settlement. Rather, it outlines the settlement's provisions, with special emphasis on the provisions that apply directly to libraries. The settlement is extremely complex (over 200 pages long, including attachments), so this paper of necessity simplifies many of its details. This paper should not be treated as legal advice, and libraries considering joining the settlement should retain counsel to advise them on the settlement's intricacies. Page references to the agreement are included in parentheses.

Basic Framework

Under the settlement, Google will continue scanning in-copyright books into its search database, and will continue to enable users to search the full content of the scanned books. The settlement creates a mechanism for Google to pay rightsholders for the right to display more of the text of books than it displays under the current program. This mechanism is the Book Rights Registry (BRR) that will distribute the payments from Google to the copyright owners. Google, in turn, will generate revenue through advertising and by selling to users the ability to see full text. Google will retain 37% of the revenue it generates under this program, and will pay the other 63% to the BRR. Additionally, Google will make an upfront payment of a minimum of \$45 million to the BRR for distribution to rightsholders whose books will have been scanned by January 5, 2009. The BRR's board of directors will consist of equal numbers of representatives of publishers and authors.

The settlement defines a book as a published or publicly distributed set of written or printed sheets of paper bound together in a hard copy. The settlement specifically excludes periodicals, personal papers (such as unpublished diaries or bundled letters), or works with more than a specified amount of musical notation and lyrics. The settlement also excludes books not registered with the US Copyright Office, unless the book was first published outside of the United States.

The settlement contemplates three categories of books:

- In-copyright, commercially available (in essence, in print or available through a print-on-demand program)
- In-copyright, not commercially available
- Public domain

By consulting with existing databases, Google will make the initial determination of whether a book is commercially available. The settlement sets forth a procedure for the

rightsholder or the BRR to challenge Google's classification. Similarly, Google will determine whether a book is in the public domain, subject to a challenge by the rightsholder or the BRR. The settlement provides Google with a safe harbor for erroneous initial classifications.

The settlement establishes default rules for what Google can do with the two categories of in-copyright books—display uses (discussed below) are turned on for books that are not commercially available and are turned off for commercially available books. (Google has complete freedom with respect to the public domain books since they are not subject to copyright.) Significantly, the settlement does not apply to books first published after January 5, 2009. Additionally, rightsholders will have the ability to opt-out of the settlement altogether, to remove specific books from Google's servers, or to vary any of the default rules with respect to specific books. Thus, as a practical matter, the settlement probably will have limited applicability to in-copyright, commercially available books; the rightsholders likely will closely manage their rights in these books rather than subject them to the settlement's general default rules.

This means that the settlement primarily focuses upon the universe of in-copyright books that are no longer commercially available. Google estimates that approximately 70% of published books fall in this category, 20% of published books are in the public domain and outside of the settlement, and 10% are in-copyright and commercially available.

Service and User Types

The settlement provides different free and fee-based services to three different but overlapping categories of users: all users; public libraries and universities; and institutions.

All Users—Free Services

- All users in the United States will have the ability to **search** Google’s entire search database for digitized books responsive to their queries.
- For a **public domain** book, Google will **display** the full text.
- For an **in-copyright, not commercially available** book, the default rule is that Google will **display** up to 20% of the book’s text. (p. 52) Currently, Google displays only three “snippets” of text per search query. The settlement, therefore, should allow a significant expansion of the amount of text users could read of an estimated 70% of published books.
- Although under this “standard preview” Google can display up to 20% of a book’s text, for most non-fiction works Google generally can display no more than five adjacent pages at a time. Thus, when a user lands on a given page from a search, the user can see four pages adjacent to that page. The user can then ask to see five other adjacent pages where the search term appears again. However, Google will block the two pages before and after any five-page display. (p. 52)
- Different default rules apply to works of fiction for the amount a user can see in response to a single command. Each time a user lands on a page of a fiction book, Google can display 5% of the book or fifteen adjacent pages, whichever is less. Google will also block the final 5%, or at least the final fifteen pages. However, the cumulative display rule of 20% still applies. (p. 52)
- Still different default display rules apply to other categories of works. No text display is allowed of anthologies of drama and fiction by multiple authors, or

collections of poetry or short stories. And for dictionaries, drug reference guides, encyclopedias, price/buyer guides, quotation books, test preparation guides, and thesauri, Google will provide only a “fixed preview”—it will display the same pages regardless of the user query, up to 10% of the book. Google will make these classifications in accordance with Book Industry Standards and Communications (BISAC) codes. (p. 53)

- For an **in-copyright, commercially available** book, the default rule is that Google will **display** only bibliographic information and front material, such as the title page, the copyright page, the table of contents, and the index. For books in this category, Google will no longer display even snippets, as it currently does, unless the rightsholder so authorizes. Hence, the settlement’s default rule for this category of books requires Google to display less than it does now.
- Users will not be able to print out or copy-and-paste any of the free displays. (p. 52)
- As noted above, a rightsholder can vary the default rules for its book. Moreover, the settlement allows the rightsholder of a work contained within another rightsholder’s book to exercise its rights under the settlement independently. The settlement recognizes “inserts,” which include: (1) text such as forewords, afterwords, essays, poems, short stories, letters, and song lyrics; (2) illustrations in children’s books; (3) musical notation; and (4) tables, charts, and graphs. Inserts do not include photographs, illustrations (other than in children’s books), maps or paintings. (p. 9) The rightsholder of an insert contained in an in-copyright, not commercially available book can choose to exclude displays of the insert, even if the rightsholder of the book itself permits Google to display the rest of the book under the default rules. (p. 29) Similarly, the rightsholder of an insert contained in a government work or a public domain book may request Google to exclude the insert when it displays the rest of the book. (p. 32)

However, unlike a book's rightsholder, an insert's rightsholder cannot insist that the insert be removed altogether from the Google Library Project. Thus, so long as a book's rightsholder does not remove the book, all inserts within the book will be searchable, even if their rightsholders exclude them from any displays.

All Users—Fee-Based Services

- Users will be able to purchase online access to the full text of in-copyright, not commercially available books through an account established with Google.
- The rightsholder can set the purchase price for the book. Google will set the price for all books not priced by the rightsholders based on a pricing algorithm designed to find the optimal price for each book to maximize the revenue for the rightsholder. Initially, books will be distributed in pricing "bins" in the following percentages: 5% of the books available for purchase will be priced at \$1.99; 10% at \$2.99; 13% at \$3.99; 13% at \$4.99; 10% at \$5.99; 8% at \$6.99; 6% at \$7.99; 5% at \$8.99; 11% at \$9.99; 8% at \$14.99; 6% at \$19.99; and 5% at \$29.99. The algorithm will place a book in a pricing bin based on aggregate data collected with respect to similar books. Google can change the price of a book in response to sales data. Google also can change the distribution of books in the pricing bins over time as the prices of individual books are adjusted based on the pricing algorithm. Google and the BRR can agree to modify the number of bins, the prices, and the distribution of books within the bins. (p. 50) Additionally, three years after the settlement takes effect, and every four years thereafter, Google or the BRR can require renegotiation of pricing structure. If the parties cannot reach agreement, the settlement provides for a dispute resolution mechanism involving binding arbitration. (p. 49)
- After purchasing the book, the user will have perpetual online access to **view** the entire book from any computer. (p. 48)

- The user will be able to **copy and paste** up to four pages of the purchased book with a single command, but, with multiple commands, can copy and paste the entire book. (p. 48)
- The user will be able to **print up** to twenty pages of the purchased book with a single print command, but, with multiple commands, can print out the entire book. Google will place a watermark on printed pages with encrypted identifying information that identifies the authorized user that printed the material. (p. 48)
- The user will be able to make **book annotations** of the purchased book. A book annotation is user-generated text that is displayed on any Web page on which a page of a book appears. (p. 48) The user can share his annotations with up to 25 other individuals who have purchased the book through this service and who have been designated by the user. (p. 39)
- A user who purchases a book will not see an insert if the insert's rightsholder chooses to exclude displays of the insert. In this situation, a purchaser (or an institutional subscriber, described below) will not have access to the complete book as published.

Free Public Access Service for Public Libraries and Universities

- Google will provide free Public Access Service (PAS) to each public library and not-for-profit higher education institution that requests PAS. (p. 60)
- A user sitting at a PAS terminal will be able to **view** full text of all books in the Institutional Subscription Database. This generally corresponds to books in the in-copyright, not commercially available category. (p. 60)
- A user can **print** pages of material viewed on the PAS terminal for a "reasonable" per-page fee set by the BRR. (p. 60) The user will not be able to copy and paste text or annotate books accessed through the PAS.

- Google can provide free PAS to one terminal in each library building in a public library system. (p. 60) A public library is a library that (a) is accessible by the public; (b) is part of a not-for-profit or government-funded institution other than an institution of higher education under the Carnegie Classification; and (c) allows patrons to take books and other materials off the premises. The settlement does not treat any library primarily funded or managed by the federal government as a public library. (p. 15)
- 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. (p. 60)
- For higher education 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. (p. 60)
- Google and the BRR can agree to expand the PAS service by making additional terminals available for free or an annual fee, but the settlement provides no further details on the terms for this expansion. (p. 60)

Institutional Subscriptions

- 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). This database will include the books in the in-copyright, not commercially available category. This access will continue only for the duration of the subscription; access will not be perpetual, in contrast to when a user purchases access to an individual book, as described above.
- Google can also offer subscriptions to subsets of the ISD that represent discipline-based collections. (p. 43)

- Through agreements with the subscribing institution, Google will limit access to ISD books to “appropriate individuals” within the institution. For educational institutions, appropriate individuals include faculty, students, researchers, staff members, librarians, personnel, business invitees, and walk-in users from the general public. For public libraries, appropriate individuals include library patrons and personnel. (p. 47)
- Each authorized user will be able to **copy and paste** up to four pages of a book in the ISD with a single command, but, with multiple commands, can copy and paste the entire book. (p. 47)
- Each authorized user will be able to **print** up to twenty pages of a book in the ISD with a single print command, but, with multiple commands, can print out the entire book. Google will place a watermark on printed pages with encrypted identifying information that identifies the authorized user that printed the material. (p. 47)
- Each authorized user may make **annotations** of books in the ISD. (p. 47)
Instructors and students in an academic course can share annotations with each other and with students enrolled in the same course the subsequent year. Also, employees of the institutional subscriber can share annotations with other employees in connection with a discrete work project for the duration of the project. (p. 40)
- Authorized users can make books in the ISD available to other users authorized by that subscription through hyperlinks or similar technology for course use such as e-reserves and course management systems. (p. 47)
- Google will not prohibit any other uses of books in the ISD that fall within the Copyright Act’s limitations and exceptions, e.g., fair use. (p. 47)
- Google can subsidize the purchase of institutional subscriptions by fully participating and cooperating libraries—categories explained below. (p. 57)

Pricing of Institutional Subscriptions

- Google and the BRR will set the price of institutional subscriptions. If they cannot agree on a price structure, the settlement provides for a dispute resolution mechanism involving binding arbitration.
- The economic terms for the institutional subscriptions will be governed by two objectives: “(1) the realization of revenue at market rates for each Book and license on behalf of Rightsholders and (2) the realization of broad access to the Books by the public, including institutions of higher education.” Moreover, “Plaintiffs and Google view these two objectives as compatible, and agree that these objectives will help assure both long-term revenue to the Rightsholders and accessibility of the Books to the public.” (p. 42)
- Google and the BRR will use the following parameters to determine the price of institutional subscriptions: the pricing of similar products and services available from third parties; the scope of the books available in the ISD; the quality of the scan; and the features offered as part of subscription. (p. 42)
- Pricing will be based on the number of full-time equivalent (FTE) users. For higher education institutions, FTE means full-time equivalent students. (p. 42)
- The FTE pricing can vary across different categories of institutions. These categories include: (1) corporate; (2) higher education institutions (which may be sub-divided based on the Carnegie Classifications for Institutions of Higher Education); (3) K-12; (4) government; and (5) public library. Only higher education institutions can have remote access without BRR approval (e.g., faculty can access the ISD from home and students from their dormitories). (p. 42)
- Google can charge a lower price for a discipline-based subset of the IDS. However, “[t]o provide an incentive for institutions to subscribe to the entire Institutional Subscription Database, Google shall design the pricing of the different versions of the Institutional Subscription such that the price for access

- to the entire Institutional Subscription Database will be less than the sum of the prices for access to the discipline-based collections.” (p. 43)
- Google will propose an initial pricing strategy consistent with the objectives outlined above that will include target retail prices for each class of institution for access to the entire ISD and the discipline-based collections, and proposed discounts for institutional consortia and early subscribers. After Google submits the initial pricing strategy to the BRR, Google and the BRR will negotiate its terms for up to 180 days. If Google and BRR do not reach agreement, the dispute will be submitted to binding arbitration. (p. 43-44)
 - FTE-based prices in the initial pricing strategy period will be based on “then-current prices for comparable products and service, surveys of potential subscribers, and other methods for collecting data and market assessment.” Google will be responsible for collecting data comparing the target retail prices with the prices for comparable products and services, and will provide this data to the BRR. (p. 45) Presumably the arbitrators will rely on this data in the event of a dispute concerning the pricing strategy.
 - The initial pricing strategy is expected to be in effect for two to three years. Google and the BRR will agree on the duration of subsequent pricing strategies. (p. 44)
 - Should Google provide other services to institutional subscribers for a fee, those services would fall within the settlement and the BRR would be entitled to a portion of the revenue if: (1) the preponderance of the value of the service is realized through access to books through the institutional subscription; and (2) the service exploits the access provided by the subscription in a manner that could not be exploited by other entities. (p. 46)

Library Types

Under the existing Google Library Project, Google has numerous partner libraries that have provided it with books to scan. In exchange, Google has provided these partner libraries with digital copies of the books. The settlement creates four categories of partner libraries with different rights and responsibilities: fully participating libraries, cooperating libraries, public domain libraries, and other libraries.

Fully Participating Libraries

- To become a fully participating library, a library must sign an agreement with the BRR. The agreement releases the library from any liability for copyright infringement for participating in the Google Library Project, and for any activity that falls within the scope of the agreement.
- A fully participating library will provide Google with in-copyright books to scan into its database, and will receive in return a digital copy of each book it provides. The set of digital copies Google provides the library is the library digital copy (LDC).
- Google can provide a fully participating library with digital copies of books in that library's collection that Google did not obtain from that library (i.e., Google obtained the book from another fully participating library). For a library with more than 900,000 books in its collection, Google can provide it with digital copies from other libraries only if Google scans more than 300,000 books from that library's collection. For a library with less than 900,000 books in its collection, Google can provide LDCs from other libraries only if it scans more than 30% of the library's collection. (For purposes of this calculation, only in-copyright books count.) However, Google can provide the library only with digital copies of books contained in that library's collection. (p. 72)

- For institutional consortia, different minimum levels of participation apply before a library can receive digital copies made from another library's collection. Google must have scanned at least 10,000 books from that library's collection. Additionally, if the consortium has more than 2,000,000 books, Google must have scanned more than 650,000 of those books; and if the consortium has less than 2,000,000 books, Google must have scanned more than 30% of the books in the consortium's collection. (p. 73)

A Fully Participating Library's Use of the LDC

The settlement specifies in detail what a fully participating library can and cannot do with its LDC.

- The library may reproduce and make technical adaptations of the LDC "as reasonably necessary to preserve, maintain, manage, and keep technologically current its LDC." (p. 73)
- The library may use its LDC to create a print replacement copy of a book in its collection that is damaged, destroyed, deteriorating, lost or stolen, or if the format in which the book is stored has become obsolete, provided that the library has determined that an unused replacement copy cannot be obtained at a fair price. (An unused replacement for a copy in print format means an unused copy that is offered for sale in print format.) (p. 75)
- The library may provide special access to books in the LDC to a user with print disabilities, i.e., a user unable to read or use standard printed material due to blindness, visual disability, physical limitations, organic dysfunction, or dyslexia. (p. 14) This access includes screen enlargement, voice output, or refreshable Braille displays. The special access cannot be provided in a way that would make a copy accessible to anyone other than the disabled user, or that would

- make the special access available longer than necessary to facilitate the special access.
- This special access is available only to a person who has provided written documentation that a “competent authority” has certified that the user has a print disability. A competent authority is a person (1) employed in a professional occupation qualified to diagnose print disabilities under federal law and regulations that govern the National Library Service for the Blind and Physically Handicapped; or (2) licensed under applicable state law to diagnose the existence of a print disability under standard and generally accepted methods of clinical evaluation. (p. 5) Additionally, a professional librarian may certify a user’s claimed print disability only if the user affirms in writing that no competent authority is available, or if the user has a print disability that is readily apparent upon physical observation of the user. The user must also provide written documentation that he or she will not reproduce or distribute books in a manner prohibited by the Copyright Act. (p. 73-74)
 - The library may develop its own finding tools that allow its users to identify pertinent material within its LDC. These tools may permit users to read or view only snippets of text from the LDC. (p. 75)
 - The library may allow users to conduct “non-consumptive research” on its LDC, provided that the library agrees to the terms of a host site of a Research Corpus. Non-consumptive research and the Research Corpus are discussed below in greater detail.
 - 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 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 must keep track of such uses and report them to the BRR in the

course of the audits required under the security provisions discussed below. (p. 76) At any time that an institutional subscription is not being offered, additional uses of books that are not commercially available may be authorized jointly by the university librarian and the university general counsel. However, such uses cannot include sale of access, interlibrary loan, e-reserves, course management systems, or any infringing uses. (p. 78-79)

- The library may allow its support personnel, archivists, information technology personnel, and legal counsel to read, print, download, and otherwise use books from the LDC as reasonably necessary to carry out their responsibilities with respect to the LDC. (p. 76-77)
- The library may authorize another fully participating library to host and store its LDC together with or separately from the hosting library's LDC. (p. 77)
- The library may authorize other third parties to exercise its rights and perform its obligations, including the hosting and storage of the LDC. However, it will be the library's responsibility that such third parties comply with the settlement, particularly the security obligations described below. (p. 78)
- The library is prohibited from using its LDC: (1) for directly or indirectly selling books or access to books; (2) for interlibrary loan; (3) for e-reserves; (4) in course management systems; and (5) any other use that would violate copyright law. (p. 78-79)

Security Obligations

A fully participating library must follow detailed procedures to protect the security of its LDC. These same procedures apply to host libraries, discussed below, as well as Google.

- A fully participating library needs to develop a security implementation plan that meets the requirements of the Security Standard, which is set forth in an attachment to the settlement agreement. (p. 94)
- The seventeen-page Security Standard addresses topics such as: (1) security management, including security awareness, designation of a security representative, and incident response; (2) identification and authentication, including user identification and authentication, and authentication and password management; (3) access controls, including account management, access approval process, and access control supervision; (4) audit and accountability, including logging and audit requirements, marking of image files, and forensic analysis; (5) network security, including electronic perimeter, network firewall, device hardening, network security testing, remote network accessing, and encryption of digitized files; (6) media protection, including media access, media inventory, media storage, and media sanitization and disposal; (7) physical and environmental protection, including physical access authorizations, physical access control, visitor control, and access records; (8) risk assessment. (p. D-i)
- The Security Standard can be revised every two years by agreement between BRR and representatives of fully participating libraries “to take account of technological developments, including new threats to security....” Disagreements between the BRR and the libraries concerning modifications to the Security Standard are subject to binding arbitration. (p. 95)
- The fully participating library must submit its security implementation plan to the BRR for approval. If disagreements between the fully participating library and the BRR as to whether the security implementation plan complies with the Security Standard cannot be resolved, they will be submitted to binding arbitration. (p. 94)

- Each fully participating library must permit a third party to conduct an annual audit of the library's security and usage to verify compliance with its security implementation plan. Google and the BRR will share in the costs of the audits. (p. 95-96)
- Upon learning of a prohibited or unauthorized access to the LDC, the fully participating library must notify the BRR of the breach and attempt to cure it, e.g., block the unauthorized access. The library must confer with the BRR on ways to prevent such breach from reoccurring, and must negotiate with the BRR or the affected rightsholder an appropriate monetary remedy. (p. 97) If the parties cannot agree on an appropriate remedy, the issue will be submitted to binding arbitration.
- The settlement establishes a schedule of monetary remedies. If a breach of the security implementation plan does not result in a prohibited access by the library or an unauthorized access by a third party, the range of the remedy is \$0-\$25,000, depending on whether the breach is inconsequential, the recklessness or willfulness of the breaching conduct, the promptness of the cure, and the number of breaches with the same root cause. (p. 100-01)
- If an inadvertent or negligent breach results in a prohibited access by the library itself, the remedy will be the actual damages, with a cap of \$300,000 for all breaches resulting from the same root cause. If the breaching conduct was reckless, willful, or intentional, the cap is \$5 million for reckless breaches and \$7.5 million for willful or intentional breaches. (p. 103)
- If a third party's unauthorized access is not the result of the library's failure to comply with the security implementation plan, then the library owes no damages. In contrast, if a third party's unauthorized access is the result of the library's failure to comply with its security implementation plan, the remedy should attempt to approximate the actual damages. The damages are capped at

\$2 million if the breaching conduct was negligent, \$3 million if the breaching conduct was reckless, and \$5 million if the breaching conduct was intentional. (p. 104)

Additional Library Categories

The settlement recognizes three other categories of libraries partnering with Google in the Library Project: cooperating libraries, public domain libraries, and other libraries.

- “Cooperating libraries” are libraries that intend to provide in-copyright books to Google for inclusion in Google Book Search. However, these libraries have decided not to retain digital copies of in-copyright books provided by Google, and therefore do not have to comply with the settlement’s security provisions. These libraries must destroy the in-copyright digital copies previously provided by Google, and in exchange receive a release from any copyright infringement liability for cooperating with Google. (p. 5) In addition, these cooperating libraries have the ability to force Google to meet certain obligations discussed below.
- “Public domain libraries” are libraries that intend to provide Google only with public domain books. In exchange for destroying any in-copyright digital copies previously provided by Google, these libraries receive a release for any past infringements, and any future inadvertent infringements, e.g., inadvertently providing Google with an in-copyright book. (p. 15)
- “Other libraries” are libraries that have agreed to provide Google books to scan, but have chosen not to participate in the settlement. Such a library presumably would retain the digital copies Google has provided it. However, a library that does not participate in the settlement in theory could find itself the target of infringement actions by the copyright owners. Going forward, Google could

continue scanning public domain books obtained from such a library, and providing the library a digital copy of these public domain books. In this event, neither Google nor the library would qualify for the settlement's safe harbor for erroneous classification of public domain materials, because this activity would not be released by the settlement.

Non-Exclusivity

- The settlement explicitly “neither authorizes nor prohibits, nor releases any Claims with respect to ... any Participating Library’s Digitization of Books if the resulting Digitized Books are neither provided to Google pursuant to this Settlement Agreement nor included in any LDC, or the use of any such Digitized Books that are neither provided to Google pursuant to this Settlement Agreement nor included in any LDC.” (p. 20-21) In other words, the settlement does not restrict fully participating, cooperating, public domain, or other libraries from engaging in other digitization projects outside of the settlement.
- Likewise, the settlement does not limit any rightsholder’s “right to authorize, through the Registry or otherwise, any Person, including direct competitors of Google, to use his, her or its Books or Inserts in any way, including ways identical to those provided for under this Settlement Agreement.” (p. 21)
- Additionally, the BRR may license rightsholders’ US copyrights to third persons to the extent permitted by law. (p. 65) As a practical matter, the BRR can grant licenses only with respect to rightholders that register with it and grant it the ability to act as its agent with respect to parties other than Google. The class action mechanism cannot bind rightsholders with respect to third parties not participating in the settlement.

Research Corpus

The settlement allows for the creation of two centers (in addition to Google) that would host the Research Corpus, the set of all digital copies made in connection with the Google Library Project. (p. 17)

- The fully participating and cooperating libraries will select the host sites. The host site could be a fully participating or cooperating library, or another institution. The host site must abide by the security procedures described above. (p. 70)
- The host sites may provide on-site and remote access to qualified users to use the Research Corpus for non-consumptive research. Qualified users must be affiliated with a fully participating or cooperating library, an accredited college or university, a not-for-profit research organization such as a museum, or a governmental agency. Additionally, an individual can become a qualified user by demonstrating to a fully participating or cooperating library that he has the necessary capability and resources to conduct non-consumptive research. (p. 15-16)
- Non-consumptive research is research involving computational analysis on books, but not research where the researcher reads and displays substantial portions of a book to understand its intellectual content. Categories of non-consumptive research include: (1) image analysis and text extraction—computational analysis of the digitized image artifact to improve the image (e.g., de-skewing) or extracting textual or structural information from the image (e.g., OCR); (2) textual analysis and information extraction—automated techniques designed to extract information to understand or develop relationships among or within books (e.g., concordance development, collocation extraction, citation extraction, automated classification, entity extraction, and natural language

processing; (3) linguistic analysis—research to understand language, linguistic use, semantics, and syntax as they evolve over time and across different genres; (4) automated translation—research on translation techniques; and (5) indexing and search—research on different techniques for indexing and search of textual content. (p. 11-12)

- The host site is responsible for oversight of the research performed on the Research Corpus, including ensuring that no person uses materials in the Corpus for purposes that involve reading portions of a book to understand its intellectual content. Qualified users may read material as reasonably necessary to perform non-consumptive research, or to explain, discuss, or verify research results. (p. 81)
- Direct, for-profit, commercial use of information extracted from books in the Research Corpus is prohibited. Qualified users may report the results of their non-consumptive research in scholarly publications, including scholarly publications sold to the academic community or the public. (p. 81) Commercial exploitation of algorithms developed when performing non-consumptive research is permitted. (p. 82) Use of data extracted from a specific book to provide services that compete with services offered by the book's rightsholder are prohibited. (p. 82)
- Prior to engaging in research, a qualifying user must file with the host site a research agenda—a document that describes the project in sufficient detail to demonstrate that it is non-consumptive research. (p. 17) Before permitting the qualified researcher to perform the research, the host site will review the research agenda to ensure that the research described is non-consumptive.
- A third party selected by the BRR will perform regular audits on the host sites to ensure that it complies with the terms of the settlement.

- The copyright owner of a commercially available book may request that the book be withdrawn from the Research Corpus. (p. 80)

Google's Obligations

- Google agrees that within five years of the effective date of the settlement, it will provide free search (including permitted displays), the Public Access Service, and institutional subscriptions for 85% of the in-copyright, not commercially available books it has scanned. If Google fails to meet this requirement, the fully participating and cooperating libraries or the BRR may seek to engage a third party to provide these services. If the libraries and the BRR cannot identify or reach agreement on a third party, the libraries may provide these services themselves, using their LDCs. (p. 84-85)
- Google must “use commercially reasonable efforts” to accommodate users with print disabilities. (p. 88) Print disability is any condition in which a user is unable to read or use standard printed material due to blindness, visual disability, physical limitations, organic dysfunction, or dyslexia. (p. 14) The accommodations include screen enlargement, voice output, and refreshable Braille displays. The objective is to accommodate “users with Print Disabilities so that such users have a substantially similar experience as users without Print Disabilities.” (p. 28) Google must make these accommodations available at no extra charge to the users. If within five years of the effective date of the settlement Google fails to offer these accommodated services, the fully participating and cooperating libraries can require Google to work with a third party to provide these services. (p. 88-89)

- The fully participating and cooperating libraries can designate a representative to enforce their collective rights under the settlement, such as the right to required services and accommodated services discussed above. (p. 90)

Process Going Forward

- Starting on January 5, 2009, plaintiffs will provide direct notice (through mail and e-mail) and summary notice of the proposed settlement in publications around the world.
- Members of the class include all persons who as of January 5, 2009, own or have an exclusive license in a US copyright in a book or an insert. Members of the class have until May 5, 2009, to opt-out of the class, or to submit to the court objections to or comments on the settlement.
- At some time after May 5, 2009, the court will conduct a hearing to consider the fairness of the settlement. The court can then accept or reject the settlement.
- A rightsholder who does not opt out of the settlement has until April 5, 2011, to request the complete removal of a specific book from the Google Library Project.
- A rightsholder at any time may request the exclusion of its book from one or more specific types of display by Google. If a rightsholder of a book that is not commercially available excludes a book from the ISD, then the book will be excluded from sale to individual customers.
- The court will retain jurisdiction over the interpretation and implementation of the settlement. (p. 131) However, the settlement specifies many categories of disputes among parties that are subject to binding, nonappealable arbitration. (p. 108)

November 13, 2008