

Terminology Services

Diane Vizine-Goetz

Senior Research Scientist

OCLC Research



A version of this presentation was given at:

- New Dimensions in Knowledge Organization Systems: A Joint NKOS/CENDI Workshop
The World Bank
Washington, DC
11 September 2008
- OCLC / ISKO-NA Preconference
Université de Montréal
Montréal, Canada
5 August 2008

Moving Vocabularies to the Network Level

Requirements:

- Expressive data structures
- Vocabularies encoded for the Web
- Access mechanisms for search and retrieval
- URI accessible content
- Use of open protocols and standards

Success will be measured by the appearance of applications that use or combine vocabulary data to create new derivative works or tools.

OCCLC Terminology Services Prototype



Employs library and Web standards to make the terms and relationships in controlled vocabularies available as Web resources.

Top-level site intended for machines

```
- <Service xmlns="urn:uuid:D30A7E67-31BF-40A3-9956-9668674FCD84" xmlns:dct="http://purl.org/dc/terms/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" rdf:about="http://tspilot.oclc.org/" xml:base="http://tspilot.oclc.org/">
- <Providers rdf:parseType="Collection">
- <Provider rdf:about="fast">
  <dct:type rdf:datatype="http://www.w3.org/2001/XMLSchema#token">REST interface provider</dct:type>
  <dct:language rdf:datatype="http://purl.org/dc/terms/ISO639-2">en</dct:language>
  <dct:format rdf:datatype="http://purl.org/dc/terms/IMT">application/x-oclc-tspilot.rest</dct:format>
  <dct:format rdf:datatype="http://purl.org/dc/terms/IMT">application/xml</dct:format>
  <dct:format rdf:datatype="http://purl.org/dc/terms/IMT">text/xml</dct:format>
  <dct:format rdf:datatype="http://www.iana.org/assignments/character-sets">UTF-8</dct:format>
  <dct:conformsTo rdf:resource="urn:uuid:D30A7E67-31BF-40A3-9956-9668674FCD84" />
</Provider>
- <Provider rdf:about="fast/">
  <dct:type rdf:datatype="http://www.w3.org/2001/XMLSchema#token">SRU/W dual interface provider</dct:type>
  <dct:language rdf:datatype="http://purl.org/dc/terms/ISO639-2">en</dct:language>
  <dct:format rdf:datatype="http://purl.org/dc/terms/IMT">application/x-oclc-tspilot.sru</dct:format>
  <dct:format rdf:datatype="http://purl.org/dc/terms/IMT">application/x-oclc-tspilot.srw</dct:format>
  <dct:format rdf:datatype="http://purl.org/dc/terms/IMT">application/xml</dct:format>
  <dct:format rdf:datatype="http://purl.org/dc/terms/IMT">text/xml</dct:format>
  <dct:format rdf:datatype="http://www.iana.org/assignments/character-sets">UTF-8</dct:format>
  <dct:conformsTo rdf:resource="http://www.loc.gov:8081/standards/sru/sru1-1archive/search-retrieve-operation.html" />
  <dct:conformsTo rdf:resource="http://www.loc.gov:8081/standards/sru/sru1-1archive/srw.html" />
</Provider>
- <Provider rdf:about="fast.sru">
  <dct:type rdf:datatype="http://www.w3.org/2001/XMLSchema#token">SRU specific interface provider</dct:type>
  <dct:language rdf:datatype="http://purl.org/dc/terms/ISO639-2">en</dct:language>
  <dct:format rdf:datatype="http://purl.org/dc/terms/IMT">application/x-oclc-tspilot.sru</dct:format>
  <dct:format rdf:datatype="http://purl.org/dc/terms/IMT">application/xml</dct:format>
  <dct:format rdf:datatype="http://purl.org/dc/terms/IMT">text/xml</dct:format>
  <dct:format rdf:datatype="http://www.iana.org/assignments/character-sets">UTF-8</dct:format>
  <dct:conformsTo rdf:resource="http://www.loc.gov:8081/standards/sru/sru1-1archive/search-retrieve-operation.html" />
</Provider>
- <Provider rdf:about="fast.srw">
  <dct:type rdf:datatype="http://www.w3.org/2001/XMLSchema#token">SRW specific interface provider</dct:type>
  <dct:language rdf:datatype="http://purl.org/dc/terms/ISO639-2">en</dct:language>
  <dct:format rdf:datatype="http://purl.org/dc/terms/IMT">application/x-oclc-tspilot.srw</dct:format>
  <dct:format rdf:datatype="http://purl.org/dc/terms/IMT">application/xml</dct:format>
  <dct:format rdf:datatype="http://purl.org/dc/terms/IMT">text/xml</dct:format>
  <dct:format rdf:datatype="http://www.iana.org/assignments/character-sets">UTF-8</dct:format>
  <dct:conformsTo rdf:resource="http://www.loc.gov:8081/standards/sru/sru1-1archive/srw.html" />
</Provider>
- <Provider rdf:about="mgpc">
  <dct:type rdf:datatype="http://www.w3.org/2001/XMLSchema#token">REST interface provider</dct:type>
  <dct:language rdf:datatype="http://purl.org/dc/terms/ISO639-2">en</dct:language>
  <dct:format rdf:datatype="http://purl.org/dc/terms/IMT">application/x-oclc-tspilot.rest</dct:format>
  <dct:format rdf:datatype="http://purl.org/dc/terms/IMT">application/xml</dct:format>
```

Human interface for everyone else



Terminology Services

Experimental Services for Controlled Vocabularies

a project of OCLC Research

Highlights

- Search descriptions of controlled vocabularies
- Search for concepts/headings in a controlled vocabulary
- Retrieve a single concept/heading by its identifier
- View relationships for a concept/heading including equivalence, hierarchical, and associative
- Retrieve concepts/headings in multiple representations including HTML, MARC XML, SKOS, and Zthes.
- Search using SRU CQL syntax

Technical Resources

- Overview ([pdf](#), [ppt](#))
- URI-patterns
- [Indexing](#)
- [Standards](#)
- [Issues](#)
- [Roadmap](#)

Vocabulary Resources

Jump to: [meta](#) [fast](#) [gsafd](#) [lcshac](#) [lcsh](#) [mesh](#) [lctgm](#) [gmgpc](#)

Name:	Controlled Vocabulary Metadata
Description:	Descriptions of controlled vocabularies
Date:	2008-03
Identifier:	meta
Links:	About SRU Interface Examples MARC Statistics

Content

Web Services

Applications

Vocabularies

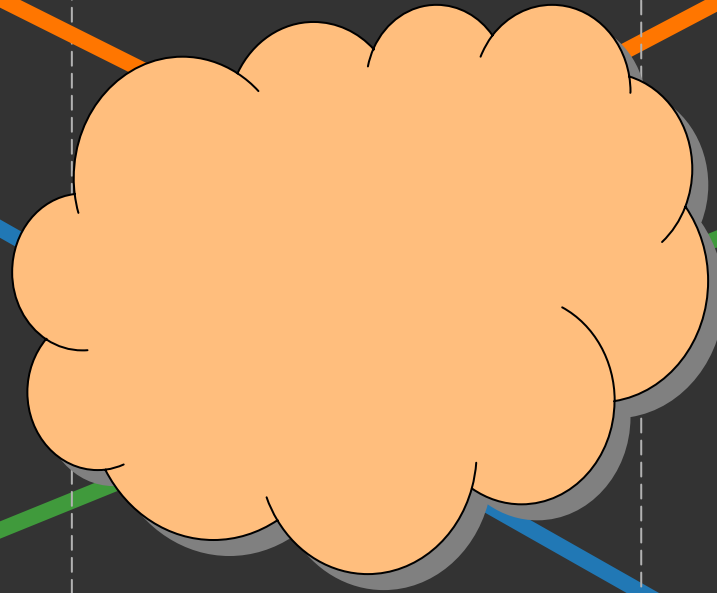
fast

gsafd

lcsh

mesh

lctgm & gmgpc



Query Expansion



Searching Heterogeneous Collections



Metadata Creation

Content

Vocabularies

fast

gsafd

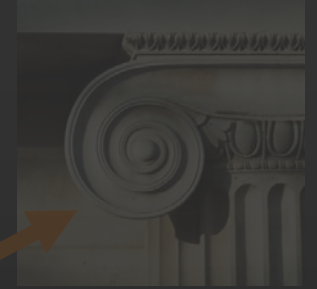
lcsh

mesh

lctgm & gmgpc

Web Services

Applications



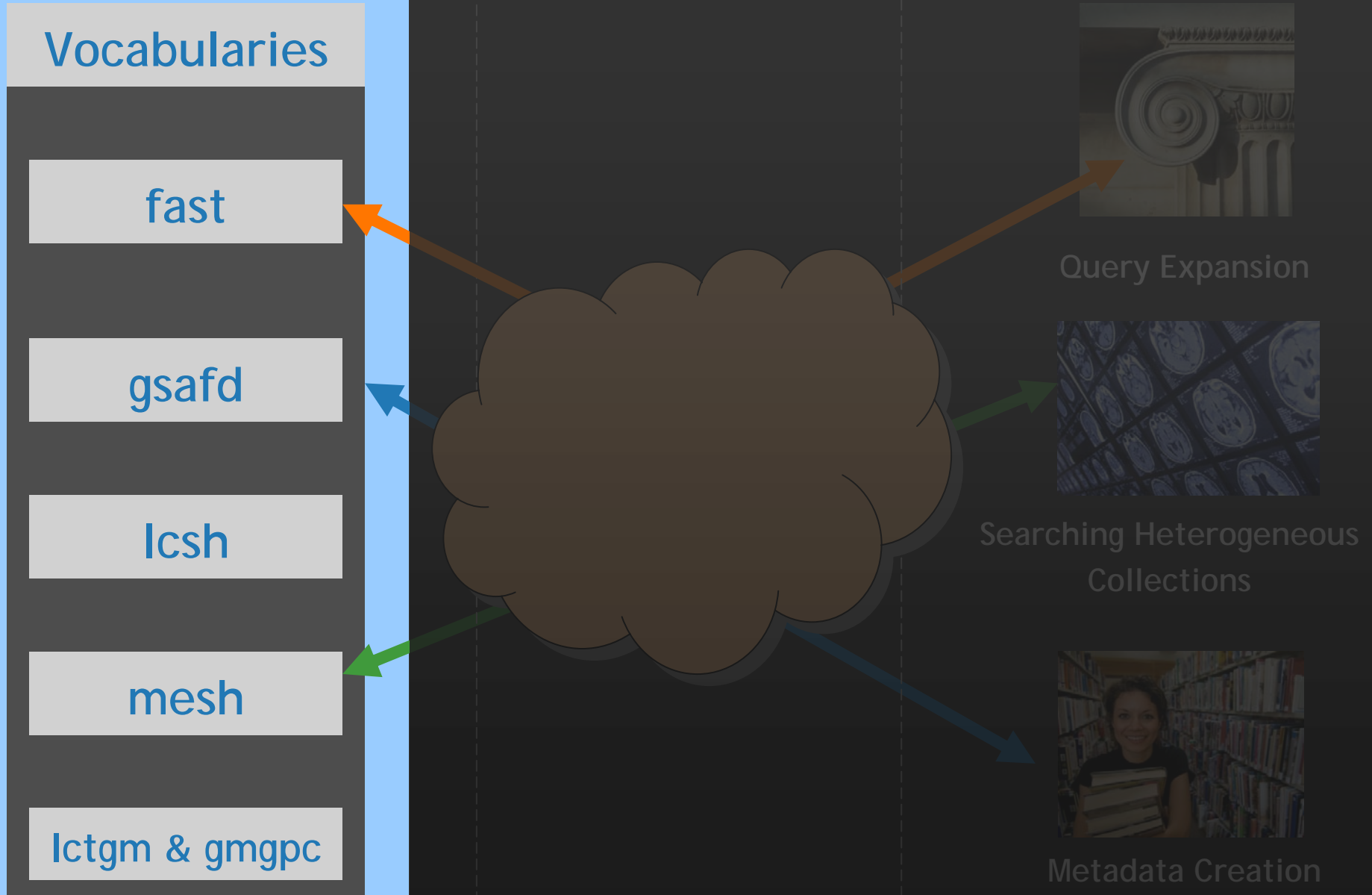
Query Expansion



Searching Heterogeneous Collections



Metadata Creation



Types of Controlled Vocabularies (Hodge 2000)



* indicates availability in Terminology Services Prototype

Term Lists

- Authority Files*
- Glossaries
- Dictionaries
- Gazetteers

Classifications and Categories

- Subject Headings*
- Classification Schemes*
- Taxonomies
- Categorization Schemes

Relationship Lists

- Thesauri*
- Semantic Networks
- Ontologies

Vocabularies (August 2008)



1. Faceted Application of Subject Terminology (fast)
2. Form and Genre Terms for Fiction and Drama (gsafd)
3. Library of Congress Subject Headings (lcsch)
4. Medical Subject Headings (mesh)
5. Thesaurus for Graphic Materials: TGM I, Subject Terms (lctgm)
6. Thesaurus for Graphic Materials: TGM II, Genre and Physical Characteristics (gmgpc)

MARC 21 Format for Authority Data

- A format for the use and exchange of information about the authorized forms of names and subjects used as access points in MARC bibliographic records.

Simple Knowledge Organization System (SKOS)

- SKOS Core is a model and an RDF vocabulary for expressing the basic structure and content of concept schemes such as thesauri, classification schemes, subject heading lists, taxonomies, 'folksonomies', [etc.]

Zthes

- A model for representing thesauri* and a specification for expressing them in XML. Zthes also provides specifications for searching Zthes compliant data using SRU/SRW or Z39.50.

Content

Vocabularies

fast

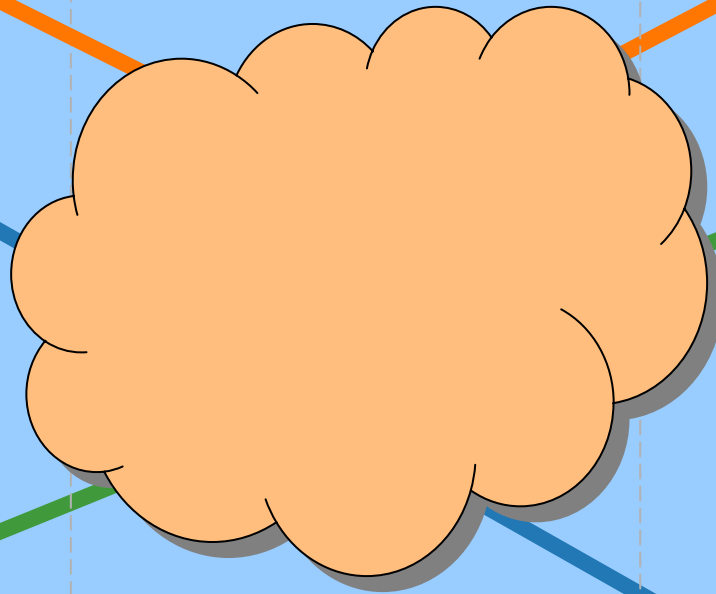
gsafd

lcsh

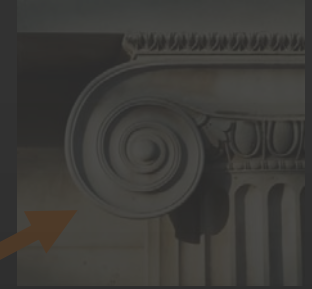
mesh

lctgm & gmgpc

Web Services



Applications



Query Expansion



Searching Heterogeneous Collections



Metadata Creation

XML (Extensible Markup Language)

- A data-interchange format for custom markup languages.

RDF (Resource Description Framework)

- A data-interchange format for the representation of graph models.

JSON (Javascript Object Notation)

- A data-interchange format based on a subset of the JavaScript Programming Language defined by the ECMA-262 3rd Edition standard.

REST (Representational State Transfer)

- A software architecture style used for building distributed systems that retrieve Web resources.

SRU & SRW (Search/Retrieve via URL)

- A standard search protocol that utilizes the Contextual Query Language (CQL) syntax to retrieve Web resources.

Content

Web Services

Applications

Vocabularies

fast

gsafd

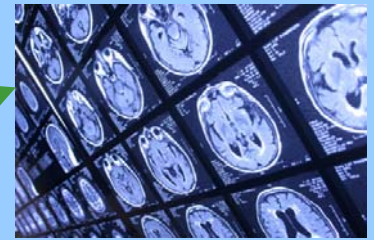
lcsh

mesh

lctgm & gmgpc



Query Expansion



Searching Heterogeneous Collections



Metadata Creation

IU Digital Library Program Collection Search

home search browse

Results for "temples":

1396ms.

The query you entered was expanded:

"temples" was expanded to search:

"[temples](#)" "[Buddhist temples](#)" "[Confucian temples](#)" "[Greek temples](#)" "[Hindu temples](#)" "[Roman temples](#)" "[Taoist temples](#)"

Suggested Search Terms:

- Instead of "temples", try searching for the broader term [Religious facilities](#)
- Instead of "temples", try searching for the related term [Churches](#)
- Instead of "temples", try searching for the related term [Pagodas](#)
- Instead of "temples", try searching for the related term [Tori](#)

Results 1 - 10 of 92

1 2 3 4 5

1 [Masonic Temple](#)



2 [Old Corinth](#)



Facets

Collection

- [archives/cushman](#) - 91
- [nw/cra/ussteel](#) - 1

Subject

- [Acropolis \(Athens\)](#) - 1
- [Acropolis \(Athens\)](#) - 2
- [Acropolis \(Athens, Greece\)](#) - 1
- [Acropolis \(Athens, Greece\)](#) - 42
- [Agora \(Athens, Greece\)](#) - 10
- [Arcades \(Architectural components\)](#) - 5
- [Arch of Hadrian \(Athens, Greece\)](#) - 5
- [Archaeological sites](#) - 76
- [Arches](#) - 15
- [Architectural decorations & ornaments](#) - 5
- [Architectural elements](#) - 1
- [Architectural sculpture](#) - 6
- [Automobiles](#) - 1
- [Balconies](#) - 3
- [Battlements](#) - 1
- [Boundaries](#) - 2
- [Bright Angel Point \(Ariz.\)](#) - 1
- [Building deterioration](#) - 1
- [Buildings](#) - 9
- [Buttes](#) - 2
- [Buttresses](#) - 2
- [Canyons](#) - 5
- [Capitals \(Columns\)](#) - 21
- [Capitol Hill \(Salt Lake City, Utah\)](#) - 1
- [Capitols](#) - 1
- [Caryatids](#) - 6
- [Ceilings](#) - 2
- [Children](#) - 1
- [Churches](#) - 5
- [Cities & towns](#) - 9
- [Cliffs](#) - 5

1

User Enters Search



2

Query Sent to TS Prototype

<http://tspilot.oclc.org/Ictgm/?query=oclcts.expandedHeading+exact+%22temples%22&version=1.1&operation=searchRetrieve>

3

MARC XML Returned

```
039__$a (DLC)Ictgm-010644
039__$a (DLC)Ictgm-10644
040__$a DLC$b eng$c OCoLC$d OCoLC$d OCoLC-O$f Ictgm$9 Ictgm
150__$a Temples$9 temples
550__$w g$a Religious facilities$0 (DLC)Ictgm008761$9 religious facilities
550__$w h$a Buddhist temples$0 (DLC)Ictgm001379$9 buddhist temples
550__$w h$a Confucian temples$0 (DLC)Ictgm002437$9 confucian temples
550__$w h$a Greek temples$0 (DLC)Ictgm004717$9 greek temples
550__$w h$a Hindu temples$0 (DLC)Ictgm004994$9 hindu temples
550__$w h$a Roman temples$0 (DLC)Ictgm008977$9 roman temples
550__$w h$a Taoist temples$0 (DLC)Ictgm010519$9 taoist temples
550__$a Churches$0 (DLC)Ictgm002048$9 churches
```

4

Client application extracts terms for query expansion

IU Digital Library Program Collection Search

home search browse

Results for "temples":

The query you entered was expanded:

"temples" was expanded to search:

["temples"](#) ["Buddhist temples"](#) ["Confucian temples"](#) ["Greek temples"](#) ["Hindu t](#)

Suggested Search Terms:

- Instead of "temples", try searching for the broader term [Religious facilities](#)
- Instead of "temples", try searching for the related term [Churches](#)
- Instead of "temples", try searching for the related term [Pagodas](#)

040__\$a DLC\$b eng\$c OCoLC\$d OCoLC\$d OCoLC-O\$f lctgm\$9
lctgm

150__\$a Temples\$9 temples

550__\$w g\$a Religious facilities\$0 (DLC)lctgm008761

550__\$w h\$a Buddhist temples\$0 (DLC)lctgm001379

550__\$w h\$a Confucian temples\$0 (DLC)lctgm002437

550__\$w h\$a Greek temples\$0 (DLC)lctgm004717

550__\$w h\$a Hindu temples\$0 (DLC)lctgm004994

550__\$w h\$a Roman temples\$0 (DLC)lctgm008977

550__\$w h\$a Taoist temples\$0 (DLC)lctgm010519

550__\$a Churches\$0 (DLC)lctgm002048

550__\$a Pagodas\$0 (DLC)lctgm007367

550__\$a Pronaoi\$0 (DLC)lctgm008289

550__\$a Torii\$0 (DLC)lctgm010868

Narrower terms



Results for "temples":

1674ms.

The query you entered was expanded:

"temples" was expanded to search:

["temples"](#) ["Buddhist temples"](#) ["Confucian temples"](#) ["Greek temples"](#) ["Hindu temples"](#) ["Roman temples"](#) ["Taoist temples"](#)

Suggested Search Terms:

- Instead of "temples", try searching for the broader term [Religious facilities](#)
- Instead of "temples", try searching for the related term [Churches](#)
- Instead of "temples", try searching for the related term [Pagodas](#)
- Instead of "temples", try searching for the related term [Torii](#)

Results 1 - 10 of 92

1 [2](#) [3](#) [4](#) [5](#)

URI accessible content



{URL for the service}/{vocabulary}/{identifier}.{format}

- <http://tspilot.oclc.org/lcsh/sh95000541.html>
- <http://tspilot.oclc.org/lcsh/sh95000541.json>
- <http://tspilot.oclc.org/lcsh/sh95000541.marcxml>
- <http://tspilot.oclc.org/lcsh/sh95000541.skos>
- <http://tspilot.oclc.org/lcsh/sh95000541.zthes>

The template shows how to link to vocabulary data in the prototype



{URL for the service}/{vocabulary}/{identifier}.{format}

1. Base URL for the service - <http://tspilot.oclc.org>
2. Vocabulary - the code for the controlled vocabulary in the MARC code list for terms
 - <http://www.loc.gov/marc/relators/relasour.html#rela6xx>
 - <http://www.loc.gov/marc/relators/relasour.html#rela655>
3. Identifier - a control number associated with a concept or term (e.g., an LCCN - Library of Congress Control Number)
4. Format - the representation of the vocabulary data

- More vocabularies
- Access to complete term hierarchies
- MADS profile
- Additional mappings

Terminology Services Prototype

- <http://tspilot.oclc.org/resources/>
- <http://tspilot.oclc.org> (machine interface)

Project page

- <http://www.oclc.org/research/projects/termservices/>