Archives’ User Studies & Archival WorldCat Records

Jennifer Schaffner
Program Officer

Jackie Dooley
Consulting Archivist

2009 Annual RLG Partnership Meeting
Boston
Archives and Special Collections Program

Overview
• Suite of OCLC Research projects supporting archives and special collections

Today’s Focus
• Discovery of archives and special collections
• Data mining of one million archival WorldCat records
Focusing our attention

Funding

- Council on Library and Information Resources (CLIR)
- Mellon Foundation
- National Historic Publications and Records Commission (NHPRC)
- National Endowment for the Humanities (NEH)

Timing

- Library of Congress *On the Record* recommendations

Transfer effort into higher-value activity. In particular, expand the possibilities for knowledge creation by exposing to more users rare and unique materials held by libraries that are currently hidden from view and, consequently, underused.

- Committee on Archives, Museums and Libraries (CALM)
- ARL Special Collections Working Group
- Continued importance to the RLG Partnership
Current Work

Managing the Collective Collection
  Shared Print Collections
  Data-mining for Management Intelligence

Research Information Management
  Support for Research Processes
  Workflows in Research Assessment

Mobilizing Unique Materials
  Archival Program
  Museum Program

Knowledge Structures
  Structure for Controlled data
  Metadata Workflows

Shared Infrastructure
  Web enablement
  Grid services
Effectively Disclose Archives and Special Collections

Our goal:
- Managing Archival Collections
  - Analyze the Archival Descriptive Practice
- Discovery Environments to Optimize User Success
  - Characterize the State of "Hidden Collections"
  - Optimize Delivery Practices
- Improve "End-to-End" Archival Workflow
  - Increase the Scale of Special Collections
  - Identify Barriers to EAD Creation
  - Improve OCLC Services for Archives & Special Collections

Effectively Disclose Archives and Special Collections
Analyze **Discovery Environments** to Optimize User Success

- Survey specialized discovery environments in which archival materials currently appear
- Synthesize user studies
- Analyze search log data from the environments to determine user behaviors and expectations
Analyze Discovery Environments to Optimize User Success
Analyze Discovery Environments to Optimize User Success


De Rosa, Cathy, et al., College Students’ Perceptions of Libraries and Information Resources, Dublin, Ohio: OCLC, 2006. Published online at: <http://www.oclc.org/reports/perceptionscollege.htm>


Freeman, Elsie T., “In the Eye of the Beholder: Archives Administration from the User’s Point of View,” American Archivist 47, no. 2 (spring 1984): 111-123.
The Metadata is the Interface: Better Description for Better Discovery of Archives and Special Collections

Jennifer Schaffner

Program Officer
OCLC Research
Analyze **Discovery Environments** to Optimize User Success

**Next steps**

- Collect logs of successful searches that lead to archival collections ("find logs")
- Compare and contrast with the results of datamining MARC records for archival materials in WorldCat
- Combine analysis to make recommendations to optimize metadata creation for discovery
- Are there possibilities for data remediation?
Data Mining of One Million Archival Records in WorldCat

• Ultimate objective: Improve discovery of archival materials
  • In all search environments, not just in WorldCat
• Specific objectives:
  • Evaluate patterns of existing practice
  • Combine with discovery analysis to optimize metadata creation
    • Are we including the words that people want to search?
    • Can we simplify record creation?
    • Are there possibilities for data remediation?
    • Determine characteristics for effective relevance ranking of searches
Data Mining Methodology

- Software developed that can ...
  - Count occurrences of tag groups, fields, subfields
  - Construct complex queries using all Boolean operators
  - Graph usage pattern within and across institutions
  - Display content of selected fields and subfields
  - Select randomized query results for analysis
  - Be extensible for use with other data sets
Data Mining Methodology

• Ask questions that reveal, for example ...
  
  • Extent to which records conform to archival standards
  • Extent to which records include access points
  • Extent to which significant fields are used, or not
  • Distribution of holding institutions across the community
  • Nature of full vs. minimal-level records
A Few Preliminary Results

- **Demographics**
  - 93% are held by U.S. institutions
  - 36% are minimal-level records
  - 57% indicate which cataloging rules were used

- **Description**
  - 72% include scope & content note
  - 36% include biographical/historical note
  - 12% have restrictions note
A Few Preliminary Results

- Access points
  - 86% have a principal creator (main entry)
    - 58% are personal names (100)
    - 28% are corporate names (110)
  - 22% have inadequate titles (Papers; Records)
  - 48% have genre/form added entries
  - 33% have personal name added entries
    - 15% have only one occurrence
    - Records exist with up to 466 occurrences!
  - 11% have corporate name added entries
    - 8% have only one occurrence
    - Records exist with up to 466 occurrences
    - 5% include an organizational subunit
Questions? Ideas? Feedback?

Jennifer_Schaffner@oclc.org

Jackie_Dooley@oclc.org