VOLUME 55 | 2010/11

MARY MURRELL

Digital + Library: Mass Book Digitization as Collective Inquiry

ABOUT THE AUTHOR: Mary Murrell is a Ph.D. candidate in anthropology at the University of California, Berkeley. This article is dedicated to the memory of Peter Lyman (1940–2007).

> *When it was proclaimed that the Library contained all books, the first impression was one of extravagant happiness.*[1]

I. INTRODUCTION

The tendency to speak of new technologies in terms of metaphors drawn from the past has a long history.[2] Metaphors help us think about new or unfamiliar phenomena by evoking at precisely the same time something we know and something we do not know, suggesting a horizon of interpretive possibilities within a specific system of meaning. As the information scholar Peter Lyman pointed out a decade ago, the term "digital library" is a metaphor that carries forward imprecise meanings from the past as a means of facing an uncertain future: "In attaching the adjective 'digital' to the noun 'library' the future seems to be reconciled with the past."[3] The term looks back in time to consider what libraries have been, and looks forward to ask what they might be in the future. In so doing, the term "digital library" transfers meanings not only from a particular historical institution (the library) and its many variations, but also from the norms and practices of the entire book system itself, opening that system up to fresh consideration, scrutiny, and reflection. In this article, I will discuss the digital library metaphor in the context of mass-digitization and, in particular, the Google Book Search Project ("Google Books").

Google Books and the controversy around the proposed Settlement of *Authors Guild, Inc. v. Google, Inc.*—the impetus for this issue of the *New York Law School Law Review*[4]—provide a rich case study through which to begin to perceive an emergent digital book system. With my attention here to the language or "poetics" around new technologies, I hope to provide a mode of analysis that is supplementary, but useful, to legal and economic perspectives.

From the moment of its official announcement in late 2004, Google's digitization project, which many describe as a digital library,[5] occasioned a tectonic collision

---

1. Jorge Luis Borges, *The Library of Babel*, *in* Labyrinths: Selected Stories and Other Writings 54–55 (Donald A. Yates & James E. Irby eds., 1964).

2. *See generally* Technological Visions: The Hopes and Fears that Shape New Technologies (Marita Sturken, Douglas Thomas & Sandra J. Ball-Rokeach eds., 2004) (containing a collection of essays on this topic).

3. Peter Lyman, *What is a Digital Library? Technology, Intellectual Property, and the Public Interest*, *in* Books, Bricks & Bytes: Libraries in the Twenty-First Century 1 (Stephan R. Graubard & Paul LeClerc eds., 1996) [hereinafter Lyman, *What is a Digital Library?*]; *see also* Peter Lyman, *Information Superhighways, Virtual Communities, and Digital Libraries: Information Society Metaphors as Political Rhetoric*, *in* Technological Visions: The Hopes And Fears that Shape New Technologies, *supra* note 2, at 201–18 [hereinafter Lyman, *Information Superhighways*].

4. Symposium, *D Is for Digitize*, 55 N.Y.L. Sch. L. Rev. 1 (2010–11).

5. *See, e.g.*, Wade Roush, *The Infinite Library*, 108 Tech. Rev. 54 (2005); Mary Sue Coleman, *Riches We Must Share . . .* , Wash. Post, Oct. 22, 2005, http://www.washingtonpost.com/wp-dyn/content/article/2005/10/21/AR2005102101451.html; Kevin Kelly, *Scan This Book!*, N.Y. Times, May 14, 2006, http://www.nytimes.com/2006/05/14/magazine/14publishing.html; *The Library of Google*, Prospect, Feb. 2007, *available at* http://www.prospectmagazine.co.uk/2007/02/thelibraryofgoogle/; Mary Sue

between the Internet and the book publishing industry. Its reverberations have moved throughout the entire book system, and possibly beyond. The most vivid aspect of this collision, *Authors Guild, Inc. v. Google, Inc.*, ultimately produced the Google Book Search Settlement—a highly elaborate plan, still unapproved, that concerns itself centrally with the provision of electronic access to out-of-print but in-copyright books.[6] In so doing, the Settlement addresses what has been understood as a key role of the library: keeping books accessible to readers after they are no longer commercially available. This attribution of "librariness" to Google Books, however imprecise or disputed it may be, has been a point of departure in the public contestation around the Settlement Agreement. The term "digital library," with its figurative intermingling of dissonant terms, is another type of collision. It poses a series of questions: What do we seek, as we assemble new digital forms, that is different from what we already have? What are we willing to give up, and not willing to give up, from the system that governs and regulates print artifacts? What bearing does format—digital or print—have on the role of a library, if any? What do we want libraries to be in the future? In fact, do we want to continue to support such institutions? How will that desired future be accomplished? What is a public library in an increasingly privatized society? Can a commercial library satisfy the social functions of the traditional public library? These are only some of the questions the term "digital library" poses.

This article, rather than answering these questions, argues that the continued relevance of the "digital library" metaphor over the past two decades—in contrast, say, to the "information superhighway"—is evidence of a persistent yet unfinished collective inquiry into the public institutions that both fashion and undergird the norms for provisioning information and knowledge in a democratic society. The initial conflict around the Google Books Library Project and the subsequent broad public hearing that the Settlement Agreement has received (through the class action notification process, court filings, conferences, formal and informal advocacy, media coverage, and legal proceedings) have combined into a remarkably robust (and unconcluded) episode in this ongoing collective inquiry.

Coleman, President, Univ. of Mich., Google, the Khmer Rouge and the Public Good, Address to the Professional/Scholarly Publishing Division of the Association of American Publishers (Feb. 6, 2006) (transcript available at http://www.battellemedia.com/archives/MSC_AAP_Google_address.pdf); *The Fight Over the Google of All Libraries: An (Updated) Wired.com FAQ*, Wired, Feb. 18, 2010, http://www.wired.com/epicenter/2010/02/the-fight-over-the-worlds-greatest-library-the-wiredcom-faq/; Frank Pasquale, *Conditions for the Digital Library of Alexandria*, Madisonian.net (Nov. 24, 2007), http://madisonian.net/archives/2007/11/24/conditions-for-the-digital-library-of-alexandria/.

6. There have been two proposed settlement agreements to the *Authors Guild v. Google* lawsuit: the original agreement of October 2008 and the amended agreement of November 2009. This article focuses on the settlement in general and not on the differences between the two versions. Therefore, references to the "Settlement" or "Settlement Agreement," unless otherwise specified, refer to what the two documents have in common rather than what distinguishes them. When I make specific references to sections of the Settlement, I reference the Amended Settlement Agreement. *See* Settlement Agreement, Authors Guild, Inc. v. Google, Inc., No. 05-CV-8136-JES (S.D.N.Y. Oct. 28, 2008), *available at* http://thepublicindex.org/documents/responses; Amended Settlement Agreement, Authors Guild, Inc. v. Google, Inc., No. 05-CV-8136-DC (S.D.N.Y. Nov. 13, 2009), *available at* http://thepublicindex.org/docs/amended_settlement/Amended-Settlement-Agreement.pdf.

This article proceeds in six parts. After this Introduction, Part II examines how two scholars have analyzed the "digital library" metaphor as a means of thinking about the contemporary destabilization of the book system. Part III provides a genealogy of the term "digital library" as it relates to notions of a "universal" library. I situate book digitization within the coming together of two specialist cultures: computer science and librarianship. Part IV provides a survey of some significant book digitization up to December 2004, when Google announced its Library Project. Part V turns specifically to Google Books both before and after the Settlement. I emphasize the switch in emphasis from "search" to "access," and indicate the need for greater attention to the deceptively simple notion of "access." Part VI concludes.

II. THE DIGITAL LIBRARY METAPHOR

In the 1990s, Peter Lyman and Mark Stefik, among others, began to examine the figurative language that emerged with the Internet. "Digital library" was prominent among these new expressions.[7] In his book *Internet Dreams*, Stefik, computer scientist and director of the Information Sciences and Technologies Laboratory at Palo Alto Research Center (PARC), interpreted the digital library metaphor as appealing to deep traditions in human history.[8] According to Stefik, it appeals to the need for community memory and cultural preservation, and draws upon the archetype of the "keeper of knowledge"—a figure in the human collective unconscious who reminds us to preserve knowledge for the future.[9] This archetype includes the wise elder, the storyteller, the bard, the curator, the scholar, and the librarian. He posits that libraries draw power from this cultural archetype and, as continuous "keepers of knowledge," they are given an "almost sacred trust" to preserve and disseminate knowledge.[10] Particular ethical values and practices develop within this sphere of trust—equal and open access, collective ownership of information, privacy, intellectual freedom, guidance and education in literacy and in evaluating information, bibliography, and librarianship.[11] In recognition of the value placed on

7. *See* Mark Stefik, Internet Dreams: Archetypes, Myths, and Metaphors (1996); Lyman, *What is a Digital Library?*, *supra* note 3. Other metaphors the authors identify are "e-mail," "information superhighway," "virtual community," "digital worlds," and "electronic marketplace." *Id.*

8. *See* Stefik, *supra* note 7, at xxii.

9. *Id.* at 4.

10. *Id.*

11. For further discussion on "library values" in the United States, see generally Rebecca Tushnet, *My Library: Copyright and the Role of Institutions in a Peer-to-Peer World*, 53 UCLA L. Rev. 977 (2006); Rory Litwin, *On Google's Monetization of Libraries*, Library Juice (Dec. 17, 2004), http://libr.org/juice/issues/vol7/LJ_7.26.html#3; American Library Association, *Library Bill of Rights* (1939), http://www.ala.org/ala/issuesadvocacy/intfreedom/librarybill/index.cfm. The *Library Bill of Rights* was amended in 1996. *Id.* For a comparative international perspective, see generally Wallace C. Koehler et al., *Ethical Values of Information and Library Professionals—An Expanded Analysis*, 32 Int'l Info. & Libr. Rev. 485 (2000).

libraries, most countries with copyright laws grant some sort of exceptions for library uses.[12]

Stefik's "archetypal" interpretation encourages us to see how libraries can function in a different temporal register than the other aspects of the book system. Library administrators concern themselves with past, present, and future. They consider their work both *in* time and *over* time in a way that, say, publishers do not. For instance, when a book goes out of print (stops selling enough to justify reprinting), a publisher does not quite forget about it, but certainly stops paying attention to it. And books do not last because paper is such a great medium for preservation—though it is that—but because institutions like libraries exist, in part, to preserve them and to keep them accessible to readers. And yet, despite the distinctiveness of libraries as institutions, Stefik understands "digital library" as conveying a newly dynamic book system.[13] It opens the entire book system to question: publishers, authors, editors, booksellers, printers, and librarians. In other words, there is no reason the *librarian* need persist as the "keeper of knowledge." For Stefik, fundamental to the digital library metaphor is the sense that roles long stabilized in the traditional system are unstable, fluid, subject to rearrangements, and up for grabs.[14]

For political scientist, information specialist, and former librarian Peter Lyman, the digital library metaphor appeals to and draws not upon a timeless archetype, but upon a historically specific institution—the public library system—one of the "most successful twentieth-century public sector innovations."[15] The innovation of the public library was that it "managed" the boundary between private property and public goods, buying information in a marketplace and subsidizing its free use by communities. For this reason, Lyman refers to the library as a boundary management institution.[16] In the 1990s, Lyman explains, "digital library" functioned as a conceptual tool for questioning and analyzing the relation between the Internet's growing digital networks, and social, economic, and public policy.[17] "Digital library," in particular, mediated between the "information superhighway," which Lyman interprets as the idea that information has economic value and must be treated as property, and "virtual community," which connotes that information is "only valuable

12. *See* World Intellectual Property Organization (WIPO), *Study on Copyright Limitations and Exceptions for Libraries and Archives*, WIPO Doc. SCCR/17/2 (Aug. 26, 2008), *available at* http://www.wipo.int/edocs/mdocs/copyright/en/sccr_17/sccr_17_2.pdf.

13. Stefik, *supra* note 7.

14. *Id.*

15. Lyman, *Information Superhighways*, *supra* note 3, at 201.

16. *See id.* at 211–14 (discussing the digital library's role as "manag[ing] the boundary between private property and public goods"); *see also* Peter Lyman, *The Poetics of the Future: Information Highways, Digital Libraries and Virtual Communities*, FreeeBay (June 30. 2009), http://www.freeebay.net/site/index.php?option=com_content&task=view&id=369&Itemid=99999999 [hereinafter Lyman, *Poetics of the Future*]. For example, the American Library Association's motto is: "The best reading, for the largest number, at the least cost." Am. Libr. Ass'n, http://www.ala.org/ala/aboutala/missionhistory/mission/index.cfm (last visited Sept. 20, 2010).

17. Lyman, *Information Superhighways*, *supra* note 3, at 205–06.

when it is learned and used by human communities."[18] In other words, we have two extremes—on one side, all information use is metered and, on the other, information wants, proverbially, to be free. Lyman saw "digital library" mediating between these extremes.[19] So, the digital library metaphor carries within it general but very important questions: Will new public institutions emerge to govern and structure the sharing and circulation of information in and through the Internet and, if so, how will they mediate between the market and free use? Moreover, how will they manage the boundary between private property and public goods?

Stefik's and Lyman's interpretations help us frame the mass-digitization of books and, in particular, Google Books and the responses to it. Stefik's archetypal approach focuses on the fundamental and enduring. However new institutions form, he seems to suggest, there will always be a "keeper of knowledge" somewhere, somehow, and, in a time of change and institutional instability, we can feel reassured by that certainty. Lyman, in contrast, does not believe that we can take anything for granted. He focuses instead on the specific, historical circumstances in which we find ourselves. To him, there is nothing inevitable about human society. Libraries, as we know them, have not existed forever. They are not archetypal; in fact, they are of relatively recent provenance. At the turn of this century, Lyman questioned whether the political climate existed in which "digital" libraries—libraries of digital objects mediated through computer networks—would have the political and societal support to emerge as powerful public institutions. This is because national information policy has favored market mechanisms over the public sector.[20] "Is there a public interest," he asked, "in providing subsidized access to a national digital library collection, or to a medium that enables participation in the creation of knowledge as well as the consumption of commercial information commodities?"[21]

It is perhaps unsurprising that the perspective of the computer scientist (Stefik) encourages enthusiasm, while that of the social scientist (Lyman) encourages skepticism. I begin with these two perspectives, however, because they represent, if perhaps too schematically, the ongoing, and often conflicting, intermingling of the traditional book system with its new technological bedfellows. This is the very space I find opened up for analysis by the digital library metaphor.

III. FROM UNIVERSAL TO DIGITAL LIBRARY

Aspirations of a universal library—the utopian dream of gathering of all human knowledge and, especially, all the books ever written in one place—can be found throughout Western history.[22] Most discussions of a "universal library" begin with

18. Lyman, *Poetics of the Future*, *supra* note 16.

19. Lyman, *Information Superhighways*, *supra* note 3, at 211.

20. Lyman, *What Is a Digital Library?*, *supra* note 3, at 3.

21. *Id.* at 24.

22. *See* Roger Chartier, *Libraries Without Walls*, Representations, Spring 1993, at 38, 42 (discussing the physical as well as "virtual" notions of the ideal universal library during the Renaissance, such as Conrad

reference to an actual library—the Library of Alexandria—which serves as a sort of prototype.[23] Founded around 300 BC, the ancient library held hundreds of thousands of scrolls at its peak.[24] The Ptolemies (Alexander's successors in Egypt) had universalistic ambitions in their collecting practices. Ptolemy III reportedly requested books from rulers from around the ancient world, promising to copy and return them.[25] Books and manuscripts were also confiscated from ships that docked in Alexandria's harbors for mandatory copying. It is reported, however, that in at least some cases the originals were kept and the copies returned.[26] The Library of Alexandria eventually ceased to be and most of its collections were destroyed or lost. Although some have contended that the destruction happened at the hands of a political opponent (either the Romans, Christians, or Arabs) it is more likely that it slowly declined through a combination of cataclysms, the passage of time, and the declining status of the city of Alexandria.[27] In their imperial ambitions, the Ptolemies sought to "create[] an institution whose history and influence would reach outward in space and endure in time,"[28] and from today's perspective, they appear to have been spectacularly successful in that regard.

The notion of gathering all knowledge together in one place persisted in Europe, but after printing spread throughout the continent, it became clear that no physical structure could ever house all books. Still, aspirations of universality did not lose their appeal and alternatives were developed. One such strategy was to telescope all knowledge into finite, manageable (printed) forms, giving rise to the Enlightenment's most representative form of knowledge: the encyclopedia.[29] Bibliographical classification schemes were another attempt at universality.[30] Through standardization and normalization, classification systems were a way to govern the population of all books, even if no individual library could manage to hold them all. At the British Museum, Anthony Panizzi developed the first universal classification system in the

Gesner's *Bibliotecha Universalis*, and Etienne-Louis Boulee's 1785 design for France's royal library); *see also* Matthew Battles, Library: An Unquiet History (2003).

23. Michael H. Harris, History of Libraries in the Western World 47 (4th ed. 1999).

24. The number varies depending on the source and how one counts. *See* Luciano Canfora, The Vanished Library: A Wonder of the Ancient World 189 (Univ. of California Press 1990) (1989); The Library of Alexandria: Centre of Learning in the Ancient World (Roy MacLeod ed., 2000).

25. But there is evidence that the originals stayed in Alexandria. The Library of Alexandria, *supra* note 24, at 4.

26. Roy MacLeod has written that "[t]he Library had little regard for intellectual property—or even for property rights *per se*." The Library of Alexandria, *supra* note 24, at 4. Historian Anthony Grafton has likened Google's permission-less scanning to the confiscatory practices in Alexandria. *See* Anthony Grafton, Worlds Made by Words: Scholarship and Community in the Modern West 293 (2009).

27. The Library of Alexandria, *supra* note 24, at 9–10.

28. *Id.* at 1.

29. *See* Peter Burke, A Social History of Knowledge: From Gutenberg to Diderot 11 (2000).

30. *See* Alex Wright, Glut: Mastering Information Through the Ages 52 (2007).

1830s.[31] Throughout the nineteenth century, various cataloging innovators (from Charles Cutter to Melvil Dewey) followed Panizzi's lead with their own innovations.[32] The standards and practices of the "industrial library" developed in that era still characterize contemporary library catalogs, even though card catalogs are now computerized.[33]

The twentieth century witnessed a revival of the universal library as both a space of extravagant speculation and a practical task. In Kurd Lasswitz's 1901 short story *The Universal Library*, the main character imagines the size of a library made from a simplified language of twenty-five symbols whose recombinations could nonetheless capture everything that could ever be expressed in any language.[34] Such a library would be so extensive as to contain any text that had ever been lost, and every text that had yet to be published. It would also contain just as much nonsense; be larger than the known universe; and prove unusable. In his 1939 essay *The Total Library,* Jorge Luis Borges considered Lasswitz's fancy a "subaltern horror: the vast, contradictory Library, whose vertical wildernesses of books run the incessant risk of changing into others that affirm, deny, and confuse everything like a delirious god."[35] A few years later, in his now famous short story *The Library of Babel,* Borges revisited the same themes.[36] The narrator, after his initial "extravagant happiness" at having all books in one place, comes to find the library to be a nightmare where having access to everything becomes worse than having access to nothing. It is a sort of torture, so disconcerting that the narrator nearly becomes suicidal. Nonetheless, he finds consolation in the possibility of a messianic librarian one day appearing who will find order in the disorder of the library.[37]

Before the digital computer or the Internet, microphotography encouraged ambitious new schemes for managing large collections of books. Photography and other reproduction processes spurred early imaginings of miniaturized libraries being made available at one's fingertips. Anticipating the current information technology discourse, M. Llewellyn Raney, the director of libraries at the University of Chicago, wrote in 1936 that "the application of the camera to the production of literature ranks next to that of the printing press."[38] Around 1934, Paul Otlet, Belgian entrepreneur and early "information scientist," speculated about a system, springing from the

31. *See id.* at 167–70.

32. *See id.* at 171–77.

33. *See id.* at 176–78.

34. Kurd Lasswitz, *The Universal Library*, *in* Fantasia Mathematica 237–43 (Clifton Fadiman ed., 1997).

35. Jorge Luis Borges, *The Total Library*, *in* Selected Non-fictions 216 (Eliot Weinberger ed., Esther Allen et al. trans., 1999).

36. *See* Jorge Luis Borges, *The Library of Babel*, *in* Collected Fictions 112 (Andrew Hurley trans., 1998).

37. *See id.* at 116.

38. Nicholson Baker, Double Fold: Libraries and the Assault on Paper 73 (2001) (quoting Microphotography for Libraries, at v, (M. Llewellyn Raney ed., 1936)).

combination of then new technologies (radio, microfilm, and television) and a massive 3" x 5" notecard system. Otlet's imagined system would make large amounts of documents (film, sound recording, radio, television, and documents, including pages from books) available to a reader for querying through a vast mechanical database and a system of screens.[39] This "radiated library" would allow the researcher to connect documents with other documents, "forming from them what might be called the Universal Book."[40] Because of these characteristics, and in addition to the fact that Otlet called his system a "web" and the connections created by users "links," some have found his speculations most prescient of the hyperlink-based World Wide Web.[41] The significant advantage of his system was that it enabled the reader not only to discover or amass books but to gain access to their textual contents.

Much better known is another imaginary, never-to-be-built contraption: Vannevar Bush's Memex. Bush drafted his famous essay *As We May Think* in 1939, the same year that Borges wrote *The Total Library*.[42] In it, Bush described an imaginary information retrieval device for personal use, comprised of microfilmed document collections and a projection screen, which would gather the "associative trails" left by a scholar's use.[43] Both Otlet and Bush's imagined contraptions were analog, dependent on microfilm, "dry photography" (xerography), and film, as well as human-performed classification. But it was Bush's famous essay, in particular, that inspired the development of automated information retrieval using digital computers. From the 1960s through the 1980s, both image processing and computer technologies (such as memory, storage, processing, software, and linear text searching) developed sufficiently so that, by the 1990s, computer scientists and engineers confidently held the conviction that the requisite technology was in place to retrospectively convert all analog information into digital form.[44]

---

39. *See* Wright, *supra* note 30, at 185–86.

40. DVD: The Man Who Wanted to Classify the World: From the Index Card to the World City, The Visionary Life of a Belgian Utopian, Paul Otlet (1868–1940) (Sofidoc Production 2002); Wright, *supra* note 30, at 186; Paul Otlet, *Transformations in the Bibliographic Apparatus of the Sciences*, *in* International Organisation and Dissemination of Knowledge: Selected Essays of Paul Otlet 148, 154 (W. Boyd Rayward ed. & trans., 1990). A good Otlet bibliography has been compiled by Michael Buckland and is available at: *Selected Writings about Paul Otlet*, U.C. Berkeley School of Information, http://people.ischool.berkeley.edu/~buckland/otletbib.html (last visited Nov. 9, 2010).

41. Wright, *supra* note 30, at 186; *see* W. Boyd Rayward, *The Case of Paul Otlet, Pioneer of Information Science, Internationalist, Visionary: Reflections on Biography*, 23 J. Libr. & Info. Sci. 135 (1991); Alex Wright, *The Web Time Forgot*, N.Y. Times, June 17, 2008, http://www.nytimes.com/2008/06/17/science/17mund.html.

42. *See* Vannevar Bush, *As We May Think*, Atlantic Monthly, July 1945, at 101, *available at* http://www.theatlantic.com/magazine/archive/1969/12/as-we-may-think/3881/.

43. *Id.* at 106–08.

44. *See generally* Michael Lesk, Practical Digital Libraries: Books, Bytes & Bucks (1997) (discussing the sense of technological readiness); Raj Reddy, *The Universal Library: Intelligent Agents and Information on Demand*, *in* Digital Libraries: Research and Technology Advances 27 (Nabil R. Adam et al. eds., 1996).

It was around this time that the term "digital library" first appeared. Mark Stefik claims its first use was in the late 1980s, around the time Robert Kahn and Vinton Cerf published their report, *An Open Architecture for a Digital Library System*.[45] Written as the Internet was being opened for broad use beyond universities, the report envisioned a distributed network of personal, public, commercial, and national "digital libraries" sharing common standards and methodologies.[46] In late 1993, three large federal funding agencies, the National Science Foundation (NSF), the Advance Research Projects Agency (ARPA), and the National Aeronautics and Space Administration (NASA), together launched the influential Digital Library Initiative (DLI) to fund research on digital libraries.[47] It was successful enough that its Phase II began in 1998.[48] In 1995 alone, there was an explosion of movement to solidify an institutional structure for thinking about and developing "digital libraries"—whatever that might come to mean. It witnessed the launch of the Digital Library Federation,[49] *D-Lib Magazine*,[50] the Making of America project,[51] JSTOR,[52] and the Library of Congress's National Digital Library Program (also known as the American Memory Project).[53] The Internet Archive, an all-digital public library, was established in

45. *Compare* Stefik, *supra* note 7, at 33 (using the term "digital library" in preference to other terms then commonly in use, such as "electronic" or "virtual" library), *with* Robert E. Kahn & Vinton G. Cerf, *An Open Architecture for a Digital Library System and a Plan for its Development*, Corp. for Nat'l Res. Initiatives (Mar. 1998), http://www.cnri.reston.va.us/kahn-cerf-88.pdf.

46. Kahn & Cerf, *supra* note 45, at 3.

47. Stephen M. Griffin, *NSF/DARPA/NASA Digital Libraries Initiative: A Program Manager's Perspective*, Mag. of Digital Libr. Res., July/Aug. 1998, *available at* http://www.dlib.org/dlib/july98/07griffin.html.

48. Google Co-Founder Larry Page developed the Web page-ranking prototype, BackRub, as a graduate student working on the Stanford Integrated Digital Library Project, a DLI-funded project. *See* David Hart, *On the Origins of Google*, Nat'l Sci. Found. (Aug. 17, 2004), http://nsf.gov/discoveries/disc_summ.jsp?cntn_id=100660&org=CISE. That project was intended to develop the "enabling technologies for a single, integrated and 'universal' library, providing uniform access to the large number of emerging networked information sources and collections." *Award abstract for the Stanford Integrated Digital Library Project*, National Science Foundation, http://www.nsf.gov/awardsearch/showAward.do?AwardNumber=9411306 (last visited Oct. 11, 2010).

49. A project of the Council on Library and Information Resources founded by sixteen institutions that sought to develop an international network of digital libraries. *See Digital Library Foundation*, Council on Libr. & Informative Resources, http://www.clir.org/dlf.html (last visited Sept. 25, 2010).

50. "D-lib" is short for "digital library." The online magazine boasts that its archive "provides much of the core literature on digital libraries." *D-Lib Alliance Participants*, D-Lib Mag., http://www.dlib.org/ dlib/alliance-participants.html (last visited Oct. 11, 2010).

51. A project of the Cornell University and University of Michigan libraries to digitize and make available nineteenth-century American books from their collections. *See Making of America*, Univ. of Mich., http://quod.lib.umich.edu/m/moagrp/ (last visited Oct. 11, 2010); *Making of America*, Cornell Univ. Libr., http://digital.library.cornell.edu/m/moa/ (last visited Oct. 11, 2010).

52. JSTOR is an acronym for "journal storage." Roger C. Schonfeld, JSTOR: A History 22 (2003). JSTOR was sponsored by the Mellon Foundation, and was largely developed at the University of Michigan, to retrospectively convert academic journals from print to digital format. *See id*.

53. The National Digital Library Program began in 1995, and is digitizing archival material from the Library of Congress to facilitate the study of American culture and history. *National Digital Library*

1996.[54] In January 1998, President Clinton, in his State of the Union address, called for a national digital library, envisioning "an America where every child can stretch a hand across a keyboard and reach every book ever written."[55] That same year, Google was incorporated.[56]

Seventy years after Vannevar Bush first imagined his Memex and Jorge Luis Borges deconstructed the "total library," those with a computer and a decent Internet connection have access to a staggering amount of information, which is growing exponentially. Prominent technologists and others see the Internet as the library of libraries.[57] Take, for instance, Brewster Kahle's remark: "I think of the Internet . . . in general, as the Library. So I'm trying to build the library. I'm not trying to build it in a little building. It's got to be a big, distributed joint thing . . . ."[58] Or Google's Dan Clancy's statement: "[T]he Library of Alexandria is this thing we call the Internet . . . we are seeing it being created today, and it will be created as a distributed entity . . . . Google Book Search . . . is a repository of some content that contributes to this broader initiative."[59] The notion that the Internet is a giant library leads some to predict the demise of libraries as we know them.

Librarians, however, insist that the Internet is not a library and can never be a library. For example, three Stanford University librarians—all deeply involved in the management of electronic resources—published an essay that posed the question *What Is a Library Anymore, Anyway?* Their answer was that libraries are collections made useful for particular communities and that "all collections are local."[60] Christine Borgman has identified "competing visions" among the groups building "digital"

Program, Library of Congress, http://memory.loc.gov/ammem/dli2/html/lcndlp.html (last visited Sept. 30, 2010); *see also* James H. Billington, *The Library of Congress's National Digital Library Program*, *in* Digital Libraries: Research and Technology Advances 17 (Nabil R. Adam et al. eds., 1996).

54. *About the Internet Archive*, Internet Archive, http://www.archive.org/about/about.php (last visited Oct. 11, 2010).

55. President William Clinton, State of the Union Address (Jan. 27, 1998) (transcript available at http://www.washingtonpost.com/wp-srv/politics/special/states/docs/sou98.htm).

56. *Google Milestones*, Google, http://www.google.com/corporate/history.html (last visited Oct. 11, 2010).

57. *Compare* George Steiner, Grammars of Creation 300 (2001) (a melancholic meditation on culture dominated by science and technology), *with* Kelly, *supra* note 5 (a paean to digitization).

58. Interview with Brewster Kahle, Co-Founder, Internet Archive (May 1, 2009).

59. Dan Clancy, Engineering Director, Google, Inc., Presentation at the 2010 American Council of Learned Societies' Annual Meeting (May 7, 2010), *available at* http://www.acls.org/publications/audio/annual_meetings/2010/default2.aspx?id=5584. Clancy is no longer the Engineering Director of Google Books. In July 2010, he moved to Google's YouTube division. *Google's Dan Clancy to Leave Books Program for YouTube*, Publishers Weekly, Jul. 23, 2010, http://blogs.publishersweekly.com/blogs/PWxyz/?p=650.

60. Michael A. Keller, Victoria A. Reich & Andrew C. Herkovic, *What is a Library Anymore, Anyway?*, First Monday (May 5, 2003), http://www.firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1053/973; *see also* Jose-Marie Griffiths, *Why the Web is Not a Library*, *in* The Mirage of Continuity: Reconfiguring Academic Information Resources for the 21st Century 229 (Brian L. Hawkins & Patricia Battin eds., 1998).

libraries.[61] The computer scientists have a narrower perspective: they seek to identify information retrieval problems and to improve access to resources in digital form.[62] The librarians, in contrast, are concerned with institutional obligations and services to research communities, which include long-term responsibilities for preservation, access, and sustained information management.[63] The computer engineer's "library" is organized content; the librarian's "library" is an institution. These different understandings of what a library is and the differing commitments it gives rise to—between technologists who approach a library as information to be managed and librarians who approach a library as "the lifeblood and substance of scholarly inquiry and endeavor"[64]—represent another facet of the collision embodied in "digital library," and serve to bring the soaring aspirations of contemporary Alexandrians closer to Earth. The incorporation of books into the networked environment of the Internet has proven to require not only technological innovations, but political, legal, and cultural transformations as well.

IV. A BRIEF HISTORY OF MASS BOOK DIGITIZATION

For Internet enthusiasts like *Wired* magazine founding editor Kevin Kelly, post-World War II computer science redeemed the "subaltern horror" of the universal library. In 1995, he wrote:

> Two decades ago nonlibrarians discovered Borges's Library in silicon circuits of human manufacture. The poetic can imagine the countless rows of hexagons and hallways stacked up in the Library corresponding to the incomprehensible microlabyrinth of crystalline wires and gates stamped into a silicon computer chip. A computer chip, blessed by the proper incantation of software, creates Borges's Library on command. . . . Pages from the books appear on the screen one after another without delay. To search Borges's Library of all possible books, past, present, and future, one needs only to sit down (the modern solution) and click the mouse.[65]

Search technology—to Kelly, a "wholly new concept"[66]—would serve as Borges's librarian-messiah to find order in the chaos of everything.[67] Kelly, however, had exaggerated the state of play in 1995 in terms of the availability of books. Though book digitization was increasing, there were very few books online at that time. In 1993, John Ockerbloom started his "Online Books Page" website to gather all freely

---

61. *See* Christine L. Borgman, *What Are Digital Libraries? Competing Visions*, 35 Info. Processing & Mgmt. 227 (1999).

62. *Id.* at 231.

63. *Id.* Although both groups share concern with "access," the term has different meanings for each group.

64. Keller et al., *supra* note 60.

65. Kevin Kelly, Out of Control: The New Biology of Machines, Social Systems, and the Economic World 263 (1995).

66. Kelly, *supra* note 5.

67. Borges, *supra* note 36, at 116.

available books on the Internet in one place.[68] Project Gutenberg, begun by Michael Hart in 1971, in particular, was one source of e-books in the early days of the Internet and continues to be.[69] Although widely practiced in the 1990s, book digitization—whether by direct keyboarding or by photographic scanning—was happening on a small scale by enthusiastic volunteers or on a prototype, experimental scale for organizational learning in research centers, libraries, humanities departments, and elsewhere.[70]

Technologists, too, were eager to have digital "content" with which to pursue computer science research into machine translation, storage formats, information retrieval, security, search, image processing, and so forth. But they hit regular roadblocks from publishers of all sorts, who feared loss of revenue, a threat to their business models, piracy, and were generally distrustful of having entities other than themselves in possession of digital copies of their books. The result was that very few books were available on the Web.

In 1995, with funding from the Andrew W. Mellon Foundation, the libraries of Cornell University and the University of Michigan—both early leaders in book and journal digitization for preservation and for access—initiated the Making of America project, "to involve research institutions and national consortia to develop common protocols and consensus for the selection, conversion, storage, retrieval, and use of digitized materials on a large, distributed scale."[71] Eventually the project would

68. The Online Books Page, http://onlinebooks.library.upenn.edu/ (last visited Oct. 11, 2010). In December 1998 it listed only 8000 books. The On-line Books Page, http://web.archive.org/web/19981205144103/www.cs.cmu.edu/books.html (last visited Oct. 11, 2010).

69. Project Gutenberg, http://www.gutenberg.org (last visited Sept. 17, 2010). Developed from the volunteer labor of contributors who scanned or typed the full text of public domain books, the Project Gutenberg website now offers 33,000 freely downloadable digitized books. *Id.* Although the number is small when compared to Google's millions, its books, which have corrected OCR and HTML coding (with the help of sister project Distributed Proofreaders), have seeded many Web e-book startups with many classic titles unencumbered by copyright. *See id.;* Distributed Proofreaders, http://www.pgdp.net/c/ (last visited Sept. 17, 2010). Project Gutenberg is arguably the first "mass" book digitization project.

70. Of course library digitization did not come out of nowhere. Microphotography efforts in libraries from the 1940s onward prefigure current large-scale book digitization. Both create surrogates through an imaging process. Like digitized books, microforms made it possible for scholars to access the works of a distant library at their home library rather than traveling, often to a foreign country. However, as has been well documented, microfilming had numerous problems and digital reformatting is now preferred. *See* Stephen G. Nichols & Abby Smith Nichols, The Evidence in Hand: Report of the Task Force on the Artifact in Library Collections (2001). Two successful and enduring book microfilm projects are ProQuest's Early English Books Online (EEBO) and Thomson Gale's Eighteenth Century Collections Online (ECCO). They are curated databases of 100,000 and 155,000 volumes, respectively, sold as subscriptions to libraries (despite the fact that the databases are comprised of public domain texts). *See* Early English Books Online, http://eebo.chadwyck.com/marketing/about.htm (last visited Sept. 23, 2010); Eighteenth Century Collections Online, http://gale.cengage.co.uk/product-highlights/history/eighteenth-century-collections-online.aspx (last visited Sept. 23, 2010). Nicholson Baker believes digitization is repeating microfilm's failures, which he attributes to technology-obsessed, short-sighted, space-saving library managers. *See* Baker, *supra* note 38.

71. *Making of America Project*, *About the Project*, Cornell Univ. Library, http://dlxs2.library.cornell.edu/m/moa/about.html (last visited Oct. 11, 2010).

digitize 10,000 books and 50,000 journal articles from the nineteenth century.[72] The University of Michigan was digitizing somewhere between 5000 and 8000 books a year before its partnership with Google.[73] The Making of America and other digitization projects represented the growing expertise around digitization, as well as the strong interest and support it was receiving from private foundations and universities. Google stood to benefit from this smaller scale digitization.

In 1999, the NSF's Division of Information and Intelligent Systems approved funding for part of the Million Book Project, a large-scale book digitization project proposed by artificial intelligence expert Raj Reddy and colleagues at Carnegie Mellon University.[74] The Million Book Project was a proposition to get books onto the Internet for "free-to-read" global distribution.[75] Supported by the Indian, Chinese, and Egyptian governments, the project had an express interest in alleviating the global digital divide.[76] Organizers of the project sought to solve the problem of digitization's high labor costs by having scanning done overseas, in India and China.[77] The NSF funding paid for scanning equipment, and the books would be borrowed from libraries in the United States and abroad, or donated.[78] Unlike Project Gutenberg, the Million Book Project did not limit itself to public domain works. Rather, it sought permission to scan in-copyright books and make them freely available.[79] After digitizing well over a million volumes, the project concluded in 2007. Its leaders consider it to have successfully accomplished a "proof of concept" for

---

72. *Making of America Project*, Univ. of Mich., http://quod.lib.umich.edu/m/moagrp/ (last visited Oct. 11, 2010).

73. Coleman, *supra* note 5, at 3.

74. For more information about the Million Book Project, see *Frequently Asked Questions About the Million Book Project*, Carnegie Mellon Libraries, http://eps.library.cmu.edu/rooms/documents/libraries-and-collections/Libraries/MBP_FAQ.html (last visited Oct. 11, 2010); *see also* Reddy, *supra* note 44.

75. *See Frequently Asked Questions*, *supra* note 74.

76. *About Us*, The Univ. Digital Library, http://www.ulib.org/ULIBAboutUs.htm (last visited Oct. 11, 2010).

> The mission [of the Million Book Project] is to create a Universal Library which will foster creativity and free access to all human knowledge. As a first step in realizing this mission, it is proposed to create the Universal Library with a free-to-read, searchable collection of one million books, available to everyone over the Internet. Within 10 years, it is our expectation that the collection will grow to 10 million books. The result will be a unique resource accessible to anyone in the world 24x7, without regard to nationality or socioeconomic background.

Id.

77. *See* John Markoff, *The Evelyn Wood of Digitized Book Scanners*, N.Y. Times, May 12, 2003, http://www.nytimes.com/2003/05/12/technology/12TURN.html?pagewanted=all.

78. The Internet Archive donated 100,000 books bought from the Kansas City Public Library for scanning in India. Interview with Brewster Kahle, Co-Founder, Internet Archive (April 16, 2009).

79. Denise Troll Covey, Acquiring Copyright Permission to Digitize and Provide Open Access to Books 38–39 (2005), http://www.clir.org/pubs/reports/pub134/pub134col.pdf.

large-scale digitization projects both in the United States and abroad.[80] Google acknowledges these projects on its Google Book Search website.[81]

At the turn of the twenty-first century, large-scale or "mass" book digitization had momentum. Commercial startups such as NetLibrary, Questia, and ebrary were working with publishers to digitize their books, and were offering them to libraries as databases.[82] They continue to offer fee-based services to libraries and other organizations, providing full access to both in-copyright and public domain books.[83] Amazon had started its own mass-digitization project in 2003—its "Search Inside the Book" service. The company made agreements with publishers to digitize their in-print books, allowing potential buyers to search through the text of the book on the Amazon website.[84]

With the possibility of even larger-scale digitization "in the air," Michael Lesk, informational retrieval expert and former chief research scientist at Bellcore, speculated as to why entire libraries had not yet been digitized despite the fact that all the technology was available and in place: "The biggest reason is that we cannot easily find $3 billion to fund the mechanical conversion of 100 million books to electronic form, plus the additional and probably larger sum to compensate the copyright owners for most of those books."[85]

Internet Archive founder Brewster Kahle, a tireless advocate for digitization of all media types, but especially books, tried to convince libraries to squeeze the money

---

80. Large-scale digitization has gathered significant governmental support in China, Japan, Holland, India, France, and Egypt. *See* Tom Imerito, *Electrifying Knowledge*, Pittsburgh Q., Summer 2009, at 96. However, surveying these projects is beyond the scope of this article. In December 2009, the French government pledged almost $1.1 billion for digitization (not only of books) as part of a national stimulus package. The partnership could include Google, with which the French government has been in talks. Scott Sayare, *France to Digitize Its Own Literary Works*, N.Y. Times, Dec. 15, 2009, http://www.nytimes.com/2009/12/15/world/europe/15france.html.

81. *About Google Books*, Google Books, http://books.google.com/googlebooks/history.html (last visited Oct. 26, 2010).

82. *See* Lisa Guernsey, *In Lean Times, E-books Find a Friend: Libraries*, N.Y. Times, Feb. 21, 2002, http://www.nytimes.com/2002/02/21/technology/circuits/21BOOK.html.

83. *See* Ebrary, http://www.ebrary.com/corp/ (last visited Sept. 21, 2010); NetLibrary, http://www.netlibrary.com/Default.aspx (last visited Sept. 21, 2010); Questia, http://www.questia.com/aboutQuestia/exploreLibrary.html (last visited Sept. 21, 2010). Questia's website states that it is "The World's Largest Online Library of Books" with three million titles in its database. *Id*.

84. *See* David D. Kirkpatrick, *Amazon Plan Would Allow Searching Texts of Many Books*, N.Y. Times, July 21, 2003, http://www.nytimes.com/2003/07/21/technology/21AMAZ.html; *see also* Lisa Guernsey, *In Amazon's Text Search, a Field Day for Book Browsers*, N.Y. Times, Nov. 6, 2003, http://www.nytimes.com/2003/11/06/technology/circuits/06amaz.html. As of September 2009, Amazon had digitized three million books for this purpose. Ryan Singel, *Amazon Scoffs at Google's Offer to Share Book Search Sales*, Wired (Sept. 10, 2009), http://www.wired.com/epicenter/2009/09/amazon-google-book-search-sales/.

85. Lesk, *supra* note 44, at 3. The research and development consortium Bell Communications Research (Bellcore) served the regional Bell companies, now known as Telecordia, and was established after the breakup of AT&T in 1982. *Telcordia Technologies, Inc. Company History*, Funding Universe, http://www.fundinguniverse.com/company-histories/Telcordia-Technologies-Inc-Company-History.html (last visited Sept. 24, 2010).

for digitization out of their acquisition budgets.[86] In April 2004, as he received the Paul Evan Peters Award,[87] Kahle announced that the Internet Archive, after years of experimentation as well as involvement with the Million Book Project, could digitize books affordably, and encouraged libraries to allow him access to their collections.[88] Only one librarian, Carole Moore at the University of Toronto, accepted Kahle's offer, and the Internet Archive began digitizing public domain books there in the summer of 2004. In early December of that year, the first fruits of that collaboration were made public.[89]

V. GOOGLE BOOK SEARCH PROJECT

A. Pre-Settlement

Google executives were among those deeply involved in tackling the various obstacles to digitization—financial, technological, and legal. As early as 2002, they were secretly involved in discussions with the library leadership at the University of Michigan, where Larry Page had been an undergraduate student, to digitize its library holdings.[90] They were also involved at an early stage with Stanford University, in a secret project known as Project Ocean, to digitize its entire library.[91]

What is now known as Google Books covers two different digitization programs: one, the "Partner Program," receives books from publishers; the other, the "Library

---

86. *See* Brewster Kahle, *Public Access to Digital Materials*, Library of Congress (Nov. 20, 2002), http://www.loc.gov/rr/program/lectures/kahle.html; *see also*, Stuart Feldman, *A Conversation with Brewster Kahle: Creating a Library of Alexandria for the Digital Age*, Queue, June 2004, at 24, *available at* http://queue.acm.org/detail.cfm?id=1016993.

87. Three nonprofit organizations (the Coalition for Networked Information, the Association of Research Libraries, and EDUCAUSE) sponsor the award to recognize "the most notable and lasting international achievements related to high performance networks and the creation and use of information resources and services that advance scholarship and intellectual productivity." *Paul Evan Peters Award*, EDUCAUSE, http://www.educause.edu/Professional+Development/EDUCAUSEAwards/PaulEvanPetersAward/852 (last visited Oct. 11, 2010).

88. Interview with Brewster Kahle, Co-Founder, Internet Archive (Apr. 16, 2009). He originally ventured he could do it for $10 a book, but the lowest figure became $.10/per page or $30 for a 300-page book. *Id.*

89. *See* Online Forum Posting About a New Internet Archive Collection, Internet Archive (Dec. 9, 2004, 1:38 AM), http://www.archive.org/post/25165/new-book-collection-color-scans-djvu-some-pdf. The University of Toronto has since digitized 250,000 public domain works through the Internet Archive Canada. *Canadian Libraries,* Internet Archive, http://www.archive.org/details/toronto (last visited Oct. 11, 2010).

90. *See* Scott Carlson & Jeffrey R. Young, *Google Will Digitize and Search Millions of Books from Five Top Research Libraries*, Chron. of Higher Educ., Jan. 7, 2005, at 37; *see also* Daniel Greenstein & Suzanne E. Thorin, The Digital Library: A Biography 52–57 (2002), www.clir.org/pubs/reports/pub109/pub109.pdf (discussing the history of the University of Michigan in the context of digitization of library materials); Schonfeld, *supra* note 52.

91. John Markoff, *The Coming Search Wars*, N.Y. Times, Feb. 1, 2004, § 3, at 1; *see also* Barbara Palmer, *Deals with Google to Accelerate Library Digitization Projects for Stanford, Others*, Stanford Report (Jan. 12, 2005), http://news.stanford.edu/news/2005/january12/google-0112.html.

Project," receives books from research libraries.[92] Google beta-tested the Partner Program, similar to Amazon's "Search Inside" program, in the fall of 2003 and then officially launched it in October 2004 at the Frankfurt Book Fair.[93] Today, the Partner Program has 30,000 publisher partners worldwide,[94] and as of June 2009, Google had scanned two million books received through the program.[95] The Partner Program has been relatively non-controversial because Google digitizes these books with publisher permission.[96] Publishers either provide Google with digital files or Google cuts the spine off the book, feeds its sheets through a sheet-fed scanner (so-called "destructive scanning"), and then processes the images through optical character recognition (OCR) software to index the book and make its text searchable.[97]

Just a couple of months after announcing the publisher program, Google announced what now makes up the other, larger part of Google Book Search: the Library Project.[98] As announced in December 2004, Google would partner with five major research libraries to digitize large numbers of their book holdings but, in the case of the University of Michigan, the plan was to digitize the *entirety* of its collection, regardless of copyright status.[99] Books would not be scanned "destructively" but would be borrowed from the libraries and scanned manually, with operators

92. *See Google Books Library Project*, Google Books, http://books.google.com/googlebooks/library.html (last visited Oct. 11, 2010); *Overview of Google's Books Partner Program*, Google Books, http://books.google.com/support/partner/bin/answer.py?answer=106167 (last visited Oct. 11, 2010). Google Print was renamed Google Book Search in November 2005. Jen Grant, *Judging Book Search by its Cover*, The Official Google Blog (Nov. 17, 2005, 2:49 AM), http://googleblog.blogspot.com/2005/11/judging-book-search-by-its-cover.html.

93. *See* John Markoff, *Google Experiment Provides Internet with Book Excerpts*, N.Y. Times, Dec. 18, 2003, at C6; Barbara Quint, *Google Print Expands Access to Books with Digitization Offer to All Publishers,* Information Today (Oct. 6, 2004), http://newsbreaks.infotoday.com/nbreader.asp?ArticleID=16357.

94. Jon Orwant, Engineering Manager, Google, Inc., Presentation at the Future of Reading Conference (June 11, 2010).

95. *See* Clancy, *supra* note 59.

96. However, publishers have tended to approach all digitization projects with considerable trepidation. *See* Edward Wyatt, *New Google Service May Strain Old Ties in Bookselling*, N.Y. Times, Oct. 4, 2004, http://www.nytimes.com/2004/10/08/technology/08book.html (stating how publishers have questioned digitization in regards to ensuring copyright protection, and voiced concern about their ability to distribute directly to consumers); Jim Milliot, *Publishers Grudgingly Cooperate with Amazon Database Effort*, Publishers Weekly, Sept. 15, 2003, at 10.

97. For more details about the processes of digitization, see Juliet Sutherland, *A Mass Digitization Primer*, 57 Libr. Trends 17 (2008).

98. Press Release, Google, Inc., Google Checks Out Library Books (Dec. 14, 2004), *available at* http://www.google.com/press/pressrel/print_library.html; *see also* Carlson & Young, *supra* note 90 (discussing Google's library project and issues involved, such as copyright issues, effects on libraries, and benefits to having increased information available via the Internet); John Markoff & Edward Wyatt, *Google is Adding Major Libraries to its Database*, N.Y. Times, Dec. 14, 2004, http://www.nytimes.com/2004/12/14/technology/14google.html (discussing the Google December 14, 2004 press release, the agreements with libraries, and other online library efforts).

99. Carlson & Young, *supra* note 90. The "Google Five" were the libraries at Oxford University, Harvard University, Stanford University, the University of Michigan, and the New York Public Library. *Id.*

turning pages as cameras photograph the pages from above.[100] Afterward, the books are returned to the libraries. The plan was to digitize all the world's books.[101] As Google's CEO Eric Schmidt wrote in the *Wall Street Journal* in 2005: "Imagine . . . all the world's books discoverable with just a few keystrokes by anyone, anywhere, anytime."[102]

The "library" in the name Library Project refers not to the project being a library, but to the provenance of the books. In its official statements, at least, Google officials have appeared reluctant to refer to Google Books as a "library"—digital, universal, or otherwise.[103] Rather, at its beginning, Google employed a different metaphor from the past to describe its project: the library card catalog. Its December 2004 press release announcing the Library Project described it as a "comprehensive, searchable, virtual card catalog."[104] Google's Eric Schmidt described it both as "one giant electronic card catalog" and as "one vast index . . . searchable . . . [and] free."[105] The Library Project website sums it up as "a comprehensive, searchable, virtual card catalog of all books in all languages that helps users discover new books and publishers discover new readers."[106]

Critics argued that Google was hiding behind the safety of something familiar and that its service was nothing like a card catalog, which does not let you search

---

100. *See* Palmer, *supra* note 91; Orwant, *supra* note 94.

101. What Google means by "all the world's books" is not clear. In June 2010, Jon Orwant estimated the total number of books in the world to be around 100 million and remarked that, at 12 million, Google was "over 10% there." Orwant, *supra* note 94. A recent post on the Google Book Search blog reported that there are 129,864,880 books in the world. Leonid Taycher, *Books of the World, Stand Up and Be Counted*, Inside Google Books (Aug. 5, 2010, 8:26 PM), http://booksearch.blogspot.com/2010/08/books-of-world-stand-up-and-be-counted.html. In July 2009, Dan Clancy said,

> I can't tell [how many books Google will scan] because I don't know exactly how many we'll put our hands on. It's probably 40–50 million or something like that. Figuring out when we've gotten the last one . . . will be a very tricky problem, and I don't know how to solve that. Luckily I've got a few years to work on that.

Interview by John C. Hollar with Dan Clancy, Engineering Director, Google, Inc., in Mountian View, Cal. (July 30, 2009), *available at* http://www.youtube.com/watch?v=Icus2RBRJ4s.

102. Eric Schmidt, Op-Ed., *Books of Revelation*, Wall St. J., Oct. 18, 2005, at A18.

103. *See* Sergey Brin, Op-Ed., *A Library to Last Forever*, N.Y. Times, Oct. 8, 2009, http://www.nytimes.com/2009/10/09/opinion/09brin.html?_r=1 [hereinafter Brin, *A Library to Last Forever*]. The title of Brin's op-ed might appear to contradict this assertion, but Brin would not have titled this piece himself. The *N.Y. Times*' editorial policy is to write all op-ed headlines. *See* David Shipley, Op-Ed., *And Now a Word from Op-Ed*, N.Y. Times, Feb. 1, 2004, http://www.nytimes.com/2004/02/01/opinion/and-now-a-word-from-op-ed.html ("[W]e write the headlines. Contributors have no say in these matters."). Brin's essay appears on the Official Google Blog. *See* Sergey Brin, *A Tale of 10,000,000 Books*, The Official Google Blog (Oct. 9, 2009, 9:30 AM), http://googleblog.blogspot.com/2009/10/tale-of-10000000-books.html [hereinafter Brin, *10,000,000 Books*].

104. *Google Books Library Project*, *supra* note 92.

105. Schmidt, *supra* note 102.

106. *Google Books Library Project*, *supra* note 92.

through the entirety of the books it indexed.[107] Because of such difficulties in describing an innovation, part of the work of metaphors from the past is to lubricate the introduction of the unfamiliar. Google's virtual card catalog combines Paul Otlet's vision of building a system that would take a reader not just *to* all the world's books, but *inside* them, with Fremont Rider's miniaturized library of "microcards," which contained bibliographical information on one side, and reduced images of every page on the other.[108] The result is what Google Books is today: a searchable database of millions of books on top of one's desk, with the hitch that you can *search* but not *read* the majority of the books. Nonetheless, the Library Project promised—and has delivered—a compelling new search and discovery tool for books.[109]

In exchange for allowing their books to be borrowed and scanned by Google, the original partner libraries received a digital copy of each book they provided.[110] This

107. *See* Allan R. Adler, Commentary, *Slouching Towards Alexandria: A Roundtable on Google's Library Project*, Bookforum, Feb./Mar. 2006, *available at* http://www.bookforum.com/archive/feb_06/roundtable.html; Nick Taylor, President, Authors Guild, Speech at The Battle Over Books: Authors & Publishers take on the Google Print Library Project Conference (Nov. 17, 2005), *available at* http://smartleydunn.com/wired/ (a discussion sponsored by the New York Public Library and *Wired* magazine regarding issues arising as a result of what was then known as the Google Print Library Project. Nick Taylor states: "If you want to create a library card catalogue why don't you just scan the card catalogues of the libraries rather than their entire contents?").

108. *Compare* Fremont Rider, The Scholar and the Future of the Research Library: A Problem and Its Solution (1944) (discussing Rider's "microcard" vision and the impact it would have on libraries), *with* Wright, *supra* note 30 (discussing Paul Otlet's vision for a "web" system enabling readers to gain access to the textual content of books).

109. The service has attracted many critics. Scholars and other commentators have reported on Google's error-ridden or confusing metadata, opaque search results, and the poor quality of its scans. *See* Paul Duguid, *Inheritance and Loss? A Brief Survey of Google Books,* First Monday (Aug. 6, 2007), http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1972/1847; Ronald G. Musto, *Google Books Mutilates the Printed Past*, Chron. of Higher Educ., June 12, 2009, at 4; Geoffrey Nunberg, Op-Ed., *Google Book Search: A Disaster for Scholars*, Chron. of Higher Educ., Aug. 31, 2009, http://chronicle.com/article/Googles-Book-Search-A/48245; Matthew Moore, *Google Book 'Finger Condoms' Cause Mirth*, Telegraph, Oct. 21, 2009, http://www.telegraph.co.uk/technology/google/6396896/Google-Book-finger-condoms-cause-mirth.html (describing poor scanning on Google's part as "finger spam"); Geoffrey Nunberg, *Google Books: A Metadata Trainwreck*, Language Log blog (Aug. 29, 2009 5:46 PM), http://languagelog.ldc.upenn.edu/nll/?p=1701 (including a long rebuttal from Google employee Jon Orwant); Robert B. Townsend, *Google Books: What's Not to Like?*, AHA Today (Apr. 30, 2007), http://blog.historians.org/articles/204/google-books-whats-not-to-like. In response to the criticisms regarding poor scanning, Google employees, at least in public, take it in stride, pledge to do better, and declare that they are always improving the quality of their product. Library partners, in Google's defense, have observed that the company's "iterative approach" to quality assurance has proved effective and the quality of the scans have improved over time. *See* Ivy Anderson, *Hurtling Toward the Finish Line: Should the Google Books Settlement Be Approved?*, Cal. Digital Library (Feb. 16, 2010), http://www.cdlib.org/cdlinfo/2010/02/16/hurtling-toward-the-finish-line-should-the-google-books-settlement-be-approved/. In 2009, the Mellon Foundation funded a pilot study by Paul Conway at the University of Michigan to develop a metric for assessing the "quality" of digital scans in the Hathi Trust database (most of these scans are so-called "library digital copies" received by contractual agreement from Google). *Mellon Grant Aids Researching Criteria for Digital Libraries*, Univ. of Mich. (Sept. 28, 2009), http://blog.si.umich.edu/2009/09/28/mellon-grant-aids-researching-criteria-for-digital-libraries/.

110. *Do the Libraries Get a Copy of the Book?*, Google Books, http://books.google.com/support/bin/answer.py?hl=en&answer=43751 (last visited Sept. 21, 2010); *see also* Cooperative Agreement between Google

detail, slow to emerge, increased publisher anxiety that they would lose control over the digital distribution of their books. However, this quid pro quo was a crucial incentive for libraries because possessing digital copies of a vast amount of one's own collection opens up many possibilities for a library's management.[111] It is for this reason that John Wilkin, Associate University Librarian at the University of Michigan, referred to December 14, 2004, as "the day the world changes."[112]

In September 2005, after frustrating negotiations and some failed concessions between publishers and Google,[113] the Authors Guild filed a class action suit against Google alleging massive copyright infringement.[114] Then, one month later, the Association of American Publishers filed its own suit seeking injunction.[115] Google's defense was that its use was a fair use under U.S. copyright law because it was only making three or so lines of text available around a search term—the so-called "snippet view."[116] Negotiations between the parties to the lawsuit began just a few months after it was filed, and would lead to a proposed Settlement Agreement almost exactly three years after the suits' filing.[117]

---

and the University of Michigan, § 4.1, at 5, *available at* http://www.lib.umich.edu/files/services/mdp/um-google-cooperative-agreement.pdf. Google and Michigan entered into an expanded agreement after the announcement of the proposed Settlement.

111. *See* Paul Courant, *On Being in Bed with Google*, Au Courant (Nov. 4, 2007), http://paulcourant.net/2007/11/04/on-being-in-bed-with-google/.

112. *Google to Scan Books from Major Libraries*, MSNBC (Dec. 14, 2004, 11:09 AM), http://www.msnbc.msn.com/id/6709342. Wilkin continued: "It will be disruptive because some people will worry that this is the beginning of the end of libraries. But this is something we have to do to revitalize the profession and make it more meaningful." *Id.*

113. *See* Adler, *supra* note 107.

114. Press Release, Authors Guild, Inc., Authors Guild Sues Google, Citing "Massive Copyright Infringement," (Sept. 20, 2005), *available at* http://www.authorsguild.org/advocacy/articles/authorsguildsuesgooglecitingmassivecopyrightinfringement.html.

115. Press Release, Association of American Publishers, Publishers Sue Google Over Plans to Digitize Books, (Oct. 19, 2005), *available at* http://www.publishers.org/main/PressCenter/Archicves/2005_Oct/Oct_03.htm. For an excellent and thorough overview of the lawsuits against Google by the Authors Guild and the Association of American Publishers, see generally Jonathan Band, *The Long and Winding Road to the Google Books Settlement*, 9 J. Marshall Rev. Intell. Prop. L. 227 (2009).

116. The legal literature on the fair use question posed by the lawsuit is extensive. *See generally* Hannibal Travis, *Google Book Search and Fair Use: iTunes for Authors, or Napster for Books?*, 61 U. Miami L. Rev. 87 (2006); Michael Warnecke, *Copyrights/Fair Use: Google's Search of Library Shelves Splits Copyright Law Experts on Fair Use*, 70 Pat. Trademark & Copyright J., Oct. 28, 2005, at 707; Jonathan Band, *The Google Library Project: A Copyright Analysis*, Policy Bandwidth, http://www.policybandwidth.com/doc/googleprint.pdf (last visited Oct. 12, 2010); Sanford G. Thatcher, Dir., Pennsylvania State Univ. Press, Panelist at NACUA's The Wired University Conference: Fair Use in Theory and Practice: Reflections on Its History and the Google Case (Nov. 10, 2005) *available at* http://www.psupress.psu.edu/news/NACUA_thatcher.pdf.

117. David Drummond, *New Chapter for Google Book Search*, Official Google Blog (Oct. 28, 2008, 7:14 PM), http://googleblog.blogspot.com/2008/10/new-chapter-for-google-book-search.html. On November 13, 2009, the Amended Settlement Agreement replaced the Original Settlement, which was announced on October 28, 2008, and is still subject to court approval as of the date of this article's publication. *See* Settlement Agreement, *supra* note 6; Amended Settlement Agreement, *supra* note 6.

B. The View from the Settlement

If approved, the Settlement Agreement will significantly fortify Google's continuing digitization project: through a broad release of liability from the millions of authors and publishers who are covered by the Agreement;[118] revenue models into the near and far futures (of which Google would receive thirty-seven percent);[119] and an all-but-insurmountable competitive advantage over any likely competing digitizer of books. Microsoft had entered the book digitization business in 2005 when it joined the Open Content Alliance.[120] It got out in 2008.[121] The Open Content Alliance, with Microsoft's support, had been lauded as a potential non-commercial alternative to Google Books.[122] It attracted hundreds of library partners,[123] and digitized millions of public domain titles (some of its members continue to scan books through the Internet Archive).[124] But the broad coalition it brought together from 2005 to 2008 has been unable to attract the financial resources required to sustain a project that could compete with Google's.[125] These competitors and other potential ones have labored under legal restrictions that the Settlement would free Google, but no one else, from.

When the original Settlement Agreement was announced in October 2008, Google revealed that it had digitized seven million books in four years.[126] Months

118. *See* Amended Settlement Agreement, *supra* note 6, § 10.2(b). Those who have chosen to opt out and those who are excluded from the settlement class (many foreign rightsholders) are still able to sue Google for copyright infringement.

119. Michael Cairns, *A Database of Riches: Measuring the Options for Google's Book Settlement Roll Out*, Personanondata (Apr. 22, 2010), http://personanondata.blogspot.com/2010/04/database-of-riches-business-model.html (projecting revenue of $260 million dollars a year (after three years) from the Institutional Subscription provided for in the Settlement).

120. Katie Hafner, *Microsoft to Offer Online Book-Content Searches*, N.Y. Times, Oct. 26, 2005, http://www.nytimes.com/2005/10/26/technology/26book.html#; *see generally* Katie Hafner, *In Challenge to Google, Yahoo Will Scan Books*, N.Y. Times, Oct. 3, 2005, http://www.nytimes.com/2005/10/03/business/03yahoo.html (discussing the Open Content Alliance).

121. Satya Nadella, *Book Search Winding Down*, Bing Comty. (May 23, 2008, 2:45 AM), http://www.bing.com/toolbox/blogs/search/archive/2008/05/23/book-search-winding-down.aspx.

122. Katie Hafner, *Libraries Shun Deals to Place Books on Web*, N.Y. Times, Oct. 22, 2007, http://www.nytimes.com/2007/10/22/technology/22library.html.

123. *Contributors*, Open Content Alliance, http://www.opencontentalliance.org/contributors/ (last visited Sept. 20, 2010).

124. *Ebook and Texts Archive*, Internet Archive, http://www.archive.org/details/texts (last visited Sept. 20, 2010).

125. *See* Andy Guess, *Post-Microsoft, Libraries Mull Digitization*, Inside Higher Ed (May 30, 2008), http://www.insidehighered.com/news/2008/05/30/microsoft ("Microsoft spent some $10 million on its book-scanning operations through the Open Content Alliance."); *see also* Declaration of Daniel Clancy in Support of Motion for Final Approval of Amended Settlement Agreement at 1, Authors Guild, Inc. v. Google Inc., No. 05-CV-8136-DC (S.D.N.Y. Feb. 11, 2010) [hereinafter Declaration of Daniel Clancy], *available at* thepublicindex.org/docs/amended_settlement/dan_clancy_declaration.pdf (acknowledging that "Google has spent hundreds of millions of dollars" on its book scanning project).

126. Drummond, *supra* note 117.

later the figure increased to ten million,[127] then to twelve million in February 2010, and fifteen in October 2010.[128] Despite the Harvard University library's alienation from the project over the course of the Settlement negotiations,[129] Google has now accumulated over forty library partners, twelve of which are foreign libraries.[130] The company continues to scan books "apace,"[131] and Google Books engineer Jon Orwant expects that, by the time his two young children enter college, "effectively" every book will be online.[132]

In the six years since Google launched its book digitization project, the arena of digital books ("e-books") has exploded. Apple introduced the iPhone in January 2007, and its e-book reading and buying applications became surprisingly popular;[133] Amazon introduced the Kindle in November 2007, and revived the dormant e-book business; Apple's iBookstore is too new to assess, but it is at least perceived as important enough that book publishers have successfully played Apple against Amazon for preferable e-book pricing models.[134] Google has indicated it will launch its own e-book selling and distribution platform, called Google Editions, in the near future.[135] One of the important outcomes of the Settlement, beyond the "copyright peace" and the new sources of income, is that it establishes Google as a counterweight to Amazon's dominance of e-book retailing. For information technology journalist Nicholas Carr the most important thing about the Google Books Search Settlement is that it "makes clear that the world's books will be digitized—and that the effort is likely to proceed quickly."[136]

---

127. Brin, *10,000,000 Books*, *supra* note 103.

128. Declaration of Daniel Clancy, *supra* note 125. James Crawford, *On the Future of Books*, Inside Google Books (October 14, 2010) http://booksearch.blogspot.com/2010/10/on-future-of-books.html.

129. *See* Laura Murviss, *Harvard-Google Online Book Deal at Risk*, Harv. Crimson, Oct. 30, 2008, http://www.thecrimson.com/article/2008/10/30/harvard-google-online-book-deal-at-risk/; *see also* Robert Darnton, The Case for Books: Past, Present, and Future (2009).

130. Dan Clancy, Engineering Director, Google, Inc., Speech at a meeting of a small group of faculty and personnel at the University of California at Berkeley (June 20, 2009). For a partial list of Google library partners, see *Google Books Library Partners*, Google Books, http://www.google.com/googlebooks/partners.html (last visited Oct. 30, 2010). Dan Clancy has since changed jobs at Google. *See supra* note 59.

131. *Alexander Macgillivray of Google on the Google Book Search Settlement*, Berkman Ctr. for Internet & Soc'y (July 29, 2009), http://cyber.law.harvard.edu/interactive/events/luncheons/2009/07/macgillivray.

132. Orwant, *supra* note 94.

133. Olga Kharif, *Readers are Devouring Apple Book Apps*, BusinessWeek (Mar. 14, 2010), http://www.businessweek.com/technology/content/mar2010/tc20100312_351841.htm.

134. Ken Auletta, *Publish or Perish; Can the iPad Topple the Kindle, and Save the Book Business?*, The New Yorker, Apr. 26, 2010, at 24, 25–26.

135. Jessica E. Vascellaro & Jeffrey A. Trachtenberg, *Google Readies its E-Book Plan, Bringing in a New Sales Approach*, Wall St. J., May 4, 2010, http://online.wsj.com/article/SB10001424052748703866704575224232417931818.html. Google has a preliminary website for the program. *See Getting Started with Google Editions*, Google Books, http://books.google.com/support/partner/bin/answer.py?answer=167975 (last visited Oct. 14, 2010).

136. Nicholas Carr, The Shallows: What the Internet is Doing to Our Brains (2010).

C. From Search to Access

As Google Books Chief Engineer Dan Clancy has explained, the Settlement Agreement, if approved, will reorient and enlarge Google Books away from "search and snippets" to "search and access."[137] Should it be approved, out-of-print but in-copyright books would no longer be made available through the severely circumscribed "snippet view." Instead, these books would be not only *searchable* but also *readable* through one of a variety of Settlement-defined licenses granted by the rightsholders, including: paid access to the entire corpus via the Institutional Subscription; a Consumer Purchase model of perpetual access to individual titles stored in "the cloud"; or free access to the entire corpus at specially designated terminals in public libraries, with fees for printing, through the Public Access Service.[138] In addition, under "preview use," up to twenty percent of all books covered by the Settlement Agreement (with a few exceptions) would be displayed by default, for free, much like today's Partner Program handles in-print books. Presumably, snippets would remain only for those books not covered by the Settlement Agreement (e.g., most foreign language books). Thus, the Settlement Agreement, in its latest version, provides for much more access to the texts of individual books than the snippet-based display that exists now. "Snippet view" of most books in the corpus is likely the most one could have expected if the case had gone to trial and Google had prevailed on fair use grounds.[139]

Still, some commentators have argued that even this considerable expansion is not enough, and that it is preferable that orphaned, out-of-print books be made available through an "Open Access" model or that access be provided through digital loaning.[140] Others believe that the full-text access available through the fee-based Institutional Subscription, the consumer purchase model, and the free public library license have "liberated" books from their imprisonment in university research libraries and "unlocked" them from the shackles of copyright law.[141]

According to Clancy, the Google Books team started out thinking only about search, but Settlement negotiations reoriented their concerns toward access:

---

137. Dan Clancy, Remarks during a law school class at the University of California, Berkeley School of Law (Apr. 8, 2010).

138. Settlement Agreement, *supra* note 6, art. IV.

139. Matthew Sag, *The Google Book Settlement and the Fair Use Counterfactual*, 55 N.Y.L. Sch. L. Rev. 19, 25 (2010–11); Quentin Hardy, *Lend Ho!*, Forbes Mag., Nov. 16, 2009, *available at* http://www.forbes.com/forbes/2009/1116/opinions-brewster-kahle-google-ideas-opinions.html.

140. *See* Pamela Samuelson, *An Academic Author's Perspective on the Google Book Settlement*, Against the Grain, June 2010, at 24; *see also* Robert Darnton, *Google and the Future of Books*, N.Y. Rev. Books, (Feb. 12, 2009), *available at* http://www.nybooks.com/articles/archives/2009/feb/12/google-the-future-of-books/.

141. *See* James Boyle, *Google Books and the Escape from the Black Hole*, The Pub. Domain (Sept. 6, 2009), http://www.thepublicdomain.org/2009/09/06/google-books-and-the-escape-from-the-black-hole/; Adam M. Smith, *Unlocking Access to Millions of Books*, Inside Google Books (Oct. 28, 2008, 6:33 AM), http://booksearch.blogspot.com/2008/10/unlocking-access-to-millions-of-books.html.

> The problem with snippets is that they always make you want more. Sometimes a snippet isn't enough to tell you if you should bother with the book. It's a far cry from what a scholar really wants. It's better than nothing, but it's not what anyone wants. The question was: how will this thing become a library which is the type of resource that everyone . . . wants, and how can that happen under copyright law?[142]

As a result of the Settlement's new focus on "access," Google Books would become more like a library than its current card-catalog/search-and-discovery version.

D. All About Access

Access to information or knowledge is generally recognized as fundamental to a democracy—to some, it is a fundamental human right.[143] Libraries are especially concerned with "access," as the University of California (UC) Vice-Provost and lead representative in UC's Google partnership, Daniel Greenstein, explained at a meeting in Berkeley, California, convened to discuss the Settlement:

> We were involved with a bunch of other libraries . . . in discussions around the Settlement advising Google, really trying to defend and protect and promote what we consider to be the public good aspect of the Settlement. We've been in this for a very long time [because] libraries are all about access . . . [P]ublic libraries, including libraries in public universities such as [UC], are all about access to information as a kind of sacred public trust because the public funded the entire corpus and all the work that goes on around it. So, the UC libraries were active in the Settlement because it promised a great deal of access.[144]

And, indeed, as Greenstein himself asked the audience before him, can anyone argue with increased access? Perhaps no one can, but it is much easier to proclaim that libraries are "all about access" than to say precisely what access is and what it, in practice, entails: Access to what, by what means, and for whom? After all, the relatively privileged UC campuses already enjoy extraordinary access to information resources and will likely stand to benefit substantially from subsidized subscriptions to Google Books through the Institutional Subscription.[145] In other words, its "access"

---

142. Clancy, *supra* note 59.

143. *See Access to Knowledge*, Consumer Project on Tech., http://www.cptech.org/a2k/ (last visited Oct. 14, 2010).

144. Daniel Greenstein, Vice Provost for Academic Planning, Programs, and Coordination, University of California, Berkeley, Remarks at Public Access and the Google Books Settlement Conference (Aug. 28, 2009), *available at* http://www.ischool.berkeley.edu/newsandevents/events/20090828googlebooksconference/video4; *see also* Daniel Greenstein, *Library Stewardship in a Networked Age*, Council on Libr. & Info. Res., http://www.clir.org/pubs/reports/pub126/green.html (last visited Oct. 14, 2010) ("Academic libraries are all about access—even where they claim a role preserving our cultural and scholarly heritage.").

145. Although the University of California has yet to finalize a revised agreement with Google, Michigan and Google signed an amended agreement in May 2010. According to that contract, Google will subsidize the entire cost of the University of Michigan's Institutional Subscription for twenty-five years in exchange, presumably, for Michigan's strong support of the project and having made available millions of books. *See The University of Michigan and Google Amended Digitization Agreement*, Mich.

is unequal to others'. Because of the immense variation across institutions and societies, it seems imperative to develop a rigorous approach to assessing modes, practices, and meanings of "access." The Open Access[146] and Access to Knowledge[147] movements have brought attention to these questions, but it seems we are only at the early stages of developing a properly rigorous approach to the notion. If "access" is of fundamental importance, it deserves just as much scrutiny as it does valorization.

Upon close examination of the Settlement Agreement, "access" is less a straightforward, transparent "democratic" principle than a fluid set of practices that regulate, limit, and/or allocate who reads what. The Settlement Agreement rations access in service to the Settlement's market orientation, which operates on a sort of "freemium" model. With regard to in-copyright books, the user gets a little something for free (the twenty-percent preview use) but has to pay for more (e.g., access to the whole book, the ability to print, etc.). As the Settlement website explains: "Preview uses are designed to serve as a marketing tool to sell the Book."[148] In other words, the Settlement's access is designed not simply as a "service" but as an enticement to buy individual books. In this commercial mode of access, "access" is something that can be sold—a commodity.

Books, of course, are commodities, but they are also more than commodities, such as when they are bought by libraries and effectively de-commodified. In December 2004, as Google's Library Project was announced, and well before the lawsuit or its proposed Settlement, librarian Rory Litwin predicted that a commercial mode of access would come to characterize Google Books, and he expressed his fear that it would commercialize the entire library system:

> [H]owever compromised it may be, in the world of scholarship and education there is a genuine culture of intellectual honesty that stems from the communal project of seeking and spreading truth for the common good. You do not see advertisements for particular historic works of literature in research libraries, or for particular publishing companies. When a work appears in a bibliography, it is there because of the independent judgment of a scholar or a librarian as to the significance and the relevance of that work; it is not there because somebody is trying to sell it and make money from it. . . . Google's . . . search results cannot be guaranteed by any public policies and could be transformed into pure e-commerce at any time.[149]

Although a large new library-like service with "unprecedented access to the world's books" may wear a veneer of democratic virtue, Litwin reminds us of Peter Lyman's

---

Digitization Project (Aug. 13, 2009), http://www.lib.umich.edu/michigan-digitization-project/university-michigan-and-google-amended-digitization-agreement.

146. *See generally* Peter Suber, *Open Access Overview,* Earlham Coll. (Sept. 15, 2010), http://www.earlham.edu/~peters/fos/overview.htm.

147. *See Access to Knowledge*, *supra* note 143.

148. *FAQs*, Google Book Settlement, http://www.googlebooksettlement.com/help/bin/answer.py?hl=en&answer=118704#q30 (last visited Oct. 14, 2010).

149. Litwin, *supra* note 11.

observation, with which I began: that the modern library has been a buffer to market forces.[150]

In contrast to Litwin's indignation, historian Carla Hesse, in discussing Google Books, has suggested flatly that those who care about libraries need to accept this commercial mode as the new way in which libraries will operate.[151] She categorizes the history of libraries in three stages: the age of private libraries (up to the French Revolution), the age of public libraries (after the French Revolution), and now the age of commercial libraries.[152] This new mode, she insists, "is not a problem; it's a condition."[153] It is part of the combustion when "digital" meets "library." But how indeed shall we think about a commercial library? Is it a private or a public good? To what authorities do we appeal for its regulation? What is reasonable to demand of a commercial library? What principles or ethics can be brought to bear?

Commentators have identified other concerns about access. Joining scholar-critics of Google's metadata, classicist James O'Donnell has identified "intellectual access" as an area of concern: "There's much more to accessibility than a price tag. Information has to be findable and, when found, usable . . . . The digital representations that Google has made available aren't yet a library—and indeed, in an important sense, they aren't even books yet."[154] Legal scholar Larry Lessig worries that the Settlement Agreement grants too much control over access to rightsholders, and essentially extends to books the "permission culture" that has made documentary films so hard to finance and to keep available:

> The deal constructs a world in which control can be exercised at the level of a page, and maybe even a quote. It is a world in which every bit, every published word, could be licensed. And what this means or so I fear [is that] we are about to make every access to our culture a legally regulated event.[155]

With similar concerns, sociologist Tarleton Gillespie, writing before the Settlement, worried that Google Books would set a bad precedent for digital culture by granting too much control to content owners not simply over the *use* of their work but also over the *access* to it.[156] From the opposite direction, Raymond Reuss, a German literary scholar and outspoken opponent of Open Access, accuses the Settlement of

---

150. *See* Lyman, *What is a Digital Library?*, *supra* note 3.

151. Carla Hesse, Dean of Social Sciences, University of California, Berkeley, Remarks at Public Access and the Google Books Settlement Conference (Aug. 28, 2009), *available at* http://www.ischool.berkeley.edu/newsandevents/events/20090828googlebooksconference/video4.

152. *Id.*

153. *Id.*

154. James O'Donnell, *A Scholar Contemplates the Google Book Settlement*, Against the Grain, June 2010, at 26.

155. Lawrence Lessig, *For the Love of Culture: Google, Copyright, and our Future*, New Republic, Jan. 26, 2010, *available at* http://www.tnr.com/article/the-love-culture.

156. Tarleton Gillespie, *Everything to Everyone*, Inside Higher Ed, Jan. 27, 2006, http://www.insidehighered.com/views/2006/01/27/gillespie.

"fetishizing" access in an attempt to undo copyright protections for authors.[157] Agitation from foreign authors like Reuss played a role in the parties amending the original Settlement Agreement, with the result that the vast majority of foreign books were removed.[158] The universal library has, for now at least, hit not only the wall of comprehensiveness (it is now limited primarily to Anglophone books), but also the wall of accessibility (it restricts access to users in the United States).

By collecting these various reactions to the terms and implications of the Settlement, I am suggesting the complexity of "access" in an attempt to encourage a perspective on the Settlement's provisions (and in general) that focuses less on access as a principle and more on access as a practice or set of practices. While the Settlement is a commendable effort to redefine how access might work in the digital future, it is limited to the resolution of one lawsuit and therefore excludes many other parties that will be affected by its terms. It is also very hard to predict how the arrangements it describes will play out over time in a rapidly changing area of innovation and how those not involved in the negotiations will be affected.

V. CONCLUSION

In 2004, writing about Project Gutenberg, Paul Duguid argued that the pretensions of mass book digitization toward a universal library fundamentally misunderstand what a book is: "There is more to books than a digital scanner can detect or the standard keyboard represent."[159] This "more" to books—their bibliographic complexity, their immense variation, their embeddedness in material systems of signification, and more—cannot be easily accommodated within automated database structures. For Duguid, the problem is not technological but social. A library is an institution populated by humans. The institution is made manifest by human knowledge and labor, and made meaningful by its community of users. Echoing the librarian-computer science divide discussed above, he argues that the Internet is not an institution but a technology.[160] It is a self-organizing storage device lacking the contextualizing editorial and curatorial practices and the "values"

157. *See* Andrew Albanese, *Frankfurt Book Fair: Europeans Play the Moral Rights Card Against Google Settlement*, Publishers Weekly (Oct. 16, 2009), http://www.publishersweekly.com/pw/by-topic/digital/content-and-e-books/article/8663-frankfurt-book-fair-europeans-play-the-moral-rights-card-against-google-settlement-.html.

158. Band, *supra* note 115, at 314 n.827; *Amended Google Deal Enters Stretch Run: Revised Settlement Narrows Scope of Agreement*, Publishers Weekly, Nov. 23, 2009, at 3, *available at* http://www.publishersweekly.com/pw/print/20091123/29413-amended-google-deal-enters-stretch-run-.html; Jessica E. Vascellaro & Jeffrey A. Trachtenberg, *New Google Book Pact Unlikely to End Flap*, Wall St. J., Nov. 16, 2009, at B1 (reporting that the Amended Settlement "would cut the number of works covered by the settlement by at least half by removing millions of foreign works").

159. Paul Duguid, *PG Tips*, Times Online Literary Supplement, June 11, 2004; *see* Robert Darnton, *The Library in the New Age*, N.Y. Rev. of Books (June 12, 2008), http://www.nybooks.com/articles/archives/2008/jun/12/the-library-in-the-new-age/?pagination=false (including similar criticisms, in a more generous spirit).

160. Duguid, *supra* note 159.

referenced above that make a library useful, trustworthy, and socially robust.[161] As his speech at the *D Is for Digitize* symposium demonstrated, Duguid has remained dogged in his criticisms of Google's application of software solutions to the complex roles that humans institutions have played.[162]

A question posed by Google Books and the lawsuit's Settlement is whether a commercial library can be an institution that draws upon the values of Stefik's archetypal "keeper of knowledge." After Sergey Brin published his 2009 op-ed *A Library to Last Forever*, legal scholar Pamela Samuelson responded with a short essay entitled *Google Books is Not a Library*.[163] There, she argued that Google Books post-Settlement could not be a library because: it monetizes out-of-print books; it has incentives to "price-gouge" universities in the sale of the Institutional Subscription; it lacks provisions for long-term commitment to the product (contrary to what Brin asserts in the op-ed); no clear conditions for succession should Google be sold, go out of business, or simply grow tired of Book Search are provided; it might sell ads inside the Institutional Subscription; and, finally, Google has been unwilling to make meaningful commitments to protect user privacy as libraries traditionally have.[164]

Samuelson draws a distinct boundary around what a "library" is or can be. At the *D Is for Digitize* symposium that occasioned this journal issue, she bluntly stated: "You create a public good this substantial, guess what? Public trust responsibilities come along with it."[165] Duguid complains that Google does not want to incorporate the hard work of human librarianship into its automated processes—that Google does not want to behave like a library.[166] Samuelson argues that Google, in exchange for the privileges that it would be granted through the Settlement, *must* be made to accept the obligations of the public's trust, whether it wants to or not.

The digital library metaphor implicitly inquires into how the library will function in networked digital environments. Recall Peter Lyman's description of the library as a boundary management institution that buys information in a marketplace and subsidizes its free use by communities.[167] Some believe Google Books may not have

---

161. *Id.*

162. Paul Duguid, Adjunct Professor, School of Information University of California, Berkeley, Presentation at D Is for Digitize Symposium Conference at New York Law School (Oct. 9, 2009), http://nyls.mediasite.com/mediasite/SilverlightPlayer/Default.aspx?peid=4f3acab65b424c17b0eeb26ec8be6bc2.

163. *See* Brin, *A Library to Last Forever*, *supra* note 103; Pamela Samuelson, *Google Books is Not a Library,* Huffington post (Oct. 13, 2009, 12:38 PM), http://www.huffingtonpost.com/pamela-samuelson/google-books-is-not-a-lib_b_317518.html.

164. Pamela Samuelson, Professor of Law and Information Management, University of California, Berkeley, Remarks at the D Is for Digitize Conference at New York Law School (Oct. 9, 2009), *available at* http://nyls.mediasite.com/mediasite/SilverlightPlayer/Default.aspx?peid=1c998d3fb46c40aaa8216b4ee7555ce9; *see also* Pamela Samuelson, *Google Book Search and the Future of Books in Cyberspace*, Minn. L. Rev. (forthcoming 2010) [hereinafter *Future of Books in Cyberspace*], *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1535067.

165. Samuelson, Remarks, *supra* note 164.

166. *See* Duguid, *supra* note 162.

167. Lyman, *Poetics of the Future*, *supra* note 16.

yet earned the distinction of "library,"[168] but Google Books will be, in fact, managing boundaries between private property and public goods. It mediates in a marketplace to subsidize access to information. But where would Google Books be situated on the continuum Lyman identified, with, at one end, all-metered use and, on the other, free use? Can Google Books, in fact, become an institution with meaningful public trust responsibilities? If so, how? Library supporters of the Settlement have focused their efforts not on what Google will or will not do, but on their own plans once they have digital copies of large portions of their collections. The digital copies that they receive back from Google are already being ingested into the Hathi Trust, a "robust shared access and preservation service" that shows more promise of becoming an institution with meaningful public trust responsibilities.[169] Still, the Settlement's rather severe restrictions on the partner libraries' use of their digital copies dictate that it be a predominantly "dark archive"—at least until the books fall out of copyright.[170]

The boundaries that the "library" manages appear to be shifting, and the discussion around the Settlement is, to some extent, about those boundaries. If the robust public response to the Google Books Settlement is properly appreciated, it should be seen as a collective asking of the hard questions Lyman found nestled in the digital library metaphor over a decade ago. A court's ruling will define some parameters for approaching these questions, but it cannot answer them definitively. These questions and their accompanying demands for vigorous scrutiny and skepticism will continue to complicate the wishes and dreams of both technologists and librarians until new, stable institutions arise for the provisioning of information in a networked environment. In the words of Peter Lyman, "[i]t is not new tools that constitute innovation but new institutions."[171]

168. *See* Duguid, *supra* note 109; Litwin, *supra* note 11; Samuelson, Remarks, *supra* note 164; Samuelson, *Future of Books in Cyberspace*, *supra* note 164; Tushnet, *supra* note 11, at 1021–23.

169. Anderson, *supra* note 109; *see generally* Hathi Trust, http://www.hathitrust.org (last visited Sept. 22, 2010).

170. *See* Amended Settlement Agreement, *supra* note 6, §§ 7.2, 2.2, 3.5(a)(i); *see also* Amended Settlement Agreement, *supra* note 6, Attachment B-1; Amended Settlement Agreement, *supra* note 6, Attachment D.

171. Lyman, *What Is a Digital Library?*, *supra* note 3, at 1.