Linked Data for Libraries (LD4L) and (Metadata) Production (LD4P)

Dean B. Krafft, Tom Cramer, and the LD4L/P Team

CNI Fall Member Meeting

December 12, 2017
Vision: Enable libraries and their users to easily create, use, and benefit from LD specifically designed for libraries and scholars, and from broader sources of LD on the web.
LD4L Labs and LD4P

• Beginning April 1, 2016, the Mellon Foundation funded two new linked data grants for two years at $1.5 million each

• LD4L Labs is a collaboration of Cornell, Harvard, Stanford, and Iowa focused on building new LD tools and services

• LD4P is a collaboration of Stanford and five partners piloting a range of projects on metadata production using linked data
LD4L Labs – Results so far

• Bibliotek-o: a model bibliographic ontology that extends and varies from BIBFRAME
• VitroLib: A prototype ontology-based instance editor for cataloging in BIBFRAME and extension ontologies
• Authority and identity lookup and reconciliation for URI-based authorities
Why bibliotek-o?

• Engage in BIBFRAME evaluation
• Provide an extension to BIBFRAME
• Demonstrate select alternative patterns
• Accommodate legacy data in real-world orientation
• **Not a competitor** to BIBFRAME, but a proposal for alternate modeling that could be folded into BIBFRAME as it evolves
Challenge: migrate existing, highly detailed and nuanced bibliographic metadata AND prepare for a future of original cataloging in RDF that captures data in meaningful and useful ways and that integrates with the broader web
What is bibliotek-o?

- Core concepts and models are BIBFRAME (e.g.: Work, Instance, Item, etc.)
- bibliotek-o ontology: which extends and supplements BIBFRAME, offering alternate models in some areas
- Defined fragments of existing ontologies, both within and outside the bibliographic domain
- An application profile specifying the recommended implementation of these ontologies in modeling data
Principles for bibliotek-o

- **Reuse and align** with existing external vocabularies to promote data exchange and interoperability.
- Conversely, **define terms broadly enough for reuse** by external data sources.
- Use **OWL axioms in moderation** to provide expressivity without overly constraining the ontology and the data it can model.
- Prefer **object properties, structured data, and controlled** over unstructured literals.
- Adopt a **single method of expressing a relationship or attribute** in order to minimize query paths.
Example: Activities in bibliotek-o
SHACL application profiles

• Want ontology driven applications, but the ontology definitions alone are not enough
• Current use case - Editor Form Specifications
  • Eventually we’ll want to address general data validation
• Current Work (in progress)
  • Hip Hop
    (https://github.com/LD4P/HipHop/tree/develop/application-profiles)
  • Moving Image (https://github.com/LD4P/moving-image/tree/develop/application_profiles)
  • Soon we will organize general shapes in the bibliotek-o github repo (e.g. Titles, Measurement, Subjects, for shared across Domains/Applications)
bibliotek-o - Documentation

Home page: http://bibliotek-o.org/ontology

OWL file: http://bibliotek-o.org/ontology.owl

Human-readable documentation: http://bibliotek-o.org/ontology.html

Visualization: https://bibliotek-o.org/overview/overview.html

GitHub repository: https://github.com/ld4l-labs/bibliotek-o/tree/v1.0.1

FAQ, pattern recommendations & RDA discussion:
https://wiki.duraspace.org/x/H5TBB
Cataloging

- Traditional practices: Authority File
  - E.g. Name Authority Files, Subject Headings, Genre Forms from LOC
  - String as unique identifier, e.g. “Mark Twain”

- Tasks and workflows
  - Identification, “Aboutness”
  - Disambiguation
  - Context and original authority record
MARC -> Linked Data

MARC RECORD

NAME AUTH FILE

Made in America
1980
Blues Brothers

BIBFRAME/ BIBLIOTEK-O ENTITIES WITH URIS

WORK

INSTANCE

AGENT/ RWO

Made in America
1980
Blues Brothers
Creating a cataloging environment with effective authority linking

- **VitroLib**
  - Prototype cataloging editor
  - Creates/uses linked data
  - Enables lookup and use of authority URIs

- **Hyrax**
  - Samvera technology stack
  - Incorporate authority URIs into institutional repository records
VitroLib Demonstration

Catalog a new resource

Statistics

80 Works
65 Instances
17 Items
96 People
50 Organizations
191 Agents
New Work

Title *
Made in America

Type *
Audio

Language *
English

Author or Other Role *
Performer

Blues Brothers
- Blues Brothers Band
- Blues Brothers (Motion picture)
- Wood Brothers (Blues group)
- Soul Brothers (Blues group)
- Blues Brothers Horns (Musical group)
- Blues Eldorados
- Blues, Thomas
- Blues Rockers
- **Blues Brothers**
- Blues revue
Audio

Language *

English

Author or Other Role *

Performer

Selected: Blues Brothers (Verify this match or change selection)

LC Subject Heading

soul music

Soul music--History and criticism
Guitar music (Soul)
Philadelphia soul (Music)
Soul--Buddhism
Soul worship
Soul mates
Soul music
Soul--Judaism
Soul musicians
Soul--Hinduism
Soul in literature
Soul--Christianity
Soul--Jainism
Soul in art
### Has Instance (RDA Manifestation)

**Title**
- Made in America

**Statement of Responsibility**
- Blues Brothers

**Type**
- `Audio disc`
- Aperture card
- Archival controled
- Audio belt
- Audio cartridge
- Audio cassette
- Audio cylinder
- Audio disc
- Audio roll
- Audio wire reel
- Audiotape reel
- Belt
- Card
- Cartridge
- Cassette
- Computer card
- Computer chip
- Computer chip cartridge
- Computer disc
- Computer disc cartridge
- Computer tape cartridge

*Select an existing Entity or create a new one.*

**Date of Publication**
Type
Audio disc

Issue Number
SD 16025

Matrix Number

UPC

EAN

Publisher
Atlantic Recording Corp
Atlantic Consultants
Atlantic jazzlore
Atlantic mystery
Atlantic Workshop

* Atlantic series

Atlantic Institute
Atlantic jazz
Atlantic Manor
Atlantic Underwriters
Atlantic romance
Type
Audio disc

Issue Number
SD 16025

Matrix Number

UPC

EAN

Selected: Atlantic Recording Corporation (Verify this match or change selection)

Date of Publication
1980

* required fields

Save or Cancel
### Made in America

**authors and other roles**

- **Blues Brothers** (Performer)

### Instance of Work

- **Made in America (Audio disc)** 1980

### Subject

- **Soul music**

### Genre Form
Manage Associated Genre Forms

Genre Form

| Video recordings | Remove |

Lookup and Add Genre Form

Search

Add Selected Term

Save  Cancel
Manage Associated Genre Forms

Genre Form

Video recordings  Remove

Lookup and Add Genre Form

animation

Add Selected Term
Save  Cancel
### Manage Associated Genre Forms

<table>
<thead>
<tr>
<th>Genre Form</th>
<th>Notes</th>
<th>Additional Info</th>
</tr>
</thead>
<tbody>
<tr>
<td>Video recordings</td>
<td>Remove</td>
<td></td>
</tr>
</tbody>
</table>

**Lookup and Add Genre Form**

Search for: **animation**

<table>
<thead>
<tr>
<th>Genre Form</th>
<th>Notes</th>
<th>Additional Info</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐ Clay animation television programs</td>
<td>No definition provided.</td>
<td>Broader:</td>
</tr>
<tr>
<td>Alternate Labels:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Claymation television programs</td>
<td></td>
<td>• Animated television programs</td>
</tr>
<tr>
<td>• Sculptmation television programs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>☐ Silhouette animation films</td>
<td>No definition provided.</td>
<td>Broader:</td>
</tr>
<tr>
<td>Alternate Labels:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Semi-object animation films</td>
<td></td>
<td>• Animated films</td>
</tr>
<tr>
<td>• Sand animation films</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Flat-figure animation films</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Stop-motion animation films</strong></td>
<td>No definition provided.</td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>------------------------</td>
<td></td>
</tr>
<tr>
<td>Alternate Labels:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Object animation films</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Frame-by-frame animation films</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Stop-action animation films</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Stop-motion animated films</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Clay animation films</strong></th>
<th>No definition provided.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alternate Labels:</td>
<td></td>
</tr>
<tr>
<td>• Claymation films</td>
<td></td>
</tr>
<tr>
<td>• Sculptmation films</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Computer animation television programs</strong></th>
<th>No definition provided.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alternate Labels:</td>
<td></td>
</tr>
<tr>
<td>• CG television programs</td>
<td></td>
</tr>
<tr>
<td>(Computer animation)</td>
<td></td>
</tr>
<tr>
<td>• Computer animated television programs</td>
<td></td>
</tr>
<tr>
<td>• CGI television programs</td>
<td></td>
</tr>
<tr>
<td>(Computer animation)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Live-action/animation films</strong></th>
<th>This heading is used as a genre/form heading for films that feature interaction between live-action and animated elements.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alternate Labels:</td>
<td></td>
</tr>
<tr>
<td>• Animated/live-action films, Hybrid media</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Narrower:</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Clay animation films</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Broader:</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Stop-motion animation films</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Broader:</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Animated television programs</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Broader:</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Animated films</td>
<td></td>
</tr>
</tbody>
</table>
## Manage Associated Genre Forms

### Genre Form

<table>
<thead>
<tr>
<th>Genre Form</th>
<th>Notes</th>
<th>Additional Info</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clay animation television programs</strong></td>
<td>No definition provided.</td>
<td><strong>Broader:</strong></td>
</tr>
<tr>
<td><strong>Claymation television programs</strong></td>
<td></td>
<td>• <a href="#">Animated television programs</a></td>
</tr>
<tr>
<td><strong>Sculptimation television programs</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Silhouette animation films</strong></td>
<td>No definition provided.</td>
<td><strong>Broader:</strong></td>
</tr>
<tr>
<td><strong>Semi-object animation films</strong></td>
<td></td>
<td>• <a href="#">Animated films</a></td>
</tr>
<tr>
<td><strong>Sand animation films</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Flat-figure animation films</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Lookup and Add Genre Form

Search for: animation [Search Button]
Clay animation television programs

From Library of Congress Genre/Form Terms

Clay animation television programs

URL(s)
- http://id.loc.gov/authorities/genreForms/qf2011028141

Instance Of
- MADS/RDF GenreForm
- MADS/RDF Authority
- SKOS Concept

Scheme Membership(s)
- Library of Congress Genre/Form Terms

Collection Membership(s)
- http://id.loc.gov/authorities/genreForms/collection_LCGFT_General

Form
- Clay animation television programs

Variants
- Claymation television programs
- Sculptmation television programs

Use For
- sh2008025507

Broader Terms
- Animated television programs

Sources
- found: Moving image genre-form guide online. January 14, 2008 (Clay animation: type of animation which generally uses plasticine or clay, such as the "Claymation" work of Will Vinton or the Wallace and Gromit series).
What just happened?

Query = animation
Translate to QA Service Request
Search LOC Genre Form data

VitroLib Search Service

Questioning Authority

LOC Genre Forms

URI: http://id.loc.gov/gf2011026141,
label: "Clay animation television programs",
altLabelList: [
  "Claymation television programs",
  "Sculptmation television programs"
].

URI: http://id.loc.gov/gf2011026141,
label: "Clay animation television programs",
context: {
  "Alternate Label": [
    "Claymation television programs",
    "Sculptmation television programs"
  ].

Demonstration: Hyrax, Questioning Authority Gem, QA Cache Server, LD Authorities (LoC, OCLC FAST, GeoNames, AGROVOC, NALT, and DBpedia)
Selected String and URI

Saves both string and URI

OCLC organization

- Cornell University

+ Add another OCLC organization

OCLC person

- Cornell, Ezra, 1807-1874
- http://id.worldcat.org/fast/409667

+ Add another OCLC person

Agrovoc keyword

- hive products
- http://aims.fao.org/aos/agrovoc/c_3655

+ Add another Agrovoc keyword

Agrovoc keyword fr
Selecting a Term using Lookup with Context

<table>
<thead>
<tr>
<th>Authoritative Label</th>
<th>Variant Labels</th>
<th>Field of Activity</th>
<th>Occupation</th>
<th>Dates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Twain, Shania</td>
<td>Twain, Eileen</td>
<td>Country Music</td>
<td>Singer</td>
<td>Birth: 19650828</td>
</tr>
<tr>
<td></td>
<td>Edwards, Eileen</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Twain, Mark, 1835-1910</td>
<td>Literature</td>
<td>Authors</td>
<td>Burt: (edt) 1835-11-30, Death: (edt) 1910-04-21</td>
</tr>
<tr>
<td></td>
<td>Tuవین, Mark, 1835-1910</td>
<td>Humor</td>
<td>Lecturers</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Twaiyn, Mǝk, 1835-1910</td>
<td>Wit and humor</td>
<td>Humorists</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T’u-wεn, Mǝk-k’a, 1835-1910</td>
<td>more...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Twain, Mark, 1835-1910</td>
<td>Writing</td>
<td>Internet Industry</td>
<td>Author</td>
<td></td>
</tr>
<tr>
<td>Twain, Charles</td>
<td>Editing</td>
<td></td>
<td>Editor</td>
<td></td>
</tr>
<tr>
<td>Twain, Jane</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Luu, Twain</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Selecting a Term using Lookup with Context
Getting more from the same authority?

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>sameas:</td>
<td><a href="http://d-nb.info/gnd/4006529-7">http://d-nb.info/gnd/4006529-7</a></td>
</tr>
<tr>
<td>sameas:</td>
<td><a href="http://cat.aii.caas.cn/concept/c_12939">http://cat.aii.caas.cn/concept/c_12939</a></td>
</tr>
</tbody>
</table>
Getting more from other authorities?

**Oclc person** Cornell, Ezra, 1807-1874

(source: [http://dbpedia.org/resource/Ezra_Cornell](http://dbpedia.org/resource/Ezra_Cornell))

Birth: 1807-01-11

Death: 1874-12-09

Ezra Cornell (January 11, 1807 – December 9, 1874) was an American businessman, politician, philanthropist and educational administrator. He was the founder of Western Union and a co-founder of Cornell University. He also served as President of the New York Agriculture Society and as a New York state Senator.

**Oclc person** [http://id.worldcat.org/fast/409667](http://id.worldcat.org/fast/409667)
Links to Code & More

- **QA Server** - Code for a small app that provides the Questioning Authority normalization layer ([https://github.com/ld4l-labs/qa_server](https://github.com/ld4l-labs/qa_server))
- **Linked Data Authorities** - Configurations that can be used with QA Server ([https://github.com/ld4l-labs/linked_data_authorities](https://github.com/ld4l-labs/linked_data_authorities))
- **LD4L Services** - UI access to Cache Server ([http://services.ld4l.org/ld4l_services/](http://services.ld4l.org/ld4l_services/))
LD4P

- LD4P = Linked Data for (Metadata) Production
- Question: How do we move from converting description to linked data to producing linked data?
  - What descriptive standards are needed?
  - What tools are needed?
  - How must cataloging workflows change?
  - How does it all work together?
LD4P Objectives

• Enable libraries to work in an open, networked environment in the construction of their metadata
• Season & extend the BIBFRAME ontology
• Spur tool development
• Engage with other strategic linked-data projects
• Engender community engagement among libraries, standards groups, software devs, commercial players
LD4P Partners

Stanford

Columbia

Cornell

Princeton

Library of Congress

Harvard
Princeton: Jacques Derrida’s personal library
Models

- Stanford
- Columbia
- Cornell
- Princeton
- Harvard
- Library of Congress

Models:
- ArtFrame
- Annotations
- BIBFRAME 2
- PMO: Performed Music Ontology
- RareMats
- Bibliotek-o
- Cartographic
- Moving Image
Tools

Stanford
Cornell
Harvard
Princeton
Columbia

CEDAR
VitroLib

LD4P

Plus...
Aliada
Kafka-Spark
Karma
Fedora 4
RML Editor
SHARE-VDE
Triple stores (various)

BIBFRAME 2 Editor and Converter
# Tool Registry

<table>
<thead>
<tr>
<th>Tool</th>
<th>Type (Whatever makes sense to you as a tool type)</th>
<th>Function</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALIADA</td>
<td>Converter, publisher, linker</td>
<td>Convert MARC to customizable RDF, save to store and link with other datasets</td>
<td>Java</td>
</tr>
<tr>
<td>Annif</td>
<td></td>
<td></td>
<td>English</td>
</tr>
<tr>
<td>Arc2 PHP Sparql and Triplestore</td>
<td></td>
<td>SPARQL endpoint and triplestore</td>
<td>PHP</td>
</tr>
<tr>
<td>bib-rdf-pipeline</td>
<td>Conversion</td>
<td>Framework for converting MARC via BIBFRAME to Schema.org, with reconciliation, work clustering etc.</td>
<td>Mixed (Makefile, marc2bibframe2, Python, shell scripts, Unix tools, Jena, SPARQL...)</td>
</tr>
<tr>
<td>Blazegraph</td>
<td>Triplestore</td>
<td>Storage/query</td>
<td>Java</td>
</tr>
<tr>
<td>CEDAR</td>
<td>Editor</td>
<td>Produce/edit RDF based on ontology-driven forms</td>
<td>Java, JS</td>
</tr>
<tr>
<td>Controlled Vocabulary Manager</td>
<td>Vocabulary Manager</td>
<td></td>
<td>Ruby/Rails</td>
</tr>
<tr>
<td>CTSASearch</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>data.world</td>
<td>Publishing</td>
<td>Collaborative platform for finding and sharing data that uses the Semantic Web as its framework</td>
<td></td>
</tr>
<tr>
<td>Fuseki</td>
<td>Triplestore</td>
<td>Storage/query</td>
<td>Java</td>
</tr>
<tr>
<td>KARMA</td>
<td>Multi-format Converter</td>
<td>Enables users to quickly and easily integrate data from a variety of data sources including databases, spreadsheets, delimited text files, XML, JSON, KML and Web APIs.</td>
<td>Java</td>
</tr>
</tbody>
</table>

4 “Tracer Bullets” for descriptive workflows
1. Acquired (MARC) data
2. Original (MARC) cataloging
3. Singleton items in digital repository
4. Bulk digital collections
LD4 Community Meeting
2nd international meeting, 24-25 Apr 2017

• Ontologies: We live in a multi-ontology world; give up on building the perfect or universal ontology.

• Workflows: 1.) MARC and RDF will co-exist for a loooong time, and 2.) It is not yet clear how we will use each other’s data

• Tools: We need more tools, smaller tools, and more dev.

• Community: We have critical interest (and mass?) in LOD, but not yet the organizing structures, interoperability or opportunities for concrete contributions.

http://meeting.ld4p.org
Biblioportal

For ontology...

• Discovery
• Visualization
• Maintenance
• Mappings
• Assessment
• Re-use

http://biblio.ontoportal.org/
What Have We Learned?

• Ontology work is time consuming
  – The process as important as the product
• More work is needed on tools
• Creating URIs is easy; linking data is harder
• Growing activity in this space
  – European adoption
  – Local authority stores
Looking Ahead

• More automation & integration
• Broad training in a cloud-based sandbox
• Focus on discovery
• Community development
More Info:

http://ld4l.org
http://ld4p.org
http://meeting.ld4p.org