The Universal Library Is Us: Library Work at Scale in HathiTrust

When major research libraries pooled resources in 2008 to launch HathiTrust—today one of the largest collaborative library initiatives in the world—they knew that they were on to something big. For more than a decade, the potential of digitization and digital technologies to dramatically enhance library access and preservation services had been moving from the periphery to the very center of library future-planning. Confronted on the one hand with increasing budgetary pressures and on the other with increasing demands for greater access to a broader range of collections, libraries quickly came to appreciate that the future would be not only digital, but largely shared as well. Collaboration would be essential if libraries, with limited resources, were to continue to provide traditional services and meet new service expectations. Libraries have a long history of collaboration, but examples of, and structures for, the kinds of collaboration that could truly impact library services at scale were few.

Enter HathiTrust (http://www.hathitrust.org/). In 2008, a group of two dozen research libraries saw the opportunity to leverage the large-scale digitization collaborations they had entered into with Google, the Internet Archive, and others and to lay the foundation for a new, shared digital collection—one increasingly comprehensive in its representation of library collections and one owned and managed collaboratively by the library partners. The libraries viewed preservation as a primary goal, but only in conjunction with access. They consequently developed a robust digital preservation infrastructure with multiple levels of redundancy and geographic distribution, as well as a highly functional access platform. From the outset, the collection was intended to be developed along with libraries’ print collections, providing an underlying preservation layer for print materials and enabling local and collective decisions about the management of print collections—in turn resulting in a more efficient use of physical library space and a stronger collective preservation infrastructure.

Today, HathiTrust continues its evolution as an “above-campus” library service. The digital repository contains more than 11 million volumes contributed by 28 libraries. More than 400 languages are represented in the deposited collections, which include materials created from medieval times to the present. Of the 11 million volumes, 3.6 million are in the public domain and available on the web. A total of 92 academic and research institutions support the infrastructure for the materials. The institutions participate in a shared governance structure and contribute to a broad variety of initiatives centered on the shared digital collection and geared toward meeting individual and collective goals for preservation and access. These initiatives include the following:

- **Copyright Review.** Since 2008, staff at HathiTrust institutions have been reviewing the copyright status of targeted bodies of material to identify works that, for reasons not apparent through bibliographic data, are actually in the public domain. More than 200,000 works previously presumed to be in copyright have been opened to the public domain as a result of this work.

- **HathiTrust Research Center (HTRC).** In 2013, HathiTrust launched a state-of-the-art research center to offer computational access to materials in the repository. The first phase of the HTFC focuses on enabling research across the public domain materials in HathiTrust. The second phase will focus on access, on a limited basis, to in-copyright materials.

- **Lawful Access to In-Copyright Materials.** Under specific conditions, HathiTrust provides access to in-copyright works for users at partner institutions. These conditions include access for users who are certified as having a print disability and access to library-owned works that are damaged or missing and out of print.

- **Government Documents.** HathiTrust is engaging in a project to expand and enhance access to U.S. federal government documents. The project will focus initially on a broad analysis of government documents records, the creation of a comprehensive registry of U.S. federal government documents, and the formation of a collaborative planning and advisory committee to guide subsequent steps.

- **Shared Print Monograph Archive.** HathiTrust is making plans to develop a distributed network of print monograph archives with holdings corresponding to the digital holdings in HathiTrust. This is part of HathiTrust’s efforts to aid libraries in reducing the long-term costs of storing and caring for print collections.

In addition to overseeing the latter two initiatives, a HathiTrust Program Steering Committee is organizing groups that will focus on the further development of HathiTrust’s collections and on issues surrounding copyright and access to works in the repository.

Through its collective activities, HathiTrust is enabling
libraries to move from vision to action by reimagining library collections, resources, and services at an unprecedented scale. These activities have tangible benefits in cost savings, broader access to materials, and enhanced user services. The laudable goal of building a “universal library” may continue to elude us, but when libraries come to be viewed as a collective whole that is greater than the sum of its parts, we shall be much closer to grasping that particular brass ring.

Note

Jeremy York (jjyork@umich.edu) is Assistant Director of HathiTrust.
Brian E. C. Schottlaender (becs@ucsd.edu) is The Audrey Geisel University Librarian at the University of California, San Diego.

© 2014 Jeremy York and Brian E. C. Schottlaender. The text of this article is licensed under the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0).