OCLC and Vocabulary Identifiers

Eric Childress
Andrew Houghton
Diane Vizine-Goetz

DC-2005 Vocabularies in Practice
13 September 2005
Madrid, Spain

Presented by

Eric Childress

Outline

- OCLC´s vocabulary activities
- OCLC Research terminology services project
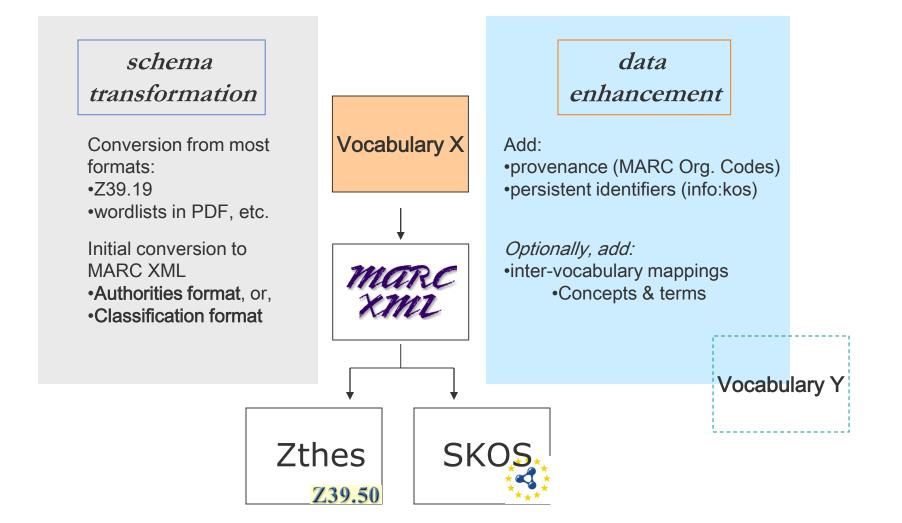- Experimental use of identifiers

OCLC vocabulary activities

- – Owner of Dewey Decimal Classification
  - Maintains English DDC file
  - Coordinates work on DDC in other languages
  - Provides DDC through various channels
  - No long-term decision on identifiers
    - – Experimental use of GUIDs
- – Various production applications
  - OCLC Connexion® cataloging interface
  - WebDewey & Abridged WebDewey
  - Loading various external files in FirstSearch, etc.

OCLC vocabulary activities (con't)

- **Standards work:**
  - NKOS/NISO/ISO, etc.
  - IFLA FRAR (Functional Requirements for Authority Records)
- **OCLC Research:**
  - Research into automatic classification
  - FAST vocabulary (faceted LCSH)
  - VIAF (Virtual International Authority File)
  - LAF (LC Authority File) – various web services
  - Terminology services project
    - Two key activities:
      - Converting, normalizing and adding value to vocabularies
      - Releasing vocabularies in a web services environment
    - Experimental use of info:kos identifier

OCLC Research
Terminology Services (TS) project

schema transformation

Conversion from most formats:
•Z39.19
•wordlists in PDF, etc.

Initial conversion to MARC XML
•**Authorities format**, or,
•**Classification format**

Vocabulary X

MARC XML

data enhancement

Add:
•provenance (MARC Org. Codes)
•persistent identifiers (info:kos)

Optionally, add:
•inter-vocabulary mappings
      •Concepts & terms

Vocabulary Y

Zthes
Z39.50

SKOS

Identifiers in TS project - MARC

- Record identifier
  - MARC 001 (#)+ 003 (agency)
- Provenance
  - MARC 040 (chain of creation/modification)
- National control number (some files)
  - MARC 010 – (OCLC transfers if known)
- URI
  - MARC 856 – experimenting with info:kos
  - A few vocabularies have native URIs

GUIDs & info identifiers

- GUID (Globally Unique Identifier)
  - Implementation by Microsoft of UUID (Universally Unique Identifier) specified by Open Software Foundation (OSF)
  - Pseudo-random number (16-byte (128-bit) number written in hexadecimal
    - 3F2504E0 4F89 11D3 9A 0C 03 05 E8 2C 33 01
- Info registry (NISO):
  - Mechanism for the registration of public namespaces that are used for the identification of information assets
- OCLC experimenting with info:kos scheme
  - Two elements in info:kos identifier:
    - scheme
    - concept
  - Structure of info:kos identifiers:
    - info:kos/scheme/«code»/«expr»/«lang»
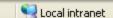   - info:kos/concept/«code»/«id»

Result of search: dc.title = "adventure"

Matches: 4

Records

Next Record(s)

Position: 1 Schema: MARC XML

```
000 00631cz a2200193n 4500
001 GSAFD000001                          Record Identifier
003 IlChALCS
005 20050503101755.2
008 000720 n anznnbabn a ana d
040 __ |aIlChALCS|beng|cIEN|dOCoLC-O|fgsafd    Provenance
155 __ |aAdventure fiction
455 __ |wnne|aAdventure stories
455 __ |aSwashbucklers
455 __ |aThrillers
555 __ |wh|aPicaresque literature|0(IlChALCS)GSAFD000099
555 __ |wh|aRobinsonades|0(IlChALCS)GSAFD000106
555 __ |aSea stories|0(IlChALCS)GSAFD000117
555 __ |wh|aWestern stories|0(IlChALCS)GSAFD000152
680 __ |iUse for works characterized by an emphasis on physical and often violent action, exotic locales, and danger, generally with little character development.
750 _1 |aAdventure and adventurers|vFiction|0(DLC)sj 96004703 |5OCoLC-O
750 _0 |aAdventure stories|0(DLC)sh 85001072 |5OCoLC-O       Mapped term Record Identifier
856 __ |uinfo:kos/concept/gsafd/GSAFD000001|yinfo:kos/concept/gsafd/GSAFD000001|zGSAFD Vocabulary info:kos URI
```

GSAFD record in MARC 21 authorities format as retrieved from an SRW server at OCLC

File   Edit   View   Favorites   Tools   Help

Sample source file: *DCMI Type*

Address | http://thompson-r.oa.oclc.org:8080/srw/search/ors-tspilot-dct?query=dc.title+%3D+%22image%22&version=1.1&operation=searchRetrieve&recordSchema=info%3Asrw%2Fsch

001	DCT000004
003	OCoLC-O

Record Identifier

005	20050810111052.4
008	040820|||a|znn|||| |a ana |d
040	|aOhDuDCMI|beng|cOCoLC-O|dOCoLC-O|fdct

Provenance

155	|aImage|9IMAGE
555	|wh|aMoving Image|0(OCoLC-O)DCT000006|5OhDuDCMI|9MOVING IMAGE
555	|wh|aStill Image|0(OCoLC-O)DCT000011|5OhDuDCMI|9STILL IMAGE
670	|aDCMI Metadata Terms: A complete historical record, viewed Apr. 18, 2005|uhttp://dublincore.org/usage/terms/history/#Image-002
680	|iAn image is a primarily symbolic visual representation other than text. For example - images and photographs of physical objects, paintings, prints, drawings, other images and graphics, animations and moving pictures, film, diagrams, maps, musical notation. Note that image may include both electronic and physical representations.
688	|aDecision on proposal to add the terms "Moving Image" and "Still Image" to the DCMI Type Vocabulary|uhttp://dublincore.org/usage/decisions/2003/2003-02.Image-terms.shtml
688	|aDate Issued: 2000-07-11; Version: Image-001
688	|aDate Modified: 2003-11-18; Version: Image-002
856	|uhttp://purl.org/dc/dcmitype/Image|yRDF/XML metadata description|zhttp://purl.org/dc/dcmitype/Image
856	|uinfo:kos/concept/dct/DCT000004|zDCMI Type Vocabulary info:kos URI

Native URI

info:kos URI

Links

- OCLC Research
  - info:kos Application Notes
    - http://www.oclc.org/research/projects/termservices/resources/info-uri.htm
  - ResearchWorks
    - http://www.oclc.org/research/researchworks/
  - Terminology Services project
    - http://www.oclc.org/research/projects/termservices/
  - Terminologies Pilot
    - http://www.oclc.org/research/projects/termservices/resources/tspilot-services.htm
- FRAR: Extending FRBR Concepts to Authority Data
  - www.ifla.org/IV/ifla71/papers/014e-Patton.pdf