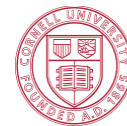


Linked Data for Libraries (LD4L) and (Metadata) Production (LD4P)

Dean B. Krafft, Tom Cramer,
and the LD4L/P Team
CNI Fall Member Meeting
December 12, 2017



Cornell University
Library



HARVARD
LIBRARY



STANFORD UNIVERSITY LIBRARIES

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

bibliotek-o Application Profile Overview

<http://id.loc.gov/ontologies/bibframe/> (bf)

<http://bibliotek-o.org/ontology/> (bib)

<http://purl.org/dc/terms/> (dcterms)

<https://www.loc.gov/standards/datetime/> (EDTF)

<http://xmlns.com/foaf/0.1/> (foaf)

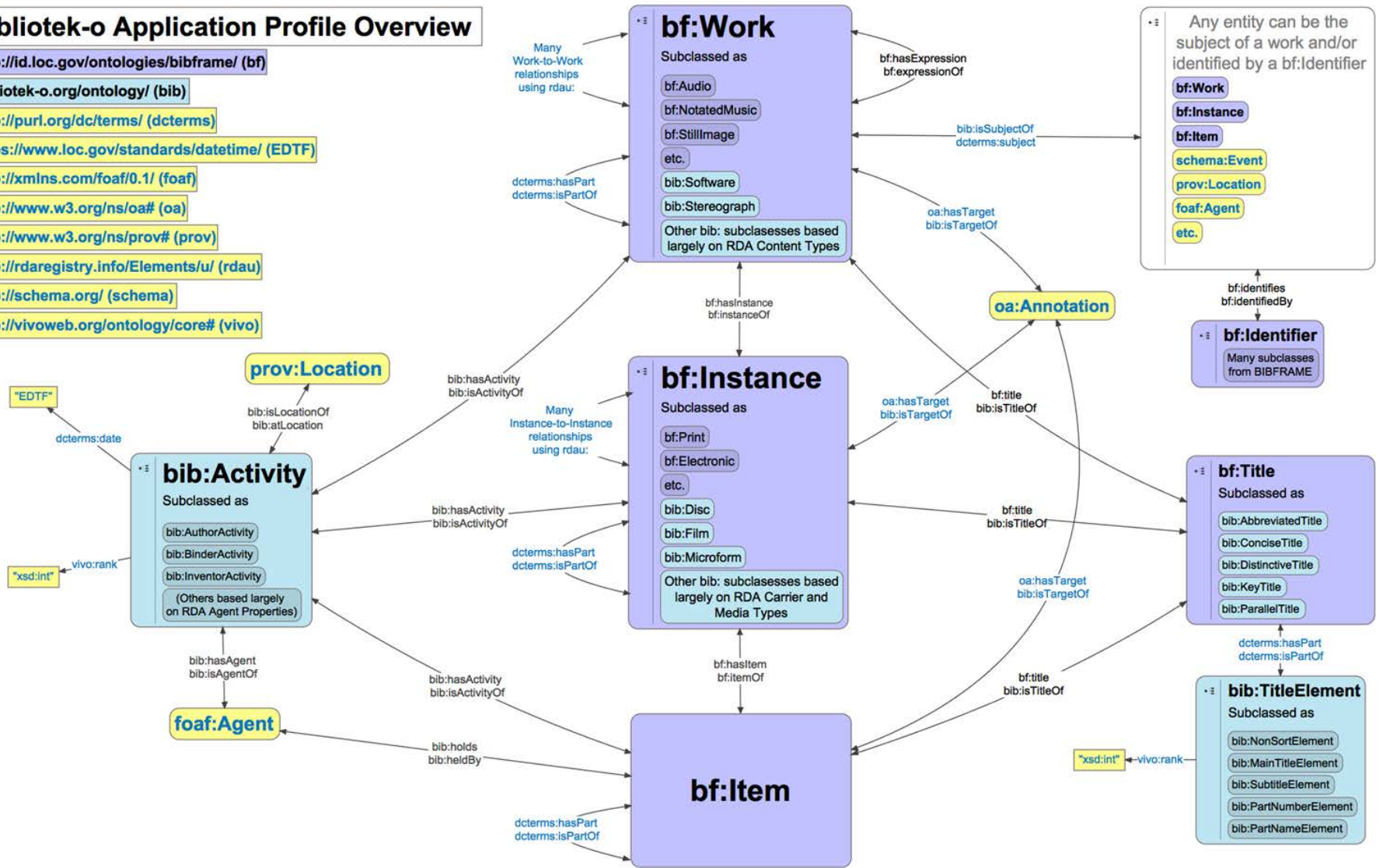
<http://www.w3.org/ns/oa#> (oa)

<http://www.w3.org/ns/prov#> (prov)

<http://rdaregistry.info/Elements/u/> (rdau)

<http://schema.org/> (schema)

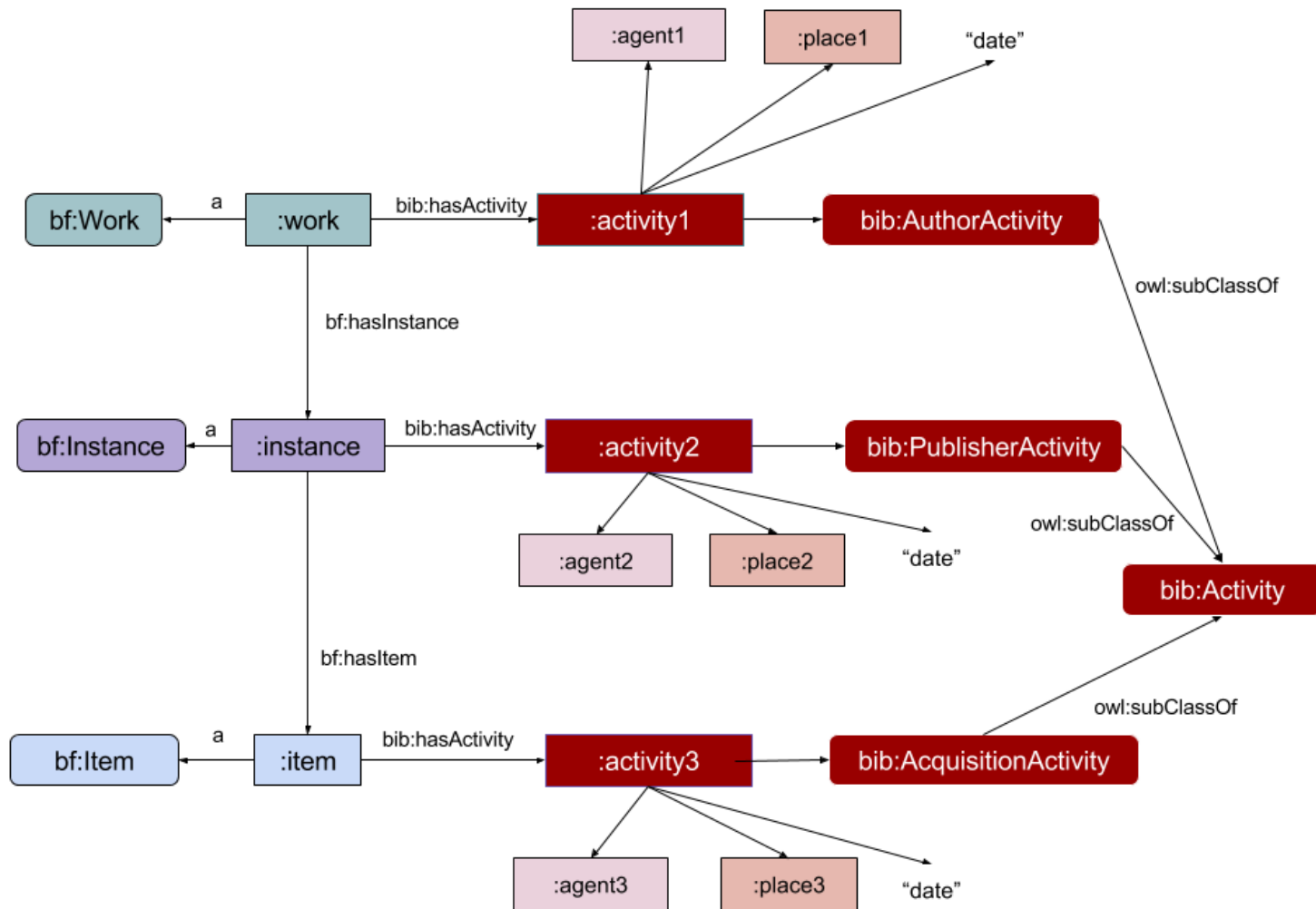
<http://vivoweb.org/ontology/core#> (vivo)



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:

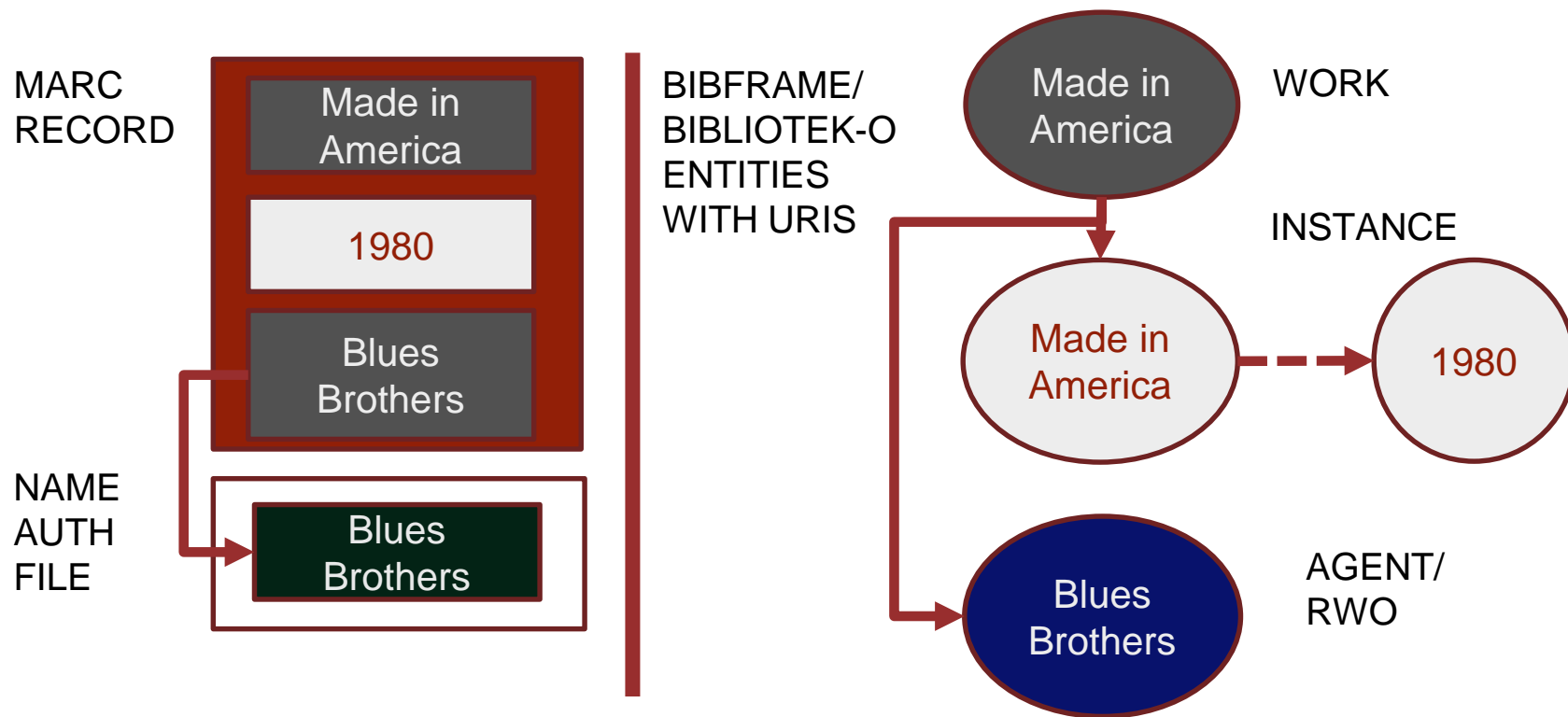
<https://wiki.duraspace.org/x/H5TBB>



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





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



Linked Data for Libraries

[Index](#) | [Site Admin](#) | [root](#) ▼

[Home](#) | [Works](#) | [Instances](#) | [Items](#) | [People](#) | [Organizations](#)

Catalog a new resource [Ⓞ]

Work +

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



Selected Subject: Soul music [\(Verify this match\)](#) or [change selection](#)

Has Instance (RDA Manifestation)

Title

Made in America

Statement of Responsibility

Blues Brothers

Type

Audio disc

- Aperture card
- Archival controlled
- 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 co

Atlantic Recording Corporation

Atlantic Consultants

Atlantic jazzlore

Atlantic mystery

Atlantic Workshop

Sa

Atlantic series

* req

Atlantic Institute

Atlantic jazz

Atlantic Manor

Atlantic Underwriters

Atlantic Underwriters

Atlantic romance

©201



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



or [Cancel](#)

** required fields*

Resource URI: <http://vitrolib.cornell.edu/individual/n2726>

[Made in America](#) |   | Audio | (work) 

authors and other roles 

[Blues Brothers](#) (Performer) |  

Instance of Work 

[Made in America \(Audio disc\)](#) 1980 |  

subject 

[Soul music](#) |  

genre form 



Linked Data for Libraries

Manage Associated Genre Forms

Genre Form

[Video recordings](#)

[Remove](#)

Lookup and Add Genre Form

[Cancel](#)

 [Search](#)

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

Genre Form	Notes	Additional Info
<input type="checkbox"/> Clay animation television programs Alternate Labels: <ul style="list-style-type: none">• Claymation television programs• Sculptmation television programs	No definition provided.	Broader: <ul style="list-style-type: none">• Animated television programs
<input type="checkbox"/> Silhouette animation films Alternate Labels: <ul style="list-style-type: none">• Semi-object animation films• Sand animation films• Flat-figure animation films	No definition provided.	Broader: <ul style="list-style-type: none">• Animated films



<input type="checkbox"/> Stop-motion animation films	No definition provided.	Narrower: <ul style="list-style-type: none">• Clay animation films
Alternate Labels: <ul style="list-style-type: none">• Object animation films• Frame-by-frame animation films• Stop-action animation films• Stop-motion animated films		
<input type="checkbox"/> Clay animation films	No definition provided.	Broader: <ul style="list-style-type: none">• Stop-motion animation films
Alternate Labels: <ul style="list-style-type: none">• Claymation films• Sculptmation films		
<input type="checkbox"/> Computer animation television programs	No definition provided.	Broader: <ul style="list-style-type: none">• Animated television programs
Alternate Labels: <ul style="list-style-type: none">• CG television programs (Computer animation)• Computer animated television programs• CGI television programs (Computer animation)		
<input type="checkbox"/> Live-action/animation films	This heading is used as a genre/form heading for films that feature interaction between live-action and animated elements.	Broader: <ul style="list-style-type: none">• Animated films
Alternate Labels: <ul style="list-style-type: none">• Animated/live-action films, Hybrid		



Manage Associated Genre Forms

Genre Form
Video recordings Remove

Lookup and Add Genre Form

Genre Form	Notes	Additional Info
<input type="checkbox"/> Clay animation television programs Alternate Labels: <ul style="list-style-type: none">• Claymation television programs• Sculptmation television programs	No definition provided.	Broader: <ul style="list-style-type: none">• Animated television programs
<input type="checkbox"/> Silhouette animation films Alternate Labels: <ul style="list-style-type: none">• Semi-object animation films• Sand animation films• Flat-figure animation films	No definition provided.	Broader: <ul style="list-style-type: none">• Animated films

Clay animation television programs

From [Library of Congress Genre/Form Terms](#)

Details


Visualization

Clay animation television programs

URI(s)

> <http://id.loc.gov/authorities/genreForms/qf2011026141>

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](#)

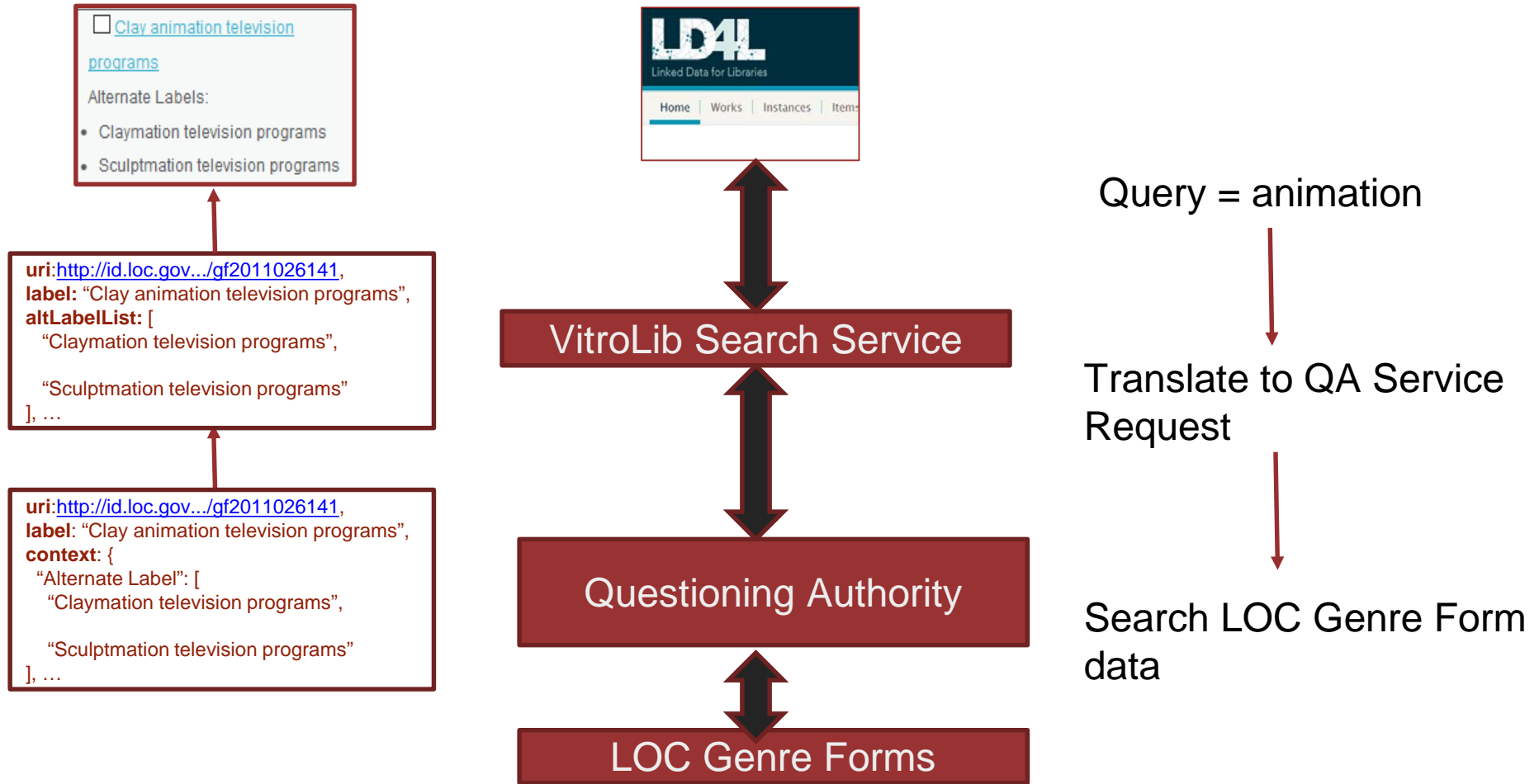
Broader Terms

>  [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?





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

Autocomplete Saving String and URI

Authority: OCLC FAST

Subauthority: PersonName

Oclc person **required**

Benson, Ezra Taft

Cornell, Ezra, 1807-1874

Ezra (Biblical figure)

Gannett, Ezra S. (Ezra Stiles), 1801-1871

Ibn Ezra, Abraham ben Meir, 1089-1164

Ibn Ezra, Moses, approximately 1060-approximately 1139

Meeker, Ezra, 1830-1928

Pound, Ezra, 1885-1972

Sims, Ezra, 1928-2015

Stiles, Ezra, 1727-1795

out [SHERPA/RoMEO](#) for specific publishers' copyright policies if you plan to patent and/or publish your Demo Work in a journal.

- Institution** Restrict access to only users and/or groups from Institution
- Embargo**
- Lease**
- Private** Only users and/or groups that have been given specific access in the "Share With" section.

I have read and agree to the **Deposit Agreement**

Selected String and URI

Saves both string and URI

Oclc organization **required**

Cornell University

+ [Add another Oclc organization](#)

Oclc person **required**

Cornell, Ezra, 1807-1874

<http://id.worldcat.org/fast/409667>

+ [Add another Oclc person](#)

Agrovoc keyword **required**

hive products

http://aims.fao.org/aos/agrovoc/c_3655

+ [Add another Agrovoc keyword](#)

Agrovoc keyword fr **required**

Selecting a Term using Lookup with Context

Library of Congress Lookup

http://localhost:3000/qa/lookup/linked_data/loc/personal_name

3 Cancel

2 Authority: Library of Congress
OCLC FAST
Cornell Local

1 twain

7 Not found?

Lookup

Add to Local

Click one to select...

Authoritative Label	Variant Labels	Field of Activity	Occupation	Dates
Twain, Shania	Twain, Eilleen Edwards, Eilleen	Country Music	Singer	Birth: 19650828
4 Twain, Mark, 1835-1910 5	Tvèn, Mark, 1835-1910 TuèTn, Mark, 1835-1910 Tuwayn, Mårk, 1835-1910 Twayn, Mårk, 1835-1910 T'u-wen Ma-k'ò, 1835-1910 more... 6	Literature Humor Wit and humor	Authors Lecturers Humorists	Birth: (edtf) 1835-11-30 Death: (edtf) 1910-04-21
Twain, Charles		Writing Editing	Author Editor	
Twain, Jane				
Luu, Twain		Internet Industry	Business Woman	

Selecting a Term using Lookup with Context

Agrovoc Lookup

http://localhost:3000/qa/lookup/linked_data/agrovoc

Authority:


7 Not found?

Click one or more checkboxes to select...

Broader	Found	Narrower	Same As
5 <input type="checkbox"/> animal products	<input type="checkbox"/> milk 6	<input type="checkbox"/> buffalo milk <input type="checkbox"/> camel milk <input type="checkbox"/> colostrum <input type="checkbox"/> cow milk <input type="checkbox"/> ewe milk <input type="checkbox"/> goat milk	http://aims.fao.org/aos/asfa/c_7251 http://cat.aii.caas.cn/concept/c_32204 http://cat.aii.caas.cn/concept/c_33165 http://d-nb.info/gnd/4039264-8 http://eurovoc.europa.eu/1565 http://linkeddata.ge.imati.cnrit.2020/resource/EARTH/55910 http://lod.nal.usda.gov/nalt/631 http://www.eionet.europa.eu/gemet/concept/5254 http://zbweu.stw/descriptor/14124-5
<input type="checkbox"/> by-products	<input type="checkbox"/> milk by-products	<input checked="" type="checkbox"/> milk products	
<input type="checkbox"/> collection	<input type="checkbox"/> milk collection		
<input type="checkbox"/> containers	<input type="checkbox"/> milk containers		
<input checked="" type="checkbox"/> animal fats	<input type="checkbox"/> milk fats		http://cat.aii.caas.cn/concept/c_38198 http://d-nb.info/gnd/4169913-0 http://eurovoc.europa.eu/1836 http://lod.nal.usda.gov/nalt/9160 http://stitch.cs.vu.nl/vocabularies/rameau/ark:/12148/cb11980632s http://cat.aii.caas.cn/concept/c_33063 http://cat.aii.caas.cn/concept/c_58677 http://lod.nal.usda.gov/nalt/13041
<input type="checkbox"/> proteases	<input checked="" type="checkbox"/> chymosin		

Getting more from the same authority?

Agrovoc Keyword


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

narrower: propolis  http://aims.fao.org/aos/agrovoc/c_15919

narrower: royal jelly  http://aims.fao.org/aos/agrovoc/c_26817

narrower: honey  http://aims.fao.org/aos/agrovoc/c_3652

narrower: beeswax  http://aims.fao.org/aos/agrovoc/c_866

narrower: honeycomb extracts  http://aims.fao.org/aos/agrovoc/c_29026

broader: animal products  http://aims.fao.org/aos/agrovoc/c_438

sameas:  <http://d-nb.info/gnd/4006529-7>

sameas:  http://cat.aii.caas.cn/concept/c_12939

Getting more from other authorities?

Oclc person [Cornell, Ezra, 1807-1874](#)

(source: 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>

uri

Links to Code & More

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

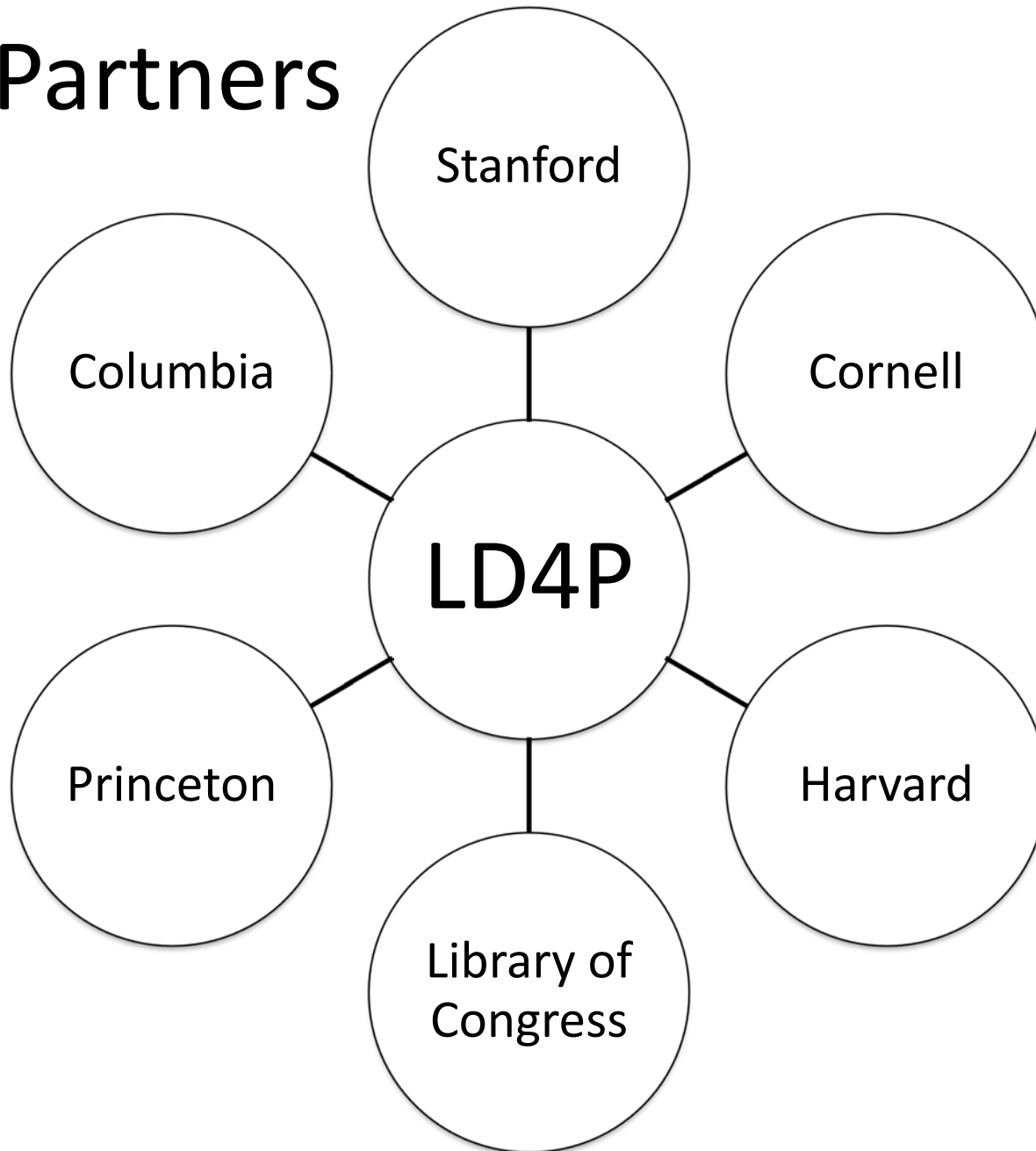
LD4P

- 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?

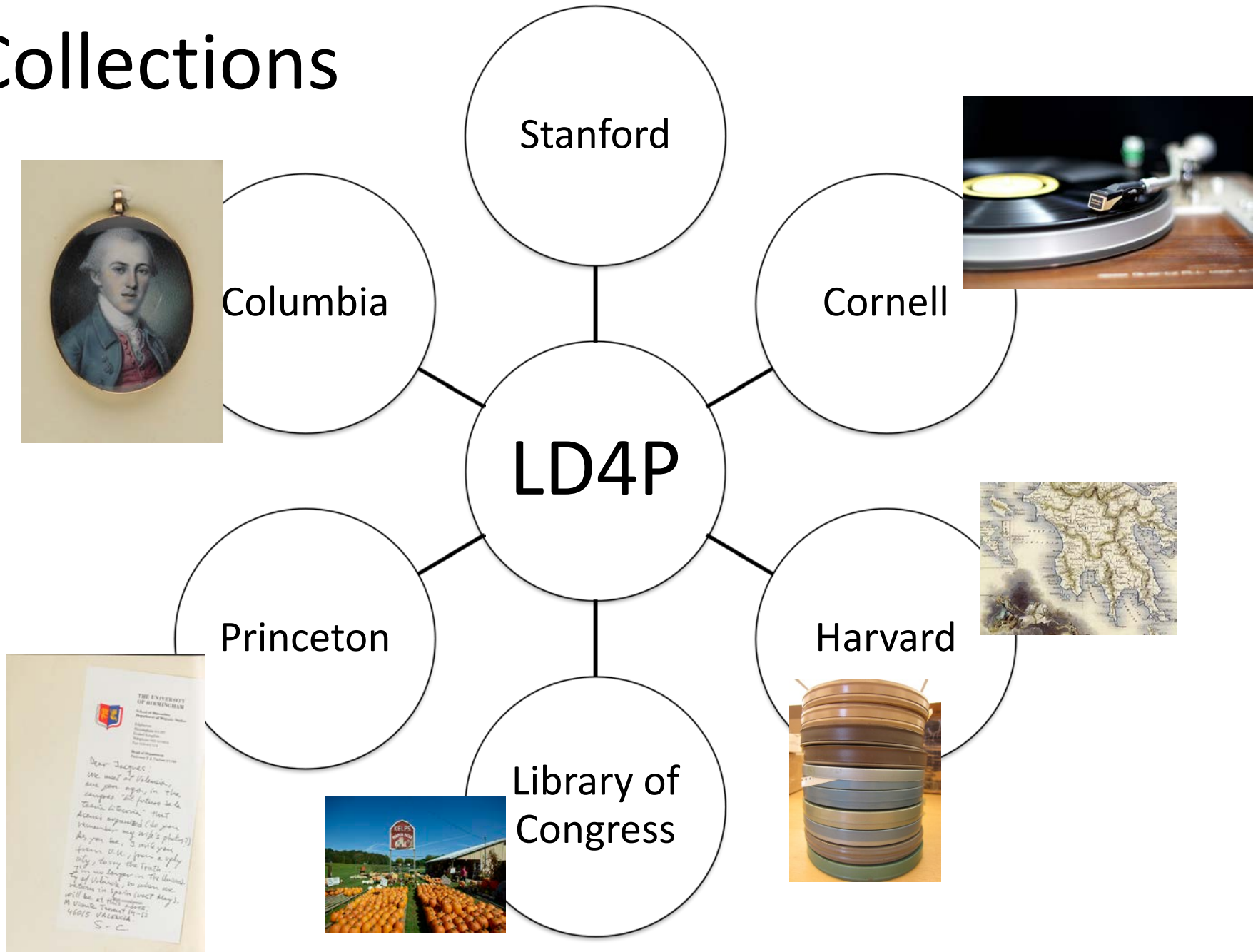
LD4P Objectives

- 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

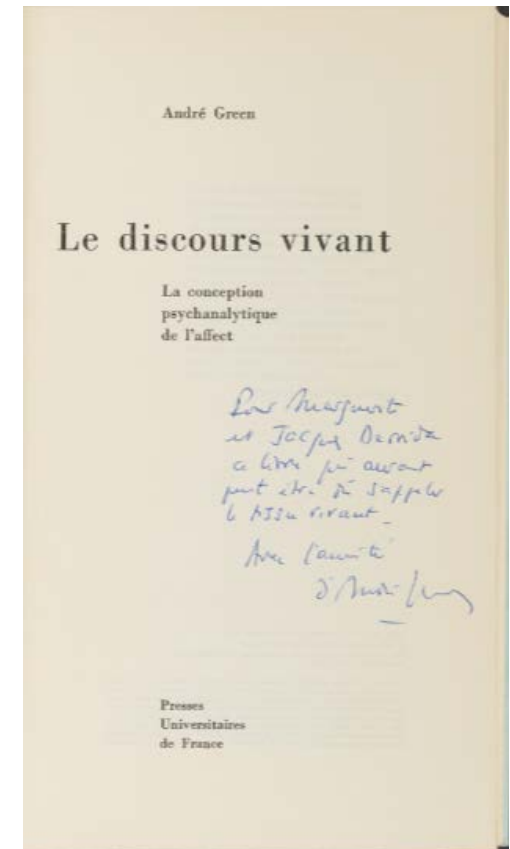
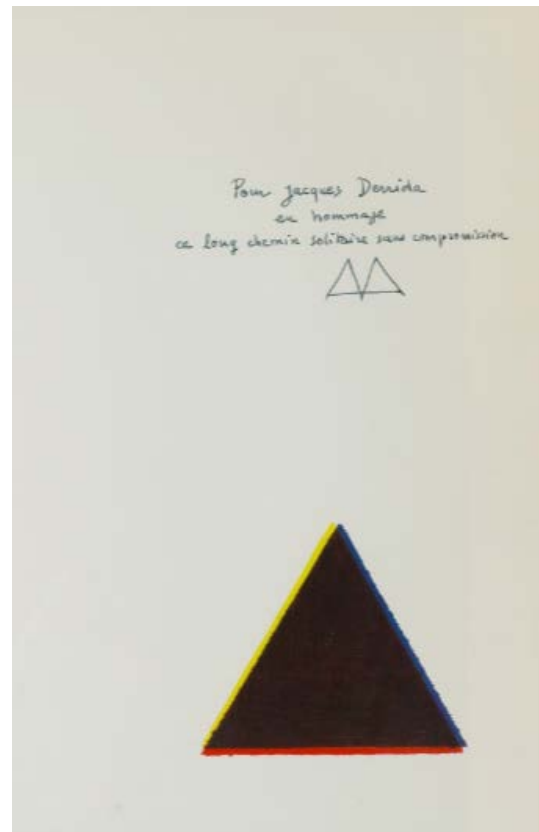
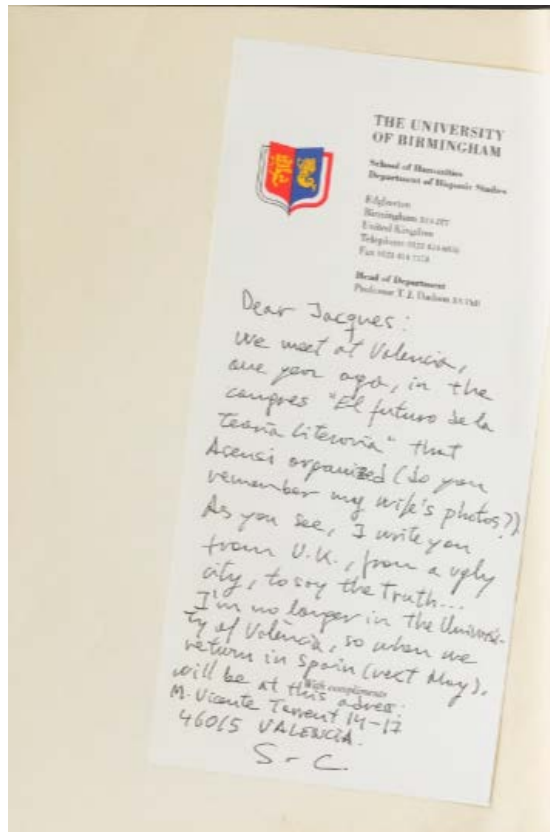
LD4P Partners



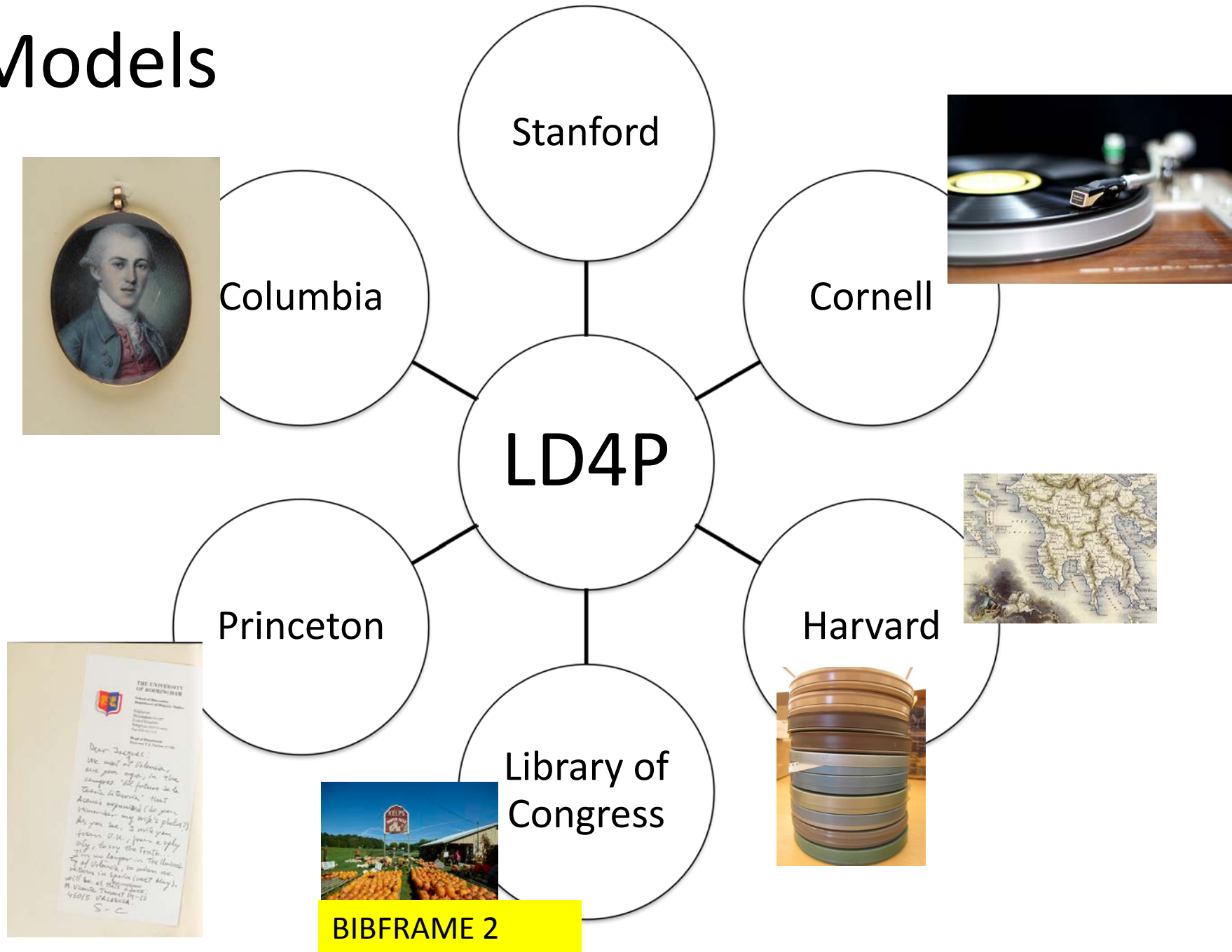
Collections



Princeton: Jacques Derrida's personal library



Models



BIBFRAME 2



Models

PMO: Performed Music Ontology

Stanford



RareMats

Cornell

Bibliotek-o

LD4P

Columbia



ArtFrame

Princeton



Cartographic

Harvard

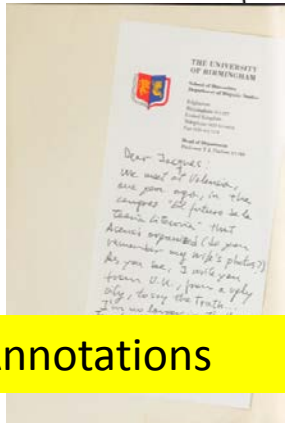


Moving Image

Library of Congress

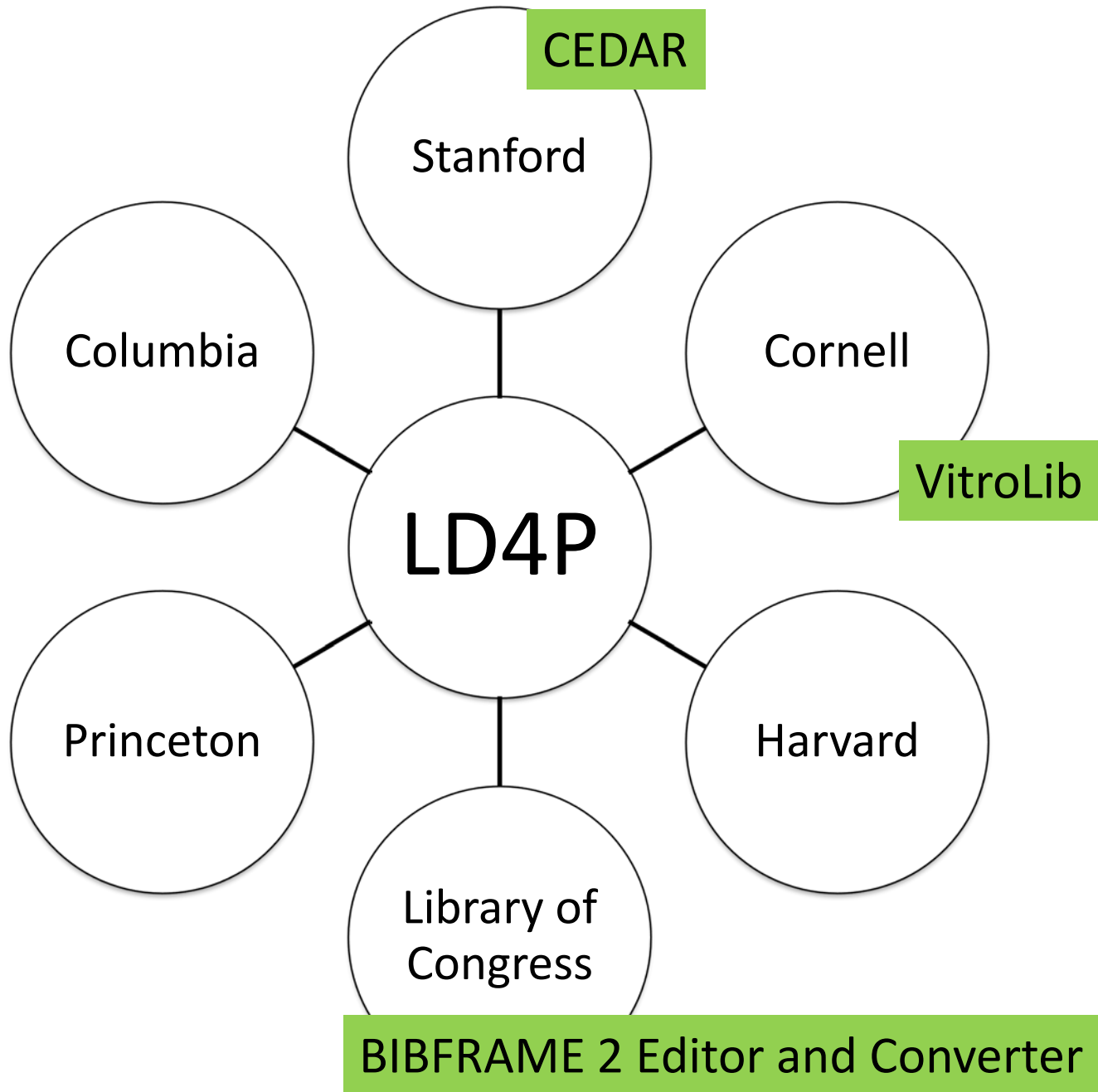


BIBFRAME 2



Annotations

Tools



Plus...

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

BIBFRAME 2 Editor and Converter

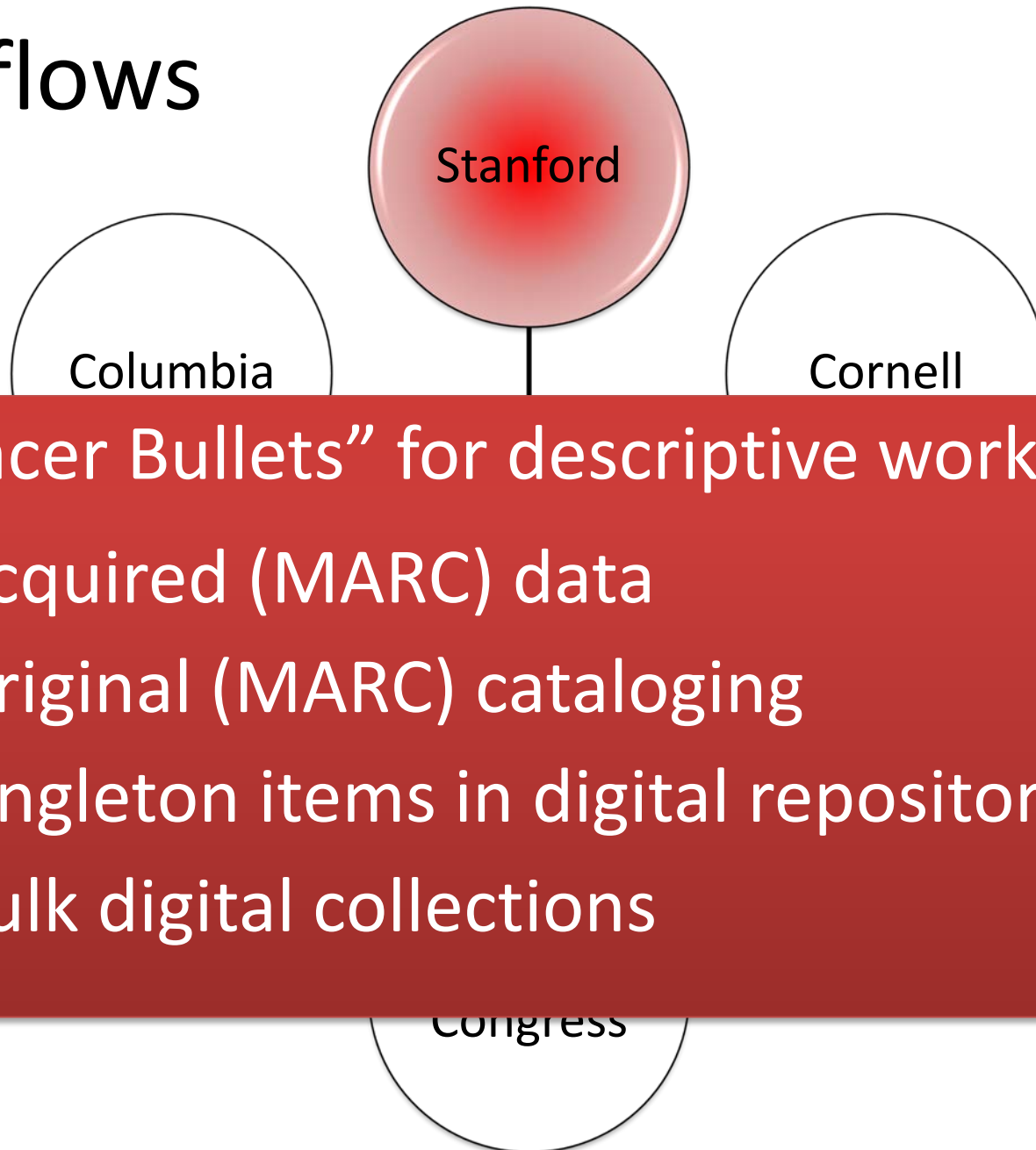
Tool Registry

	A	B	C	D
1	This is a shared, open sheet of tools for working with Linked Open Data in GLAM - Galleries, Libraries, Archives, Museums, and			
2	Put whatever tools you find helpful for the above. Link to this open spreadsheet (feel free to share out wherever): http://bit.ly/GL			
3	These will be collated/tidied up/moved to https://wiki.duraspace.org/display/LD4P/Registry+of+Tools - Open to anyone to add (fi			
4	Tool	Type (Whatever makes sense to you as a tool type)	Function	Language
5	ALIADA	Converter, publisher, linker	Convert MARC to customizable RDF, save to store and link with other datasets	Java
6	Annif			English
7	Arc2 PHP Sparql and Triplestore		SPARQL endpoint and triplestore	PHP
8	bib-rdf-pipeline	Conversion	Framework for converting MARC via BIBFRAME to Schema.org, with reconciliation, work clustering etc.	Mixed (Makefile, marc2bibframe2, Python, she scripts, Unix tools, Jena, SPARQL...)
9	Blazegraph	Triplestore	Storage/query	Java
10	CEDAR	Editor	Produce/edit RDF based on ontology-driven forms	Java, JS
11	Controlled Vocabulary Manager	Vocabulary Manager		Ruby/Rails
12	CTSASearch			
13	data.world	Publishing	Collaborative platform for finding and sharing data that uses the Semantic Web as its framework	
14	Fuseki	Triplestore	Storage/query	Java
15	KARMA	Multi-format Converter	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.	Java

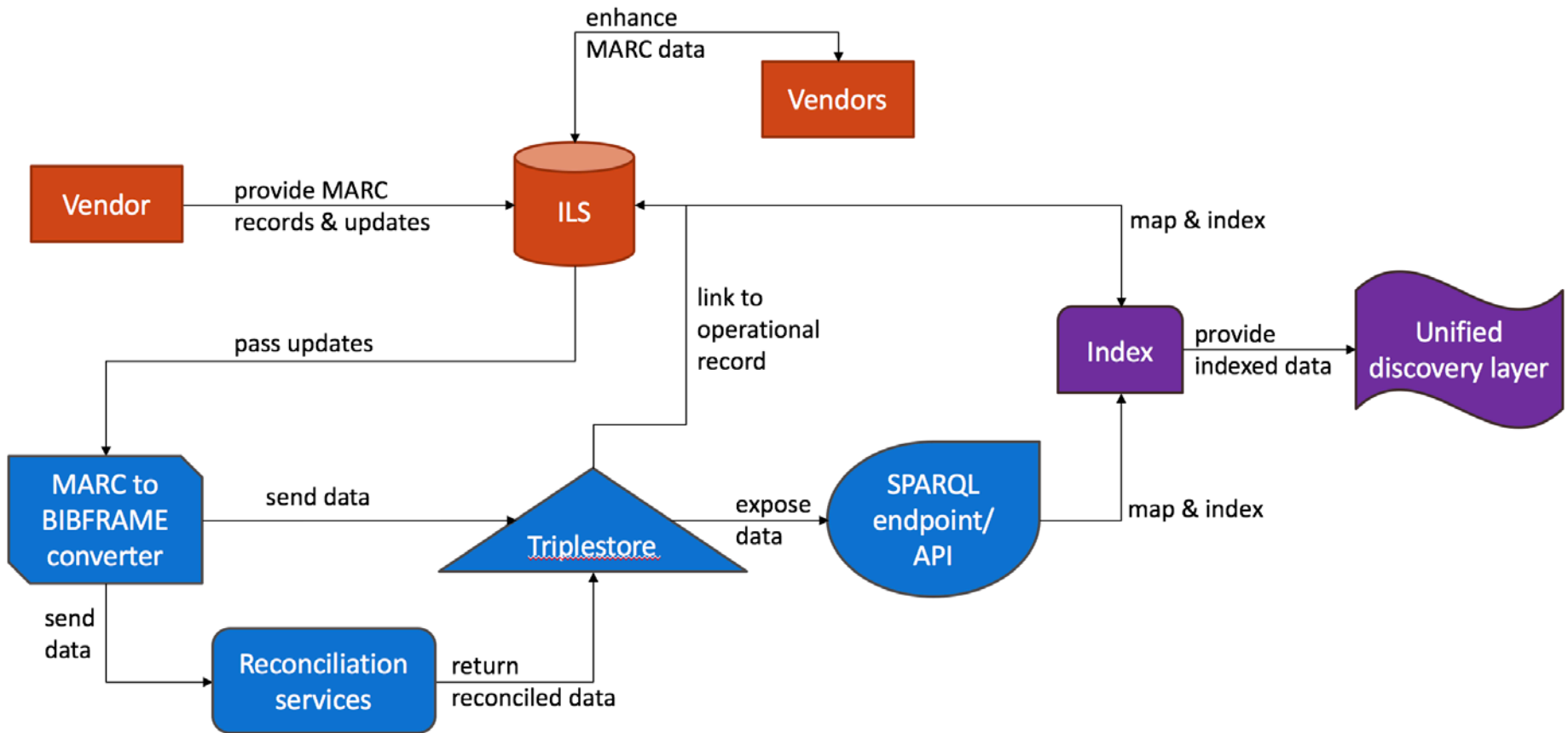
<http://bit.ly/GLAMLodToolsRegistry>



Workflows



Tracer Bullet 1



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.

Biblioport

For ontology...

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

The screenshot shows the Biblioport website interface. The browser address bar displays "biblio.ontportal.org". The page features a red header with the Biblioport logo and navigation links for "Login", "Tools", and "Support". The main content area is titled "Welcome to Biblioport" and includes several functional sections:

- Search for a class:** A search box with the placeholder text "Enter a class, e.g. Work" and a search button. Below it is a link for "Advanced Search".
- Find an ontology:** A search box with the placeholder text "Start entering ontology name, e.g. BIBFRAME, then choose from list" and a search button. Below it is a link for "Browse Ontologies".
- Ontology Visits (November 2017):** A bar chart showing visits for various ontologies. The x-axis ranges from 0 to 5,000. The y-axis lists ontologies: OA, PMO, BIBFRAME, BIBLIOTEK-O, and SCHEMA. OA has the highest number of visits, followed by PMO, BIBFRAME, BIBLIOTEK-O, and SCHEMA. A "More" link is provided at the bottom left of the chart.
- Biblioport Statistics:** A table showing the total number of ontologies and classes.

Category	Count
Ontologies	15
Classes	1,934

At the bottom of the page, the footer text reads: "OntoPortal Appliance 2.5 RC3 : Powered by NCBO BioPortal : Projects".

<http://biblio.ontportal.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>