Investing in Collection Representation for More Useful Repositories

Carole L. Palmer

Center for Informatics Research in Science and Scholarship (CIRSS)

Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign

ASIS&T Annual Meeting
19-24 October 2007, Milwaukee, Wisconsin
How do we get more value from the growing body of digital content?

• Investment in decades of opportunity-driven “projects”

• Not yet realized the collective value of the many, often specialized, distributed collections
  – What content is complementary?
  – How to improve our ability to use collective digital resources?

• Integrated access to digital collections one viable strategy
1. IMLS Digital Collections and Content Project (DCC)

   Investigating process and problems of aggregating digital materials with a registry-repository / harvesting approach

   Collections remain important, as collections, not just as aggregations of items

3. Shift from critical mass to “contextual mass” in collecting collections

6. Key role of collection level representation for enhancing development and use
Development aim: integrated access

Digital content from IMLS National Leadership Grant program (& some LSTA projects)

Collection registry
202 collections from libraries, museums, archives, historical societies, etc. funded from 1998 -

Metadata repository
Harvested metadata - 328,210 item-level records

Assistance for projects to develop shareable metadata.
Research aim: investigate “aggregating”

Metadata:

• Range and evolution of practices & interoperability issues

  Tension between local practices / needs and the more global potential of digital collections

• How to best represent items and collections to meet the needs of service providers and diverse user communities

Collections:

• Role of individual collections within a federation
Critical mass and usability are not enough

Important gains:  http://imlsdccc.grainger.uiuc.edu/

Centralized base of unique cultural heritage resources

Integration of materials from smaller institutions—
museums, historical societies, public libraries, archives, botanical
gardens, etc.—
with more numerous university based special collections.

More awareness of metadata best practices, quality, sharing

Collection description schema based on DC and RSLP

As we will see, not yet adequate

But, as it aggregation grows it becomes more nebululous as a “collection”

What’s in it?  What’s it good for?
4 core problems of scale and granularity

1. Lack of cohesion
   - IMLS-funded content not adequate as criteria for inclusion

   • Flat representation of items
     - all items equal, strengths of concentrations not evident
     - small window into large, diverse accumulation of content

4. Diminished “intentionality”
   - identity of individual, purposeful collections not evident enough

6. Low functioning metadata relationships
   - Normalization at item level and refinement of collection level, but item/collection metadata relationships not understood, fully exploited

Solutions in traditional and emergent collection principles
1) Cohesion - strategic remediation

Adhere to collection development fundamentals:

Conspectus-like assessment to determine strengths and potentials

Selection criteria based on potentials in terms of

1) aims of institution
   – to build significant national cultural heritage resource

2) needs of user groups
   – academic libraries / scholars primary intended audience

Inclusion of complementary non-IMLS content
   - made more difficult by lack of access to collection level descriptions
“Contextual mass” approach

First identified in CLIR/DLF study of humanities scholars (Brockman, Neumann, Palmer, & Tidline, 2001)

- Pull and value of traditional library subject collections
- Evidence, “lead-to-lead” driven nature of personal collecting
  Conceivably becoming more valuable to researchers than collections found at many large libraries
  (e.g., Blake Archive or Monuments and Dust for cultural study of Victorian London)

• Size is not a priority

• Emphasis on principled selection and integration of sources that work together to support research area or community of researchers:

  Aim is multiple “working” scholarly collections
2) Representation - intermediate units

Operationalize Lee’s (2000) collection (aggregate) as information seeking context. Will require making explicit related and emergent collections, subcollections.

Subject strengths **by items:**
- United States
- people
- songs with piano
- trees
- archeology of the United States
- Work Progress Administration
- cities & towns
- Women
- photographers
- mountains
- men
- archaeological site
- insects
- bodies of water
- shrubs

Subject strengths **by collection:**

Social Studies (80% of collections):
- U.S., state, world history
- U.S. government
- urban studies
- anthropology
- geography …

Arts (46% of collections):
- visual arts
- photography
- popular culture
- architecture
- music
- history of art

Very different views, neither adequate
3) Intentionality – retain and optimize

Numerous large collections providing raw materials with aim of leaving interpretation to other services and users (Lynch, 2002)

We aim to retain and optimize interpretations inherent in collectors’ acts of collocation
- DCC collections include “exhibits”, “tours”, “events”

Collection descriptions show purposeful design:

Further enable materials to function as evidence (Buckland, 1999)
- like secondary sources, already processed and refined

“explore”, “demonstrate”, “provide insight into”
“record of Lincoln's career”
“document distinctly American approach to natural science”
“detail how housing policy changes the cities we live in”
4) Metadata relationships - formalization

- Collection metadata can establish scholarly significance of an item: But many properties irreducible & non-inducible

  aspects of completeness, uniqueness, representativeness (of a period or style), developed according to some systematic method (or not), heterogeneous with respect to genre or type of object, etc.

Working toward what can be propagated automatically
Renear (2007) conjectures:

- Many collection level features can’t be inherited or converted to item level features – (paintings vs. comprehensive)

- Nor can collection level features such as “comprehensiveness” be induced in any simple way from features of the items
References


Acknowledgements

DCC project team members:

University Library Co-PIs:  Tim Cole
Sarah Shreeves
Bill Mischo

GSLIS Co-PIs  Allen Renear
Mike Twidale

Project Coordinator:  Amy Jackson

Research Assistants:  Oksana Zavalina
Richard Urban

This research has been funded by IMLS, NLG grant LG-02-02-0281

Questions and comments always welcome: clpalmer@uiuc.edu
End

Thank you