

Recommendation: OSU philosophically supports the mission and goals of the HathiTrust to preserve digitized books in the public domain; however, the organization does not currently provide enough tangible benefits at this time to warrant the significant cost of membership.

Over the past decade, Oregon State University (OSU) Libraries has developed and executed a thriving digitization program. Through these efforts, OSU Libraries has sought to provide greater access to material uniquely available at OSU and provide a path for long-term preservation for digitized materials. As OSU's digitization program has developed, so too has the libraries' methods for access and preservation evolved to meet the current needs and technologies available to the Libraries. It is in this spirit of ongoing discovery that we undertake this process of investigating and evaluating the HathiTrust Digital Library and the potential benefits and opportunities membership in that organization may bring to the OSU Libraries.

What is the HathiTrust Digital Library?

In some respects, the HathiTrust Digital Library is a confusing organization simply because of the many roles that the organization expects to play in the future. On the one hand, the HathiTrust Digital Library currently represents the best example of a national digital library, in that organizations with digitization programs have the ability to contribute their content to the HathiTrust Digital Library allowing their content to be universally viewable and discoverable through the organization's searching and discovery tools. On the other hand, the HathiTrust Digital Library is a group of organizations that have organized around a shared vision for long-term preservation of digitized items. In creating a consortial organization concerned about the long-term preservation of content, the HathiTrust Digital Library has developed a rigorous set of metadata and standards necessary to upload materials onto the network. These standards ensure a level of homogeneity, enabling better long-term resource discovery.

But what is the HathiTrust Digital Library? To understand the genesis of the organization, one needs to only look back to 2008 and the large-scale digitization partnerships entered into by Google and select public/private libraries around the United States. The HathiTrust and the materials currently preserved and available through the HathiTrust come as a direct result of those initial partnerships. While those partnerships resulted in millions of print documents being digitized (and continuing to be digitized) -- the partnerships with Google spurred other media organizations like Microsoft and cultural heritage organizations like Internet Archive, to begin their own digitization programs resulting in tens of thousands of more digitized documents from library collections around the world. The HathiTrust started as an organization originally dedicated to serving the original thirteen member libraries with large-scale Google digitization partnerships. The HathiTrust was developed to facilitate the bringing together of these

documents as well as a vehicle for sharing best practices related to the long-term preservation of the digitized content.

Today, the HathiTrust has evolved to resemble the initial characteristics of a national digital library. Expanding beyond the original membership, the HathiTrust now includes organizations utilizing homegrown digitization programs, as well as libraries and organizations wishing to provide funding to sustain the growth and development of the project. The digital library includes only materials in the public domain, predominantly pre-1923 publications.

What would it cost?

Until recently, membership in the HathiTrust required that an institution contribute material to the archive. Costs were based on the amount of content contributed.

HathiTrust partners pay the infrastructure costs for the content they deposit. The infrastructure is made up of five elements: storage, data centers, tape backup, servers and miscellaneous hardware, and staff to oversee and maintain these elements. [1]

A new membership model was recently established for institutions that wish to support the work of the HathiTrust as “sustaining members” but have no significant digital content to provide. Utah State, Dartmouth and the University of Washington are examples of members that are not expecting to contribute large amounts of content. This model, as described in a report written by Ann Miller for the University of Oregon (UO) Libraries,

requires interested partners to establish “sharedness” by identifying works held (now or in the past) in common with the HathiTrust. Two separate formulas would establish the cost per member. Those formulas establish the support cost of public domain material and in-copyright material based on the overlap in holdings. [2]

Without doing time-consuming holdings comparisons, it is not possible to provide an accurate estimate of how much it would cost OSU to become a member under this model. It may be sufficient to know that the UO estimates the annual cost to their library of joining the HathiTrust would be nearly \$25,000.

What would we contribute?

If OSU were to join, it would do so as a sustaining member rather than as a content contributor. None of the books that OSU has already digitized meet the HathiTrust scanning requirements. HathiTrust requires files at the page level in the form of TIFFs. Since 2009 OSU has not produced page level TIFFs for documents when a print copy that serves as a master archival copy is owned.

Also, most of the documents that the Digital Production Unit has digitized and has plans to digitize -- unique OSU publications, theses and dissertations and relatively rare natural resources books (e.g. Williams Collection monographs) -- do not meet HathiTrust copyright guidelines because they were published after 1923. Of course, all of these resources are already available open access via the ScholarsArchive@OSU and are already searchable via Google, 1Search, the online catalog, Worldcat and Summit.

There will likely be some expectation that sustaining members contribute to the curatorial work of the HathiTrust, either through committee involvement or by engaging in the work of curating content; however, of the 3 member libraries that we contacted (Dartmouth, University of Washington, Utah State), none are currently involved in committees or curatorial work.

What would we get out of membership?

All of the accessible content contained within the HathiTrust -- approximately 1.7 million books -- is public domain and already freely available online via Google Books, Internet Archive, other digital collections and/or via the HathiTrust database itself. In terms of access benefits, members can receive bibliographic records for the materials to download to their local catalog. Members can also download those freely available resources in their entirety whereas non-members can only read the works online page by page.

Another membership benefit follows from the HathiTrust's mission to "coordinate shared storage strategies among libraries" by allowing member libraries to withdraw books that are in the public domain because they are available digitally and well-curated, saving money for members by reducing the need for large amounts of print storage. However, the HathiTrust only contains digital content for public domain works at this point so there would be no immediate benefits for withdrawing in-copyright works.

Perhaps the biggest, if less immediately tangible benefit for members, is the ability to contribute to the HathiTrust as a public good. According to Tim Jewell, who was instrumental in the University of Washington joining, UW, "thought [the HathiTrust] represented a potentially very important development [for libraries]." They also "wanted a seat at the table when the work of organizing for the long haul started up this year." He refers to the October 31, 2010 deadline for participants to be part of the constitutional convention to be held October 2011. The libraries we spoke with all agree that the HathiTrust represents an important development for libraries as a way to ensure long-term access to important resources, especially since Google, "as a public company [may] eventually have to retire [Google Books] if it doesn't continue to perform financially." (email from Elizabeth Kirk, Dartmouth)

Summary

We agree with the libraries with whom we spoke: The primary benefit of supporting the HathiTrust Digital Library as a member is to monetarily support the work as a public good. We

do not feel that the tangible benefits – ability to download books in their entirety rather than page by page and the ability to reliably withdraw pre-1923 materials that have been digitized by other member libraries – are, in and of themselves, compelling enough reasons for us to join, especially given the relatively high cost of membership and the ways we might otherwise spend that money.

[1] HathiTrust Digital Library – Cost. <http://www.hathitrust.org/cost>

[2] Ann Miller, HathiTrust: An assessment of costs and benefits to participating. Internal report presented to University of Oregon Libraries administration, September 2010.