The OCLC Metadata Switch Project

Jean Godby, Thomas Hickey,
Diane Vizine-Goetz
OCLC Office of Research

Digital Library Federation
May 14, 2003
Outline of this talk

- The changing library collection
- Web services for the digital library
- A detailed look at three Web services in the OCLC’s Metadata Switch project
- Recurring themes
The management of resources in the digital library

Some priorities:

- Updated versions of proven solutions
  - Metadata
  - Knowledge organization
  - Authority control
  - Automated processing

- Interoperability
The “Metadata Switch” project

A research project at OCLC that...

- Acknowledges the investment of libraries in metadata.

- Leverages the investment by adding value.

- Decomposes the functionality of large systems into independent, reusable modules.
### Some examples

<table>
<thead>
<tr>
<th>A module that accepts:</th>
<th>...and returns:</th>
</tr>
</thead>
<tbody>
<tr>
<td>A record in one format (Dublin Core)</td>
<td>A record in another format (MARC)</td>
</tr>
<tr>
<td>A Web page</td>
<td>A class number or subject heading</td>
</tr>
<tr>
<td>A name</td>
<td>A set of variant names for the same entity</td>
</tr>
<tr>
<td>A class number</td>
<td>An associated subject heading</td>
</tr>
</tbody>
</table>
A Web service is...

- A Web page with functions.
  - stock quotes, traffic conditions, calculators, weather

- A Web-accessible program that performs a single task...

- …using standard Web protocols.
  - XML syntax, HTTP packaging
    - REST: Representational State Transfer
    - SOAP: Simple Object Access Protocol
Web services for the digital library

- Register
- Search
- Resolve
- Navigate
- Decompose
- Enhance
- Transform
A closer look at three Web services
Example 1: The XWork ISBN Web Service

- A resolution service

- Associates related works using new research results in cataloging theory
  - Builds on John Udell’s work on library lookup
  - Adds a reference to a FRBRized collection of OCLC’s WorldCat
  - Bridges library collections with a bookstore

by Chuck Musciano, Bill Kennedy

List Price: $39.95
Price: $27.97 & This item ships for FREE with Super Saver Shipping. See details.
You Save: $11.98 (30%)
Availability: Usually ships within 24 hours
Also available for in-store pickup.

9 used & new listings from $23.78

Edition: Paperback
<?xml version="1.0" encoding="UTF-8" ?>

- <idlist>
  <isbn>0596000382</isbn>
  <isbn>0596000026</isbn>
  <isbn>2841771326</isbn>
</idlist>
Database Name: Cornell University Library
Search Request: Guided Keyword = (059600382x OR 059600026x OR 2841771326)[in ISBN]
Search Results: Displaying 1 through 2 of 2 entries.

<table>
<thead>
<tr>
<th>#</th>
<th>ISBN</th>
<th>Title</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>059600382X</td>
<td>HTML and XHTML, the definitive guide / Chuck Musciano and Bill Kennedy.</td>
<td>2002</td>
</tr>
<tr>
<td>1</td>
<td>Library Location: Engineering Library (Carpenter Hall)</td>
<td>Call Number: QA76.76.H94 M87x 2002 Status: Not Charged</td>
<td></td>
</tr>
<tr>
<td></td>
<td>059600026X</td>
<td>HTML and XHTML, the definitive guide / Chuck Musciano and Bill Kennedy.</td>
<td>2000</td>
</tr>
</tbody>
</table>
Example 2: A collection of terminology services

- Libraries expend significant time and effort creating and managing terminology resources.

- The value and usefulness of this work can be extended by...
  - Validating a subject heading or class number
  - Navigating hierarchies in thesauri or classification schemes
  - Enhancing/augmenting metadata descriptions or queries
Levels of services

**Basic**
- Implement protocol for navigating/querying thesauri
  - DDC, ERIC thesaurus, LCSH, MeSH

**Enhanced**
- Exploit unique resources that do not fit thesaurus model

**Experimental**
- Most complex, outcomes less certain
- Automated classification/categorization
XML Translation of DDC Summaries:

```xml
<?xml version="1.0" encoding="utf-8" ?>
  - <caption>
    <lang>de</lang>
    <data>Technik</data>
  </caption>
  - <caption>
    <lang>en</lang>
    <data>Technology</data>
  </caption>
  - <caption>
    <lang>es</lang>
    <data>Tecnología</data>
  </caption>
  - <caption>
    <lang>fr</lang>
    <data>Technologie</data>
  </caption>
  - <caption>
    <lang>it</lang>
    <data>Tecnologia (Scienze applicate)</data>
  </caption>
  - <caption>
    <lang>ru</lang>
    <data>Технология</data>
  </caption>
</ArrayOfCaption>
```
Retrieve genre terms

<xml version="1.0" encoding="utf-8" ?
  <string>Adventure fiction</string>
  <string>Adventure stories</string>
  <string>Allegories</string>
  <string>Alternative histories</string>
  <string>Apocalyptic fantasies</string>
  <string>Arthurian romances</string>
  <string>Autobiographical fiction</string>
  <string>Bildungsroman</string>
  <string>Bildungsromans</string>
  <string>Biographical fiction</string>
  <string>Black humor (Literature)</string>
  <string>Christian fiction</string>
  <string>Detective and mystery stories</string>
  <string>Detective and mystery stories; English</string>
  <string>Didactic fiction</string>
  <string>Domestic fiction</string>
  <string>Dystopias</string>
  <string>Epic literature</string>
  <string>Epistolary fiction</string>
  <string>Erotic fiction</string>
  <string>Erotic stories</string>
  <string>Fables</string>
  <string>Fantastic fiction</string>
  <string>Fantasy fiction</string>
  <string>Fantasy fiction—Juvenile</string>
  <string>Ghost stories</string>
  <string>Gothic fiction</string>
  <string>Graphic novels</string>
  <string>Historical fiction</string>
  <string>Horror fiction</string>
  <string>Horror tales</string>
  <string>Humorous fiction</string>
  <string>Humorous stories</string>
  <string>Legal stories</string>
</ArrayOfString>
Retrieve subject types

```xml
<?xml version="1.0" encoding="utf-8" ?>
<ArrayOfString xmlns:xsd="http://www.w3.org/2001/XMLSchema"
    <string>Characters</string>
    <string>Place/Setting</string>
    <string>Topics</string>
    <string>People</string>
</ArrayOfString>
```
Retrieve place/setting for Science fiction

<?xml version="1.0" encoding="utf-8" ?>
- <ArrayOfString xmlns:xsd="http://www.w3.org/2001/XMLSchema"
  <string>Mars (Planet)—Fiction</string>
  <string>Dune (Imaginary place)—Fiction</string>
  <string>Pern (Imaginary place)—Fiction</string>
  <string>Venus (Planet)—Fiction</string>
  <string>New York (N.Y.)—Fiction</string>
  <string>Amber (Imaginary place)—Fiction</string>
  <string>Los Angeles (Calif.)—Fiction</string>
  <string>Outer space—Exploration—Fiction</string>
  <string>Darkover (Imaginary place)—Fiction</string>
  <string>London (England)—Fiction</string>
  <string>San Francisco (Calif.)—Fiction</string>
  <string>France—History—14th century—Fiction</string>
  <string>California—Fiction</string>
  <string>Moon—Fiction</string>
  <string>Earth—Core—Fiction</string>
  <string>Antarctica—Fiction</string>
  <string>Arizona—Fiction</string>
  <string>Australia—Fiction</string>
  <string>Dorsai (Imaginary place)—Fiction</string>
  <string>England—Fiction</string>
  <string>Jupiter (Planet)—Fiction</string>
  <string>South America—Fiction</string>
  <string>Africa—Fiction</string>
  <string>Chicago (Ill.)—Fiction</string>
  <string>Geneva (Switzerland)—Fiction</string>
</ArrayOfString>
Example 3: Metadata translation

The environment:

- Multiple formats (MARC, Dublin Core, GEM) and a need for interoperability
- Lots of ad-hoc or duplicated effort
- Evolving standards and practices
High-level design

Crosswalk Registry client

A crosswalk

Crosswalk registry

Transformed record

Record translation client

Record transformer

A record to transform
Two pathways to translation

- **Structural transforms**
  Simple, lightweight, fully automatic, capitalize on current trends, appropriate when differences are purely structural

  But: Hardcoded

- **Semantic translations**
  More general, closer to true interoperability

  But: Not fully automatic
A working client
Recurring themes in the Metadata Switch project

- Relevance to the management of metadata in the digital library
- Appropriateness of the Web services model
- Levels of granularity
- Research vs. development
For more information

The Metadata Switch Project at OCLC

http://www.oclc.org/research/projects/mswitch/