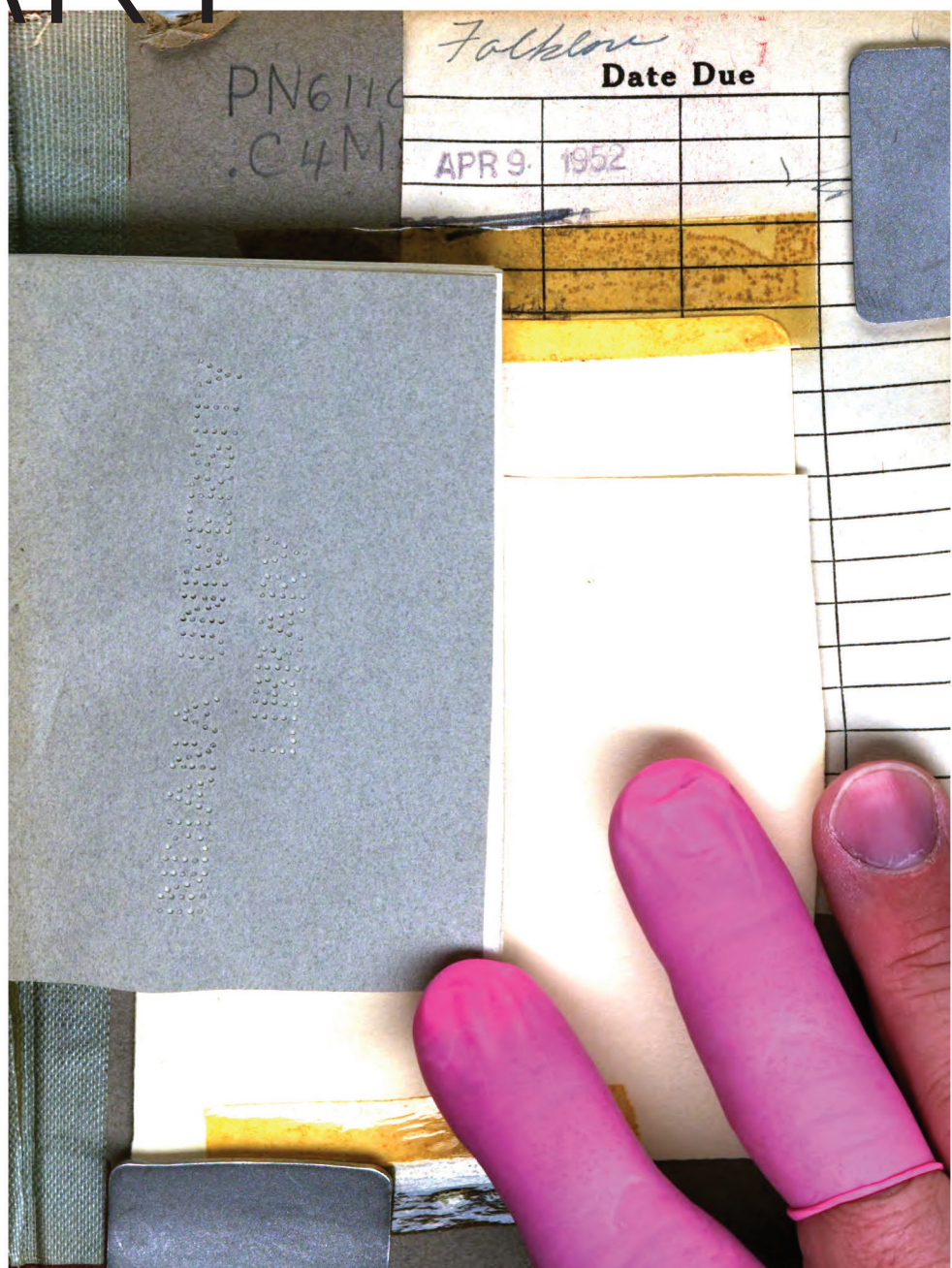


# UNPACKING GOOGLE'S LIBRARY

Google wanted to digitize all the world's books but eventually abandoned that goal.

**Mary Murrell** explores the rise and fall of one utopian library project and the emergence of new ones in its wake.

ANDREW NORMAN WILSON, MOTHER GOOSE'S MELODY-164. FROM HIS 2012 SCANOPS EXHIBITION.



**IN DECEMBER 2004 GOOGLE REVEALED** its Library Project, a hugely ambitious plan to digitize “all books in all languages” through partnerships with some of the largest research libraries in the world—Harvard, Stanford, Oxford, the New York Public Library, and the University of Michigan, to begin—and to make those books accessible online. The news astonished interested observers, eliciting both fear and excitement. Enthusiasts found in it an intoxicating combination of the humanistic and the technoscientific: a new and improved Library of Alexandria, a generation’s moonshot, a humanistic complement to the Human Genome Project.<sup>1</sup> The mass digitization of library collections promised to give a future to the past currently “imprisoned” in print form. In so doing it would also give new life to research libraries, open up new lines of scholarly inquiry and practice, and vastly expand people’s access to library holdings. Opponents embraced these possibilities, too, but they also feared Google’s motives in the project, its will to power, and its increasing control over access to knowledge. Others accused the company of undermining the central tenets of copyright. Brought together in a decade-long saga, these and other contentions around the Library Project swelled into what might be considered an archival fever: one ambitious total archive ramifying into new ones.

In crucial respects, the Library Project has been remarkably successful. The company has scanned, page by page, more than 25 million books in more than 400 languages.<sup>2</sup> Although it is hard to know for certain how many books exist to be digitized, 25 to 30 million certainly represent a significant percentage. (As a point of comparison, its closest competitor, the Open Content Alliance/Internet Archive has digitized roughly 2 million books.) The Library Project has also survived legal challenges against it. In 2005, authors and publishers sued Google alleging copyright infringement (*Authors Guild et al. v. Google*), and in 2011 the Authors Guild sued Google’s library partners over their possession and use of Google’s scans (*Authors Guild v. HathiTrust*). In both cases, judges found scanning the entirety of an in-copyright book, for circumscribed uses, to be a fair use under U.S. law—in both lower courts and on appeal.

And yet, despite these successes, Google has quietly forsaken its Library Project, despite being far short of its original outsized goal (“all books in all languages”) as well as its pledge to digitize the entirety of the University of Michigan’s libraries, its principal partner. After the proposed settlement to *Authors Guild et al. v. Google* was rejected in March 2011, its commitment tapered off significantly. The settlement would have set aside legal differences between copyright owners and Google by opening the Library Project up to extensive commercialization (see Samuelson 2011).



Andrew Norman Wilson, *The Inland Printer* — 164. From his 2012 *ScanOps* exhibition.

Without that potential for revenue generation, the costly project appears to have been deemed too dear even for deep-pocketed Google: scanning capacity was drastically cut in 2011; the Google Books blog was discontinued in 2012; its Twitter feed went silent in 2013; and its staff left or was reassigned. Although the company continues to scan books from libraries, according to partner libraries, it stopped scanning in-copyright books back in 2011, limiting itself now to books in the public domain. This about-face returns to the state of affairs circa 2004, when the announcement of Google’s project made such a splash as a bold move forward. By 2011, the Web had changed too. It was no longer in need of high-quality content as it had been in the early 2000s when mass digitization seemed worth the company’s investment (Edwards 2011). The moonshot, in short, fell back to earth.

Based on these developments, it is not unreasonable to wonder whether the company might allow its books platform to languish in light of shifting priorities (see Biao 2015; cf. Lemov, this issue). Nonetheless, the momentum around mass digitization has shifted to successor projects such as, in the U.S., the Hathi Trust and the Digital Public Library of America (DPLA)—both of which grew out of the Library Project.

The Hathi Trust began in 2008 as a collaboration among research libraries to pool the digitized books that Google provided as part of their contractual arrangements. It has since grown to include books from other digitizers such as the Internet Archive and from libraries’ own scanning initiatives, but its core remains the Google-digitized books. Like a traditional research library or archive (and unlike Google), the Trust’s mission is to steward “the cultural record long into the future,” with all that that entails (HathiTrust n.d.; see also Christenson 2011). But like Google, it too pursues a totality—a different totality. The Hathi Trust’s specific operative aspiration is not the scholar’s dream of a “universal library” but rather the technologist’s dream of effecting a crucial tipping point, from a print-dominated intellectual infrastructure to an electronic one. By creating one total archive of all books held by its network of research libraries (“curation at scale”), libraries can identify and eliminate the redundancies between their collections, drastically reduce their print holdings, and thus cut out the costs associated with maintaining large and underused print collections (Wilkin 2015). Solving “the print problem” will enable the reallocation of scarce resources to new areas of library activity: institutional repositories, publishing initiatives, redesigned library spaces, data curation, digital preservation, and so on. This explains University of

1 For two examples of the many comparisons to the HGP, see Vaidhyanathan 2012 and Michel et al. 2010. On the HGP as an archive, see also Reardon p. 72 in this issue.

2 The number is now no doubt considerably higher.

Michigan librarian John Wilkin's declaration that December 14, 2004 (the day that Google announced the Library Project) was "the day the world changes" (Associated Press 2004). At last, libraries could look forward to moving beyond the immense burden of their print collections. The problem now, of course, is that Google did not complete the digitization, and it is unclear who will.

Whereas the Hathi Trust emerged through direct collaboration with Google, the DPLA developed in direct critical reaction to the Library Project and, in particular, Google's failed attempt to settle its differences with the publishers and authors. One of the leading opponents of that settlement was Harvard University Librarian Robert Darnton, whose experience working with Google had convinced him that the company's interests were antithetical to those of libraries and the "public interest" (Darnton 2009). In the course of seeking the settlement's rejection, Darnton proposed an alternative "national digital library" which later became articulated as "an open, distributed network of comprehensive online resources" (DPLA n.d.). Officially launched in 2013 with start-up funding from philanthropies and government agencies, the DPLA is not a collection—it has no holdings—but rather a platform that connects dispersed library collections (including the Hathi Trust's). It is more diverse in intention than the Hathi Trust, involving a wider range of organizations (not just elite university libraries) and more diverse types of content (not just books), but it is also more ambitious. Like the Hathi Trust, the DPLA aspires to be yet a different "total archive"—one with a more spatial inflection. In language strongly evocative of early twentieth-century utopians and visionaries, such as H. G. Wells, Paul Otlet, and Robert C. Binkley, who were convinced that microfilm technologies would enable superior scholarly infrastructures, the DPLA seeks ultimately to become a "worldwide network that will bring nearly all the holdings of all libraries and museums within the range of nearly everyone on the globe" (Darnton 2013). To this end, its technical infrastructure was designed to interoperate with Europeana, the European Union-funded Web portal that launched in 2008—and which was yet another response to Google's Library Project.<sup>3</sup>

When, in the early 2000s, libraries forged their awkward partnership with Google over the problem of the printed book, they had sought to manage and to "rationalize" print accumulations. Those attempts, at least so far, appear not to have eased a burden but only to have ramified their responsibilities and increased their accumulations. Library book accumulations have proven themselves to be, more than ever, part of that maddening "universe of things that cannot be disposed of and that keep spawning new things" (Povinelli 2011). Google's Library Project now seems, oddly, if not small at least smaller. Its successor projects appear to carry more capacious hopes, more intractable obligations—most of which seem remarkably out of proportion to what library leaders understand to be an "era of constrained circumstances" (Wilkin 2015). ■

---

**MARY MURRELL** is currently an honorary fellow in the Department of Anthropology at the University of Wisconsin, Madison. She is writing a book entitled *The Open Book: An Anthropologist in the Digital Library*.

- 
- 3 It is worth noting that, as a project born in response to Google's Library Project, the DPLA, for the time being, involves no book scanning. That job falls to a patchwork of enterprises with varying modes of access: underground "shadow" libraries (see Bodo p. 19 in this issue), some national libraries (though notably not the Library of Congress), the Internet Archive's scanning, boutique library projects, in addition to Google's continued public domain scanning.

## BIBLIOGRAPHY

- Associated Press. 2004. "Google to Scan Books from Major Libraries." MSNBC.com, December 14. <http://www.nbcnews.com/id/6709342#.Vlt3Qd-rSV4>
- Biao, Andy. 2015. "Never Trust a Corporation to do a Library's Job." *Medium*, January 28. <https://medium.com/message/never-trust-a-corporation-to-do-a-libraris-job-f58db4673351>
- Christenson, Heather. 2011. "HathiTrust." *Library Resources & Technical Services*, 55(2):93-102.
- Darnton, Robert. 2013. "The National Digital Public Library Is Launched!" *New York Review of Books*, April 25, <http://www.nybooks.com/articles/2013/04/25/national-digital-public-library-launched/>
- . 2009. *The Case for Books: Past, Present, and Future*. New York: PublicAffairs.
- Digital Public Library of America (DPLA). n.d. "History." <http://dp.la/info/about/history/>
- Edwards, Douglas. 2011. *I'm Feeling Lucky: The Confessions of Google Employee Number 59*. New York: Houghton Mifflin Harcourt.
- HathiTrust. n.d. "Welcome to the Shared Digital Future." <https://www.hathitrust.org/about>
- Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, Joseph P. Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak, and Erez Lieberman Aiden. 2010. "Quantitative Analysis of Culture Using Millions of Digitized Books." *Science* 331(6014):176-182.
- Povinelli, Elizabeth. 2011. "The Woman on the Other Side of the Wall: Archiving the Otherwise in Postcolonial Digital Archives." *differences* 22(1):146-171.
- Samuelson, Pamela. 2011. "Why the Google Book Settlement Failed—and What Comes Next," 54 Comm. ACM 11, 29-31. November.
- Vaidhyanathan, Siva. 2012. *The Googlization of Everything (And Why We Should Worry)*. Berkeley: University of California Press.
- Wilkin, John. 2015. "Meanings of the Library Today," in *The Meaning of the Library: A Cultural History* ed. Alice Crawford, 236-53. Princeton, NJ: Princeton University Press.