Budget pressures being experienced by all parts of higher education have hit research libraries in a distinctive way, challenging those institutions in their efforts to create comprehensive collections and in their attempts to devote reasonable levels of resources to the work of curating their collections. Although many areas of library work will benefit from an “above the campus” approach,1 coordinated action is the only viable solution for collecting and preserving specialized publications from many parts of the globe.

North American research library directors have thrown down the “collective action” gauntlet: the Association of Research Libraries (ARL) Spring 2010 Membership Meeting in Seattle, Washington, focused on challenges ranging from shared collection development to shared print storage. Nevertheless, like other parts of higher education including information technology, libraries have been far more successful at identifying problems that call for collective solutions than mustering the energy to solve them. Against this background of a desire for collective action, the lessons from HathiTrust (http://www.hathitrust.org) provide an example of deep collaboration across many institutions to solve practical problems.

What Is HathiTrust?
When the floodgates of large-scale digitization with Google, the Internet Archive (http://www.archive.org/), and others opened, several dozen research libraries created HathiTrust, a shared effort designed to aggregate these resources and capitalize on the collective expertise of their institutions. HathiTrust is attempting to create a comprehensive repository of published literature, beginning with literature captured through digitization. HathiTrust holds more than six million volumes and will probably reach eight million volumes by the end of 2010. In the holdings today there are more than 3.5 million books and over 100,000 serial titles. The current collection spans several centuries and hundreds of languages, drawn from the collections of many of the more than thirty HathiTrust partner libraries. In addition to its early focus on digitized content, HathiTrust is actively exploring models for preserving new publications in wholly digital formats.

HathiTrust is about collections, writ large, not simply about Google digitization. Although Google digitization comprises the bulk of the content online, HathiTrust probably contains more non-Google content from local efforts and Internet Archive digitization than any other digital library effort. HathiTrust is also collaborating with several university presses in an effort to put new books and book backlists online as “open access” and will extend that model. From the outset, the partners have seen HathiTrust’s mission as developing a digital collection in conjunction with print collections: many of the partners hope to reduce the physical space used to store print collections by exploiting this shared digital initiative.

HathiTrust’s first priority is long-term preservation of its digital content, but preservation is meaningless without access. Although legal restrictions keep large parts of the repository inaccessible, the partners have leveraged their collective resources to make available over one million volumes to anyone in the world; either those volumes are in the public domain, or rights holders have granted their permission to provide broad access. HathiTrust also stores an increasingly complex set of rights attributes for content it holds. Even where the materials are in copyright and their use is restricted, HathiTrust provides full-text search and page references to aid in discovery. Moreover, the initiative seeks to provide all means of lawful access and so has begun making some preservation-related use of in-copyright materials and provides comprehensive access for a limited number of users with disabilities.

HathiTrust partners are creating new printed books from the collection, with several hundred thousand books held in HathiTrust made available through Amazon and other retailers.

HathiTrust takes the business of sustainability seriously and does so with regard to governance, finances, and technology. HathiTrust has a strong, multi-institutional Executive Committee and a broadly representative Strategic Advisory Board. The technology used is extraordinarily robust, involving several layers of redundancy at each site as well as geographic redundancy and a third instance of tape backup. The Center for Research Libraries is now reviewing HathiTrust for compliance with the Trustworthy Repositories Audit and Certification (TRAC) Checklist. HathiTrust finances receive regular scrutiny from the Executive Committee, and its budgets include built-in replacement costs along with all reasonable elements of keeping the bits alive. And the collaboration and business model
was designed to scale with regard to the number of partners from the beginning. It is also notable that HathiTrust was launched using the collective will of colleges and universities and was not underwritten by a grant.

How Can HathiTrust Make a Difference?
Much of HathiTrust’s work is devoted to ensuring the future viability of this unique cultural heritage institution—the library—through incremental steps.

Digital curation for the long term poses a continuing challenge for research libraries. HathiTrust has quickly demonstrated something that might seem intuitive: collective action drives down costs and enables the discovery of solutions to key problems that face many libraries. Economies of scale have reduced the cost of storage for each library. Moreover, together the partners are able to improve the strength of their archiving efforts; for example, the partners are validating content in ways too expensive to afford in isolation, thus finding and resolving problems in Internet Archive and vendor content. The simple consolidation of talent has been a boon as well. Instead of pursuing the same tools and methods in isolation from each other, the partners share in the job of making decisions about formats and quality, of solving copyright problems, and of resolving bibliographic ambiguity (e.g., supplying missing dates or correcting records). The partners need to solve these problems only once, and the solution can be extended to the benefit of all. The most important collective benefit is basic to the mission of libraries. These benefits are substantial but pale in comparison with the basic user benefit of improved discoverability through consolidation.

HathiTrust can also be used to help develop more effective and efficient print curation. HathiTrust is exploring use of the repository as a means by which the partner libraries can register all of their print holdings. Most of the print volumes held by partners will soon correspond to digital volumes in HathiTrust, and a combination of manual and automated processes should make it possible to link those print holdings records to corresponding HathiTrust records. If the partners can perform record-keeping in a coordinated way, first giving attention to creating reliable and reliably described digital representations, they may be able to coordinate their efforts to store print. Having fewer print copies, stored in better preservation conditions, can help reduce library costs and free physical space.

These subsidiary benefits of bringing content together can fundamentally reshape key areas of library work as HathiTrust partners learn to operate a large-scale, above-campus service. Pooling content and effort affords partners the opportunity to devote collective attention to determining copyright, seeking rights from rights holders, developing methodologies to manage orphan works, and defining paths for fair uses and preservation decisions. Through HathiTrust, members of the library community will be able to quantify problems that have eluded them, such as determining the number of volumes in the public domain or the number of unique titles in research libraries.

Library Work at Scale
Writing about the question of the scale at which library work should be done, Lorcan Dempsey has laid out a three-level taxonomy of “scale.” First, at the institution-scale, “activity is managed within an institution with a local target audience.” Second, at the group-scale, “activity is managed within a supra-institutional domain” (typically geographic in his examples), and the “audience is correspondingly grouped.” And third, at the web-scale, “activity is managed at the network level” where “the activity is potentially all web users.” He explains that these activities can be supplied through services that are “sourced” in ways that are institutional, collaborative, and third party.

The scale of library collaboration is changing. It is changing with economic pressures as libraries can no longer afford to do in isolation what they can do more cost-effectively together. It is changing with unforeseen opportunities as technological advances enable libraries to craft new models of collaborative collection development. It is changing with new priorities as libraries turn their attention to increasingly intensive partnerships with the communities of which they are a part and away from those isolated and isolating activities that occupied them in the past.

Libraries have an unprecedented opportunity to bring together preservation and access in a way that can change the scale at which they work. What our culture has called the “Universal Library” is now more realistic than ever, and the opportunity comes at a critical time for higher education institutions. Through deep collaboration previously only imagined, libraries can now create and maintain a comprehensive digital collection and a well-coordinated and shared print collection. The cost for doing this work can be significantly reduced compared with previous ways of managing collections, allowing libraries to devote increasing amounts of attention to building better services and more dynamic relationships with the teaching and research staff at their institutions. Although the models for multi-institution, above-campus services are still being developed, HathiTrust provides one contemporary example for how that aspiration can be achieved.

Notes

Paul N. Courant (pnc@umich.edu) is the University Librarian and Dean of Libraries at the University of Michigan. John Wilkin (jpwilkin@umich.edu) is Associate University Librarian for Library Information Technology at the University of Michigan and serves as the Executive Director of HathiTrust.

© 2010 Paul N. Courant and John Wilkin. The text of this article is licensed under the Creative Commons Attribution 3.0 Unported License (<http://creativecommons.org/licenses/by/3.0/>).