HATHITRUST

A Shared Digital Repository

Defining and Identifying the GovDocs Corpus: the HathiTrust Registry

Federal Depository Library Conference
May 1, 2014
Valerie Glenn, HathiTrust Government Documents Registry Analyst,
University of Michigan

Overview

- What are we trying to accomplish?
- Why a Registry?
- Activities to-date
  - Scope
  - Metadata
  - Functional requirements
  - More!

What are we trying to accomplish?

- 2011 Constitutional Convention – Ballot Initiative #4
  - Resolved: “that HathiTrust facilitate collective action to create a comprehensive digital corpus of U.S. federal publications including those issued by GPO and other federal agencies”
  - Resolved: that “HathiTrust initiate and carry out a planning process to coordinate operational plans and a business model to further sustain coordinated digitization, ingest, and display of U.S. federal publications including those issued by GPO and other federal agencies”
  - Resolved: “that HathiTrust develop a process of catalog record review to ensure accurate and full display of U.S. federal publications including those issued by GPO and other federal agencies”

Why a Registry?

- A registry of **metadata** for the comprehensive corpus of US federal documents.
- Envisioned as a stand-alone resource, separate from the HTDL, and open to all

Scope

- “The Registry is intended to include metadata for the comprehensive corpus of U.S. federal documents. This will include materials produced at U.S. government expense, **in all formats, at the item level**, from 1789 to the present.”

Metadata Analysis

- Sources
  - Libraries, agencies, publications
- Schemas
  - MARC; Dublin Core; MODS; Custom XML
- Minimum required for a registry record:
  - Title
  - Agency OR SuDoc number
  - Publication date

Metadata Comprehensiveness

- Comprehensiveness
  - Agency list
  - Current HTDL content (sets, agencies)
- Relationship detection
- Gap detection

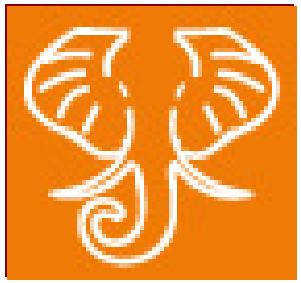
Functional requirements

- Use cases
- Focus groups

Future Plans

- Continue working to automate ways to identify relationships between items, along with gaps in metadata coverage
- Build the registry (hire a developer!)
- Identify ways in which the community can help & find/build platforms to support that

Thank you! Questions?



How to find out more

- HathiTrust US Federal Government Documents Initiative: http://www.hathitrust.org/usgovdocs
- Creating a Registry of US Federal Government Documents: http://www.hathitrust.org/usgovdocs_registry
- Contact me: valglenn@umich.edu