DLF-DCC Beta Sprint

and Linked Open Data (Appendix)

Richard J. Urban

IMLS Digital Collections and Content Project

Graduate School of Library and Information Science

University of illinois at Urbana-Champaign

Open Licenses and OAI-PMH Data

- The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) provides a mechanism to provide license information about a metadata set.

- However, according to Foulonneau & Shreeves, 2004 (http://hdl.handle.net/2142/80), less than 20% of published OAI sets included a setDescription. This research did not examine how many sets included license information.

- IMLS DCC currently takes a hands-on approach to this problem by directly contacting data providers before harvesting metadata. This is a time consuming task that increases the costs of aggregating metadata.

Implications for DPLA

- In order to reduce costs, new automated methods of sharing licenses are needed, especially when an open license is assumed.

- Proposed 4-star classification for linked-open cultural metadata. http://goo.gl/KZxcF

- Need risk analysis of using data when licenses are not available. How much good-faith effort is required for sets missing licenses?

Collection/Item Relationships

- Traditionally IMLS DCC has stored collection descriptions and item descriptions separately. Linked Data provides the opportunity to merge collection-level and item-level graphs into a single contiguous graph.

- Organizing item descriptions through a collection graph can preserve contextual information that augments our understanding of items, improves search, and builds on "linked" data principles.

Implications for DPLA

- DPLA infrastructures should be designed to take advantage of collection descriptions that serve as organizing nodes in a linked network.

- Institutions and other organizational structures are also part of the provenance and trust of cultural heritage metadata. Currently libraries, archives and museums are inconstantly represented as entities in the linked data cloud. DPLA infrastructures should support full and complete representation of the organizations that are the backbone of a national collection.

Metadata Quality

- A major theme throughout the IMLS DCC project has been the value of quality metadata.
    - Metadata that uses literal values from a controlled vocabulary is easier to reconcile with linked data representations of that vocabulary.
    - Uncontrolled, but consistent, values are easier to reconcile with a variety of sources.
    - Metadata values that conform to a value encoding scheme are easier to represent as typed literals

Metadata Quality

- The semantics of the RDF model impose new requirements for logical coherence that were only loosely imposed by OAI-PMH XML schemas. Additional work is needed to understand how to preserve the implicit semantics of source metadata.
  - Which Identifiers are best used as for linked data?
  - What are the optimal translations of colloquial OAI-PMH XML into more formal RDF models?
  - Support RDF models tailored to communities of practice.

Implications for DPLA

- The DPLA Technical Workstream needs to provide leadership in several areas
    - OAI-PMH development has grown stagnant and is not keeping pace with current developments. What is the next-generation aggregation infrastructure for a national digital library?
    - How can we build on the existing OAI-PMH installation base?
    - How can the DPLA provide motivation for participating organizations, vendors, and community service providers to move forward?

Implications for DPLA

- Should DPLA and other aggregators augment contributed/harvested data in order to produce Linked (Open) Data?
  - Normalize values?
  - Reconcile with existing Linked Data
  - Contribute augmented metadata as Linked Data
  - Open licenses should propagate through augmentation
- The DPLA should support tool-builders who contribute applications that reduce the costs of this work.

Beta Sprint Partners

GRADUATE SCHOOL OF **LIBRARY AND INFORMATION SCIENCE**
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

CIRSS Center for Informatics Research in Science and Scholarship

DLF
Program of the Council on Library and Information Resources

http://cirss.lis.illinois.edu

http://www.diglib.org

And the many libraries, archives and museums that participate in IMLS DCC

Thanks to the support of:

The
Andrew W. Mellon
Foundation

INSTITUTE of
Museum and Library
SERVICES