Final Draft 2009-11-16

HathiTrust Vision for Discovery Interface and Services:

-- for discussion at the November 20, 2009 HathiTrust-OCLC Joint Meeting in Chicago

---

Executive Summary

Over the past year, the HathiTrust has been working with OCLC to develop an instance of WorldCat Local that will function as HathiTrust's online catalog. This instance provides a strong foundation for HathiTrust's goals of providing access and discovery services to HathiTrust's content. The longer-term vision for HathiTrust's catalog is that it *must* integrate or interoperate with HathiTrust services (and open services provided by others), in addition to facilitating discovery and access to its content.

I. Vision for HathiTrust Discovery Interface and Services

The following statements define the high-level goals and vision for the creation of the HathiTrust discovery interface.

The HathiTrust seeks to:

- Create a tailored interface to HathiTrust, one that is contemporary, fully-featured, and clearly branded for discovering content in HathiTrust.

- Create an interface that will not only help the user discover the rich body of resources held in the HathiTrust preservation repository, but will also help partner institutions gain a clear sense of the content preserved in HathiTrust and, over time, its relationship to the collective print holdings of HathiTrust members and participants. We believe we will benefit considerably from building a view into the HathiTrust on top of WorldCat, through which we will have a comprehensive sense of library holdings.

- Forge collaborations (such as with OCLC and others) that will help shape a future that brings together key library services at the network level, not otherwise possible without collaboration or the large-scale aggregation of data.

- Enable a cohesive user experience across all HathiTrust services -- with specific attention to creating near-seamless user flows between interoperating service environments, including interface conventions, authentication protocols, and other "border-crossing" transition points.

- Seize opportunities to go beyond the functionality of a next-gen bibliographic catalog interface to that of a entry point to an entirely digital library, containing content that originally existed in many formats which likely should include such services as an embedded book viewer, the facilitation of full-text search of, and results from, the volume directly from the record, and integration of information about the preservation status of the resource in each record.

- Pursue requirements for an openness to the catalog interface and architecture that allows a flexibility for customer integration of new and existing services, some of which may be identified now and others not yet known.

II. Caveats and Cautions

We would like to offer a few comments about possible points of tension in the attempts to coalesce the roadmaps of WorldCat Local and the HathiTrust:

- Unlike other subscribers to WorldCat Local, HathiTrust will be providing digital access and enhanced services for all of the materials within the collection. This will create some challenging decisions about the user interface, i.e., what to do with the Google Viewer, or other affiliate links. HathiTrust must be part of the discussion about which options are provided within a HathiTrust view.

- As HathiTrust's holdings grow, the means by which these materials are made available will become more important. It becomes critical that WorldCat Local not function as a silo, but as a facilitator for various uses and methods for discovery, including 1) MARC records for HathiTrust holdings should be made available to those who want to put them in their catalogs; 2) HathiTrust materials should be discoverable via WorldCat APIs; 3) HathiTrust materials should be accessible via web discovery methods (e.g., crawlable by seach engines); and 4) HathiTrust materials should be accessible via library discovery methods (e.g., metasearch and z39.50).

III. HathiTrust Future Functional Priorities

HathiTrust exists within a rapidly evolving dynamic environment which necessitates a relatively flexible and nimble approach to development planning. Our principles and functional objectives provide a high-level perspective on directions in which HathiTrust is likely to be heading over the next several years. At the lower, more concrete level of exactly *how* we will be pursuing those objectives, it is difficult to project beyond a couple of years out. When looking at 3, 4, and 5 years out, choices regarding specific directions and approaches will, necessarily, be significantly affected by developments and changes in our environment. Noting this, the following future priorities represent, in part, selected items from the HathiTrust Short- and Long-Term Functional Objectives http://www.hathitrust.org/objectives *and* additional thoughts from the HathiTrust Discovery Interface team regarding priorities for evolving the HathiTrust digital repository platform towards a massively-scaled digital library system:

- **Significant change in the way in which HathiTrust stores and manages bibliographic and item level descriptive metadata:** while there are no specific plans at present, it is probable that we will be migrating that metadata storage and management functionality to a new, dedicated system that is tuned more specifically to meet the needs of HathiTrust. Details of such a new system, from

platform(s) used, to data structure(s) employed, to specific content to be stored and managed, do not yet exist, but we know that such a system would need to integrate more directly with OCLC services. Regardless of the directions those decisions take, however, it seems inevitable that the new system would introduce new and different options for managing the relationship(s) between metadata stored in HathiTrust systems and the metadata describing HathiTrust materials that is stored in WorldCat.

- **Development of an open service definition to make it possible for partner libraries to develop other secure access mechanisms and discovery tools:** We believe that the great wealth of resources that HathiTrust now makes available can only be effectively exploited through the creation of an open service definition that makes it possible for others to create new tools and approaches to access. As a first step, we intend to create a parallel production system that does not compromise the content in the repository, and gives developers access to the functions of the HathiTrust repository system. We hope that the availability of this development sandbox will make it possible for partner institutions to collaborate in creating new services through, for example, new or expanded APIs. The HathiTrust Data API is an example of this. A draft functional specification of the API has been completed, and is available now for public comment at http://www.hathitrust.org/data_api.

- **Development of APIs that will allow partner libraries to access information and integrate it into local systems individually:** The HathiTrust partners identified the need for a mechanism by which a bibliographic identifier (e.g., an ISBN or OCLC number) can be submitted to a HathiTrust Rights API and resolved as a persistent URL with information about levels of access (e.g., full text or search only). A preliminary version of such an API has been released and is in use in the online catalogs of several partners. For more information, see http://www.hathitrust.org/rights_api. A more fully-featured and robust version of this API is currently in active development. A second API, known as HathiTrust Data API, is available to provide secure access to HathiTrust data and metadata resources. Making these resources available to client applications (examples of current applications are the HathiTrust Collection Builder and Pageturner) will enable the creation of additional services and uses of repository materials. Other similar APIs will be developed as needed in the future. It is important to note here that we are not seeking a requirement for adding the Collection Builder (or other) functionality to WorldCat Local itself. Rather, a prospective requirement would be to make it possible for HathiTrust to do integration work to facilitate this service as HathiTrust may support it locally.

- **Development of data mining and other tools for HathiTrust and use by HathiTrust of other analysis tools from other sources:** Because of the vast bodies of content held by HathiTrust, an important function of the HathiTrust repository will be to support data mining and other forms of large-scale analysis. What then is the relationship between the discovery and repository layers as relates to external services to provide post-discovery data manipulation, analysis and management support?

- **Page turner mechanism:** A page turner has been deployed for all content in HathiTrust. We hope to report soon on a strategy to re-engineer the current page turner application so that it provides access to materials in HathiTrust through an API. The intention is to provide a wider variety of functions or modes of access to the collections than are currently available. It is important that the HathiTrust's

primary discovery interface leverage this API to allow for an embedded viewing experience.

- **Support for formats beyond books and journals:** HathiTrust's first "content" priority is support for digitized books and journals, but we believe that HathiTrust must expand its support to other formats (particularly born-digital publications) and materials. This is an area of future work.

- **Integration of Full-text Searching:** As HathiTrust releases large-scale cross-repository full-text searching, questions are raised about the integration or orchestration of full-text search with the bibliographic discovery mechanism that is baseline to the implementation today. Might these be presented as discrete, well-understood services? Might they be presented in intelligent ways that are sensitive to user needs and workflow?

- **Access mechanisms for persons with disabilities:** HathiTrust has deployed an interface for visually impaired users (optimized for use with JAWS and other screen readers). This interface presents to the user the entire text version, with navigation, on one screen. Staff members at the University of Michigan are currently working with UM School of Information interns to optimize this interface for use with screen readers, as well as the general accessibility of the pageturner. For in-copyright resources, access is currently limited to authorized users at the University of Michigan. We plan to add Shibboleth support to the HathiTrust repository so that resources like access mechanisms for persons with disabilities can tie into the authentication environments of our partner institutions.

- **Full Control over User Interface:** The unique content profile of the HathiTrust changes many of the assumptions of current library catalogs and many discovery interfaces. Implicit in the HathiTrust vision is to push the envelope of current interfaces, and services to craft a new, holistic vision for how to provide access to the contents of the HathiTrust. This extends far beyond crafting metadata requirements, and basic interface configuration, but, for example, replacing existing search tools with ones capable of searching the contents of books; adjusting the display of results to highlight full-text contents; integrating full-text annotation, collection management, or other custom services into the interface; or simply focusing the user on HathiTrust content.

- **Ability to publish virtual collections:** Vast bodies of digital content benefit from methods to gather together subsets into "collections" that can be searched and browsed. HathiTrust has created an early release of a Collection Builder that permits individuals to create public (i.e., shared) and private collections. We will turn our attention to creating mechanisms by which persons such as bibliographers can create and share collections with a more formal identity (cf. imagine having full text resources associated with classic bibliographies such as the Wing or Pollard and Redgrave short title lists). We are now performing intensive usability review on the Collection Builder. Although the Collection Builder's authentication and authorization now relies on the University of Michigan "friend" system (see http://www.itd.umich.edu/itcsdocs/s4316/), we will work to add Shibboleth support to the HathiTrust repository so that resources like the Collection Builder can tie into different authentication environments.

- **Commitment to ongoing collaborative assessment**: As user needs change and technology evolves, collaborative ongoing assessment of discovery services and user needs is critical.