The Hathi Trust Research Center and tool builders

John Unsworth (with Beth Plale, Scott Poole, Robert McDonald, and others)
Project Bamboo
Corpora Space Workshop II
Maryland Institute for Technology in the Humanities
1-1:20 pm, June 6, 2011
Our raison d'etre

- Phase I: starting Apr 2011 and going for 18 mos.
- Phase II: starting Fall 2012 and going for ...
- Goal: enable strong computational research and education on a collection that has not been amenable to computational exploration EVER before!
More formally

- HTRC is founded as a joint venture between Indiana University and the University of Illinois Urbana-Champaign, aimed at solving the difficult challenges of increasing computational access to the public domain and copyrighted material in HathiTrust.
HTRC will:

• Maintain repository of text mining algorithms and retrieval tools available on-line for human and programmatic discovery. Also register derived data sets, indexes, and versions in registry repository.

• Be a user-driven resource, with an active advisory board, and a community model that allows users to share algorithms and tools.

• Support interoperability across collections and institutions, through use of inCommon SAML identity.
Non-consumptive research

• One of HTRC’s unique research challenges is support for non-consumptive research. Google Book Settlement defines non-consumptive research as “research in which computational analysis is performed on one or more books, but not research in which a researcher reads or displays.”

• In practice, we believe this will mean HTRC users should not be able to export copyrighted materials from the research environment.

**Long-term goals**

- Support *innovation in cyberinfrastructure* to deliver optimal access and use of HathiTrust corpus.

- *Implement “Non-consumptive” research*: a technical and intellectual challenge

- *Identify and host existing data analysis, text mining and retrieval tools* that are of interest to the community.

- *Stimulate development of new analytical methods and tools*. We hope that the scale of the HTRC will promote new levels of collaboration in tool development.