



# 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
  - Increase the Scale of Special Collections Digitization
    - Identify Barriers to EAD Creation
  - Improve OCLC Services for Archives & Special Collections
- Improve "End-to-End" Archival Workflow**

# 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

Online Archive of California - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.oac.cdlib.org/



gapingvoid: "cartoons ..."

## About the OAC

- Learn more about the OAC
- Virtual collections
- List of contributing

Northwest Digital Archives - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://nwda.wsulibs.wsu.edu/

gapingvoid: "cartoons ..."



home

NWDA provides digital reproductions of primary sources available in some cases.

search NWDA

what will I find?

search

more search options

University of Montana-Missoula  
R. H. McKay with camera on tripod, undated.

12

13

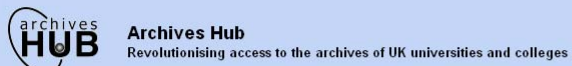
Transferring data from nwda.wsulibs.wsu.edu...

Archives Hub - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.archiveshub.ac.uk/

gapingvoid: "cartoons ..."



Helpdesk email [archiveshub@mimas.ac.uk](mailto:archiveshub@mimas.ac.uk)  
Phone +44 (0)161 275 6789  
Please fill in our [feedback form](#)

**Archives Hub**  
Revolutionising access to the archives of UK universities and colleges.

Quick Search  
[Browse Indexes](#)  
[Subject Finder](#)

A national gateway to [descriptions of archives in UK universities and colleges.](#)

[frames](#) | [Help](#)

**Manchester Histories Festival**

very first [Manchester Histories Festival](#), an exploration of Manchester's past. The Festival is a one day public event held in Manchester Town

butors are involved in the Festival. This special feature highlights how includes links to their collections on the Hub.

skyline. Public domain image sourced from [Wikimedia Commons](#).

city Library are temporarily unavailable. We apologise for any

[Advanced Search](#)  
[Preferences](#)  
[Language Tools](#)

calames



Google Search

I'm Feeling Lucky



Catalogue en ligne des archives et des manuscrits de l'enseignement supérieur

Mot(s) clé(s) :

Bibliothèque(s) :

- Toutes les bibliothèques
- ANGERS, BU Lettres et Sciences

# Analyze Discovery Environments to Optimize User Success

- Cox, Richard, "Researching Archival Reference as an Information Function: Observations on Needs and Opportunities," *RQ* 31, no.3 (spring 1992): 387-397.
- Cox, Richard J., "Revisiting the Archival Finding Aid," *Journal of Archival Organization* 5, no. 4 (2007): 5-32.
- De Rosa, Cathy, et. al., *Perceptions of Library and Information Resources*, Dublin, Ohio: OCLC Research, 2005. Published online at: <  
<http://www.oclc.org/reports/2005perceptions.htm>>
- 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>>
- Dempsey, Lorcan, "The Library Catalogue in the New Discovery Environment: Some Thoughts," *Ariadne* 48 (July 2006). Published online at: [www.ariadne.ac.uk/issue48/dempsey/](http://www.ariadne.ac.uk/issue48/dempsey/).
- DeRidder, Jody L., "Googlizing a Digital Library," *The Code4Lib Journal* 2 (March 24, 2008). Published online at: <http://journal.code4lib.org/articles/43/>
- Dowell, Erika, "Web Site Usability for Rare Book and Manuscript Libraries," *RBM* 9, no. 2 (fall 2008): 168-181.
- Dowler, Lawrence, "The Role of Use in Defining Archival Practice and Principles," *American Archivist* 51 (winter/spring 1988): 74-95.
- Dooley, Jackie M., "Subject Indexing in Context," *American Archivist* 55 (spring 1992): 344-354.
- Dryden, Jean, "Do We Care What Users Want? Evaluating User Satisfaction and the LibQUAL+™ Experience," *Journal of Archival Organization* 4, no. 4 (2004): 83-88.
- Duff, Wendy et al., "Archivists' Views of User-based Evaluation: Benefits, Barriers, and Requirements," *American Archivist* 71, no. 1 (spring/summer 2008): 144-166.
- Duff, Wendy M. and Johnson, Catherine A., "Accidentally Found on Purpose: Information-Seeking Behavior of Historians in Archives," *Library Quarterly* 72, no. 4 (2002): 472-496.
- Duff, Wendy and Stoyanova, Penka, "Transforming the Crazy Quilt: Archival Displays from the Users Point of View," *Archivaria* 45 (spring 1998): 44-78.
- 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



A publication of OCLC Research

**Analyze  
Discovery  
Environments to  
Optimize User  
Success**

# 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](mailto:Jennifer_Schaffner@oclc.org)

[Jackie\\_Dooley@oclc.org](mailto:Jackie_Dooley@oclc.org)

