

Pre-Conference #2 – Sharing Images and Data: Making Access to Collections Easier and Better

Search Interoperability, OAI PMH, and Metadata: an introduction to the OAI protocol for metadata harvesting - Sarah Shreves, University of Illinois at Urbana-Champaign

A partial solution: The OAI Protocol for metadata harvesting provides a relatively low barrier means for integrated access to the metadata describing items in these collections.

Why share metadata?

- Can't expect users will know about your collection and remember to visit it
- Users expect to go to one place

Search interoperability: “the ability to perform a search over diverse sets of metadata records and obtain meaningful results” (Priscilla Caplan)

2 models – (1) federated search for sharing metadata – searches distributed databases directly (Z39.50, SRU/SRW) – can take a long time to get results; (2) data aggregation model: the user searches a pre-aggregated database of metadata from diverse sources – then goes to re-harvested – (search engineers, union catalogs, OAI PMH) ***can use both models together

OAI-PMH is a Tool – moves metadata from a data provider to a service provider; a set of rules that defines communication between two systems (like FTP and HTTP); facilitates aggregation of metadata (like a union catalog); developed in 2001 out of the e-print/pre-print community (see <http://gita.grainger.uiuc.edu/registry>)- became widely used protocol

How it works: Databases of metadata (XML files, databases, digital management system, etc.), then OAI data provider works uniquely on this metadata for each institution, software sends a request over HTTP (only 6 options are allowed), response comes back from data provider as XML, does some processing with it based on standards, creates aggregated metadata, and services are overlaid on this for user access.

OAI Data Provider Software

- included in digital content management systems (DSpace, Digital Commons, eprints, etc.)
- open source OAI data providers in different flavors (java, PHP, Perl, etc.)
- static repository option (flat xml that is accessed through a gateway)

Examples of OAI PMH Service Providers:

- OAIster
- CIC Metadata Portal
- IMLS Digital Collections and Content
- Also, Google will use OAI PMK to index your content; “google oai sitemap”

OAI-PMH and Dublin Core

- DC (15 element dataset) is OAI PMH's lowest common denominator BUT OAI PMH supports and encourages use of other community-driven metadata schemas (MODS, MARC XML, etc.)

OAI is not a metadata format, not a search tool, not a database, is open access
You control...what you share, how you share, who you share with (can put behind authentication system)

Metadata challenges: "obtain meaningful results" is challenge for aggregators – the metadata record may not give enough information to identify what the material really is.

Shareable Metadata

- is quality metadata
- promotes search interoperability
- is human understandable outside of its local context
- is useful outside of its local context (can we build something off of it?)
- preferably is machine processable!

Create metadata for interoperability – see "Best Practices for Shareable Metadata"

RECAP

- OAI PMH is a tool
- OAI PMH is easy – metadata can be hard
- Better metadata = better interoperability

Making Access to Collections Easier and Better - Erin Coburn, J. Paul Getty Museum

Created CDWA Lite – an XML schema to describe core records for works of art and material culture based on CDWA and the CCO; intended for contribution to union catalogs and other repositories using the OAI harvesting protocol.

RESULT: ARTstor accesses Getty collection through a minimal amount of metadata and then links back to Getty and a variety of materials and exhibitions including videos that they have available on their website related to the material.

Reality Check #1 – you need data (new acquisitions reports may map almost one to one to metadata schema – use what you already have)

Reality Check #2 – getting your information out of your system and into XML in accordance with the CDWA Lite Spec (they are working on this to make it easier)

Reality Check #3 – making your CDWA Lite XML records available for harvesting (Getty took OAICat and sent to OCLC for others to use)

Reality Check #4 – how good is your data? (good cataloging practice is necessary – subject headings, authorities, etc.)

If we make our images easily harvestable, then there will be new modes of artistic expression based on access to those images. You can limit access if needed. But, this can bring traffic back to your collection and to new audiences.

Sharable Metadata for Non-Bibliographic Materials: Implications for Libraries and Archives – Murtha Baca, J. Paul Getty Museum

It's not all- and not only- about MARC (or EAD, etc.)

- For art images/cultural works, MARC may not be most appropriate
- So, non-standard schema have been developed over time

Some Emerging Trends in Metadata Creation

- “schema-agnostic” metadata (consistent data can express it in different schema)
- Shareable and re-purposeable
- Harvestable (OAI-PMH)
- Non-exclusive/cross cultural metadata (okay to combine standards)
- Importance of authorities and controlled vocabularies
- Need for practical, economically feasible approaches to metadata creation
- “cross-cultural” use of standards

Solutions to Vocabulary Problem?

- Include variants, broader, and narrower terms
- Service providers/aggregators employ the appropriate controlled vocabularies and thesauri as “search assistants” (promising but not now a reality)
- Can OCLC’s Terminologies Service help?

The Way Forward?

- Service providers should be more demanding (i.e. require data providers adhere to certain standards and use certain vocabularies, require “pre-washed” metadata)
- Data providers should consistently use appropriate standard schemas in their local systems.
- Service providers should consider “adding value” via services like vocabulary mapping, query expansion, vocabulary-assisted searching, user-added metadata, post-harvest sub-setting, metadata enhancement, etc. – see Roy Tennant article

Case Studies: New Implementations of OAI-PMH – Gunter Waibel, RLG Programs, OCLC

- Need to think of cataloging standards not as community specific (library, museum or archive), but what is suited to a specific material type
- OCLC Programs and Research: (1) leverage schema transformation software to create CDWA Lite XML (tool for transforming data); (2) Leverage OAICat: share collection (create museum version of OAI software for sharing or aggregating data (open source)