714: Metadata

Metadata Services

Margaret E.I. Kipp - kipp@uwm.edu
https://pantherfile.uwm.edu/kipp/public/courses/714
Metadata Services

• "the systems and tools supporting the creation and maintenance of metadata schemes and records."

• "record creation requires human catalogers in addition to the support of tools, procedures, and policies that ensure efficient and effective production."
Metadata Services 2

- "maintain mappings between name attributes for data items and other descriptive metadata attributes and respond to queries about these mappings." (Zheng and Qin p. 211)
Metadata Services 3

• metadata registries
  – information about all metadata elements, schemas, and application profiles, crosswalks, and a search and browse interface

• metadata repositories
  – provide metadata records, database schemas, XML schemas, taxonomies and XSL transformations
Metadata Services 4

- metadata development and production services
  - authentication of users, authorisation of metadata publication and metadata creation/design tools
    - e.g. http://libraries.mit.edu/metadata/
Metadata Services
Infrastructure

Level 1 (lowest): Networks, Servers, Databases and Applications

Level 2: XML

Level 3: XML Schema Language

Level 4: Metadata Registry Standard ISO11179, OAI-PMH

Level 5: Metadata Models and Schemas

Level 6: Policies and Procedures (Best Practices)

Level 7: Metadata Services
Metadata Registries

- centralised server maintaining the metadata schema and related standards

- example:
  - Dublin Core Metadata Initiative
    http://dublincore.org/
    - DCMI Terms
      http://dublincore.org/documents/dcmi-terms/
  - MODS http://www.loc.gov/standards/mods/
  - CORES Project http://www.cores-eu.net/
    - CORES Registry http://cores.dsd.sztaki.hu/
Metadata Registries (MDR)

• goal of MDRs
  - allow reuse and standardisation of metadata schemas

• users of metadata registries
  - standards creators, record creators, data curators and service providers, software developers, other registries
Functional Requirements for Metadata Registries

- support discovery of terms and vocabularies
- verify status of terms and vocabularies (still active, deprecated, etc)
- access machine readable XML descriptions of terms and vocabularies
- provide location of related resources
- navigate relationships between terms and vocabularies and between terms
Functional Requirements for CORES Schema Creation

- schema creation tool must support definition and identification of:
  - Element Sets and Elements; Application Profiles;
  - Encoding Schemes and Value Standards;
  - Agencies who own/create/maintain these resources
  - Commentaries (contextual annotations) outlining deployment of the Element Sets, Application Profiles, and Schemes
  - Links to User Guidelines for the Element Sets, Application profiles or Schemes
Types of Metadata Registries

- **cross domain and cross schema**
  - CORES Registry [http://cores.dsd.sztaki.hu/](http://cores.dsd.sztaki.hu/)
  - contains many different element sets, APs, etc

- **domain specific**
  - IEMSR Registry [http://www.ukoln.ac.uk/projects/iemsr/](http://www.ukoln.ac.uk/projects/iemsr/)
  - educational metadata, generally based on LOM or DC
Types of Metadata Registries 2

• project specific
  – European Library (TEL)
    http://www.dlib.org/dlib/february04/vanveen/02vanveen.html
  – http://theeuropeanlibrary.org/ [organisation -- projects]

• schema specific
  – DCMI http://dublincore.org/dcregistry/

• format oriented
  – SchemaWeb http://www.schemaweb.info/
  – RDF based schemas
ISO 11179 Information Technology Metadata Registries

- guidelines:
  - framework
    - elements, value standards, etc.
  - classification
    - registering and administering the namespace
  - metamodel and attributes
    - conceptual model of registry

- formulation of data definitions
  - creating a data dictionary or AP

- naming and identification principles
  - specifying formal naming/identification of metadata

- registration
  - quality/provenance of metadata
Metadata Repositories

• focuses on metadata schemas and records
• may belong to a digital library
• must comply with the Open Archive Initiative Protocol for Metadata Harvesting (OAI-PMH) standard
  http://www.openarchives.org/OAI/openarchivesprotocol.html

• example:
  – National Science Digital Library
    http://nsdl.org/
  – E-LIS http://eprints.rclis.org/ and DLIST Arizona
    http://arizona.openrepository.com/arizona/handle/10150/105067
Metadata Repository Features

- may be composed of metadata datasets from many sources
- may contain records created using different metadata schemas and practices
- records may be created by trained metadata creators or by authors/depositors of material
- may harvest freely accessible metadata from other repositories
- e.g. http://digital2.library.ucla.edu/sheetmusic/
Metadata Harvesting Models

• Z39.50
  - maintained by LC and implemented in many OPACs
  - newer systems also support SRU
  - http://www.loc.gov/z3950/

• OAI-PMH - Open Archives Initiative Protocol for Metadata Harvesting
  - for eprint repositories, intended to optimise the discovery and retrieval of document like objects
  - supports multiple metadata formats
  - http://www.openarchives.org/
OAI-PMH Model

- facilitates two kinds of processes:
  - data providers use OAI-PMH to provide access to metadata and thus to documents
  - service providers use OAI-PMH harvested data to provide additional services

- metadata from an OAI-PMH compliant repository is collected via a query
  - user submits a query or request
  - repository sends a response
Seven Key Definitions

- Harvester: client application issuing OAI-PMH requests
- Repository: network accessible server, able to process OAI-PMH requests correctly
- Resource: object the metadata is "about"
- Item: component of a repository from which metadata about a resource can be disseminated
- Record: metadata in a specific metadata format
- Identifier: unique key for an item in a repository
- Set: optional construct for grouping items in a repository

http://www.oaforum.org/tutorial/english/page3.htm#section3
OAI PMH Structure

http://www.oaforum.org/tutorial/english/page3.htm#section3
OAI-PMH Commands

• six commands:
  • Identify
  • ListMetadataFormats
  • ListSets
  • ListIdentifiers
  • ListRecords
  • GetRecord

• Example:
  • retrieves a single record from a repository

• http://www.openarchives.org/OAI/openarchivesprotocol.html
OAI-PMH Record Components

• header
  – unique identifier of the item and additional properties for harvesting

• metadata
  – full metadata record for the item

• about
  – data about the metadata creation, may also contain rights and provenance information
<?xml version="1.0" encoding="UTF-8"?>
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">

<responseDate>2009-11-18T02:49:06Z</responseDate>

:request verb="GetRecord" identifier="oai:arXiv.org:cs/0112017"
metadataPrefix="oai_dc">http://export.arxiv.org/oai2</request>

<GetRecord>

<record><header>
<identifier>oai:arXiv.org:cs/0112017</identifier>
<datestamp>2007-05-23</datestamp>
<setSpec>cs</setSpec>
</header><metadata>
Using Structural Metadata to Localize Experience of Digital Content

Dushay, Naomi

Computer Science - Digital Libraries

2001-12-14

http://arxiv.org/abs/cs/0112017
Other OAI-PMH Examples

  - provide information about the repository

  - list all available metadata formats

  - lists available sets in the repository
Other OAI-PMH Examples 2

  - retrieve metadata record for listed item

  - retrieve records from a specific set for a specific time frame
SRU

• SRU (Search and Retrieval via URL)
• search and retrieval protocol which uses CQL (Common Query Language)
• consists of a SRU base URL and associated parameters
• http://www.loc.gov/standards/sru/
• http://www.loc.gov/standards/sru/specs/cql.html
Building SRU Queries using CQL

- e.g. http://z3950.loc.gov:7090/voyager?version=1.1&operation=searchRetrieve&query=dinosaur&startRecord=2&maximumRecords=5&recordSchema=mods

- query consists of the base URL with CQL terms
  - operation=searchRetrieve
  - query=dinosaur (your search term)
  - startRecord/maximumRecords (where to start, how many records to return)
  - recordSchema=mods (output format)
  - http://www.loc.gov/standards/sru/resources/servers.html
Encoding CQL and More Complex Queries

- CQL statements must be URL encoded
- e.g. http://z3950.loc.gov:7090/voyager?version=1.1&operation=searchRetrieve&query=%22title%3Ddinosaur%22
  - %3D is the code for =
  - the query term is title=dinosaur
  - %22 are quotes
  - http://www.loc.gov/standards/sru/sru1-1archive/search-retrieve-operation.html

- other ideas for query terms:
- http://zing.z3950.org/cql/intro.html
Metadata Retrieval and Discovery

• union catalogues
  – http://worldcat.org/

• federated search systems
  – http://www.deepwebtech.com/

• search engines
  – http://www.google.com/

• open access repositories
  – http://www.opendoar.org/

• exposing OAI-PMH data (or other metadata e.g. RSS) to search engines and other repositories
In Class Exercise: Using Metadata Services

- Using metadata services may involve entering instructions for construction of a query into a form or building a URL based on specific rules.
  - A. Search one of the Z39.50 links (http://www.loc.gov/z3950/) for a topic of your choice.
  - B. Search the LOC SRU service (http://z3950.loc.gov:7090/voyager?) for the same topic.
  - C. Search the LOC OAI-PMH service (http://memory.loc.gov/ammem/oamh/oai_request.html)

- Compare the systems and harvesting methods.
Example: OAI PMH and Omeka (Repository)

- **OAI-PMH Base URL:**
  http://mysois.uwm.edu/omeka/oai-pmh-repository/request?
  - e.g. http://mysois.uwm.edu/omeka/oai-pmh-repository/request?verb=Identify
  - http://mysois.uwm.edu/omeka/oai-pmh-repository/request?verb=ListSets
  - http://mysois.uwm.edu/omeka/oai-pmh-repository/request?
    verb=ListRecords&metadataPrefix=oai_dc
Example: OAI PMH Harvesting with Omeka

• The LOC American Memory Project has an OAI-PMH service

• OAI-PMH Harvesting in Omeka is handled by the OAI-PMH Harvesting plugin (see the tab on Omeka)

• using the OAI-PMH base URL, you can harvest data from another repository and enter it into Omeka

• e.g. http://memory.loc.gov/cgi-bin/oai2_0