Library of Congress Controlled Vocabularies as Linked Data: http://id.loc.gov

Clay Redding

Library of Congress

Network Development &
MARC Standards Office

Credits

- Ed Summers, LC Office of Strategic Initiatives: leading developer and creator
- Antoine Isaac, Vrije Universiteit Amsterdam & TELplus project for RAMEAU mappings
- Rebecca Guenther, LC NDMSO: assistance with previous slides and presentations

About SKOS

- Simple Knowledge Organization System
- RDF application used to express knowledge organization systems such as classifications, thesauri, taxonomies, and the concepts within.
- Allows distributed, decentralized management of KOS through Linked Data-inspired application.
- All instances of SKOS classes are resources and their descriptions require a URI.

About SKOS: Data Model (Classes)

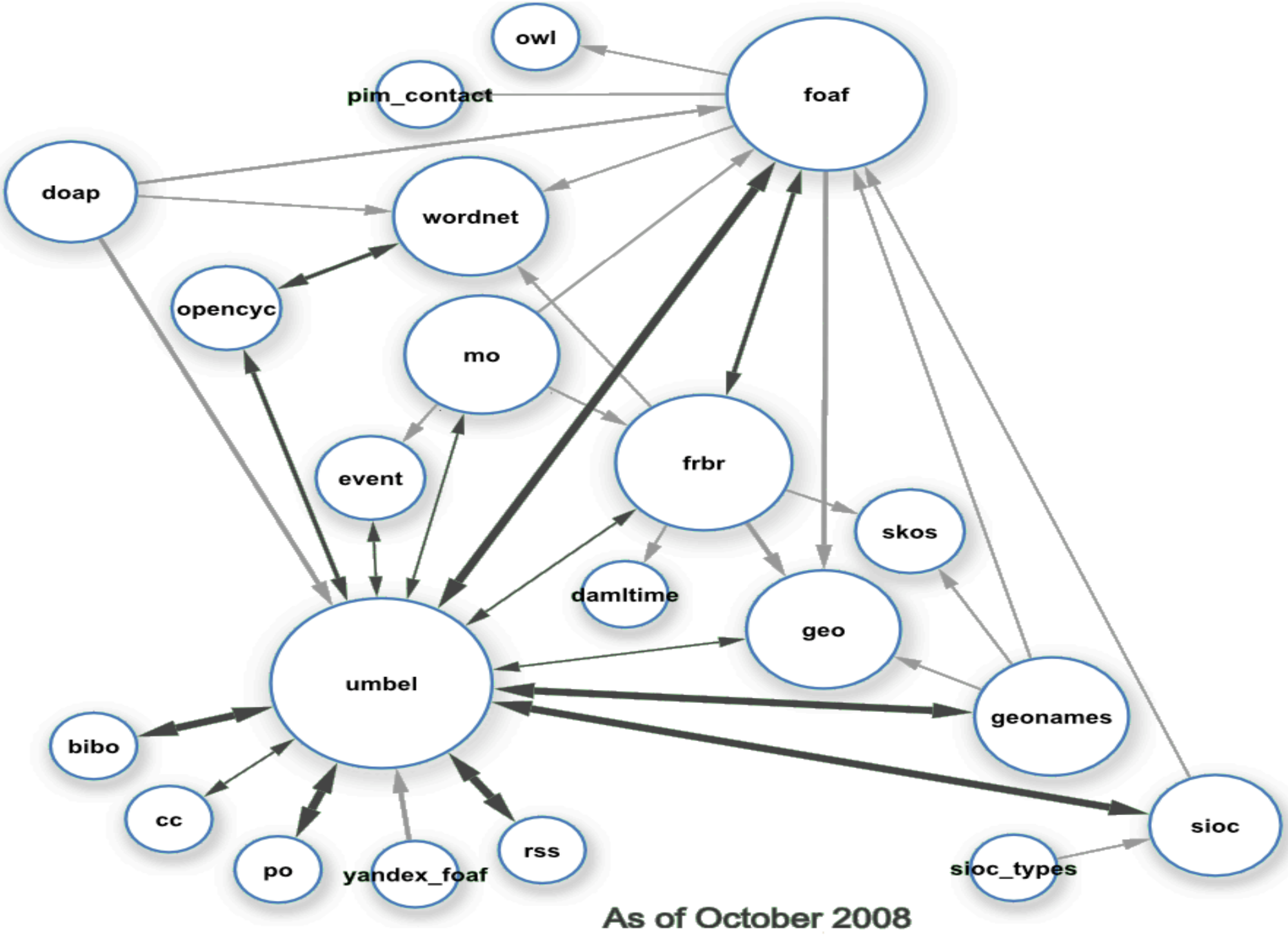
- ConceptSchemes (e.g., published vocabularies, thesauri, code lists, etc.)
- **Concepts** (individual entries or terms within the larger vocabulary)
- Collections (logical groupings of Concepts)
- Disjoint: Concepts can be in ConceptSchemes and be members of Collections, but are not contained by them. Linked by URIs instead.

About SKOS: Concepts

- Labeling properties: **prefLabel, altLabel, hiddenLabel, notation**
- Annotation properties: **note, historyNote, scopeNote, changeNote, editorialNote, example, definition**
- Associative properties: **broader, narrower, related, broadMatch, narrowMatch, closeMatch, exactMatch, minorMatch, majorMatch** (match properties go to Concepts in external ConceptSchemes)

About Linked Data

- Wikipedia: "Method of exposing, sharing, and connecting data via dereferenceable URIs."
- A feature of the Semantic Web where links are made between resources using RDF.
- Users can "follow their noses" by traversing links to find similar resources; inferencing two or more RDF descriptions on similar resources can aggregate all assertions into one to create new knowledge.



Source: Wikipedia

Introducing id.loc.gov

- Library of Congress Authorities & Vocabularies service: http://id.loc.gov
- Allows both human-oriented and programmatic access to LC-promulgated authorities and vocabularies.
- First offering is Library of Congress Subject Headings, but more to come: Thesaurus of Graphic Materials, ISO 639 code lists, MARC code lists, etc.

Introducing id.loc.gov

- Offers bulk data downloads in several RDF serializations (likely more to come).
- Goals
  - Convey a clear, concise policy regarding access, usage and distribution.
  - Provide continuous updates to keep the data sets fresh.

Introducing id.loc.gov

- Provides resolvability by assigning RESTful URIs. Each vocabulary and data value within it possesses a resolvable URI.
- Only serves data values: authority and vocabulary data, not bibliographic.
- Promotes these resources to first-class, dereferenceable web resources
- Influenced by the Linked Data movement; implements SKOS, REST, and HTTP content negotiation

Features of id.loc.gov

- Supports OpenSearch for browser-driven search and Atom feeds for change notices.

- Known-label searches: use when you know the label but not the identifier/LCCN

  http://id.loc.gov/authorities/label/orchids,
  http://id.loc.gov/authorities/label/orchidaceae

  - HTTP Redirects to the proper URI

Features of id.loc.gov

Linked Data to RAMEAU via the TELplus Project.

Linkage between LCSH and RAMEAU by Multilingual Access to Subjects Project (MACS)

Future: NAL Thesaurus? OCLC Terminology Services? TGM, etc.

```
<dcterms:source xml:lang="en">Random House (grass family)</dcterms:source>
<skos:closeMatch rdf:resource="http://stitch.cs.vu.nl/vocabularies/rameau/ark:/12148/cb11931581x"/>
<skos:closeMatch rdf:resource="http://stitch.cs.vu.nl/vocabularies/rameau/ark:/12148/cb12328041w"/>
<skos:broader rdf:resource="http://id.loc.gov/authorities/sh85035165#concept"/>
```

Features of id.loc.gov

- Interaction with any given individual term and vocabulary with through its URI.
  - Default serialization for is RDFa + XHTML, which can be transformed by RDFa tools
  - RDFa is a W3C-driven, standards-based approach to microformats.
- Visualizations
- Content negotiation on Concept and Concept Scheme URIs

HTTP hooks

- Content negotiation on the "Accept" request header.
  - For use with HTTP-aware libraries in your programming language of choice
  - Console utilities like cURL and wget
- Response header "X-PrefLabel" returns a URL-encoded (percent-encoded) representation of the preferred label or heading on GET and HEAD HTTP requests.

Rethinking our metadata

- Moving from data value strings to URIs enables true authority references.
- RDF: strings/literals still prevalent. When URIs are used instead, most tools will resolve to grab labels/text as needed for inclusion in metadata.
- XML: move from plain text nodes to a XLink+XQuery or XInclude approach
- HTTP-aware XQuery modules in eXist and Zorba engines can do content negotiation.

Next steps

- Advocacy, user feedback, etc.
- Implement update mechanism for processing changes processed from LC CDS.
- Expand system to allow more vocabularies and Linked Data relationships
- MADS OWL Schema to enable identification of facets within name and subject authorities:
  Aeronautics--Soviet Union--History