Terminology Services in a Digital Repository

Michael Durbin

Outline

o Introduction and Background

o Service Architecture

o Application Enhancements

DIGITAL LIBRARY PROGRAM

INDIANA UNIVERSITY
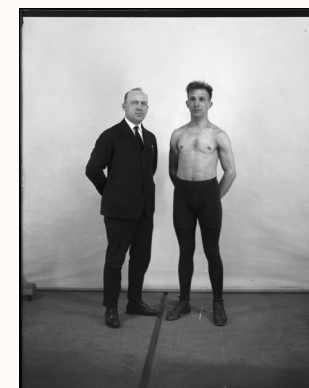
Libraries/University Information Technology Services

About Controlled Vocabularies

o List of terms that represent concepts

o Lacks the ambiguity of simple text

o Maintained by authorities

- Library of Congress

- Getty

- Local researchers/experts

o Shared to describe different resources

vs

Standardizing vocabularies and protocols

o Standard vocabularies have been in use for years

o 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
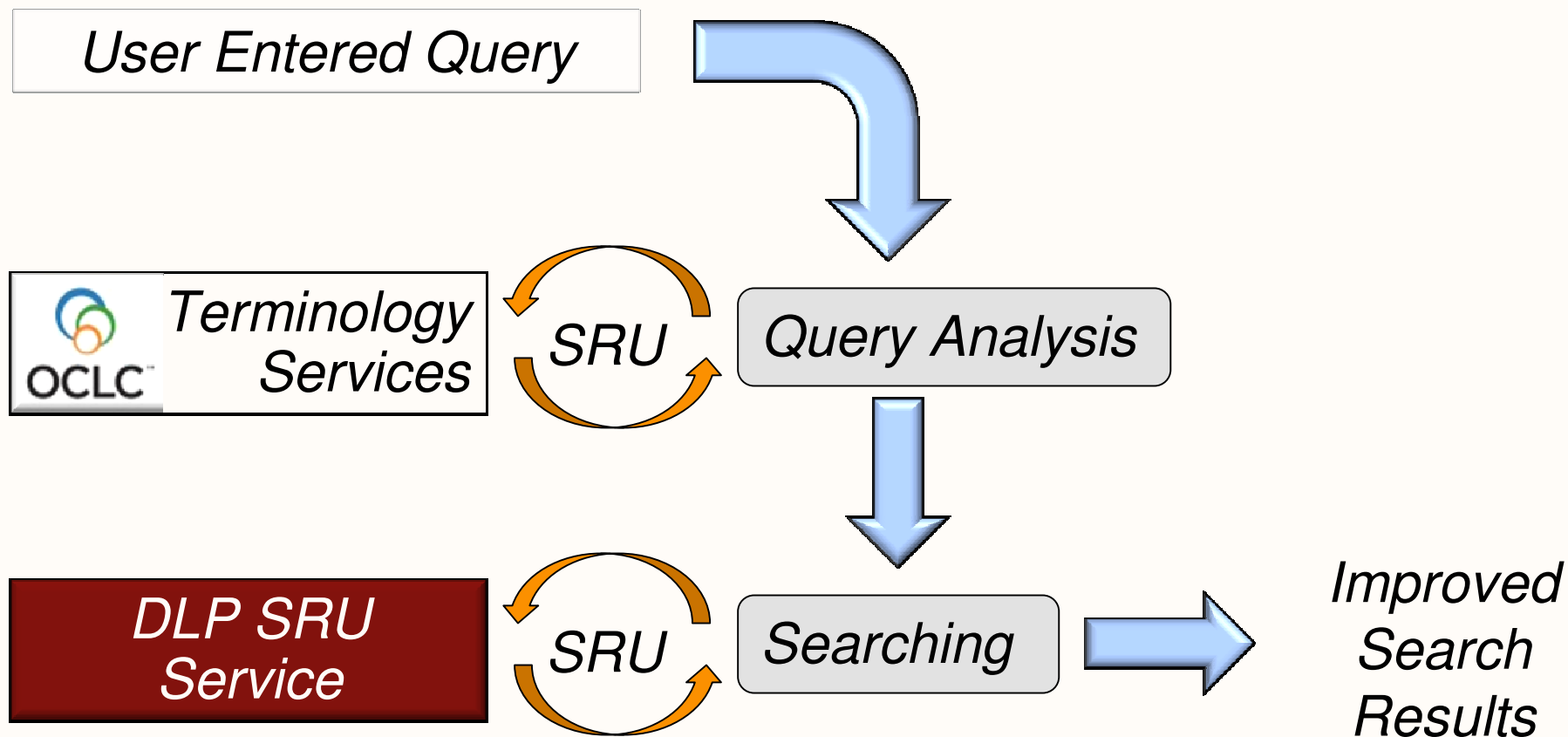Libraries/University Information Technology Services

Prototype

o Goals

- HTTP access to controlled vocabularies
- XML response
- Ability to lookup terms and then query relationships
- Use of standard technologies
- 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/)

DIGITAL LIBRARY PROGRAM

INDIANA UNIVERSITY
Libraries/University Information Technology Services

Design/**Algorithm**

Local Production Services

o 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

- One-time import from source

o Automatic Process

- Store vocabularies in FEDORA repository
- Update index in response to JMS messages

DIGITAL LIBRARY PROGRAM

INDIANA UNIVERSITY
Libraries/University Information Technology Services

Accessing Vocabularies

o SRU access point

- Search

- Disambiguate (manual, if necessary)

- Access term information or broader, narrower or related terms

o JSON wrapper to SRU

APPLICATION ENHANCEMENTS

DIGITAL LIBRARY PROGRAM

INDIANA UNIVERSITY
Libraries/University Information Technology Services

Metadata Editor: PhotoCat

o Descriptive Metadata Editor

o Primarily for photographs, using MODS schema

o Updates objects in the repository

o Some fields should be "linked" to CV sources

More information was provided at a talk at last year's Fedora User Group Meeting, at OR09 **http://hdl.handle.net/1853/28530**

DIGITAL LIBRARY PROGRAM

INDIANA UNIVERSITY

Libraries/University Information Technology Services

PhotoCat: current and future CV integration

o Auto-complete linked fields

o Thesaurus Lookup

o *Browse hierarchically ＊*

o *Validate and suggest terms＊*

o *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

Field	Value
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
Newspaper industry
Newspaper vendors
Newspapers
Newsrooms
Television news anchors
Underground newspapers

City:
State:
Other Place Name:

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 conside
* There might be a short delay before y

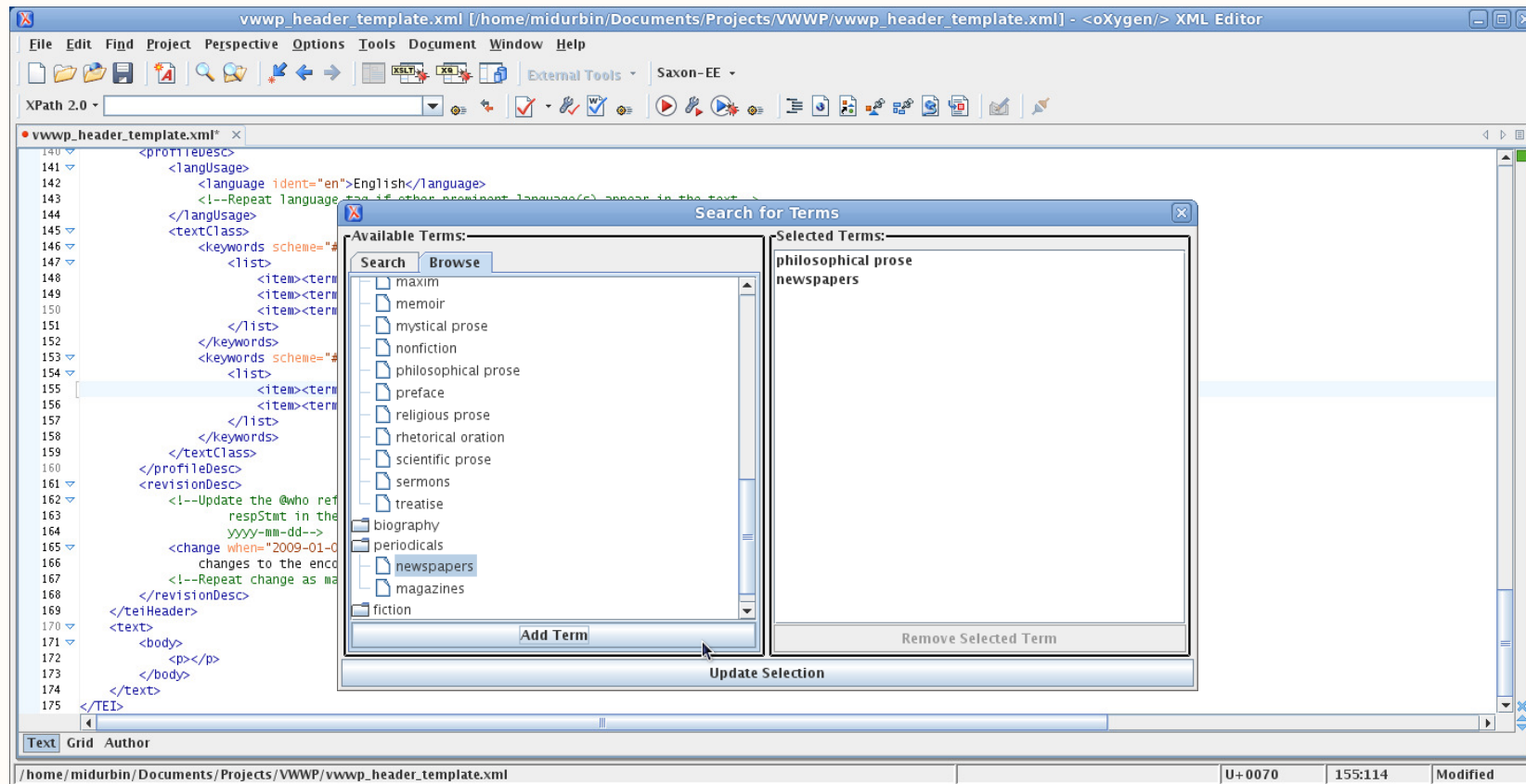# Oxygen Plug-in

o Oxygen

- Commercial XML editor
- Used in some of our workflows
  - TEI
  - EAD

DIGITAL LIBRARY PROGRAM

INDIANA UNIVERSITY

Libraries/University Information Technology Services

Oxygen Plug-in

Search Application

o Expand Queries

o Suggest alternate queries

- Related queries

- Broader queries

o Implementation

- Server side or Client side

IU DLP Image Search

Results for "*vineyards*":

The query you entered was expanded:

"vineyards
    "vine

Results 1

1. A N

au

su

IU DLP Image Search

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 2

1. Pet raccoon is fed a grape by master.

author
    Cushman
subject headings

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

Facets

Indexer

o Feature

- Relationship based term expansion
- Inclusion of notes or other term information in index

o Implementation

- XSLT
- Java
- Anything…

DIGITAL LIBRARY PROGRAM

INDIANA UNIVERSITY

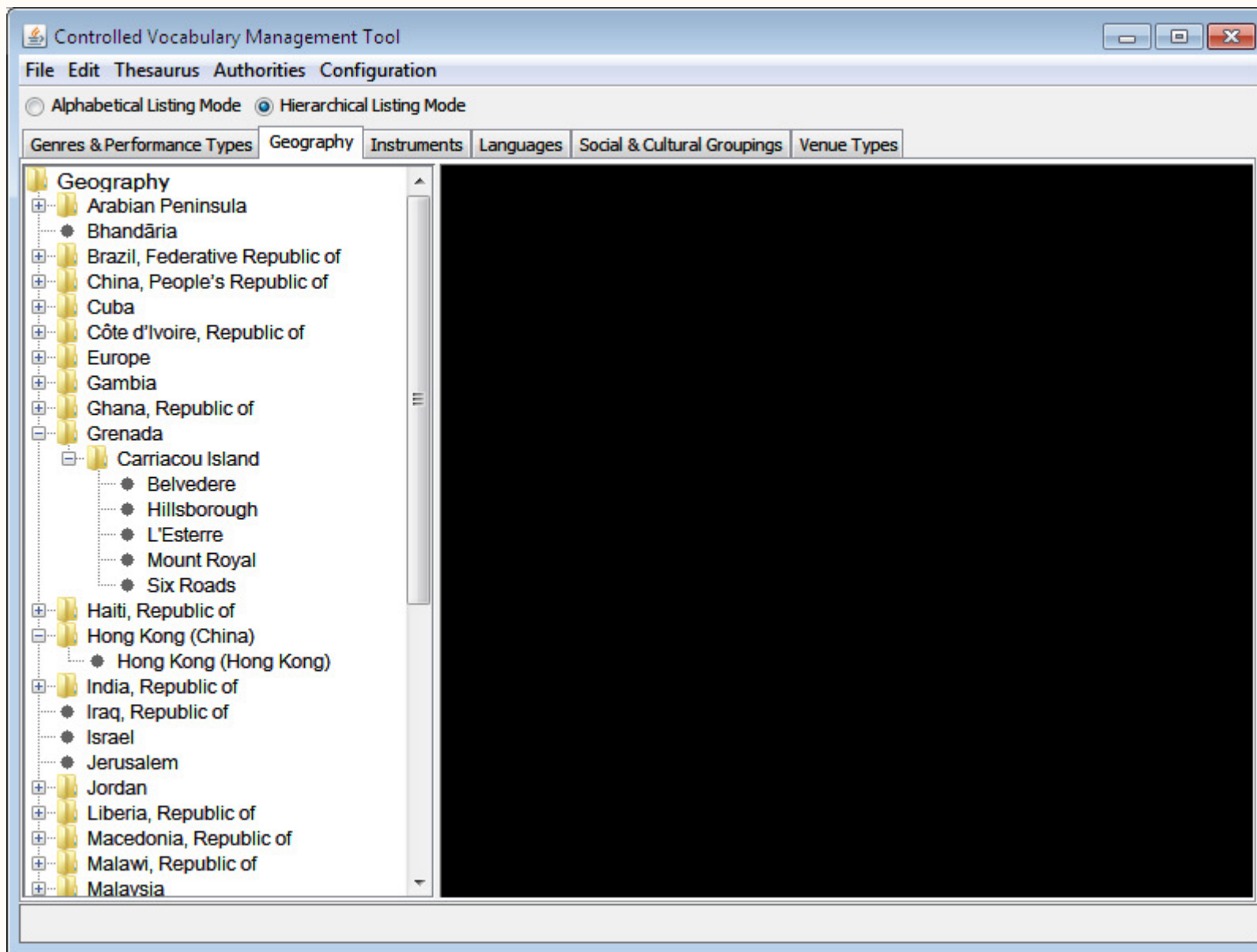Libraries/University Information Technology Services

CV Management Tool

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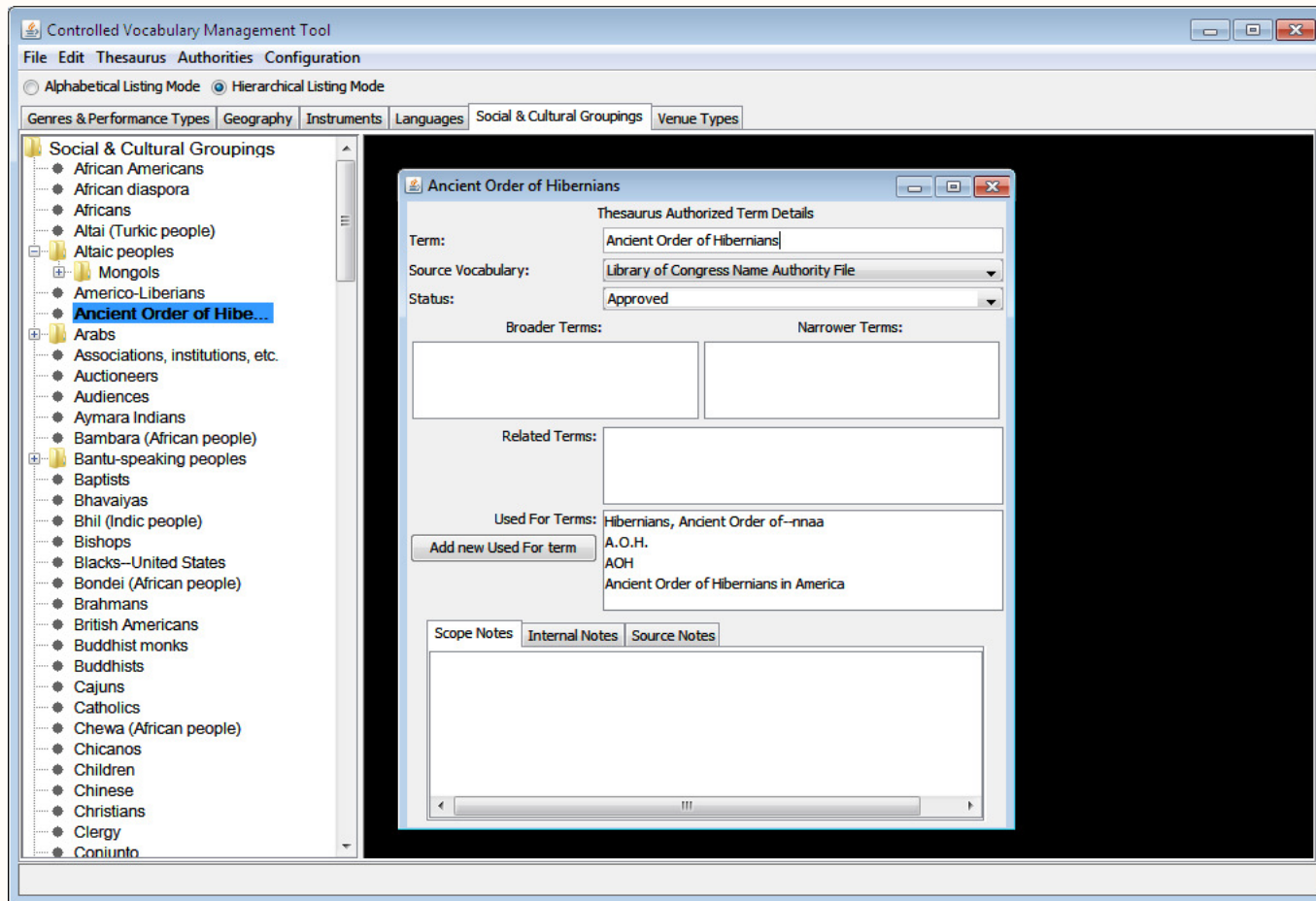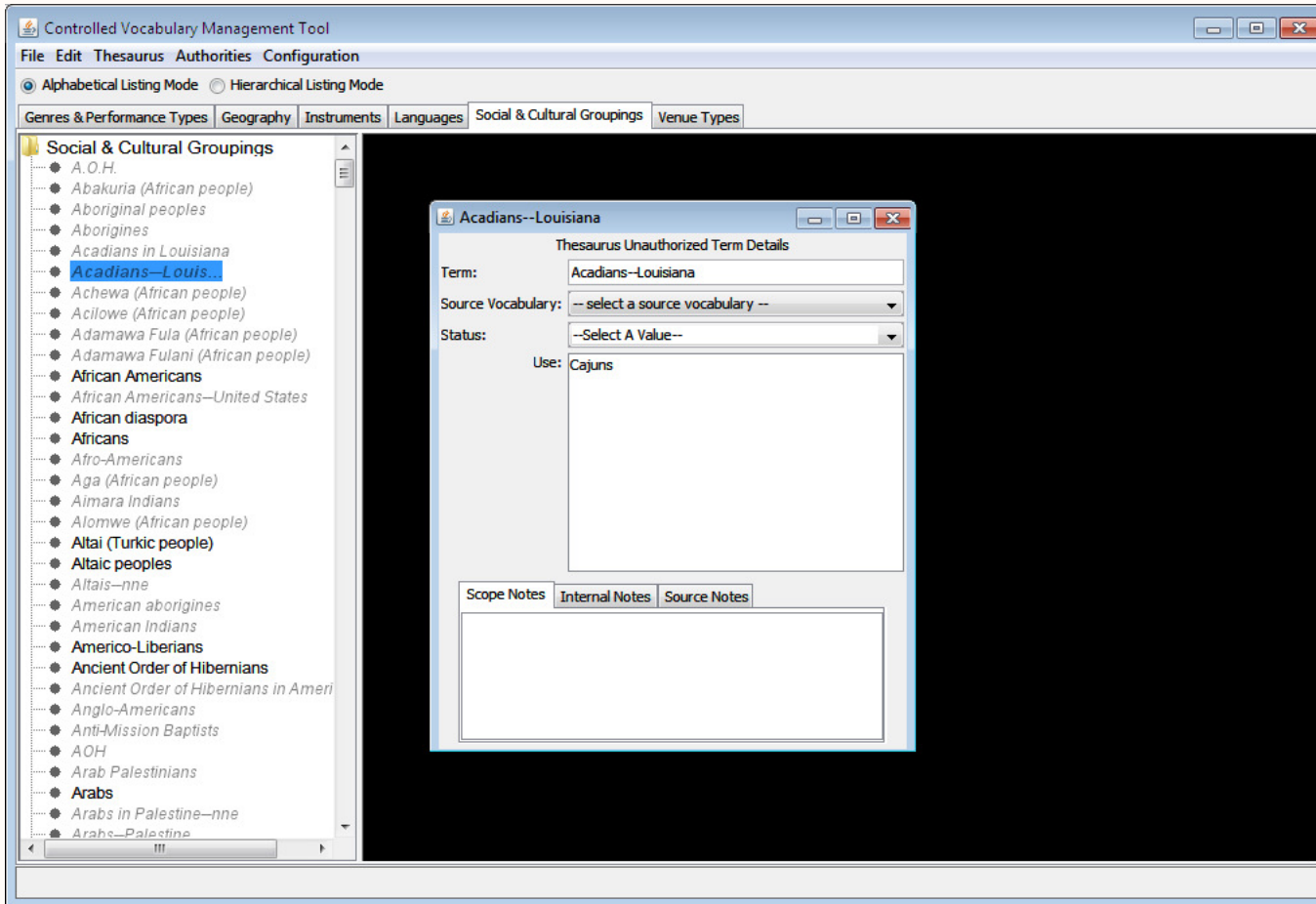- Drag and drop Importing
- Drag and drop relationship building

DIGITAL LIBRARY PROGRAM

INDIANA UNIVERSITY
Libraries/University Information Technology Services

Controlled Vocabulary Management Tool

File   Edit   Thesaurus   Authorities   Configuration

○ 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
- 📁 Malaysia

Upgrade Validation

o 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

DIGITAL LIBRARY PROGRAM

INDIANA UNIVERSITY

Libraries/University Information Technology Services

Next Steps

o More advanced/intuitive interactions

- Hierarchical browsing

o Publishing/Sharing local vocabularies

DIGITAL LIBRARY PROGRAM

INDIANA UNIVERSITY
Libraries/University Information Technology Services

What You Can Do

o Experiment with public services

o Bring up your own service

- SOLR

- JSON

- Something else

o Share your vocabularies

Thanks! Questions?

Contact
Michael Durbin - midurbin@indiana.edu

Links
OCLC Research Terminology Services -
http://www.oclc.org/research/activities/termservices

DIGITAL LIBRARY PROGRAM
INDIANA UNIVERSITY
Libraries/University Information Technology Services