

HathiTrust Strategic Advisory Board Meeting
Friday, October 23, 2009
10:00 a.m. – 4:00 p.m.
California Digital Library Office

Present: John Butler (recorder) , Patricia Cruse, Bernie Hurley, Bruce Miller, Ed Van Gemert (chair), John Wilkin (ex-officio)

Not Present: Sarah Pritchard, Paul Soderdahl

1. Introductions

- a. Ed Van Gemert welcomed Bernie Hurley, University of California, Berkeley, to the Board. He is replacing Robin Dale, who has left the University of California, Santa Cruz.
- b. E. Van Gemert asked that meeting attendees be mindful of members not present, especially around decision- making.

2. Google Partner Summit

- a. SAB members John Wilkin, E. Van Gemert, and B. Hurley attended the Summit, which was held in Palo Alto earlier this week. A Project Leads meeting preceded general meeting. Noted was the attendance of many more institutional representatives – new partners – than in past meetings.

3. Development and operations updates (J. Wilkin)

- a. As background to the development and activities of HathiTrust to date, J. Wilkin noted early agreement to leverage the work at Michigan as a starting point for establishing HathiTrust operations. As such the *HathiTrust Short- and Long-Term Functional Objectives* (<http://www.hathitrust.org/objectives>) should be viewed as just that -- a starting point from the SAB develops goals and a roadmap.
- b. Referring to the *Objectives*, Wilkin noted that considerable progress has been made in most targeted areas, including those development areas considered long-term. Highlights he noted included:

- i. Large-scale full-text search, which uses Apache Lucene Solr technology, is ready for public launch
- ii. HathiTrust page turner application (noting additional development underway at the University of California)
- iii. Public HathiTrust catalog (temporary version using the VuFind software)

One of the objectives listed as short-term – Non-Google Ingest – has presented underestimated complexities and difficulty, and little progress has been made. To address pressing needs for non-Google ingest, J.Wilkin noted the operations staff will “brute force methods” for now, deferring the development of tools to support a distributed pre-ingest file validation methodology to a later time.

Regarding size of operations, J. Wilkin noted that there are approximately 10 Michigan staff members on UM (i.e., not HathiTrust) budgets with varying levels of FTE commitment.

4. Discussion: Understanding the Strategic Advisory Board’s role and prioritizing work

A number of key questions were posed at the beginning of the discussion:

- a. How do we identify and prioritize requirements for development, in a *collective* context? Are principles needed to guide roadmap development?
- b. What is the relationship between the HathiTrust SAB and the HathiTrust Executive Committee (EC), noting that J. Wilkin serves on the EC and E. Van Gemert serves in an ex-officio capacity on the EC. It was noted that the EC's main focus is on fiduciary and partnership/membership responsibilities, and for general matters of executive oversight for HathiTrust.
- c. What are the current and forecasted needs for SAB attention and input?
 - i. J. Wilkin provided a list of areas that he would raise to the SAB's attention:
 1. Development of models for growing and maturing HathiTrust, laying groundwork for the long-term growth and sustainability.
 2. Quality filtering and assurance issues
 3. Duplicates in the repository (5% level now)
 4. Policy and access issues for users with disabilities (advancing opportunities to provide access under a legal framework)
 5. Evaluation of HathiTrust in time for the Constitutional Convention, scheduled for March 2011.
 6. Transitioning HathiTrust operations from a service relationship model to a partnership model where the work is done collectively with collective priorities?
 - ii. In an open discussion, these additional needs were identified:
 1. Reporting mechanisms
 2. Further development and articulation of technical metadata standards
 3. Audit mechanisms for gauging health of content
 4. Collection development strategy (is there a need for a collections working group?)
 5. Enhancing the case for institutions to begin shifting efforts from the "local build" to collaborative development
 6. Establishing a way for the SAB to work together, making decisions, and taking action (are principles for setting priorities needed?)
- d. What is the current set of operational working groups?
 1. Storage
 2. Computational Research
 3. HathiTrust-OCLC Discovery Interface Development
 4. Third Instance
 5. Development Sandbox
 6. Error Rate

5. Lunch Discussion with Staff of the California Digital Library

An open discussion about the growth, development, and possible futures of the HathiTrust took place including points on the importance of the effort bringing distinctive value to the larger environment, guiding by principles and strategy. Value proposition areas discussed included preservation services and user services. Specific points made on these areas included:

- a. Leveraging the opportunities that HathiTrust provides for scaling up (and out) integrative action, including collective print/digital collection management strategies (i.e., with a focus on scholarly curation)
 - i. Can we build a collection management framework around print and digital holdings of the HathiTrust partners? ? It is a priority of the CDL to think collectively about collection development within the context of HathiTrust. What are the principles for growing the collection?

- b. Reducing silos at many levels
- c. Integrating full text search services into discovery workflows
- d. Increase efforts to aggregate and deliver public domain holdings at U.S. research institutions
- e. Explore additional service interoperabilities with Google

6. SAB Work Agenda Development

The Board synthesized the input from the earlier discussion and framed a work agenda. Four rubrics of interest and concern were identified. Under each, specific objectives or activities were listed. With this framework and initial detailing in place, the Board took a non-binding straw poll with each member given multiple votes to distribute. The result of this overall work, including the polling) follows:

- a. Managing and curating the cultural record, including preservation**
 - i. Quality filtering (*18 votes*)
 - ii. Auditing (health of content) (*14*)
 - iii. Coordinated inventory control of print and digital holdings associated with HathiTrust; e.g., items for S. 108 defense, etc. (*14*)
 - iv. Duplication – (*13*)
 - 1. Policy may be needed
 - v. Administrative reporting (inventory and origination) (*13*)
 - vi. Capturing more of the cultural record – growing the collection (*11*)
 - vii. Copyright management of the repository (see other listing) (*11*)
- b. Building distinctive and valued services that leverage the unique aggregated content of HathiTrust**
 - i. Services for users with print disabilities (*17 votes*)
 - ii. Non-consumptive research (*12*)
 - iii. Provide electronic access to materials that qualify for S. 108 exemptions (*11*)
- c. Understand the diverse scholarly environment and how to integrate HathiTrust's content and services to satisfy scholarly requirements.**
 - i. Strategic integration of HathiTrust into WorldCat and other environments – (*16 votes*)
 - ii. Scholarly and behavioral assessments (*14*)
- d. Governance / trustworthy organization**
 - i. How do we work collectively, as an alternative to operating under a service relationship (*17 votes*)
 - ii. Convention (*15*)
 - iii. Moving from local builds to collectively built environment (*15*)
 - iv. Copyright outreach to attract more resources to the public domain (*12*)

In a review of our discussions, framework development, and polling, we identified the following omissions that may need incorporation:

- Work with institutions to establish policy, legal, and implementation readiness for specific authorizations in the HathiTrust environment (e.g., full access for qualifying sight impaired users)
- Identification and prioritization of core end-user resources and services (e.g., the value of a union catalog for HathiTrust and/or the provision of cross-repository full-text search.

- e. Metadata priorities, standards, and best practices

7. Next Steps

- a. Form a working group, consisting of UC and CIC members with requisite knowledge, to define and inventory the foundational elements of HathiTrust
- b. Form a working group to define a process/framework for fostering and enacting joint development
- c. Pilot process for addressing missing foundational elements.
- d. Meeting face-to-face was recognized as valuable to certain aspects of the SAB work and, therefore, we committed to scheduling face-to-face meetings once or twice per year, with the possibility of coinciding with ALA attendance or another national meeting, in addition to monthly conference calls.

8. Specific Action Items

- a. J. Butler will send out the charge to the HathiTrust-OCLC Discovery Interface Development
- b. J. Wilkin will bring forth to the EC a proposal to have the HathiTrust-OCLC Discovery Interface Development working group report to the SAB.

9. The meeting was adjourned