

ID.LOC.GOV

Sharing Vocabulary and Authority Data at the Library of Congress

<http://id.loc.gov>

Kevin Ford
NDMSO, Library of Congress
kefo@loc.gov

12 August 2010
Federal Data Architecture Subcommittee
Arlington, VA

Presentation Outline

Overview of ID.LOC.GOV

- Goal

- Authorities and Vocabularies in LIS Community

- Service offerings

- How it's used

Recipe for Success (?)

- Historical context

- Timeline events

- Data format select

- Data additions and updates

- Service offerings

- Community engagement

- Promotion

ID: What is it?

**ID makes LC owned or maintained
authorities and vocabularies
available as Linked Data.**

ID: What is it?

Available Authorities and Vocabularies

- Library of Congress Subject Headings (LCSH)
- Thesaurus of Geographic Materials
- MARC Code List for Relators
- Cryptographic Hash Functions
- Preservation Events
- Preservation Level Role

Coming in September (I hope):

- Countries
- Geographic Areas
- Organizations (less certain)
- Controlled Lists used in Preservation (also pending)

ID: What is it?

Art, Renaissance

From Library of Congress Subject Headings

Details Visualization Suggest Terminology

Art, Renaissance

URI
<<http://id.loc.gov/authorities/sh85007860#concept>>

Type
Topical Term

Alternate Labels

- > Art, Renaissance--Europe
- > Renaissance art

Narrower Terms

- > [Little masters \(Artists\)](#)

Created
2002-10-29

Modified
2002-12-04 10:03:10

Similar concepts from other vocabularies

- > <http://stitch.cs.vu.nl/vocabularies/rameau/ark:/12148/cb11952475c> ↗

Instance Of

- > [SKOS Concept](#) ↗

Alternate Formats

- > [RDF/XML](#)
- > [N-Triples](#)
- > [JSON](#)

Controlled Vocabs

Controlled Vocabularies & Authorities

Control values that occur in metadata

Reduce ambiguity

Control synonyms

Document and publish for reuse

Establish formal relationships among terms (where appropriate)

650 _0 |a Art, Renaissance |z Italy.

650 _0 |a Competition (Psychology)

600 00 |a Michelangelo Buonarroti, |d 1475-1564.

ID: What's Offered

Getting the data

Bulk downloads of each authority and vocabulary

Download Current Version

LCSH RDF/XML (May 26 2010; 37.6 MB) Zip Compressed

Individual resources via content negotiation

XHTML/RDFa

RDF/XML

N-Triples

JSON

Alternate Formats

> [RDF/XML](#)

> [N-Triples](#)

> [JSON](#)

ID: What's Offered

Other features

Dereferenced URIs

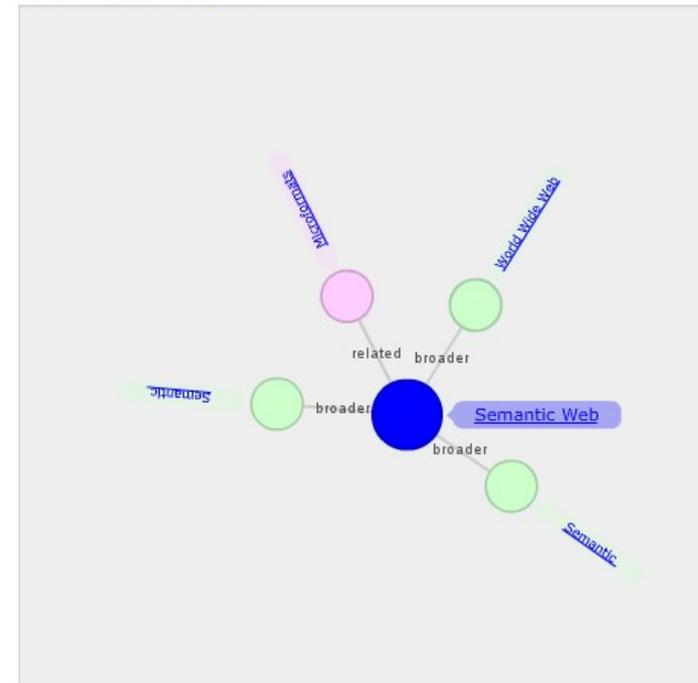
UI search and discovery (all or specific vocabularies)

Type-ahead for LCSH

ATOM feed of updated LCSH terms

Visualization

Semantic Web



ID: How it's used

John Ocklerbloom, University of Pennsylvania Online Books

Used LCSH bulk download, manipulated for local needs
Integrated with OPAC records
Broadens user search/discovery experience

Medicine, Botanic

Here are entered works on a 19th century system of medicine developed by Samuel Thomson and based on the use of plant remedies.

Broader terms:

- [Alternative medicine](#)
- [Medicine](#)

Used for:

- Botanic medicine
- Thomsonianism

Filed under: [Medicine, Botanic](#)

① [A Guide to Health, Being an Exposition of the Principles of the Thomsonian System of Practice, and Their Mode of Application in the Cure of Every Form of Disease \(1846 edition\)](#), by Benjamin Colby (text and PDF with commentary at swsbm.com)

① [Life and Medical Discoveries of Samuel Thomson, and a History of the Thomsonian Materia Medica, As Shown in "The New Guide to Health" \(1835\), and the Literature of That Day \(1909\)](#), by John Uri Lloyd and Samuel Thomson (PDF at swsbm.com)

Items below (if any) are from related and broader terms.

Filed under: [Alternative medicine](#)

① [Natural Liberty: Rediscovering Self-Induced Abortion Methods \(Creative Commons licensed edition, c2008\)](#), by Sage-Femme Collective (Javascript-dependent Flash at scribd.com)

ID: How it's used

Ethan Gruber, University of Virginia

Data used in cataloging application

Indexed with Solr (Lucene)

XForms application with autosuggest feature

Data loaded locally – harvested weekly

EADitor

0 records received from id.loc.gov.

[< Return](#)

Manage Controlled Vocabularies

You can use this form to scrape controlled vocabulary terms from all EAD guides in the eXist database and add the terms into the Solr vocabularies index. Eventually one will be able to create custom vocabularies and edit existing ones.

[Update LCSH](#) [Import](#) [Create/Edit](#) [Purge](#) [Optimize](#)

Update LCSH terms from Atom Feed

This function enables querying the Library of Congress's LCSH service at id.loc.gov to extract terms from the Atom feed that have been created or updated since the previous date this function was run. This may take several minutes.

Date of last update: 2010-06-23T14:15:55.754-04:00

[Update Terms](#)

EADitor ©2010 Ethan Gruber. [License](#).

ID: How it's used

RAMEAU

Part of MACS (Multi-lingual Access to Subjects) project

Determined skos:closeMatch resources between RAMEAU and LCSH

Both RAMEAU and LCSH cross linked

Concept information

URI	http://stitch.cs.vu.nl/vocabularies/rameau/ark:/12148/cb11932889r	
prefLabel	x-notation	FRBNF119328898
	fr	Oiseaux

Similar concepts from other vocabularies

> <http://stitch.cs.vu.nl/vocabularies/rameau/ark:/12148/cb11932889r> ↗

Mappings (simple SKOS statements)

Mapping Relation	Concept
closeMatch	http://id.loc.gov/authorities/sh85014310#concept

Mappings (OAEI cells)

Concept as left hand side

URI	Link to details	prefLabel	Mapping Relation
http://id.loc.gov/authorities/sh85014310#concept	details	http://id.loc.gov/authorities/sh85014310#concept	http://www.w3.org/2004/02/skos/core#closeMatch
http://id.loc.gov/authorities/sh85014310#concept	details	http://id.loc.gov/authorities/sh85014310#concept	http://www.w3.org/2004/02/skos/core#closeMatch

ID: How it's used

National Library Sweden

Unreleased/published Linked Data component of catalog
Links to LCSH

Swedish subject heading: "Mödrar"

Broader terms

[Föräldrar](#)

Narrower terms

[Ensamstående mödrar](#)

[Lesbiska mödrar](#)

[Soldatmödrar](#)

[Styvmödrar](#)

[Surrogatmödrar](#)

[Unga mödrar](#)

[Yrkesarbetande mödrar](#)

Related terms

[Gravida kvinnor](#)

[Mor-barnrelationer](#)

LCSH:

[Mothers](#)

misc.

[text/rdf+n3](#)

[application/rdf+xml](#)

[OpenLink RDF browser](#)

<http://id.loc.gov/authorities/sh85087526>

Recipe for Success

Stipulating that ID.LOC.GOV has been a success for the Library of Congress, what has contributed to its success?

Recipe for Success

Ingredients

- History (and culture) of making data available
- Timeline events – smart and subtle planning, sprinkle of luck
- Embrace of open, accepted, well-supported metadata standards
- Data refreshing and additions
- Added services
- History of strong user support
 - Continued relationship with community
- Promotion
- Significant internal interest, and support

Historical Context

History of Making Data Available

What:

Bibliographic records

As cards: 1900s-1980s

In MARC format: 1960s-present

In MARCXML: 2007-present

In MODS: 2007-present

In Dublin Core: 2007-present

Authority records

As cards: 1900s-1980s

In MARC format: 1960s-present

Historical Context

History of Making Data Available

How:

Cataloging Distribution Service (CDS)

catalog.loc.gov (OPAC)

Individual bibliographic records in MARC

Individual bibliographic records in text format

authorities.loc.gov

Individual authority records in MARC

Individual authority records in text format

Search interfaces

Z39.50

SRU

OAI-PMH

Historical Context

Bottom Line...

We want to make our data available
We like making our data available

Timeline Events

- 2006 – Barbara Tillett and Corey Harper publish article *Library of Congress controlled vocabularies and their application to the Semantic Web*
- 2007 – Ed Summers & Clay Redding appointed to W3C SWDG
- 2008 – lcsh.info - Ed Summers tests feasibility of making LCSH available as Linked Data
- 2009 (May) - With Ed's assistance, Clay Redding migrates lcsh.info, ID.LOC.GOV established
Rebecca Guenther designated Project Lead
- 2009 (Dec) – Open Government Directive released
- 2010 – New authorities added, service updates

Metadata Decisions

ID.LOC.GOV uses SKOS

- Simple Knowledge Organization System
- W3C Recommendation (9 August 2009)
- An RDF application specifically created for **publishing and sharing** thesauri, authority, and vocabulary data **for use as Linked Data** and **within the Semantic Web** framework.

Example:

L Rice	→	skos:prefLabel
UF Paddy	→	skos:altLabel
BT Cereals	→	skos:broader
BT Plant products	→	skos:broader
NT Brown rice	→	skos:narrower
RT Rice straw	→	skos:related

Metadata Decisions

Challenges and imperfections using SKOS

SKOS intentionally *Simple*

Our data is more complicated

SKOS is lossy

Augmented with terms from other RDF vocabularies

Now employing SKOSXL (SKOS eXtensions for Labels)

Addressing these issues

Developing a RDF vocabulary better suited to LIS needs

MADS XML → MADS RDF

MADS XML designed for LIS data

Primary motivation: Better support for complex headings

Will be mapped to SKOS to ensure interoperability

Will continue to provide data in SKOS RDF

Data Refreshment

LCSH data at ID.LOC.GOV updated at least monthly

Good:

- LCSH terms updated or added regularly
- Generates return visits

But:

- LCSH data from CDS updated weekly

New vocabularies

Good:

- Successfully added, but inconsistent schedule

But:

- Additions often require new ID.LOC.GOV functionality
- No defined schedule or plan

Tips:

- Automate as much as possible (already do)

- Allocate more staff time (naturally)

Added Value, Services

Offer as many serialization formats as possible

Offered: JSON, N-triples, RDF/XML, RDFa

Good:

- Provides selection to users

- Generally easy to implement

Downside:

- More moving parts

Updates, update services and notification

Offered: ATOM feed, community announcements

Good:

- Keeps service active

- Provides users with automated way to acquire data

- Reason to re-advertise service

Downside:

- More moving parts

Added Value, Services

Downloads

Offered: All data available in N-triples and RDF/XML formats

Good:

Easy developer access to data, especially for local integration

Downside:

More moving parts

Web service:

Offered: Search and discovery, data visualization

Good:

Can be used by non-specialists, non-techies

Downside:

More moving parts

Odds and ends: Type ahead service, XSD snippets

Good:

Widget-like: Type ahead service can be used in web form

Widget-like: XSD snippets can be added to XSDs (coming)

Added Value, Services

But, it is all the **moving parts**
that help to make ID successful.

ID & the Community

Preconditions

LIS world:

- Strong and engaged developer community
- Metadata/cataloging specialists very well represented
- Ed Summers's lcsh.info was an overnight success in the community

Position of LC within LIS community:

- Leader: But, if not a leader, certainly a very influential guide
- “When the LC sneezes, the library world catches a cold.”
- General considerations for any LC (NDMSO) project:
 - Benefit to the LC
 - Impact of decisions/services on the greater LIS community
 - Commitment and sustainability

ID & the Community

Community Engagement

LIS community was already engaged by time ID.LOC.GOV was released

- Fantastic!
- How to leverage this to everyone's advantage?

Established listserv for ID.LOC.GOV

- Point of discussion for those interested in ID.LOC.GOV
- Point of discussion for similar community/Linked Data endeavors
- Publish announcements regarding ID.LOC.GOV

Listserv proven absolutely vital forum for us (NDMSO)

- Service problems/bugs can be reported
- Ideas, service requests can be shared
- Many of these have been or will be integrated
- Service questions can be asked by users
 - And answered (we hope)

ID & the Community

Promotion

Major service announcements

- Published on ID.LOC.GOV listserv
- Other LIS community listservs

Talks/Presentations

Last 3-4 months:

- Semantic Technology Conference, San Francisco, CA
- American Library Association Annual Conference, Washington, DC
 - LC Booth
 - Saco-at-large meeting
 - ALCTS Metadata Interest Group
- Linked Data/Semantic Web Mtg, London

ID in the Library

Feedback/Suggestion form for all LCSH terms

- Suggestions forwarded to LCSH maintainers for consideration
- Wider community input in LCSH development and maintenance
- Fosters collaboration

Internal staff have recognized ID's application and popularity

- Want to add more authority information
- Want to share more information
- Promote international coordination with other national libraries
 - Offering LCSH in non-English languages

ID – Thank You

Thank you.

Questions? Comments?

Kevin Ford
kefo@loc.gov