

CIC metadata aggregation: A collaborative initiative to enhance metadata sharing

Timothy W. Cole (t-cole3@uiuc.edu)

Muriel Foulonneau (mfoulonn@uiuc.edu)

Grainger Engineering Library
University of Illinois at Urbana-Champaign

*Charlottesville, VA
8 November 2005*



The project

- Sept. 2003 – June 2006
- All 13 CIC institutions are participating
 - Currently 19 data providers with 2 more repositories in test
 - Aggregation contains 520,000 resource descriptions
- Objectives
 - Support CIC mission of regional collaboration
 - Investigate issues surrounding metadata shareability
 - Test OAI – PMH as technical infrastructure for metadata sharing
 - Provide Web interface(s) to CIC aggregated material

Context of & motivations for collaboration

- Follow-up to Mellon-funded experiments with OAI
 - Had shown need for collaborations to build competencies & capacity and serve as catalyst to encourage institutions to implement OAI (M. Halbert et al., ECDL 2003)
 - 2 CIC institutions (Illinois & Michigan) had been involved
 - Homogeneity of CIC institutional missions provided opportunity to explore high level of interoperability
- Participants shared need to enhance discoverability and reusability of library-published digital & digitized content
- Attractiveness of shared funding model



Key Accomplishments

- Technical Validation of CIC OAI-PMH implementations
- Advances in metadata sharing
 - ➔ created incentive for data providers to modify the way they share metadata by showing the direct impact possible
 - ➔ increased understanding of DP & SP roles; demonstrated (again) dependence of each on the other
 - ➔ provided input to broader community initiatives – e.g., DLF-NSDL best practices for OAI and shareable metadata
 - ➔ demonstrated critical importance of retaining context and providing reliable, direct linkages to resources



Evolution in CIC metadata sharing

- Aggregation item count has grown by 50% so far during project
- Broader adoption of optional protocol features
 - More metadata formats
We reprocess UDC, QDC and MODS records for items;
MARC and DC-Collection records for collections
 - Increased use of OAI sets & <setDescription> element
Inclusion of collection-level descriptions within <setDescription>
Now harvesting set descriptions and transforming them as
collection level descriptions
- More consistent & robust links to resources being described
 - Enhanced user experience
 - Value-added services (e.g., automated capture of thumbnails)



CIC collaboration as a test bed for DPs

- Data providers are interested in impact and importance of what they do or don't do
 - Do I need to use resumption tokens?
 - How might including collection-level descriptions enhance the discoverability of my content?
 - What are the benefits of exposing metadata in multiple formats?
 - Which URL(s) for the resource should be shared in the metadata records?

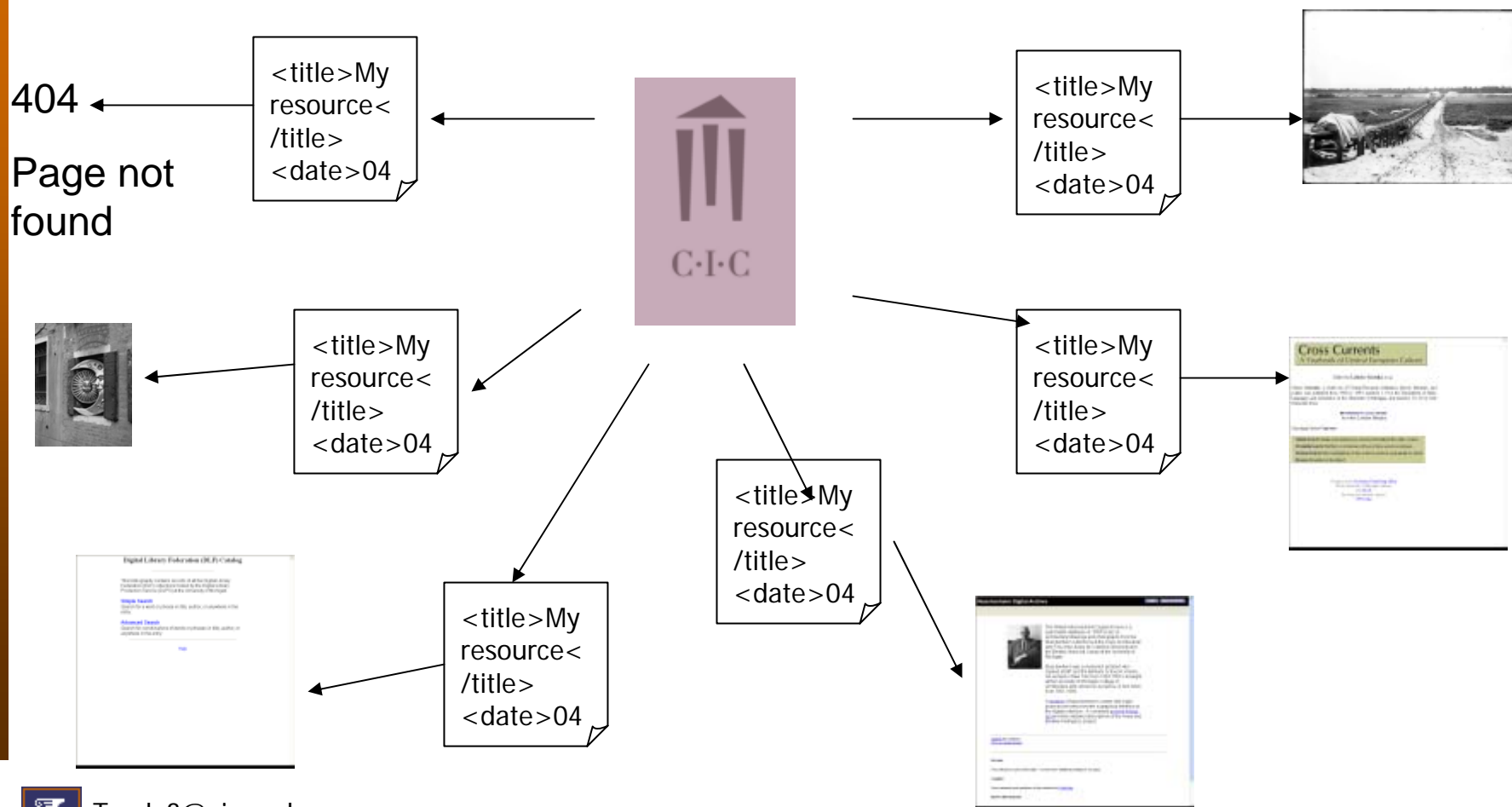


How much normalization can SP do?

Concept	% of records in the repository
Type	99%
Format	66%
Language	60%
Collection	100%
Resource URL	72%
Temporal Coverage	71%
Spatial Coverage	21%

These results may or may not be generalizable for other SPs

Examining resource URLs embedded in metadata



Generating or including thumbnails

CIC-OAI Metadata Search Portal

Home Search View Institutions Help

Your search was in **all fields** for **"russian"**. The search was limited to the **"University of Illinois Library DIMTI"** archive.

You found **83 records**.

* [Revise your search](#) to retrieve fewer records.

* View your results, starting with **records 1 to 25 of 83**.

Results by Sort by [Next 10 Records](#)

Jump to Records: [1](#) | [26](#) | [51](#) | [76](#)

Indiana University:
Indiana University Digital Library Program
[90 records](#)

Michigan State University:
Michigan State University Digital and Multimedia Center
[52 records](#)

Ohio State University:
Ohio Link Electronic thesis and dissertations
[10 records](#)


Pennsylvania State University:
Pennstate Libraries Image Collections
[3 records](#)
Pennstate thesis and dissertations
[2 records](#)

University of Chicago:
University of Chicago Library
[4 records](#)

University of Illinois at Urbana-Champaign:
Engineering Documents Center Collection
[1 record](#)
University of Illinois Finding Aids
[49 records](#)
University of Illinois Sheet Music
[24 records](#)
University of Illinois Library DIMTI
[83 records](#)

Record 1 of 83

Title



Dievotchka pokupaet yarenye grushi. Une petite fille a chetant des poires. Ein kleines mÃdchen kauft Birnen.

Author/Creator (unknown)

Description (unknown)


Image was captured using Epson ### scanner and SilverFast import utility in Adobe Photoshop 6. Scanned at 42->24 Bit Colour at 300 dpi. No archival images were created, but access images were created as JPEGs with quality of 0.

URL http://images.library.uiuc.edu:8081/RussianPublics/image/97752201452003_fonar60-5.jpg

Collection [Early 19th Century Russian Readership & Culture](#)

Record 2 of 83

Title



Dvornik i Trubochist. Le portier et le ramoneur. Der Hausknecht und der Schornsteinfeger.

Author/Creator (unknown)

Description (unknown)

Menu

[search](#)

Partner section

Show all results

More info

More info

More info

9541 matches like this ...View 🧐🧐

[More info](#)

title	Arquing (What Else?) about Jewish-American Poetry
creator	Laurence Goldstein
type	Book Review

and 1288 items matching

and 475 items matching

and 173 items matching

The Database of African-American Poetry covers the works of 54 African-American poets writing in the eighteenth and



What we've not been able to do (yet)

- Adequate involvement of end-users – will require
 - Better integration with traditional & commercial resources
 - Development of focused portals --> manual classification or automatic clustering across institutional boundaries
 - Normalization / cross-walking across vocabularies; i.e., repurposing metadata for new use
- Bring collaborators to the same level of sophistication
 - Wide range of technical solutions / dependence on vendors
 - Wide range of descriptive practice & richness (considerable static content not being further developed)



References

- Halbert, M., Kaczmarek, J., Hagedorn, K., Findings from the Mellon Metadata Harvesting Initiative, ECDL 2003
<http://www.springerlink.com/app/home/contribution.asp?wasp=6590eb7adfc74a07acf9a25e6c518fcd&referrer=parent&backto=issue,7,47;journal,780,2099;linkingpublicationresults,1:105633,1>
- Foulonneau, Muriel and Timothy W. Cole. Strategies for reprocessing aggregated metadata. *ECDL 2005*
- Foulonneau, Muriel, Timothy W. Cole, Thomas G. Habing, Sarah L. Shreeves (2005). Using Collection Descriptions to Enhance an Aggregation of Harvested Item-Level Metadata. In JCDL 2005
- CIC metadata portal <http://cicharvest.grainger.uiuc.edu>
- DLF-NSDL best practices for OAI and shareable metadata wiki