Building a Public Research Center for the HathiTrust Digital Library

@hathitresearch | @hathitrust

http://www.hathitrust-research.org

Robert H. McDonald
Associate Dean for Library Technologies and Digital Libraries
Associate Director-Data to Insight Center, Pervasive Technology Institute
Indiana University

May 17, 2011
DPLA Global Interoperability and Linked Data Workshop
HathiTrust Digital Library History

- To contribute to the common good by collecting, organizing, preserving, communicating, and sharing the record of human knowledge.
- Launched in October 2008
  - University of Michigan
  - Indiana University
- Used Google Books Repository at Michigan as model
- Expanded to include content from
  - CIC Member Libraries
  - UC System Libraries
  - University of Virginia
- Now includes more than 50 partner institutions
Towards an HT Research Center

- Started in response to proposed Google Settlement - June 2009
  - Specific Funding set aside by Google to build a public research center
  - Worked to identify key stakeholders from HT institutions to collaborate and write RFP
- Developed specific RFP for HathiTrust to solicit proposals – Summer/Fall 2009
  - HTRC RFP Working Group
- RFP Released – Winter 2010
HTRC RFP Working Group

http://www.hathitrust.org/documents/hathitrust-research-center-rfp.pdf or

- Kat Hagedorn-HathiTrust
- Jeremy York-HathiTrust
- Melissa Levine-HathiTrust
- John Wilkin-HathiTrust
- Steve Abney-Michigan
- Jack Bernard-Michigan
- Geoffrey Fox-Indiana
- David Goldberg-California-Irvine
- Robert H. McDonald-Indiana
- Qiaozhu Mei-Michigan
- Shawn Newsam-California-Merced
- John Ober-California Digital Library
- Beth Plale-Indiana
- Scott Poole-Illinois
- Sarah Shreeves-Illinois
- John Unsworth-Illinois
HathiTrust Research Center Today

- HTRC is dedicated to the provision of access to a comprehensive body of published works for scholarship and education for computational research purposes.
- Lightweight Organization
  - Executive Committee
    - Beth Plale, Indiana
    - Scott Pool, Illinois
    - Robert H. McDonald, Indiana
    - John Unsworth, Illinois
  - Advisory Board
    - TBD
  - HT Executive Committee Sponsor
    - Laine Farley, California Digital Library
HathiTrust Research Center Today

- $250K in funding for initial 18 month startup
- Creating Themed Collections for early Use Cases
  - Astronomy – Victorian Literature - Influenza
- Ingest and Replication Mechanisms Between HT and HTRC
  - Full-text
  - SOLR indexes
  - Data Capsule integration
  - Karma integration
- Integration with SEASR SOA services at NCSA
  - Alignment with Bamboo Technology Project
  - Possible alignment with DPLA? or Europeana?
- Establishing long-term non-consumptive research methodologies
HTRC Proposed Technical Architecture

Courtesy IU Data to Insight Center – Beth Plale/Yiming Sun
HathiTrust Research Center Futures

- HTRC Kickoff Event at Digital Humanities Conference 2011
  - Stanford University - June 20, 2011
- HTRC at JCDL 2011
  - Ottawa, Canada – June 13-17, 2011
- Working on models for collaborative research
  - NEH/JISC/NSF Digging into Data Round 2
    - http://www.diggingintodata.org/
- Working on Collaborations and Integrations for the HT Corpus
Network Graph of those on twitter connected to topic of Big Data

By mrflip (Phillip Kromer) on twitter, Attribution-ShareAlike 2.0 Generic (CC BY-SA 2.0)