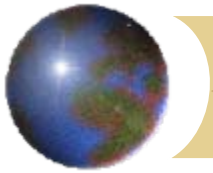


The OCLC Metadata Switch Project

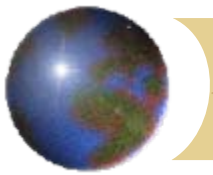
Jean Godby, Thomas Hickey,
Diane Vizine-Goetz
OCLC Office of Research

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

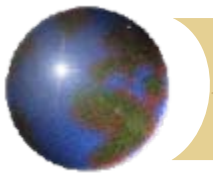


The management of resources in the digital library

Some priorities:

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

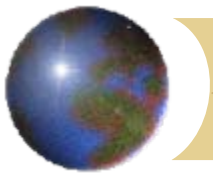
- Interoperability



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

A module that accepts:

**A record in one format
(Dublin Core)**

A Web page

A name

A class number

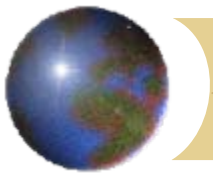
...and returns:

**A record in another format
(MARC)**

**A class number or
subject heading**

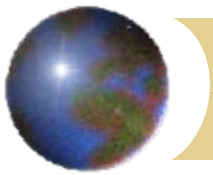
**A set of variant names for
the same entity**

**An associated subject
heading**



A Web service is...

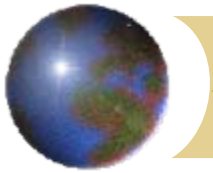
- **A Web page with functions.**
 - stock quotes, traffic conditions, calculators, weather
- **A Web-accessible program that performs a single task...**
- **...using standard Web protocols.**
 - XML syntax, HTTP packaging
 - REST: Representational State Transfer
 - SOAP: Simple Object Access Protocol



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

Amazon Exclusive!!
Order a Segway now!
It's only at Amazon

amazon.com

VIEW CART | WISH LIST | YOUR ACCOUNT | HELP



Thomas's Gold Box

WELCOME THOMAS'S STORE BOOKS APPAREL & ACCESSORIES ELECTRONICS TOYS & GAMES DVD BABY SEE MORE STORES

SEARCH BROWSE SUBJECTS BESTSELLERS MAGAZINES CORPORATE ACCOUNTS E-BOOKS & DOCS BARGAIN BOOKS USED BOOKS

Super Saver Shipping **FREE** on orders over \$25 [Restrictions apply](#)

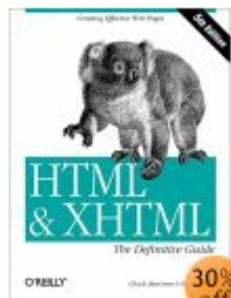
SEARCH

Books

GO!

HTML & XHTML: The Definitive Guide, Fifth Edition

by [Chuck Musciano](#), [Bill Kennedy](#)



[see larger photo](#)

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

READY TO BUY?

[Add to Shopping Cart](#)

or

[Sign in](#) to turn on 1-Click ordering.

MORE BUYING CHOICES

[9 used & new listings](#) from **\$23.78**

Available for in-store pickup now

from: **\$39.95**

Price may vary based on availability

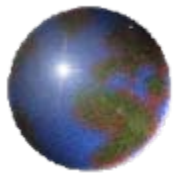
Enter your ZIP Code

43085

[Choose a store](#)

RECENTLY VIEWED ITEMS

[Dynamic HTML: The Definitive](#)



F:\junk.xml - Microsoft Internet Explorer provided by OCLC, Inc.

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites History Print Copy Paste

Address [F:\junk.xml](file:///F:/junk.xml) Go

Links [Customize Links](#) [Extractor](#) [Free Hotmail](#) [Windows Media](#) [Windows](#)

```
<?xml version="1.0" encoding="UTF-8" ?>
- <idlist>
  <isbn>0596000382</isbn>
  <isbn>0596000026</isbn>
  <isbn>2841771326</isbn>
</idlist>
```

Done Local intranet

Cornell University
LIBRARY CATALOG

[Library Gateway](#) | [e-Reference Collection](#) | [e-Journal Titles](#) | [My Library](#) | [Ask a Librarian](#) | [Individual Libraries](#)

[NEW SEARCH](#) | [RETURN TO TITLES LIST](#) | [SEARCH HISTORY](#) | [PATRON INFO](#) | [REQUESTS](#) | [PREFERENCES](#) | [SAVED SEARCHES](#) | [BOOKBAG](#) | [INTERLIBRARY LOAN](#) | [BORROW DIRECT](#)

[HELP](#) | [END SESSION](#)

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.

[previous](#) [next](#)

Sort by:

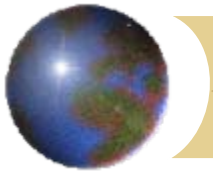
Post Limit

#	ISBN	Title	Date
<input type="checkbox"/> [1]	059600382X	HTML and XHTML, the definitive guide / Chuck Musciano and Bill Kennedy.	2002
	<i>Library Location: Engineering Library (Carpenter Hall) Call Number: QA76.76.H94 M87x 2002 Status: Not Charged</i>		
<input type="checkbox"/> [2]	059600026X	HTML and XHTML, the definitive guide / Chuck Musciano and Bill Kennedy.	2000



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**

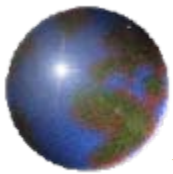
- Implement protocol for navigating/querying thesauri
 - DDC, ERIC thesaurus, LCSH, MeSH

● **Enhanced**

- ▣ Exploit unique resources that do not fit thesaurus model

● **Experimental**

- Most complex, outcomes less certain
- ▣ Automated classification/categorization

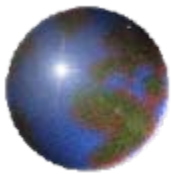


DDC Summaries: Translations

```
http://houghton/WebServices/DeweySummaries/DeweySummaries.asmx/RetrieveDeweySummaryCaptions?ddc - Micr...
File Edit View Favorites Tools Help

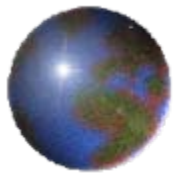
<?xml version="1.0" encoding="utf-8" ?>
- <ArrayOfCaption xmlns:xsd="http://www.w3.org/2001/XMLSchema"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="urn:research-oclc-
  org:WebServices.DeweySummaries">
- <caption>
  <lang>de</lang>
  <data>Technik</data>
</caption>
- <caption>
  <lang>en</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>
```

Done Local intranet



Retrieve genre terms

```
http://houghton/WebServices/GsafSubjectHeadings/GsafSubjectHeadings.asmx/RetrieveAvailableGen - Microsoft Internet Explorer
File Edit View Favorites Tools Help
<?xml version="1.0" encoding="utf-8" ?>
- <ArrayOfString xmlns:xsd="http://www.w3.org/2001/XMLSchema" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance" xmlns="urn:research-oclc-org:WebServices.GsafSubjectHeadings">
  <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>Bildungsromane</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>
```

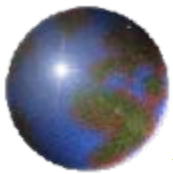



Retrieve subject types

A screenshot of a Microsoft Internet Explorer browser window. The address bar shows the URL: http://houghton/WebServices/GsafdSubjectHeadings/GsafdSubjectHeadings.asmx/RetrieveAvailableSub. The browser's menu bar includes File, Edit, View, Favorites, Tools, and Help. The main content area displays XML data with the following structure:

```
<?xml version="1.0" encoding="utf-8" ?>
- <ArrayOfString xmlns:xsd="http://www.w3.org/2001/XMLSchema"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="urn:research-oclc-
  org:WebServices.GsafdSubjectHeadings">
  <string>Characters</string>
  <string>Place/Setting</string>
  <string>Topics</string>
  <string>People</string>
</ArrayOfString>
```

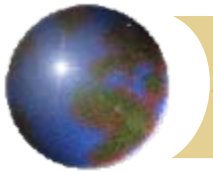
The status bar at the bottom shows "Done" on the left and "Local intranet" on the right.



Retrieve place/setting for Science fiction

The screenshot shows a web browser window with the address bar containing the URL: `http://houghton/WebServices/GsafdSubjectHeadings/GsafdSubjectHeadings.asmx/RetrieveGenreSubject`. The browser's menu bar includes File, Edit, View, Favorites, Tools, and Help. The main content area displays XML data in a monospaced font. The XML starts with a declaration: `<?xml version="1.0" encoding="utf-8" ?>`. It then contains an `<ArrayOfString>` element with various attributes, including `xmlns:xsd="http://www.w3.org/2001/XMLSchema"` and `xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"`. The array contains 25 string elements, each representing a science fiction setting followed by `-Fiction`. The settings listed are: Mars (Planet), Dune (Imaginary place), Pern (Imaginary place), Venus (Planet), New York (N.Y.), Amber (Imaginary place), Los Angeles (Calif.), Outer space-Exploration, Darkover (Imaginary place), London (England), San Francisco (Calif.), France-History-14th century, California, Moon, Earth-Core, Antarctica, Arizona, Australia, Dorsai (Imaginary place), England, Jupiter (Planet), South America, Africa, and Chicago (Ill.). The browser's status bar at the bottom shows "Done" and "Local intranet".

```
<?xml version="1.0" encoding="utf-8" ?>
- <ArrayOfString xmlns:xsd="http://www.w3.org/2001/XMLSchema"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="urn:research-oclc-
  org:WebServices.GsafdSubjectHeadings">
  <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>Genova (Switzerland)-Fiction</string>
```

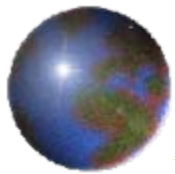


Example 3:

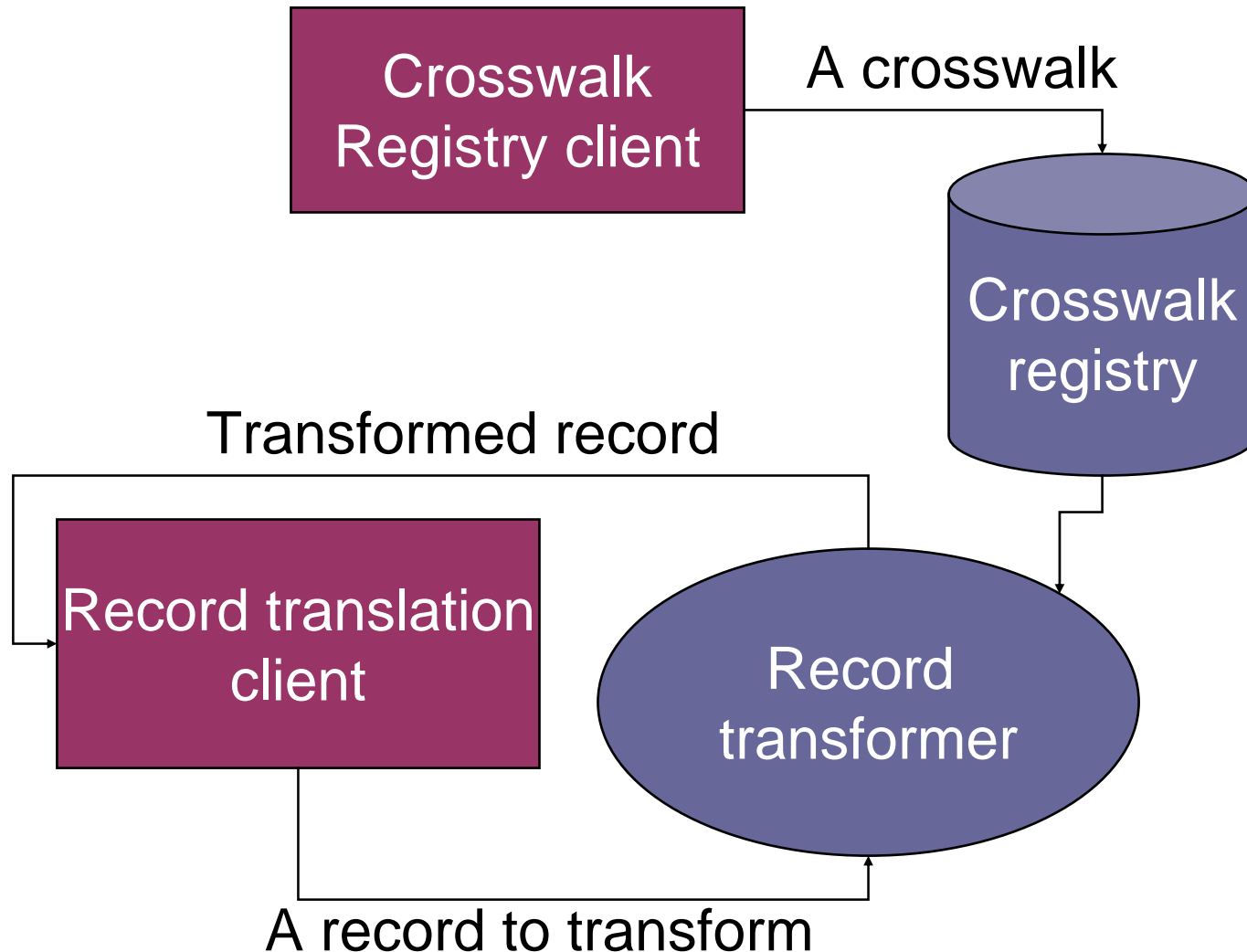
Metadata translation

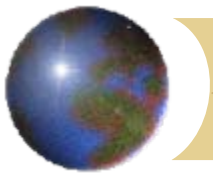
The environment:

- 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





Two pathways to translation

- **Structural transforms**

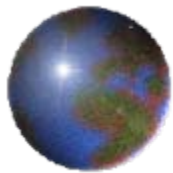
Simple, lightweight, fully automatic, capitalize on current trends, appropriate when differences are purely structural

But: Hardcoded

- **Semantic translations**

More general, closer to true interoperability

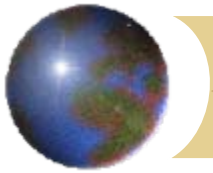
But: Not fully automatic



A working client

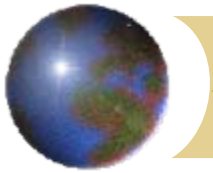
The screenshot shows a web application interface with a toolbar at the top containing icons for navigation and document operations. Below the toolbar is a row of five buttons: "List Schemas", "List Transformations", "Load A Schema", "Load A Transformation", and "Record Transform". The main content area features a text input field containing the URL "http://smithde.dev.oclc.org/devon/marcRecord.xml" with the label "(URI of Record)". Below this is the text "(Overrides File Upload)" and "OR". There is a "Browse..." button next to an empty text field, followed by the label "Record (XML)". A dropdown menu is open, showing a list of XSLT transformation paths:

- /LOC//DOCUMENT XSLT MarcXML to DC 2002//EN
- /LOC//DOCUMENT XSLT MarcXML to DC 2002//EN
- /OCLC//DOCUMENT XSLT Dummy Transform 20021021//EN
- /OCLC//DOCUMENT XSLT Marc21slim to Dublin Core Record Set 20020919//EN
- /OCLC//DOCUMENT XSLT Marc21slim to RDF Dublin Core 20021122//EN



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/>