Terminology Services in a Digital Repository

Michael Durbin



DIGITAL LIBRARY PROGRAM

INDIANA UNIVERSITY Libraries/University Information Technology Services

Outline

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



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

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



October 16, 2010

SERVICE ARCHITECTURE

DIGITAL LIBRARY PROGRAM

INDIANA UNIVERSITY Libraries/University Information Technology Services

τĪτ

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



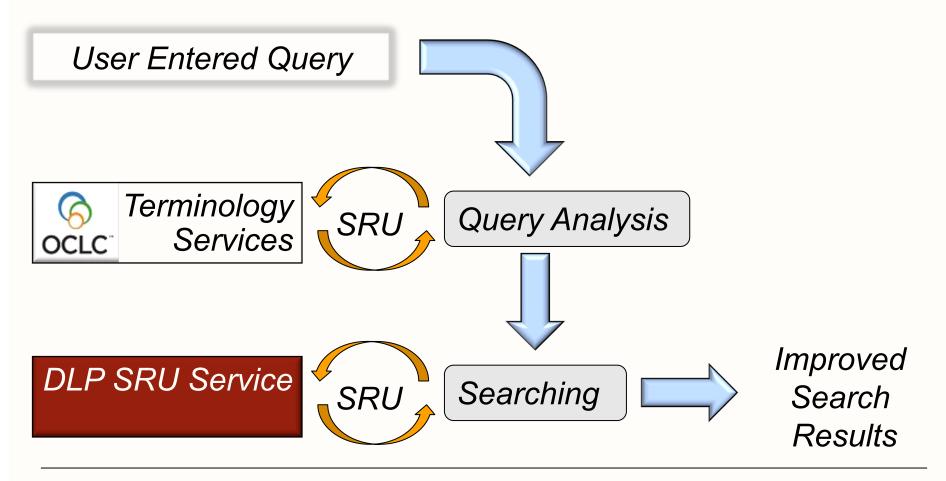
Proof Of Concept Experiment

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





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/

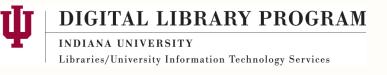
IJ DIGITAL LIBRARY PROGRAM

INDIANA UNIVERSITY Libraries/University Information Technology Services

Importing Vocabularies

o Manual Process

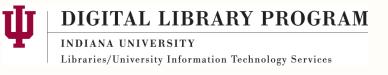
- Acquire vocabulary data
- Process it into an index
- Put that index into place
- Restart the search service
- Repeat as necessary



Importing Vocabularies

o Automatic Process

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



Vocabulary Objects in Fedora

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



Accessing Vocabularies

o 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/tgm1

o JSON wrapper to SRU

DIGITAL LIBRARY PROGRAM INDIANA UNIVERSITY Libraries/University Information Technology Services

October 16, 2010

APPLICATION ENHANCEMENTS

DIGITAL LIBRARY PROGRAM

INDIANA UNIVERSITY Libraries/University Information Technology Services

τĪΤ

Metadata Editor: PhotoCat

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



PhotoCat: current and future CV integration

- o Auto-complete linked fields
- o Thesaurus Lookup
- o Browse hierarchically *
- 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



Photographer/Studio:	Photographic Services	•		
Donor's Image No.:	8139			
Accession/Collection No.:	2000/003	•		
Copyright Owner:	Indiana University			
Use Restrictions:				
Personal Name:		Ð		
Location of Individuals in Picture:]
				el
Building Name:		? +		
Corporate Name:	Indiana Daily Student			
Event Name:				
Event Name:		+		
Topic:	News	?	The	topical content of the
	Newspaper carriers	?	phot	ograph, including significant
	Newspaper editors	? +		s, objects, animals, classes ersons, architectural features
City	Newspaper headlines		orpe	general types of events.
City:	Newspaper industry	1	Inclu	ide one topic per field, and
State:	Newspaper vendors	1		at this this field as necessary. e value from TGM I.
Other Place Name:	Newspapers	÷	Take	
	Newsrooms			
* indicates fields that should be conside	Television news anchors			
* These wight be a short delay before y	Underground nowenanors			

* There might be a short delay before y Underground newspapers

Oxygen Plug-in

o Oxygen

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



 DIGITAL LIBRARY PROGRAM

 INDIANA UNIVERSITY

 Libraries/University Information Technology Services

Oxygen Plug-in

]	vwwp_heade	r_template.xml [/home/midurbin/Documents/Project	ts/VWWP/vwwp_header_t	emplate.xml] - <oxygen></oxygen> XM	L Editor	_ D X
<u>F</u> ile <u>E</u> dit Fi <u>n</u> e	d <u>P</u> roject Pe <u>r</u> spective <u>O</u> ption:	s <u>T</u> ools Do <u>c</u> ument <u>W</u> indow <u>H</u> elp				
🗅 📂 🤌 🖡	👔 🔍 😥 🖊 🔶 🤿	External Tools - Saxon-EE -				
	• j 🔺 j 🔨 🖾 j 🥐 🗸					
XPath 2.0 -		💶 💿 🎋 📝 - 🗞 🕎 💩 🖒 🖧 🎭 🐲	. 📜 💿 🚼 🛫 📽 💁 🖢	1 🗹 🗡		
vwwp_header_	template.xml* ×					< ▷ ■
140 🗸	<protilevesc></protilevesc>					
141 🗸	<1 angUsage>					
142 143	<language ident="en
<!Repeat language</td><td><pre>">English</language> tag if other prominent language(s) appear in the text >					
145		X Search	for Terms	R		
145 😎	<textclass></textclass>	Available Terms:	Selected Terms:			
146 🔝	<keywords #<="" scheme="#</td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>147 🔽</td><td><list></td><td>Search Browse</td><td>philosophical prose</td><td></td><td></td><td></td></tr><tr><td>148</td><td><item><tern</td><td></td><td>newspapers</td><td></td><td></td><td></td></tr><tr><td>149
150</td><td><item><tern
<item><tern</td><td>🛛 — 🗋 memoir</td><td></td><td></td><td></td><td></td></tr><tr><td>151</td><td></list></td><td>- nystical prose</td><td></td><td></td><td></td><td></td></tr><tr><td>152</td><td></keywords></td><td>- nonfiction</td><td></td><td></td><td></td><td></td></tr><tr><td>153 🤝</td><td><keywords scheme=" td=""><td></td><td></td><td></td><td></td><td></td></keywords>					
154 😎	<list></list>	– 🗋 philosophical prose				
155	<item><tern< td=""><td>— 🗋 preface</td><td></td><td></td><td></td><td></td></tern<></item>	— 🗋 preface				
156 157	<item><term </term </item>	🗕 🗋 religious prose				
158		- 🗋 rhetorical oration				
159		- 🗋 scientific prose				
160						
161 🤝	<revisiondesc></revisiondesc>	- C sermons				
162 🗸	Update the @who ref</td <td>🗆 🗋 treatise</td> <td></td> <td></td> <td></td> <td></td>	🗆 🗋 treatise				
163 164	respStmt in the yyyy-mm-dd>	🗂 biography				
165 -	<change <="" td="" when="2009-01-0"><td>🚍 periodicals</td><td></td><td></td><td></td><td></td></change>	🚍 periodicals				
166	changes to the enco	- newspapers	11			
167	Repeat change as ma</td <td>— 🗋 magazines</td> <td>11</td> <td></td> <td></td> <td></td>	— 🗋 magazines	11			
168						
	teiHeader>	fiction 💌				
170 マ <te 171 マ</te 	<body></body>	Add Term	Remove	Selected Term		
172		Ľ	11			
173		Update	Selection			
	text>					
175						_ _
						•
Text Grid Aut	hor					
ome/midurbin	n/Documents/Projects/VWWP/vw	www.haadar.tampiata.ymi			U+0070 155:	114 Modified
one/muulbii	i, bocaments/ riojects/ vwwP/ v	wwp_neaver_template.ann			1040010 155:	114 Joounned



Search Application

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



 $\Psi_{\rm bloomington}^{\rm indiana\ university}$

IU DLP Image Search

about search

Results for "vineyards":

Results fo	r "vineyards":	Facets	
The query ye	ulentered was expanded:		
"vineyards "vine			
ome	IU DLP Image Search		
Results 1	about search		
1. A N	Results for "grapes":		Facets
au	Suggested Search Terms:		Collection
su	 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 		 archives/cushman - 17 lilly/hohenberger - 3
	Results 1 - 10 of 20	12	Subject
	1. Pet raccoon is fed a grape by master.		 Agricultural facilities - 1 Alexander Valley (Calif.) - 4 Balconies - 2 Barbed wire - 1 Barbed wire - 1
	author Cushman subject headings	Re -	 Barnes, Wash - 1 Barrels - 2 Bays (Bodies of water) - 1

Indexer

o Feature

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

UIGITAL LIBRARY PROGRAM INDIANA UNIVERSITY Libraries/University Information Technology Services

CV Management Tool

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

J DIGITAL LIBRARY PROGRAM INDIANA UNIVERSITY Libraries/University Information Technology Services

🙆 Controlled Vocabulary Management Tool					
File Edit Thesaurus Authorities Configuration					
Alphabetical Listing Mode OHIERARCHICAL Listing Mode					
Genres & Performance Types Geography Instruments Languages Social & Cultural Groupings Venue Types					
Geography Arabian Peninsula Arabian Peninsula Brazil, Federative Republic of China, People's Republic of Colta China, People's Republic of Colta China, People's Republic of Colta China, Colta					

₹)	Controlled	Vocabulary	Management	Тоо
----	------------	------------	------------	-----

File Edit Thesaurus Authorities Configuration

Alphabetical Listing Mode O Hierarchical Listing Mode

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

 African Americans African diaspora African diaspora African diaspora African diaspora African s African conservation Atlaic (Turkic people) Ancient Order of Hibernians Source Vocabulary: Library of Congress Name Authority File Source Vocabulary: Library of Congress Name Authority File Status: Approved Narrower Terms: <li< th=""><th>Social & Cultural Groupings</th><th><u>~</u></th><th></th><th></th><th></th><th></th><th></th></li<>	Social & Cultural Groupings	<u>~</u>					
 Africans Africans Africans Africans Atlai (Turkic people) Atlai (peoples) Ancient Order of Hibernians Source Vocabulary: Library of Congress Name Authority File Ancient Order of Hibernians Source Vocabulary: Library of Congress Name Authority File Status: Approved Approved Ancient Order of Hibernians Banbara (African people) Bantu-speaking peoples Baptists Bhavaiyas Bhil (Indic people) Bishops Banderi (African people) Brathmans Bordei (African people) British Americans Buddhists Cajuns Catholics Chewa (African people) Chicanos Children Chinese Christans 			A	61 PL - 1			
 Altai (Turkic people) Altaic peoples Mongols Ancient Order of Hibernians Broader Terms: Narrower Terms:	•		Ancient Orde	r of Hibernians			
Altaic peoples Mongols Ancient Order of Hibe Americo-Liberians Ancient Order of Hibe Arabs Source Vocabulary: Library of Congress Name Authority File Arabs Associations, institutions, etc. Audiences Aymara Indians Bambara (African people) Bantu-speaking peoples Baptists Bhavaiyas Bishops Bishops Bishops Brahmans Brahmans Buddhists Cajuns Catholics Chicanos Children Chinese Chinese		=		The	saurus Authorize	ed Term Details	
Attaic peoples Ancient Order of Hibe Arabs Arabs Arabs Avabs Avabs Aucioneers Audiences Audiences Audiences Audiences Aymara Indians Bantu-speaking peoples Baptists Baptists Bishops Bishops Bishops Bishops Braitish Americans Buddhists Cajuns Chewa (African people) Buddhists Cajuns Chewa (African people) Chicanos Children Chirstians			Term:	Ar	ncient Order of H	libernians	
 Americo-Liberians Arabs Arabs Arabs Associations, institutions, etc. Auciences Aymara Indians Bambara (African people) Bantu-speaking peoples Baptists Bhavaiyas Bhil (Indic people) Bishops Bishops Bishops Braiths Americans Buddhists Catholics Chewa (African people) Chicanos Chicanos Chirstians Chirstians 						•	
Ancient Order of Hibe Arabs Arabs Arabs Arabs Arabs Banbara (African people) Bantu-speaking peoples Used For Terms: Hibernians, Ancient Order of-nnaa A.O.H. AOH Add new Used For term A.O.H. AOH Ancient Order of Hibernians in America Scope Notes Internal Notes Scope Notes Internal Notes Scope Notes Chidren Chidren Christians			Source Vocabulary	/:	orary of Congres	is Name Authority File	▼.
Arabs Arabs Associations, institutions, etc. Audiences Audiences Aymara Indians Bambara (African people) Bantu-speaking peoples Baptists Bhavayas Bhil (Indic people) Bandu-speaking peoples Bayists Bayists Bayists Bantu-speaking peoples Bayists Buddhist monks Buddhists Buddhists Catholics Chicanos Chidren Chicanos Chidren Chinese Christians			Status:	Ar	proved		_
 Alabs Associations, institutions, etc. Audiences Related Terms: Banburspeaking peoples Bantu-speaking peop	1997		Br	oader Terms:		Narrower Ter	ms:
 Auctioneers Audiences Aymara Indians Bambara (African people) Bantu-speaking peoples Baptists Bhavaiyas Bhil (Indic people) Bishops Bishops Blacks-United States Bondei (African people) Brahmans Brahmans Brahmans Brahmans Brahmans Brahmans Cigiuns Catholics Chewa (African people) Chicanos Children Chinese Christians 							
 Audiences Aymara Indians Bambara (African people) Bantu-speaking peoples Baptists Bhavaiyas Bhavaiyas Bhavaiyas Bhavaiyas Bishops Bishops BlacksUnited States Bondei (African people) Brahmans British Americans Buddhists Cajuns Catholics Chewa (African people) Chinaps Children Chinaps Chinaps Chinaps Chinaps Christians 							
 Aymara Indians Aymara Indians Bambara (African people) Bantu-speaking peoples Baptists Bhavaiyas Bhil (Indic people) Bishops Blacks-United States Bondei (African people) Brahmans British Americans Buddhists Cajuns Catholics Chewa (African people) Chinaps Chinaps Chinaps Chinaps Christians 							
 Bambara (African people) Bantu-speaking peoples Baptists Bhavaiyas Bhavaiyas Bhil (Indic people) Used For Terms: Hibernians, Ancient Order ofnnaa A.O.H. AOH Ancient Order of Hibernians in America Scope Notes Internal Notes Source Notes Catholics Chinanos Chinanos Chinanos Chinase Christians 							
 Bantu-speaking peoples Bantu-speaking peoples Baptists Bhavaiyas Bhil (Indic people) Bishops BlacksUnited States Bondei (African people) Brahmans British Americans Buddhists Cajuns Catholics Chewa (African people) Chicanos Children Chinese Christians 			Del	atad Torma			
 Baptists Bhavaiyas Bhil (Indic people) Bishops Bishops BlacksUnited States Bondei (African people) Brahmans British Americans Buddhists Cajuns Catholics Chewa (African people) Chicanos Children Chinese Christians 			Ke	ateu remis:			
 Bhavaiyas Bhil (Indic people) Bishops Blacks-United States Bondei (African people) Brahmans British Americans Buddhist monks Buddhists Cajuns Catholics Chewa (African people) Chicanos Children Christians 							
 Bhil (Indic people) Bishops Blacks-United States Bondei (African people) Brahmans British Americans Buddhist monks Buddhists Cajuns Catholics Chewa (African people) Chicanos Chinese Christians 							
 Bishops BlacksUnited States Bondei (African people) Brahmans British Americans Buddhist monks Buddhists Cajuns Catholics Chewa (African people) Chicanos Chicanos Children Chinese Christians 	-		Lice	Eor Termer		1 Order of seven	
 Blacks-United States Bondei (African people) Brahmans British Americans Buddhists monks Buddhists Cajuns Catholics Chewa (African people) Chicanos Children Chinese Christians 						t Order ofnnaa	
 Bondei (African people) Brahmans British Americans Buddhist monks Buddhists Cajuns Catholics Chewa (African people) Chicanos Children Chinese Christians 	-		Add new Used				
 Brahmans British Americans Buddhist monks Buddhists Cajuns Catholics Chewa (African people) Chicanos Children Chinese Christians 						·	
 British Americans Buddhist monks Buddhists Cajuns Catholics Chewa (African people) Chicanos Children Chinese Christians 				An	ident Order of H	idernians in America	
 Buddhist monks Buddhists Cajuns Catholics Chewa (African people) Chicanos Children Chinese Christians 							
 Buddhists Cajuns Catholics Chewa (African people) Chicanos Children Chinese Chinese Christians 			Scope Notes	Internal Notes	Source Notes		
 Cajuns Catholics Chewa (African people) Chicanos Children Chinese Christians 							
 Catholics Chewa (African people) Chicanos Children Chinese Christians 							
 Chewa (African people) Chicanos Children Chinese Christians 	-						
 Chicanos Children Chinese Christians 							
 Children Chinese Christians 							
Christians							
Christians							
			•		111		+
	Clergy						
Conjunto		-					

File Edit Thesaurus Authorities Configuration Adphabetical Listing Mode Herminical String Mode Social & Cultural Groupings - A.O.R. - A.D.R. - A.M.A.M.RICANS - A.M.A.M.RICANS - A.M.A.M.RICANS - A.M.A.M.RICANS - A.M.A.M.RICANS - A.N. - A.M.A.M.RICANS - A.M.M.A.M.RICANS - A.D.R. - A.M.M.K.SLONG - A.M.M.SISON Baptists - A.D.R. - A.D.R. <	🖆 Controlled Vocabulary Management Tool					
Gerres & Performance Types Geography Instruments Languages Social & Cultural Groupings A AO.H. A.O.H. A.O.H. A.O.H. A Abakuna (African people) Aboriginal peoples Image: Column of the colum	File Edit Thesaurus Authorities Configuration					
Gerres & Performance Types Geography Instruments Languages Social & Cultural Groupings A AO.H. A.O.H. A.O.H. A.O.H. A Abakuna (African people) Aboriginal peoples Image: Column of the colum	Alphabetical Listing Mode Mierarchical Listing Mode					
 Social & Cultural Groupings A. O.H. A. A. O.H. A. A. D.H. A. A. D.H. A. A. D.H. A. A. D.H. A. Abariginal peoples A. Acadians in Louislana A. Cadians in Louislana A. Acadians - Louisiana A. Acadians - Louisiana A. Cadians - Louislana Source Vocabulary - Source Vocabulary - Source Vocabulary - Source Vocabulary - Status: Select A Value Valica Nameicans A. Anter. Anglo-A. Maricana people) A. Ataic peoples A. Anter. Oliginas A. Anglo-A. Mericans A. Anglo-A. Mericans<						
 A O.H. A bakuria (African people) A boriginal peoples A cardians Louisiana Cardians Louisiana A cardians Louisiana A cardians Louisiana Source Vocabulary: select a source vocabulary Adamawa Fulian (African people) A damawa Fulian (African people) A frican Americans A frican diaspora African andians A frican andians A falt nne A merican indians Ancient Order of Hibernians in Ameri Anglo-Americans A cardb Paistina <li< th=""><th>Genres & Performance Types Geography Instruments Languages Sodar & Culu</th><th>Venue Types</th><th></th></li<>	Genres & Performance Types Geography Instruments Languages Sodar & Culu	Venue Types				
Arahs-Palestine	 Social & Cultural Groupings A.O.H. Abakuria (African people) Aboriginal peoples Acadians in Louisiana Adamawa Fula (African people) African Americans African Americans African Indians Ander American so Anti-Kiran peoples Antiai Corder of Hibernians Ancient Order of Hibernians in Ameri Anglo-Americans Andi-Americans Andi-Americans Andi-Mission Baptists AOH Arabs 	ouisiana Thesaurus Unauthorized Term Details AcadiansLouisiana ry:select a source vocabularySelect A Value se: Cajuns				
	Arahs-Palestine					

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

o More advanced/intuitive interactions

- Hierarchical browsing in all cataloging applications
- o Publishing/Sharing local vocabularies
- Creating/adapting vocabulary management interfaces



October 16, 2010

What's Next?

o Your Ideas...

DIGITAL LIBRARY PROGRAM

INDIANA UNIVERSITY Libraries/University Information Technology Services

Thanks!

Contact Michael Durbin - <u>midurbin@indiana.edu</u>

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

