

# Terminology Services in a Digital Repository

---

*Michael Durbin*



**DIGITAL LIBRARY PROGRAM**

INDIANA UNIVERSITY

Libraries/University Information Technology Services

# Outline

---

- Introduction and Background
- Service Architecture
- Application Enhancements
  
- Discussion of current needs and future development



**DIGITAL LIBRARY PROGRAM**

INDIANA UNIVERSITY

Libraries/University Information Technology Services

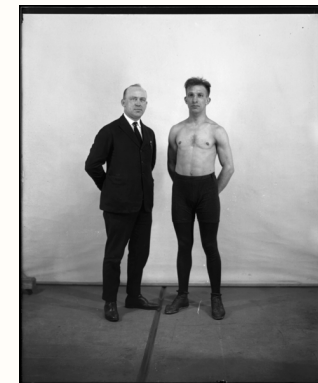
# About Controlled Vocabularies

---

- List of terms that represent concepts
- Lacks the ambiguity of simple text
- Maintained by authorities
  - Library of Congress
  - Getty
  - Local researchers/experts
- Shared to describe different resources



VS



**DIGITAL LIBRARY PROGRAM**

INDIANA UNIVERSITY

Libraries/University Information Technology Services

## Standardizing vocabularies and protocols

---

- Standard vocabularies have been in use for years
- Standard integration of vocabulary access is less commonplace
  - Technical Integration
  - User Experience Integration



**DIGITAL LIBRARY PROGRAM**

INDIANA UNIVERSITY

Libraries/University Information Technology Services

---

# SERVICE ARCHITECTURE



**DIGITAL LIBRARY PROGRAM**

INDIANA UNIVERSITY

Libraries/University Information Technology Services

# Definitions

---

## o Service-Oriented Architecture

*Service-orientation requires loose coupling of services with operating systems, and other technologies that underlie applications. SOA separates functions into distinct units, or services, which developers make accessible over a network in order to allow users to combine and reuse them in the production of applications. These services and their corresponding consumers communicate with each other by passing data in a well-defined, shared format, or by coordinating an activity between two or more services.*

*Wikipedia, Sept 29, 2010*

## o Fedora

*The Fedora Repository Project (i.e., Fedora) implements the Fedora abstractions in a robust open source software system. Fedora provides a core repository service (exposed as web-based services with well-defined APIs). In addition, Fedora provides an array of supporting services and applications including search, OAI-PMH, messaging, administrative clients, and more. Fedora provides RDF support and the repository software is integrated with semantic triple store technology, including the Mulgara RDF database. Fedora helps ensure that digital content is durable by providing features that support digital preservation.*

*<http://www.fedora-commons.org/about>*



**DIGITAL LIBRARY PROGRAM**

INDIANA UNIVERSITY

Libraries/University Information Technology Services

## More Definitions

---

### o SRU/SRW

*Search and Retrieval via URL*

*Search and Retrieval via Web Service*

*SRU is a standard XML-focused search protocol for Internet search queries, utilizing CQL (Contextual Query Language), a standard syntax for representing queries.*

*<http://www.loc.gov/standards/sru/index.html>*



**DIGITAL LIBRARY PROGRAM**

INDIANA UNIVERSITY

Libraries/University Information Technology Services

# Proof Of Concept Experiment

---

## o Goals

- HTTP access to controlled vocabularies
- XML response
- Ability to lookup terms and then query relationships
- Use of standards
- Eventually be able to host or mirror locally

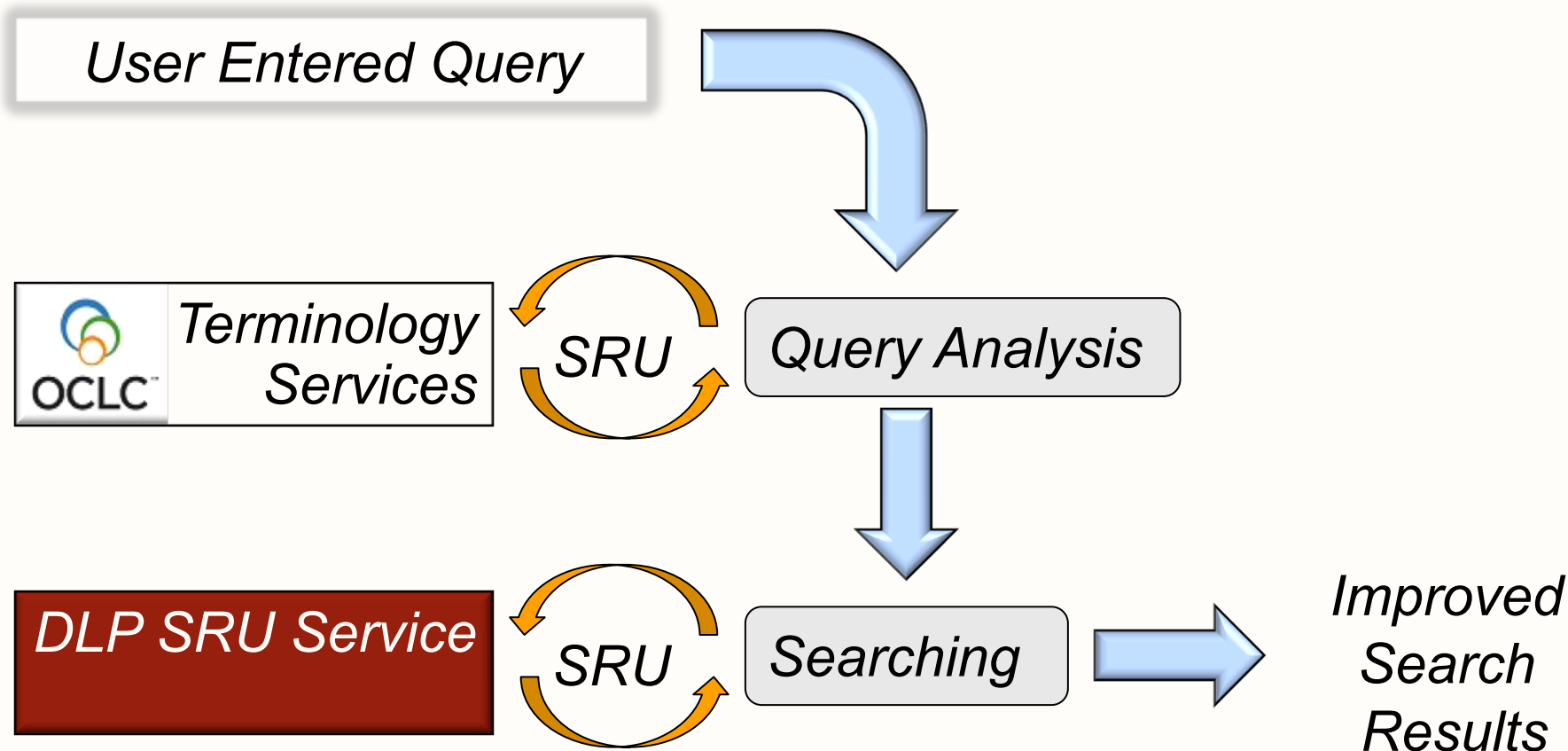
## o Outcomes

- OCLC Research's Terminology Services (<http://www.oclc.org/research/activities/termservices/>)
- Prototype discovery application (<http://webapp1.dlib.indiana.edu/imagesearch/>)





# Design/Algorithm



DIGITAL LIBRARY PROGRAM

INDIANA UNIVERSITY

Libraries/University Information Technology Services

## Local Production Services

---

- OCLC SRU/W server + local customizations
  - Lucene-based search
  - ZThes context set

*<http://zthes.z3950.org/cql/>*

*<http://www.oclc.org/research/activities/srw/>*



**DIGITAL LIBRARY PROGRAM**

INDIANA UNIVERSITY

Libraries/University Information Technology Services

## Importing Vocabularies

---

### o Manual Process

- Acquire vocabulary data
- Process it into an index
- Put that index into place
- Restart the search service
- Repeat as necessary



# Importing Vocabularies

---

- Automatic Process
  - Store vocabularies in Fedora repository
  - Update that object in fedora as necessary
    - Index updates will be automatically triggered



## Vocabulary Objects in Fedora

---

- One object contains an entire thesaurus
- Working format is stored and versioned
- Standard, normalized versions are made available
  - Through the SRU service
  - As disseminations from the object\*



**DIGITAL LIBRARY PROGRAM**

INDIANA UNIVERSITY

Libraries/University Information Technology Services

## Accessing Vocabularies

---

- SRU access point
  - Search
  - Disambiguate (manual, if necessary)
  - Access term information or broader, narrower or related terms

<http://fedora.dlib.indiana.edu:8080/SRW/search/tqm1>

- JSON wrapper to SRU



DIGITAL LIBRARY PROGRAM

INDIANA UNIVERSITY

Libraries/University Information Technology Services

---

# APPLICATION ENHANCEMENTS



**DIGITAL LIBRARY PROGRAM**

INDIANA UNIVERSITY

Libraries/University Information Technology Services

## Metadata Editor: PhotoCat

---

- Descriptive Metadata Editor
- Primarily for photographs, using MODS schema
- Updates objects in the fedora repository
  
- Some fields should be “linked” to CV sources





## PhotoCat: current and future CV integration

---

- Auto-complete linked fields
- Thesaurus Lookup
- *Browse hierarchically \**
- *Validate and suggest terms\**
- *Many more ...*

<https://wiki.dlib.indiana.edu/confluence/display/INF/CV+integration+methods+for+image+cataloging+application>

*\*not yet implemented*



**DIGITAL LIBRARY PROGRAM**

INDIANA UNIVERSITY

Libraries/University Information Technology Services

Photographer/Studio: Photographic Services +

Donor's Image No.: 8139 +

Accession/Collection No.: 2000/003 +

Copyright Owner: Indiana University

Use Restrictions:

Personal Name: +

Location of Individuals in Picture:

Building Name: ? +

Corporate Name: Indiana Daily Student +

Event Name: +

Topic: News| ?

Newspaper carriers ? +

Newspaper editors ?

Newspaper headlines ? +

City: Newspaper industry

State: Newspaper vendors

Other Place Name: Newspapers +

Newsrooms

Television news anchors

Underground newspapers

The topical content of the photograph, including significant foods, objects, animals, classes of persons, architectural features and general types of events. Include one topic per field, and repeat this this field as necessary. Take value from TGM I.

\* indicates fields that should be considered

\* There might be a short delay before you

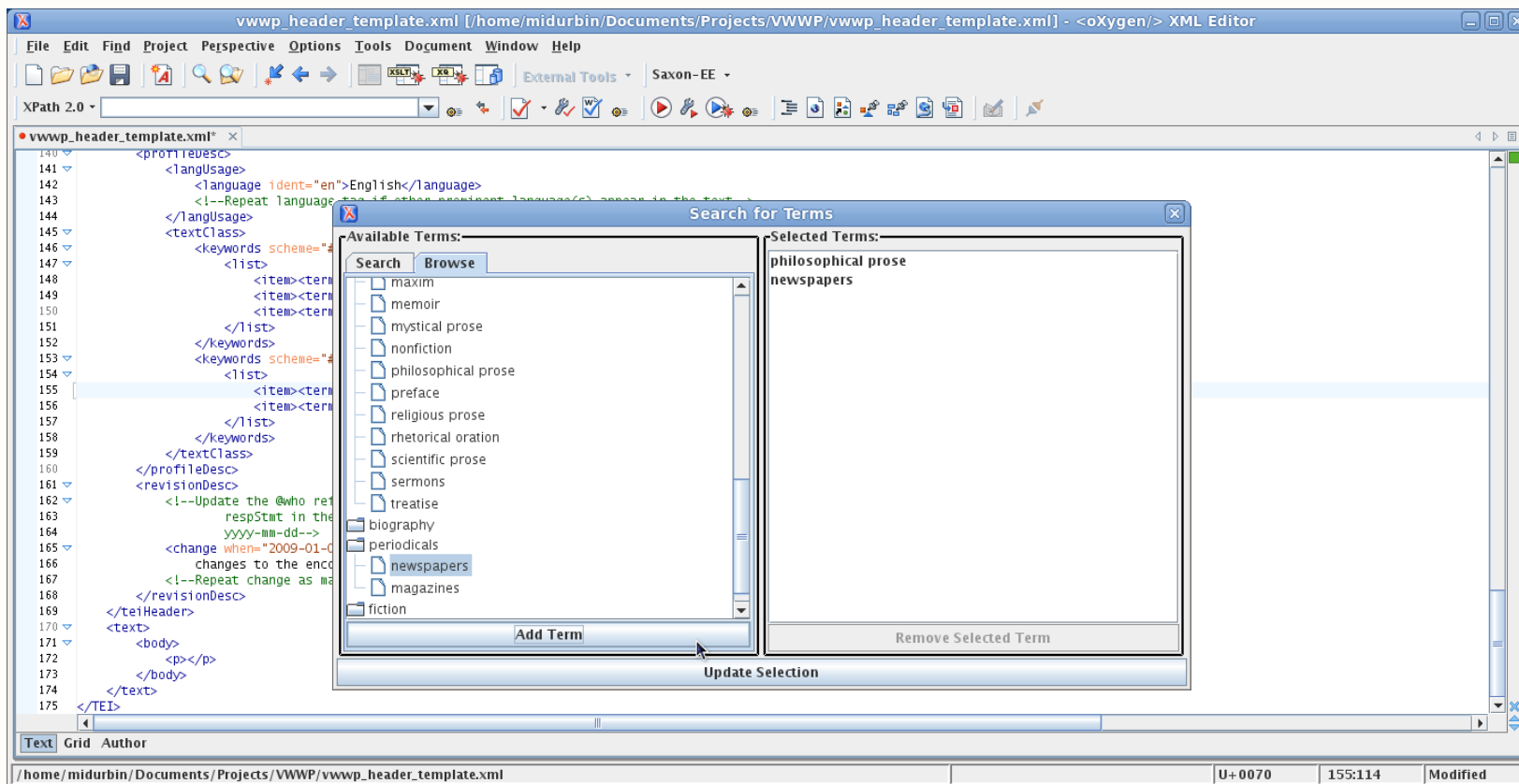
## Oxygen Plug-in

---

- Oxygen
  - Commercial XML editor
  - Used in some of our workflows
  - Easy, but somewhat limited interface for plug-in development



# Oxygen Plug-in



## DIGITAL LIBRARY PROGRAM

INDIANA UNIVERSITY

Libraries/University Information Technology Services

# Search Application

---

- Expand Queries
- Suggest alternate queries
  - Related queries
  - Broader queries
- Implementation
  - Server side or Client side



# IU DLP Image Search

about search

## Results for "vineyards":

The query you entered was expanded:

"vineyards"  
"vine"

# IU DLP Image Search

Results 1

about search

1. A N

## Results for "grapes":

### Suggested Search Terms:

- Instead of "grapes", try searching for the related term [Fruit](#)
- Instead of "grapes", try searching for the related term [Grapevines](#)
- Instead of "grapes", try searching for the related term [Wine industry](#)

Results 1 - 10 of 20

### 1. Pet raccoon is fed a grape by master.

author  
Cushman  
subject headings



1 2

## Facets

## Facets

### Collection

- [archives/cushman](#) - 17
- [lilly/hohenberger](#) - 3

### Subject

- [Agricultural facilities](#) - 1
- [Alexander Valley \(Calif.\)](#) - 4
- [Balconies](#) - 2
- [Barbed wire](#) - 1
- [Barnes, Wash](#) - 1
- [Barrels](#) - 2
- [Bays \(Bodies of water\)](#) - 1

# Indexer

---

- Feature
  - Relationship based term expansion
  - Inclusion of notes or other term information in index
- Implementation
  - XSLT
  - Java
  - Anything...



# CV Management Tool

---

- Written for a prior project (EVIADA)
- Features
  - Drag and drop Importing
  - Drag and drop relationship building





Alphabetical Listing Mode  Hierarchical Listing Mode

Genres & Performance Types **Geography** Instruments Languages Social & Cultural Groupings Venue Types

- Geography
  - + Arabian Peninsula
  - Bhandāria
  - + Brazil, Federative Republic of
  - + China, People's Republic of
  - + Cuba
  - + Côte d'Ivoire, Republic of
  - + Europe
  - + Gambia
  - + Ghana, Republic of
  - Grenada
    - Carriacou Island
      - Belvedere
      - Hillsborough
      - L'Esterre
      - Mount Royal
      - Six Roads
  - + Haiti, Republic of
  - Hong Kong (China)
    - Hong Kong (Hong Kong)
  - + India, Republic of
  - Iraq, Republic of
  - Israel
  - Jerusalem
  - + Jordan
  - + Liberia, Republic of
  - + Macedonia, Republic of
  - + Malawi, Republic of
  - + Malavsia



Alphabetical Listing Mode  Hierarchical Listing Mode

Genres & Performance Types Geography Instruments Languages Social & Cultural Groupings Venue Types

Social & Cultural Groupings

- African Americans
- African diaspora
- Africans
- Altai (Turkic people)
- Altaic peoples
  - Mongols
- Americo-Liberians
- **Ancient Order of Hibe...**
- Arabs
- Associations, institutions, etc.
- Auctioneers
- Audiences
- Aymara Indians
- Bambara (African people)
- Bantu-speaking peoples
- Baptists
- Bhavaiyas
- Bhil (Indic people)
- Bishops
- Blacks—United States
- Bondei (African people)
- Brahmans
- British Americans
- Buddhist monks
- Buddhists
- Cajuns
- Catholics
- Chewa (African people)
- Chicanos
- Children
- Chinese
- Christians
- Clergy
- Coniunto

**Ancient Order of Hibernians**

Thesaurus Authorized Term Details

Term:

Source Vocabulary:

Status:

Broader Terms:	Narrower Terms:
<input type="text"/>	<input type="text"/>

Related Terms:

Used For Terms:

Scope Notes Internal Notes Source Notes

Alphabetical Listing Mode  Hierarchical Listing Mode

Genres & Performance Types Geography Instruments Languages Social & Cultural Groupings Venue Types

- Social & Cultural Groupings
  - A.O.H.
  - Abakuria (African people)
  - Aboriginal peoples
  - Aborigines
  - Acadians in Louisiana
  - Acadians--Louis...**
  - Achewa (African people)
  - Acilowe (African people)
  - Adamawa Fula (African people)
  - Adamawa Fulani (African people)
  - African Americans**
  - African Americans--United States
  - African diaspora**
  - Africans**
  - Afro-Americans
  - Aga (African people)
  - Aimara Indians
  - Alomwe (African people)
  - Altai (Turkic people)**
  - Altaic peoples**
  - Altai--nne
  - American aborigines
  - American Indians
  - Americo-Liberians**
  - Ancient Order of Hibernians**
  - Ancient Order of Hibernians in Ameri
  - Anglo-Americans
  - Anti-Mission Baptists
  - AOH
  - Arab Palestinians
  - Arabs**
  - Arabs in Palestine--nne
  - Arabs--Palestine

Acadians--Louisiana

Thesaurus Unauthorized Term Details

Term: Acadians--Louisiana

Source Vocabulary: -- select a source vocabulary --

Status: --Select A Value--

Use: Cajuns

Scope Notes Internal Notes Source Notes

## Upgrade Validation

---

- Over time, even perfectly cataloged items may need to be updated:
  - Vocabularies change
  - Practices change
  
- Scripted validation and migration utilities can easily be written using information provided through the terminology service



## Next Steps

---

- More advanced/intuitive interactions
  - Hierarchical browsing in all cataloging applications
- Publishing/Sharing local vocabularies
- Creating/adapting vocabulary management interfaces



# What's Next?

---

- o Your Ideas...



**DIGITAL LIBRARY PROGRAM**

INDIANA UNIVERSITY

Libraries/University Information Technology Services

# Thanks!

---

## **Contact**

Michael Durbin - [midurbin@indiana.edu](mailto:midurbin@indiana.edu)

## **Links**

OCCLC Research Terminology Services -  
<http://www.oclc.org/research/activities/termservices>



**DIGITAL LIBRARY PROGRAM**

INDIANA UNIVERSITY

Libraries/University Information Technology Services