

IMLS Digital Collections and Content

Grant LG-06-07-0020



Interim Performance Report 3 **1 October 2008 – 31 March 2009**

Submitted by:

Carole Palmer, Principal Investigator
cpalmer@illinois.edu

Amy Jackson, Project Coordinator
(217) 244-7809
amyjacks@illinois.edu

April 2009

University of Illinois at Urbana Champaign

**Graduate School of Library and Information Science
Center for Informatics Research in Science and Scholarship
University of Illinois at Urbana-Champaign
501 E. Daniel Street, MC-493
Champaign, IL 61820-6211
(217) 333-1981**

**Grant LG-06-07-0020
Interim Performance Report 3
1 October 2008 – 31 March 2009
Submitted by
Carole Palmer, Principal Investigator
Amy Jackson, Project Coordinator
April, 2009**

General Project Activities

The performance period of October 1, 2008 through March 31, 2009 began with the release of Opening History: U.S. History Resources from Libraries, Archives, and Museums (<http://imlsdcc.granger.uiuc.edu/history>). The portal was seeded with 161 collection records for resources relevant to U.S. History from the IMLS Digital Collections and Content portal (<http://imlsdcc.granger.uiuc.edu>), and approximately 300,000 items from these collections. Additional collections and items were added during the performance period with a total of 531 collection records and 918,037 item records by the end of March 2009.

In early October 2008, Carole Palmer and Tim Cole travelled to Jackson, Mississippi to attend the Chief Officers of State Library Agencies (COSLA) meeting and introduce the state librarians to the IMLS DCC and Opening History projects. State librarians were encouraged to participate and contribute to the resource to assist in building a resource of national scope. State librarians from Georgia, Utah, Idaho, Ohio, New York, Louisiana, Indiana, and Illinois responded to the call for participation, and work progressed to incorporate digital collections and items from each of these states.

The project team held a two hour conference call with Advisory Board members on February 20, 2009. The purpose of this call was to update board members on current project activities and solicit feedback about future directions. Each of the project's subgroups led discussion areas during the call, and advisory board members provided useful feedback and advice. Subgroup topics can be found on the project blog at <http://imlsdcc.wordpress.com/>. Subgroups corresponded with advisory group members on key topics in advance of the meeting and increased overall interaction during the meeting through the use of web conferencing software.

Digital Collection Evaluation (DiCE) subgroup

The DiCE subgroup (Carole Palmer, Oksana Zavalina, Amy Jackson, and Richard Urban) held regular meetings to discuss collection development and evaluate the subject concentrations of the evolving collection.

The chart below shows evolution of the collection subjects:

Subject	March 2008 – IMLS DCC Portal* (202 collections)	February 2009 – Opening History (499 collections)
Asian Americans	10 collections	40 collections
Military history	25	50
Native Americans	12	32
Universities and colleges	14	34
Education	24	43
Transportation	11	28
African Americans	12	27
Exploration and travel	4	19
Religion	12	26
Economics	22	35
Mining	5	17
U.S. Presidents	3	14
Architecture	19	29
Cities and towns	21	31
Agriculture	11	20
Women	7	16

*does not include 2007 NLG collections

As of February 2009, geographic coverage changed as below:

- Utah – 47 new collections
- Georgia – 33 new collections
- Nevada – 18 new collections
- Idaho – 15 new collections
- Illinois – 11 new collections
- Southwest U.S. (general region) [no specific state information] – 8 new collections
- Mountain Region U.S. (general region) [no specific state information] – 6 new collections
- D.C. – 5 new collections
- Tennessee – 5 new collections
- Virginia – 5 new collections
- California – 4 new collections
- Oregon – 3 new collections
- Pacific Coast U.S. (general region) [no specific state information] – 3 new collections
- Colorado – 2 new collections
- New Mexico – 2 new collections
- South Carolina – 2 new collections
- Southern U.S. (general region) [no specific state information] – 2 new collections
- Arizona – 1 new collection
- Connecticut – 1 new collection
- Hawaii – 1 new collection

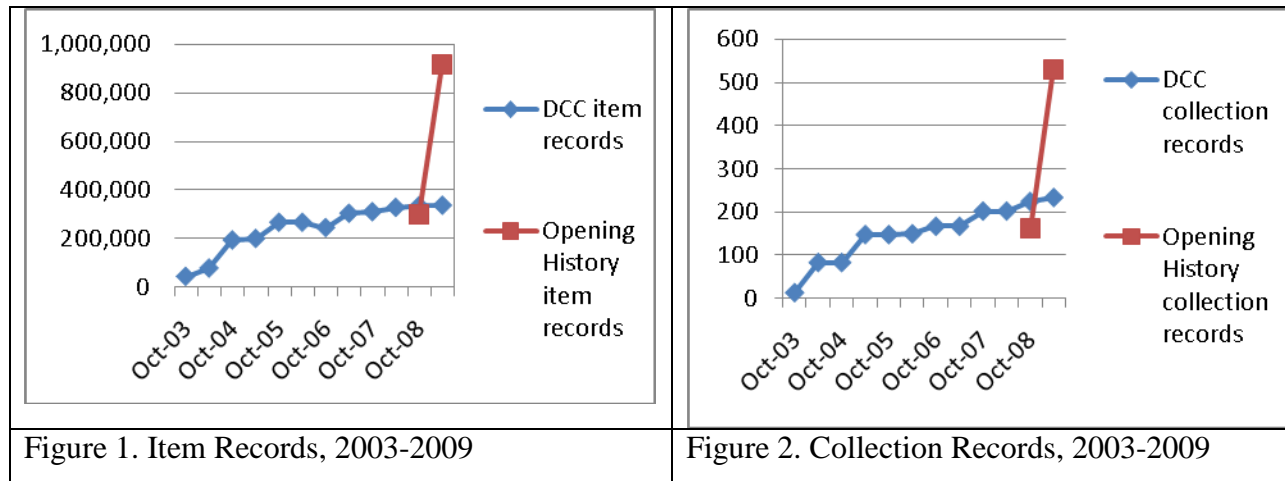
- New York – 1 new collection
- North Carolina – 1 new collection
- Rhode island -- 1 new collection
- West Virginia – 1 new collection

At the close of the performance period, Opening History contained 531 collection records and 918,037 items. An additional 36 collection records are password protected for vetting by collection administrators.

The DiCE subgroup also drafted a collection development policy for Opening History and asked the advisory board for feedback. The final policy is posted on the website (<http://imlsdcc.grainger.uiuc.edu/docs/CollectionDevelopmentPolicy.pdf>) and contains the following mission statement:

Opening History (OH) provides organized access to digital resources of value for research on United States history and culture. It aggregates a range of distributed and complementary cultural heritage collections from libraries, museums, and archives to increase their visibility and to enhance the value and usefulness of individual collections by integrating them with related collections, providing search and browse functions across collections, and linking to related resources outside of OH. The OH initiative encourages sharing of digital resources in open access formats and promotes coordinated access to regional, state, and local collections to support the creation of a digital cultural heritage aggregation of national scope.

The following figures show the number of item and collection records in DCC and Opening History from October 2003 through March 2009.



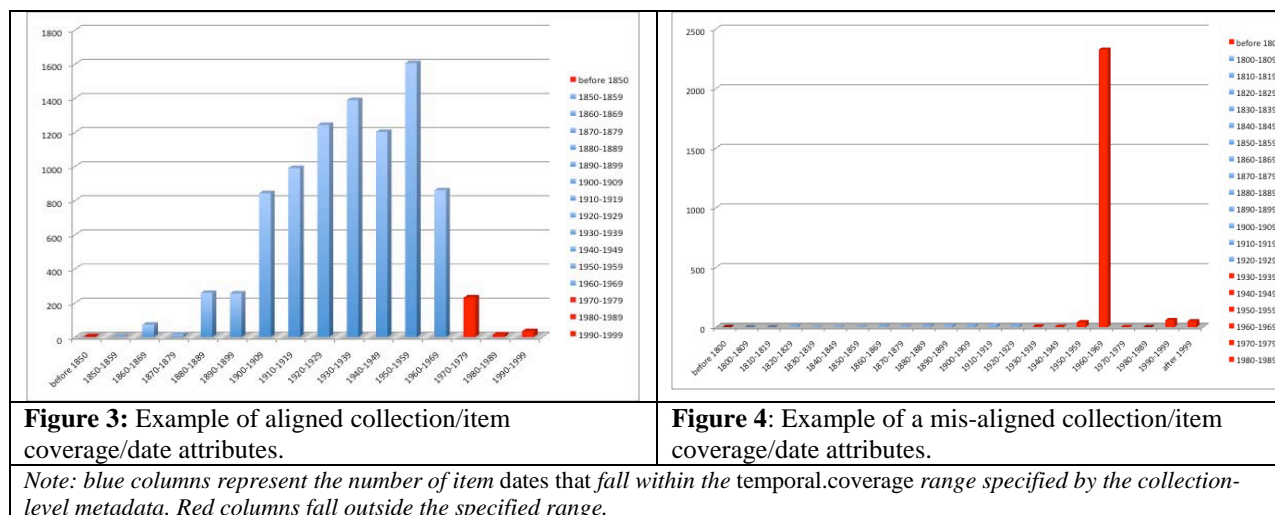
Collection and Item Metadata Relationships (CIMR) subgroup

The CIMR subgroup, led by Co-PI Allen Renear, held weekly meetings through the fall and winter. Graduate research assistants Richard Urban (coordinator), Karen Wickett, and Wu

Zheng continue to play key roles in this work. Other regular contributors included Tim Cole, Dave Dubin, Amy Jackson, Michael Twidale, Oksana Zavalina, and other members of the DCC research team.

During this period our research focused on looking for evidence confirming conjectured metadata *propagation relationships* (Renear, et al., 2008) within instances of IMLS DCC collection-level and item-level metadata. Based on research conducted by the DiCE research group, we placed special attention on spatial and temporal metadata elements. Most of the cases examined in fact were consistent with the conjectured relationships. For instance, Figure 1 presents a histogram showing that the values of item-level *date* statements closely matched the date range of the corresponding collection-level record, which has a *temporal.coverage* of 1850-1969.

However we also discovered metadata distributions inconsistent with conjectured relationships. In Figure 2, a different collection-level record with a *temporal.coverage* of 1800-1929 results in a histogram of item-level *dates* with a large number of items found with an item-level *date* of 1960-1969. There are several possible explanations for variations of this sort. One is that our understanding of the intended semantics of DCMI metadata is wrong, although in the cases inspected this seems unlikely. Another is that metadata creators misunderstand the intended semantics of DCMI metadata elements, or apply these elements inconsistently. In other cases, the problem appears to be that projects have difficulty maintaining and current and accurate collection-level metadata as the collection grows. In some cases, the provider failed to refine OAI-PMH set specifications associated with collections when new materials were added to a repository. Further investigation to discover which of these explanations apply, in what circumstances, and in what proportions, is underway.



Zheng (2009) also explored hidden connections that exist between item-level metadata records by adapting language-based discovery methods currently used to identify “hidden collections” in

medical literature. While only applied to a single collection, this work shows promise for identifying other unexpected relationships among items and collections.

In addition Karen Wickett has confirmed that the level of expressiveness required to implement propagation rules will require Semantic Web Rule Language (SWRL) rules as well as OWL modeling. (Wickett, 2009).

In March 2009, the CIMR group began to lay the foundations for a testbed that allows us to explore a more formal set of inferencing rules based on the previously indicated categories as well as better continue our empirical studies of how actual metadata use conforms to those rules. A subset of DCC collection-level and item-level metadata will be translated into RDF triples that enables the use of OWL-based ontologies and SWRL rules. GSLIS Research Programmer Amit Kumar will be joining the CIMR group for these activities.

Metasearch

Metasearch functionality was integrated into the Opening History portal during the performance period. Although access to the databases is only available from campus IP addresses, users outside of campus can see databases that have results relevant to their query. Findings from the IMLS grant, “Demonstration of Portal Mechanisms for Enhanced Resource Integration in the Academic Information Environment” at UIUC will be integrated into the Opening History metasearch application during the next performance periods. Additionally, conversations about differences between searching images and full-text resources have resulted in brainstorming additional access points for metasearch results, such as the browse results screen or individual item screens. This feature would be similar to “If you like this item you may like this other item” found on many e-commerce sites, and may be integrated into the portal in future iterations.

Interface development

The interface subgroup (Sarah Shreeves, Michael Twidale, Tim Cole, Richard Urban, Amy Jackson, Hong Zhang, Oksana Zavalina, and Wu Zheng) met several times a month during the performance period to discuss updating the look and feel of the user interface. The subgroup explored interfaces for related projects and discussed features and functionality that we would like to see incorporated into the new design. The Research Programmer transformed the site from embedded HTML tables to CSS DIVs so that future design elements will be more modular. Additionally, Michael Twidale and Richard Urban developed a patchwork prototyping model that will be demonstrated at the 2009 Museums and the Web conference (see http://www.archimuse.com/mw2009/abstracts/prg_335002112.html), the UIUC Graduate School of Library and Information Science Research Showcase, and the HASTAC III Conference held at UIUC in 2009.

Registry/Repository maintenance

At the end of the performance period Opening History contained 531 collection records and 918,037 item records. IMLS Digital Collections and Content contained 232 collection records and 337,016 item records. Data was transferred during this period to a Microsoft SQL 2008 server for enhanced reliability and speed. The structure of the item level database was reworked to accommodate the three-fold increase in item-level metadata available in Opening History.

Dissemination

Presentations and Publications, October 2008-March 2009

Zavalina, Oksana, Carole L. Palmer, Amy S. Jackson, Myung-Ja Han. 2008. Evaluating descriptive richness in collection-level metadata. In *Journal of Library Metadata*. Vol. 8 no.4, 263-292.

Wickett, Karen. 2009. Logical Expressiveness of Semantic Web Languages for Bibliographic Modeling.

Displayed at iConference, 2009. University of North Carolina, Chapel Hill, February 8-11, 2009.

Zheng, Wu. 2009. Exploring Hidden Connections among Historic Images. Displayed at iConference, 2009. University of North Carolina, Chapel Hill, February 8-11, 2009.

Palmer, Carole, Timothy W. Cole. 2008. IMLS Digital Collections and Content. Presented at Chief Officers of State Library Agencies Fall 2008 meeting. Jackson, Mississippi, Oct. 6-8, 2008.

Upcoming/Submitted Presentations and Publications

Urban, Richard, Michael Twidale, Amy Jackson. 2009. Patchwork Prototyping a Collections Dashboard. Abstract for a demonstration at Museums and the Web, 2009. Indianapolis, Indiana, April 18, 2009. Available at http://www.archimuse.com/mw2009/abstracts/prg_335002112.html.

Han, Myung-Ja, Christine Cho, Timothy W. Cole, Amy S. Jackson. 2009. Metadata for Special Collections in CONTENTdm – How to Improve Interoperability of Unique Field through OAI-PMH. *Journal of Library Metadata* (submitted for review).

Urban, Richard, Mike Twidale. 2009. Patchwork Prototyping a Collections Dashboard. Presented at the University of Illinois Graduate School of Library and Information Science Research Showcase, Champaign, IL, April 6, 2009.

Twidale, Michael. 2009. Patchwork Prototyping an IMLS DCC Collection Dashboard. Presented at HASTAC III, Champaign, IL, April 21, 2009.

Project publicity

IMLS DCC Project blog: <http://imlsdcc.wordpress.com>

Opening History Marketing blog: <http://sowingculture.wordpress.com>

Zavalina et al. (2008) DCMI presentation was mentioned in:

M. Curran. 2009. Cataloging News. *Cataloging & Classification Quarterly*, 47(2), 171-207.

Palmer et al. (2006) "Collection definition in federated digital resource development" was cited in: B. Stvilia, and C Jörgensen. 2009. User-generated collection-level metadata in an online photo-sharing system. *Library and Information Science Research*, 31(1), pp. 54-65.

Renear et al. (2008) "Collection/Item Metadata Relationships" will be cited in: Greenberg, J. (In press). Automatic Metadata Propagation and Co-curation in the Dryad Repository. *Cataloging & Classification Quarterly*. (47)4.

Jackson et al. (2008) "Dublin Core Metadata Harvested through OAI-PMH" was required reading for L&I SCI 714 (Metadata) at the University of Wisconsin at Milwaukee School of Information Studies.

Related activities

We continued to hold the metadata roundtable study group with a frequency of 2-4 times a month. Recent roundtable topics have included PBCore metadata; IMLS DCC usability studies; and Folktales, facets, and FRBR. The website, which includes a full listing of the metadata roundtable topics and background readings, can be found at: <http://www.isrl.uiuc.edu/~dcc/mdrt.html>.

References:

Renear, A.H., Urban, R.J., Wicket, K.M. Dubin, D. and S.L. Shreeves. (2008). Collection/Item Metadata Relationships - Allen H. Renear, Karen M. Wickett, Richard J. Urban, David Dubin, Sarah L. Shreeves. Presented at DC 2008 International Conference on Dublin Core and Metadata Applications, Berlin, Germany, September 24, 2008.