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

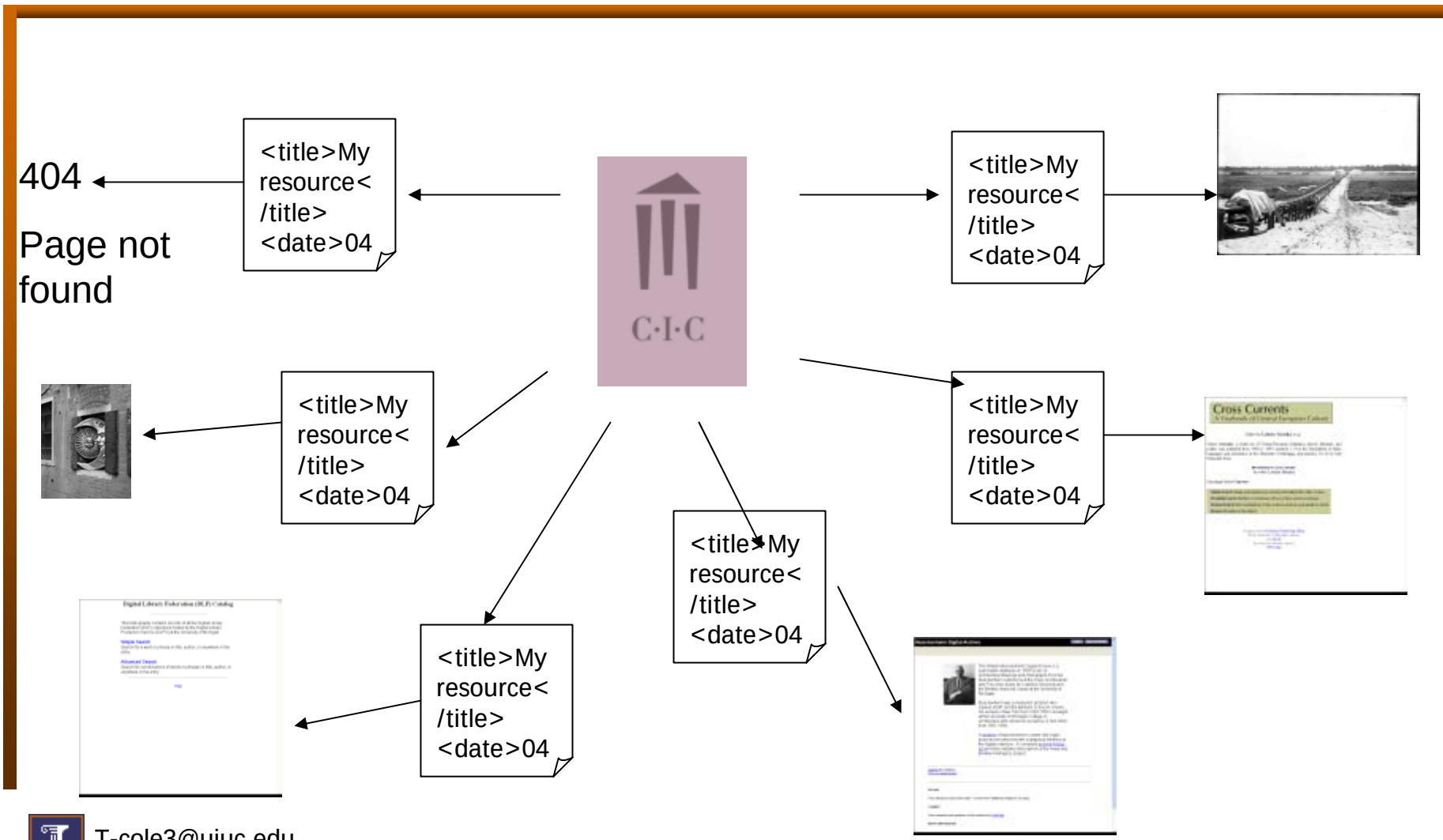
- Data providers are interested in impact and importance of what they do or don't do
  - Do I need to use resumption tokens?
  - How might including collection-level descriptions enhance the discoverability of my content?
  - What are the benefits of exposing metadata in multiple formats?
  - Which URL(s)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

How much normalization can SP do?

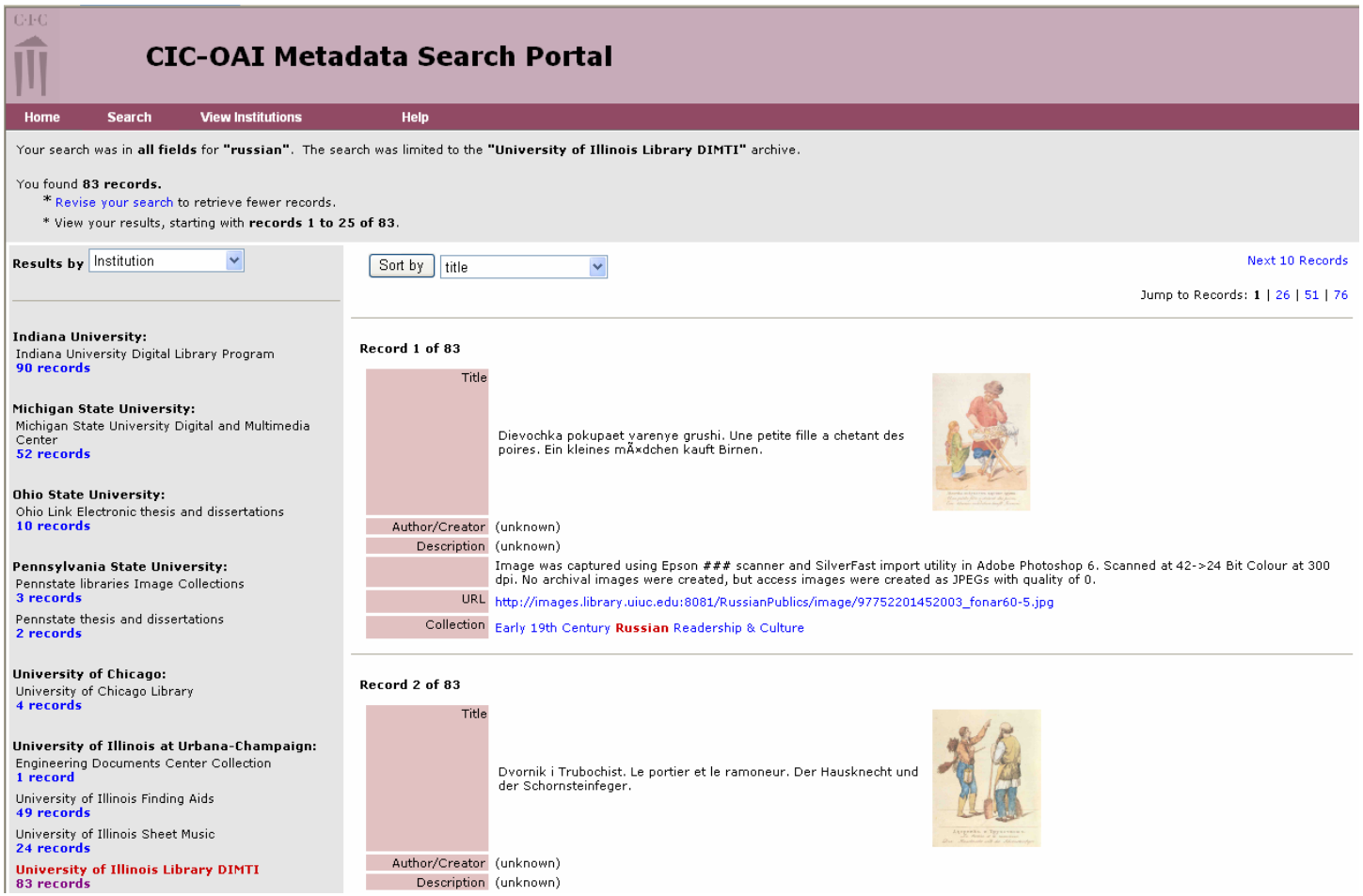
Concept	% of records in the repository
Type	99%
Format	66%
Language	60%
Collection	100%
Resource URL	72%
Temporal Coverage	71%
Spatial Coverage	21%

These results may or may not be generalizable for other SPs

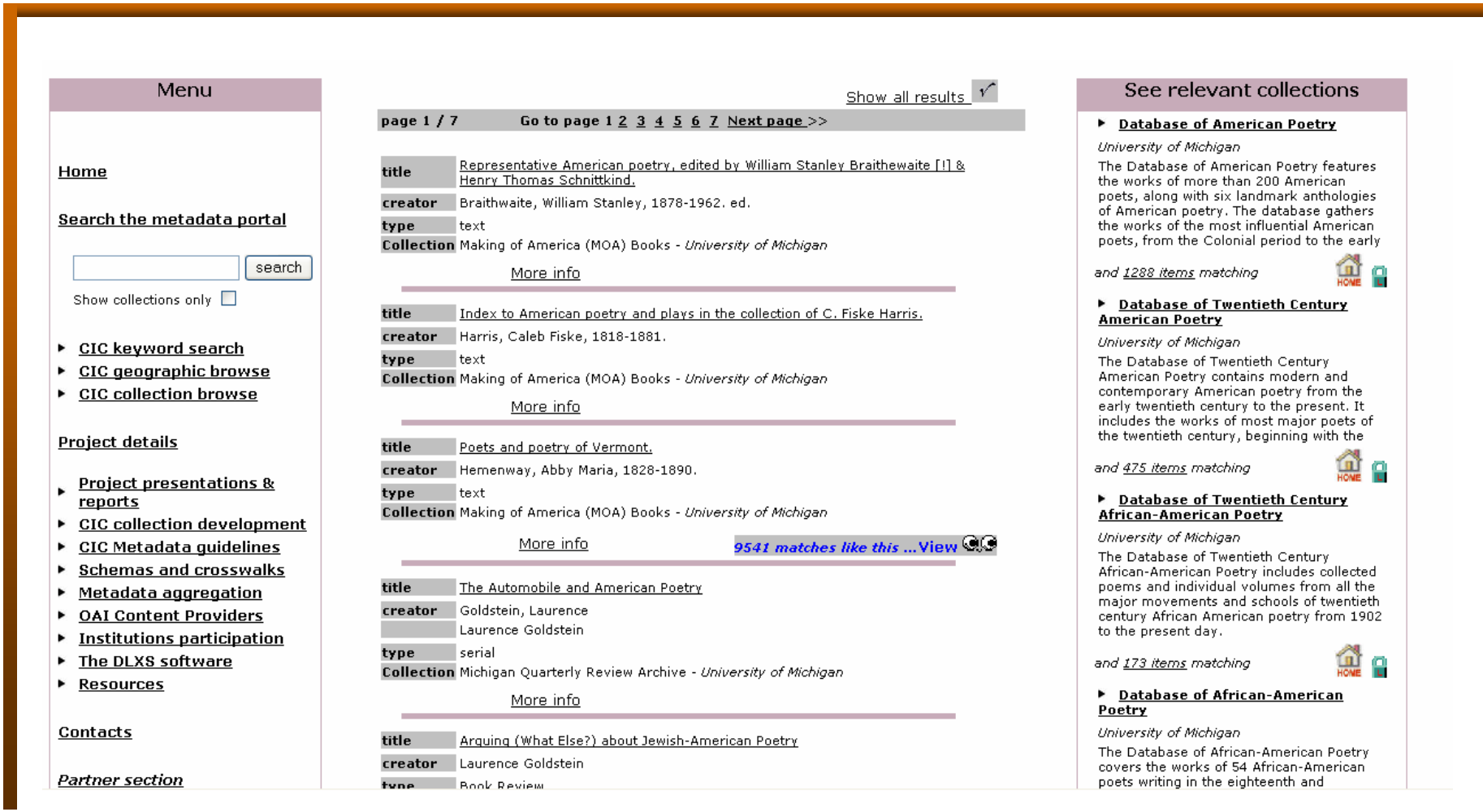
Examining resource URLs embedded in metadata



Generating or including thumbnails



Using CLD to enhance discoverability



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