The Linked Data for Libraries (LD4L) Project: A Progress Report

Dean B. Krafft, Cornell University Library

Tom Cramer, Stanford University Libraries

CNI Fall Meeting, Dec. 8-9, 2014

Cornell University Library

HARVARD LIBRARY

STANFORD UNIVERSITY LIBRARIES

Linked Data for Libraries (LD4L)

- On December 5, 2013, the Andrew W. Mellon Foundation made a two-year $999K grant to Cornell, Harvard, and Stanford starting Jan ‘14
- Partners will work together to develop an ontology and linked data sources that provide relationships, metadata, and broad context for Scholarly Information Resources
- Leverages existing work by both the VIVO project and the Hydra Partnership

Vision: Create a LOD standard to exchange all that libraries know about their resources

Courtesy National Gallery of Art, Washington

“The goal is to create a Scholarly Resource Semantic Information Store model that works both within individual institutions and through a coordinated, extensible network of Linked Open Data to capture the intellectual value that librarians and other domain experts add to information resources when they describe, annotate, organize, select, and use those resources, together with the social value evident from patterns of usage.”

Collections

Andrew D. White Architectural Photographs Collection
The Andrew Dickson White Architectural Photographs document a wide range of 19th and early 20th century architecture of Europe, the Middle East, and the Americas, including structures, panoramas and habitats that have vanished due to wars and urban...

Collection record: http://rdc.library.cornell.edu/index.php?identifier=Cornell:24

arXiv.org
ArXiv.org is a fully automated electronic archive and distribution server for research papers in physics, computer science, mathematics, and other scientific communities. Physicist Paul Ginsparg, the creator of the arXiv, joined the Cornell faculty in...

Collection record: http://rdc.library.cornell.edu/index.php?identifier=Cornell:48

Bernard Kassoy Teacher News Cartoons
Editorial cartoons created by artist Bernard Kassoy (1914-2008) document a time when many US policy makers were enthralled by the search for Communist Party members and sympathizers. Pithy drawings and captions produced for the New York Teacher News...

Collection record: http://rdc.library.cornell.edu/index.php?identifier=Cornell:1250

Beyond the Taj: Architectural Traditions and Landscape Experience in South Asia
The Beyond the Taj database is a collection of images depicting Indian architecture and culture. Primarily photographed by the late professor Robert D. MacDougall, the images are currently owned by Cornell University. This database is a joint project...

Collection record: http://rdc.library.cornell.edu/index.php?identifier=Cornell:46

Billie Jean Isbell Andean Collection: Images from the Andes
Billie Jean Isbell Andean Collection: Images from the Andes is derived from Professor Billie Jean Isbell's years of research in the Andes, primarily in the southern Andean department of Ayacucho and specifically in the village of Chuschi, Peru, and...

Collection record: http://rdc.library.cornell.edu/index.php?identifier=Cornell:27

Bolivian Digital Pamphlets Collection
This collection of 715 digitized works comes principally from a donation made to Cornell by the Bolivian bookseller, Alfredo Montalvo, who has supplied the university with library materials for over a quarter century. The pamphlets document a century...

Collection record: http://rdc.library.cornell.edu/index.php?identifier=Cornell:1111

Feminist, Gender, and Sexuality Studies: A Research Guide

Tags: feminism, gender, research_guides, sexuality_studies, women

A guide to discovering the extensive FGSS resources available in the Cornell University Library and beyond. Click on the PURPLE TABS to access each section in this guide.

Last Updated: Apr 1, 2014 | URL: http://guides.library.cornell.edu/fgss | Print Guide | RSS Updates | Email Alerts

Encyclopedias, Dictionaries and Almanacs

American Masculinities: A Historical Encyclopedia. Ed. Bret E. Carroll. Thousand Oaks, CA: Sage Publications, 2003.
(online and Olin Reference HQ 1090.3 .A453x 2003)
"[R]eference guide and an introduction to the many ways in which men have defined, imagined, and experienced male identity in the social, cultural, and political contexts of the United States." [Introduction]

Andermahr, Sonya. **Concise Glossary of Feminist Theories**. New York: Arnold, 1977.
(Olin Reference HQ 1115 .A543 1997)
"The main aim of the glossary is to identify the concepts that have structured feminist theory in the period of the second wave, from about 1968, indicating their provenance in travelling theory from this time.... The glossary aims to trace the variety of ways in which these concepts have been transformed as they were critiqued and appropriated." [Introduction] Liberally cross-referenced. Extensive bibliography of works cited.

Dictionary of Feminist Theologies. Ed. Letty M. Russell and J. Shannon Clarkson. Louisville: Westminster John Knox Press, 1996.
(Olin Reference BT 83.55 .D53x 1996)
"The *Dictionary* covers standard topics of Christian theological studies, as well as topics of other religions that are important for the continuing feminist dialogue among many different religious, cultural, and racial groups [e.g., see Islam]. There are also entries on topics of particular interest to those engaged in feminist studies. There are no entries on persons...." [Preface] The Bibliography of Works Cited is comprehensive for English-language publications on feminist theologies in the U.S. context.

Encyclopedia of Feminist Literary Theories. Ed. Elizabeth Kowaleski-Wallace. New York: Garland, 1997.
(Olin Reference PN 98 .W64 E53x 1997)
"The mission of this encyclopedia is to map out the vast intellectual territory that has arisen mostly since the 1970s, due to the efforts of feminist literary scholars and critics in the United States and Great Britain." [Preface] Definitions of important terms, summaries of critical ideas, and essays on theoretical movements and ideas in feminist literary scholarship and thought.

Encyclopedia of Feminist Theories. Ed. Lorraine Code. New York: Routledge, 2000.
(Olin Reference HQ 1190 .E63x 2000)
Succinct entries on major individuals, theories, developments, and terms in feminist history and thought. Bibliographies.

Encyclopedia of Gender and Society. Ed. Jodi O'Brien. 2 volumes. Thousand Oaks, CA: Sage Publications, 2009.
(online and Olin Reference HQ 1115 .E54 2009 +)
"Gender is so significant in the ways it shapes individuals and societies that it isn't possible to cover every aspect of gender scholarship in a single venue. Our intent with this encyclopedia is to provide users with a "gender lens" on society by focusing on significant gender scholarship within commonly recognized areas of social

Clark Physical Sciences Library

Classic Texts

Database Finder

eBook Collections

eJournal Finder

Use our feedback forum to discuss these tools and others you'd like to see.

Announcements

- ASCE ebooks – new in the Engineering eLibrary
- Math Library Exhibit – History of the Mathematics Department

Librarians

Dianne Dietrich VIVO
(Physics & Astronomy Librarian)

Leah McEwen VIVO
(Chemistry Librarian)

Library Guides

- Resources in the physical sciences
- Finding chemical properties and data

Essentials

- Passkey Off-campus Access
- Annex Request Annex/ILL
- Classic Catalog Classic Catalog
- SciFinder Chemical Abstracts
- Reaxys Beilstein & Gmelin
- Web of Science Scientific Abstracts
- PubMed Biomedical Abstracts
- arXiv Physics Preprints
- ADS Astrophysics Data System

Classic Texts

All
Astronomy
Chemistry
Physics

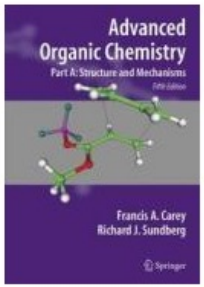
Alphabetical by Author Alphabetical by Title

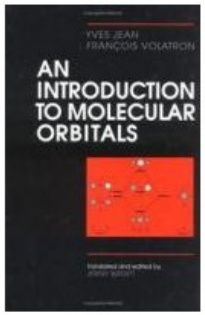
Grid View List View

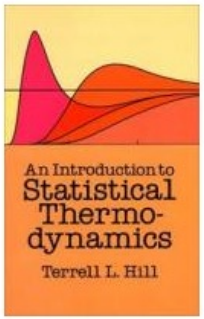
Advanced inorganic chemistry

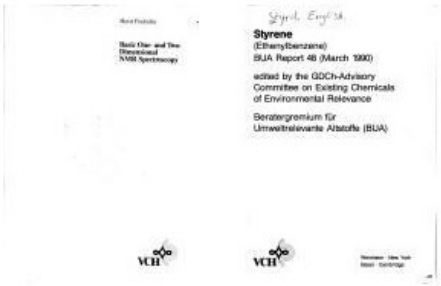
Advanced inorganic chemistry
5th edition
Cotton, F. Albert

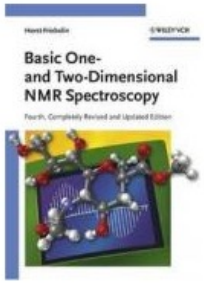
Advanced inorganic chemistry : a comprehensive text
4th edition.
Cotton, F. Albert



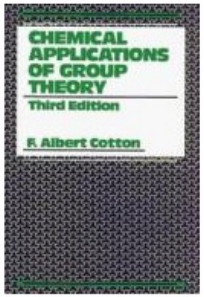








Chemical applications of group theory
2nd edition.
Cotton, F. Albert



Chemical applications of group theory
1st edition.
Cotton, F. Albert

1 2 3 4 5 6 7 next › last »

Comstock Entomology Collection

Albert R. Mann Library

Search the collection

Databases

- Journals & eJournals
- Books & eBooks
- Reference Collection
- Cornell Theses & Dissertations

PassKey Off Campus Access

New Books

Community dynamics and nutritional benefits of ohe Drosophila gut microbiota

Wong, Chun Nin Adam **2013**

Amino acid metabolism in maternal bacteriocytes of the pea aphid, Acyrthosiphon pisum

Russell, Calum W **2013**

Investigations of the biology of the pest aphid Aulacorthum solani (Kaltenbach) (hemiptera : aphididae), and of biological control agents for multi-species aphid outbreaks in greenhouse floriculture crops

About

The Comstock Memorial Library of Entomology was established in 1914, when the students of John Henry Comstock created a book fund in his honor. Later, Professor Comstock increased the fund by personal contribution and donated his books to the library. The primary subjects of the collection are general and applied entomology; related areas include parasitology, medical entomology, ecology, zoological nomenclature, and allied orders of arthropods. The collection has an extensive number of reprints, departmental theses, and rare books. In an effort to sustain the accessibility of this world-class collection in line with the strategic planning process of Cornell University and the Cornell University Library, the collection and services were integrated into the main life sciences collection housed at Albert R. Mann Library following the closing of the Comstock Memorial Library of Entomology in June 2012.

Comstock Entomology Collection

Albert R. Mann Library

Search results **do not include journal articles**. For entomology related articles select a recommended database. For articles outside of entomology browse databases by discipline. To find books and journals outside the field of entomology, please use the Library Catalog or Ask a Librarian.

Reference Collection

Search ... Search Retain filters

34 results

Sort by: Title A-Z 10 per page

1 2 3 4 next › last »

Publication Year

1862 2007

Format

- book 28
- database 3
- ebook 1
- ejournal 1
- journal 1

Type

- reference 34
- encyclopedia 6
- catalog 3
- dictionary 3
- handbook 3
- current journal 2
- classification 1

Bibliotheca entomologica. Die litteratur über das ganze gebiet der entomologie, bis zum jahre 1862

Hagen, Hermann August **1862**

Publisher: W. Engelmann

Availability

Mann Library Microfiche E-11 no.577

Biology data book

Altman, Philip L **1964**

Publisher: Federation of American Societies for Experimental Biology

Availability

Mann Library QH310 .A46
Library Annex QH310 .A46
Geneva Experiment Station Library QH310 .A46
Kroch Library Rare & Manuscripts (Request in advance) QH310 .A46

VIVO | Research & Expertise Across Cornell

Index | Log in

Search

Home | People | Organizations | Research | Events

Brittain, Charles | Professor

Positions

- Department Chairperson, College of Arts and Sciences, College of Arts and Sciences
- Professor/Classics/Philosophy, Classics (CLASS), College of Arts and Sciences

Charles Brittain is a Professor of Classics and Philosophy, specializing in ancient philosophy.

His research is primarily concerned with Hellenistic philosophy (especially epistemology and ethics), Cicero, Augustine, and the Platonic tradition from Plato to Simplicius.

Networks

Co-author Network

Map of Science

Websites

Classics profile
Philosophy profile

Affiliations | Research | Publications | Teaching | Service

selected publications

academic article

Posidonius' theory of predictive dreams 2011

Cicero: On Academic Scepticism. *Anzeiger für die Alte*

The New Academy's Appeals to the Presocratics . *Phro*

book

Philo of Larissa: The Last of the Academic Sceptics. Oxford ; New York: Oxford University Press. 2001

philo of larissa | All Fields | Search

Advanced Search | Search Tips

or Start over

All Fields: philo of larissa ✖ | Format: Book ✖

1 - 5 of 5

20 per page

1. Philo of Larissa : the last of the academic sceptics 2001

Brittain, Charles

Book Oxford ; New York : Oxford University Press, 2001. English

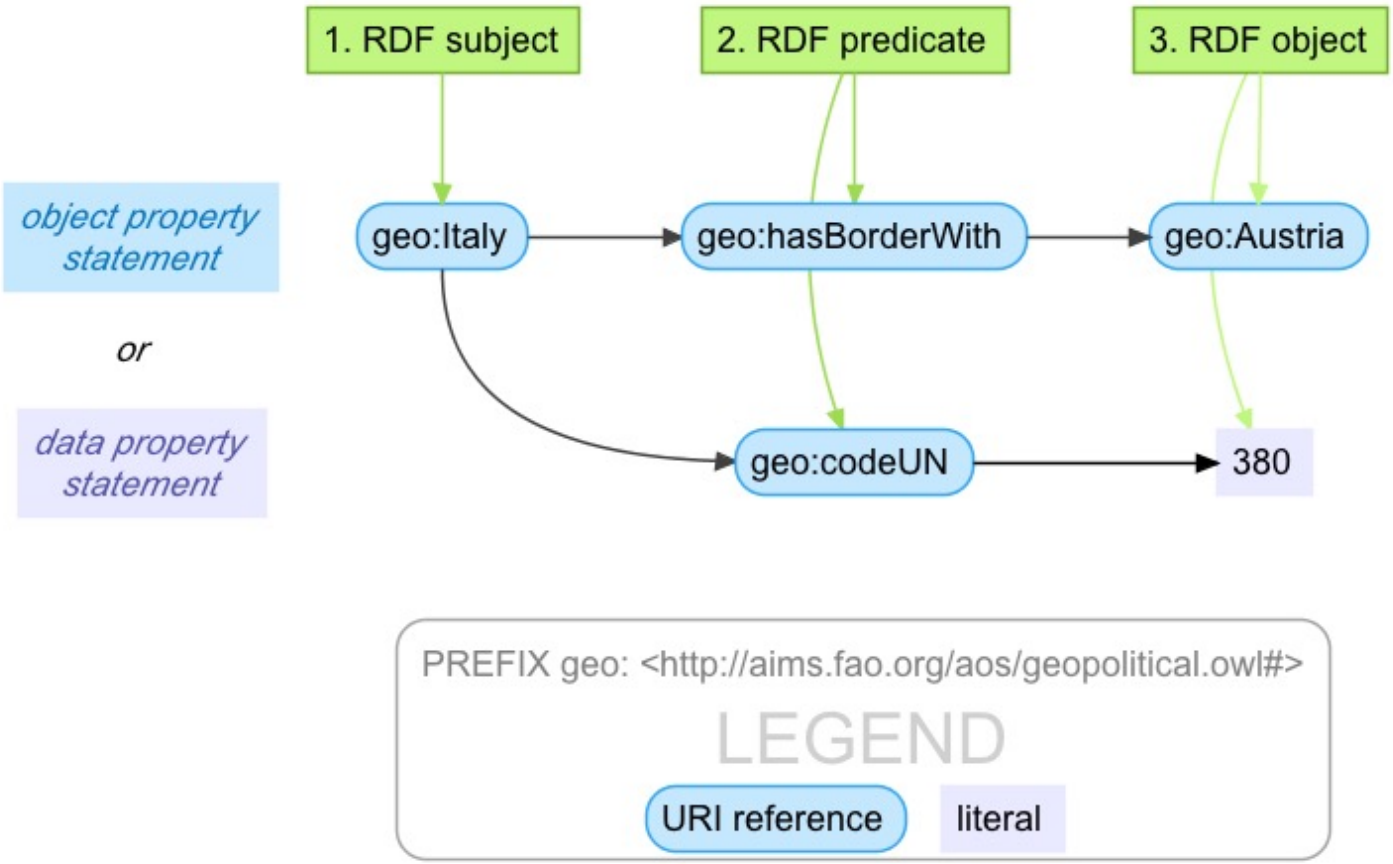
Text in English, with some Greek.

Olin Library B598.P34 B758x 2001

What is Linked Open Data?

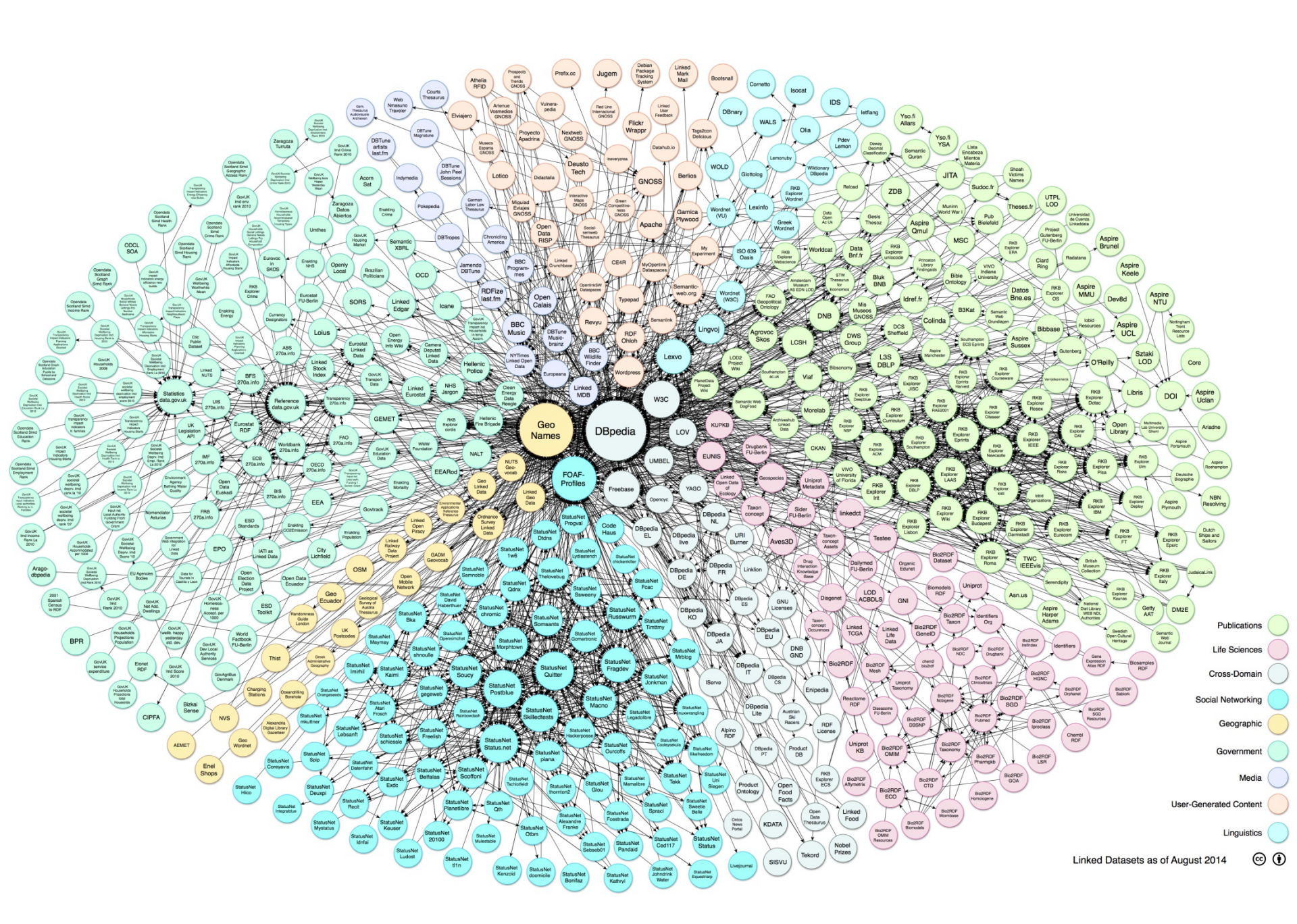
- Data
  - Structured information, not just documents and text
  - A common, simple format
- Open
  - Available, visible, mine-able
  - Anyone can post, consume, and reuse
- Linked
  - Directly by reference
  - Indirectly via common references and inference

RDF “triples”


1. RDF subject
2. RDF predicate
3. RDF object
object property statement
or
data property statement
geo:Italy
geo:hasBorderWith
geo:Austria
geo:codeUN
380
PREFIX geo: <http://aims.fao.org/aos/geopolitical.owl#>
LEGEND
URI reference
literal

Why Linked Data

- It is a flexible and extensible framework that libraries can use to describe, organize, and relate scholars, scholarship, and the scholarly context
- There are a wide range of tools, systems, ontologies, and vocabularies already available
- It is a growing ecosystem of developers, standards, and sources of relevant Linked Open Data



Linking Open Data cloud diagram 2014, by Max Schmachtenberg, Christian Bizer, Anja Jentzsch and Richard Cyganiak. http://lod-cloud.net/"

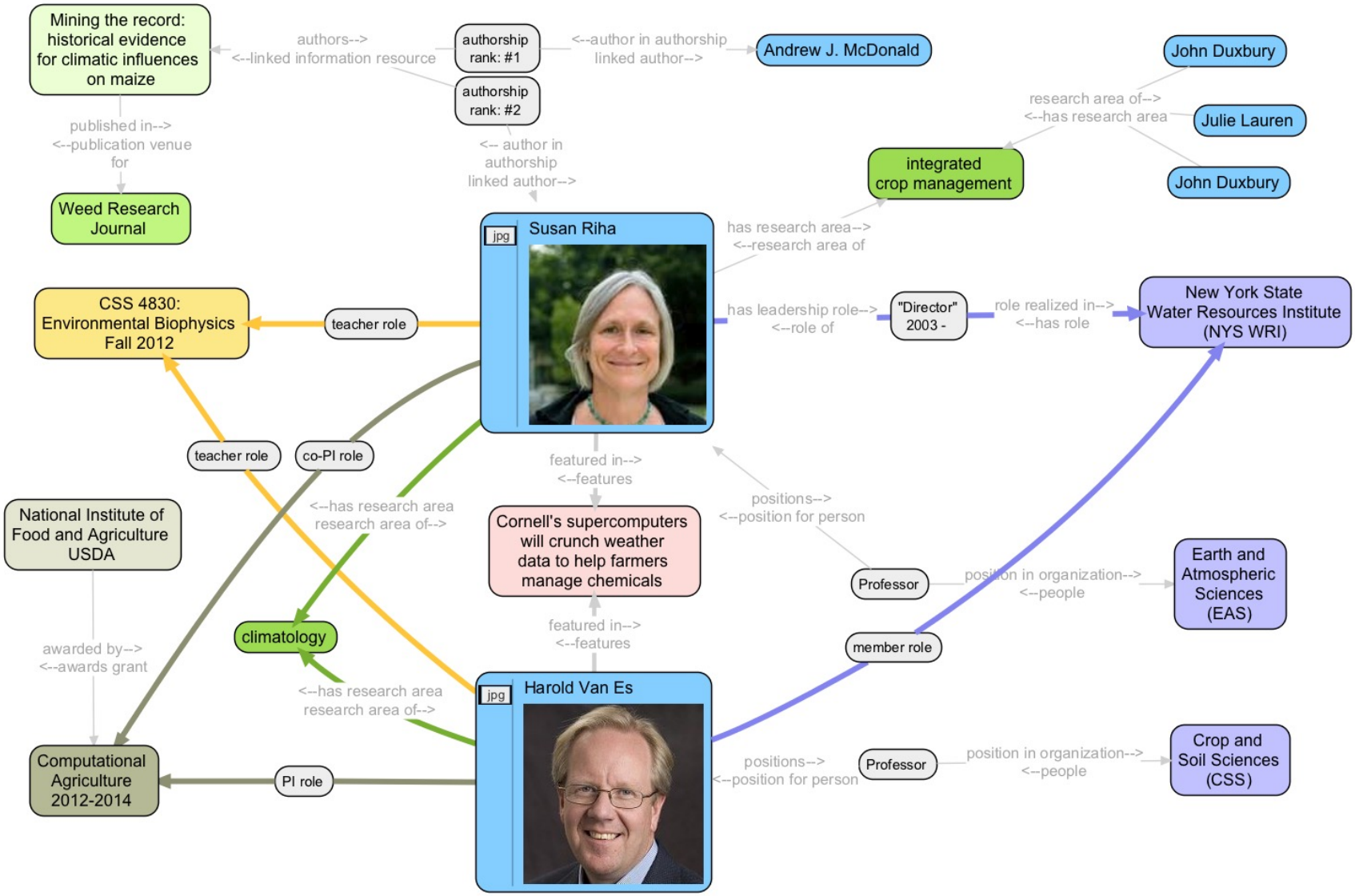
What is VIVO?

- Software: An open-source semantic-web-based researcher and research discovery tool
- Data: Institution-wide, publicly-visible information about research and researchers
- Standards: A standard ontology (VIVO data) that interconnects researchers, communities, and campuses using Linked Open Data
- Community: An open community with strong national and international participation

VIVO is a Semantic Web application

- Provides data readable by machines, not just text for humans
- Provides self-describing data via shared ontologies
  - Defined types
  - Defined relationships
- Provides search & query augmented by relationships
- Does simple reasoning to categorize and find associations
  - Teaching faculty = any faculty member teaching a course
  - All researchers involved with any gene associated with breast cancer (through research project, publication, etc.)

VIVO connects scientists and scholars with and through their research and scholarship



How LD4L builds on VIVO

- LD4L brings the relationship and identifier-based architecture of VIVO to mainstream library use cases and applications
- The LD4L ontology will draw on VIVO-ISF ontology design patterns (among others)
- Vitro as a semantic web browser will play a role in LD4L infrastructure along with more specialized, purpose-built tools
- The multi-institution LD4L demonstration search will be an adaptation of VIVOsearch.org
- LD4L will link to existing VIVO data in addition to Harvard Faculty Finder and Stanford CAP data

BIBFRAME

- The Library of Congress (LoC) has developed the BIBFRAME ontology as an (eventual) replacement for MARC, the current cataloging standard for library resources
- Both LoC and Zepheira, a contractor, have developed converters that produce BIBFRAME RDF from MARC XML

http://bibframe.org/

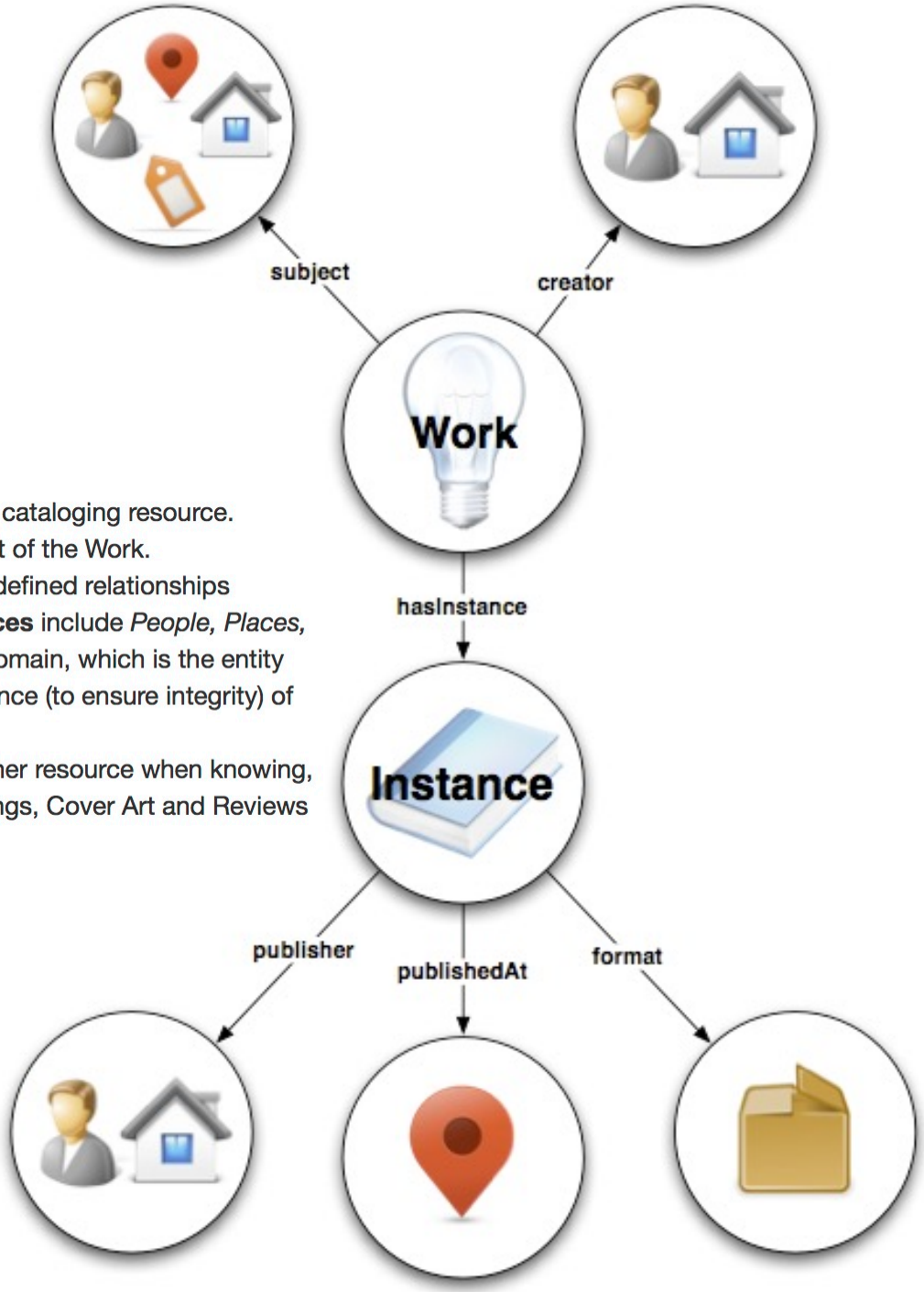
Why BIBFRAME?

- The Library of Congress BIBFRAME initiative "provides a foundation for the future of bibliographic description, both on the web, and in the broader networked world."
- We are academic libraries
- We are building from existing library catalog data in MARC and want to mainstream the use of identifiers and linked data in our library workflows

bibframe.org

BIBFRAME basic entities and relationships

- **Creative Work** - a resource reflecting a conceptual essence of the cataloging resource.
- **Instance** - a resource reflecting an individual, material embodiment of the Work.
- **Authority** - a resource reflecting key authority concepts that have defined relationships reflected in the Work and **Instance**. Example of **Authority Resources** include *People, Places, Topics, Organizations,* etc. One important concept in Authority is domain, which is the entity taking responsibility for the recognition, organization and maintenance (to ensure integrity) of the authoritative resources.
- **Annotation** - a resource that enhances our knowledge about another resource when knowing, minimally, 'who' is doing the annotating is important. Library Holdings, Cover Art and Reviews are examples types.



http://bibframe.org/vocab-model/

Adding identifiers

- Translating MARC records into RDF will not in and of itself make useful linked data
- Identifiers are essential
  - Local identifiers for statements made by an institution, both local authority information and annotation
  - Global identifiers for people, organizations, and places where they can be reliably determined
    - OCLC Work URIs for shared works across institutions
    - VIAF, ORCID for people
    - Evaluating Dbpedia for place linkages

Goal: from strings to things

- People
- Organizations
- Places
- Subjects
- Events
- Works
- Datasets

Image: designyoutrust.com

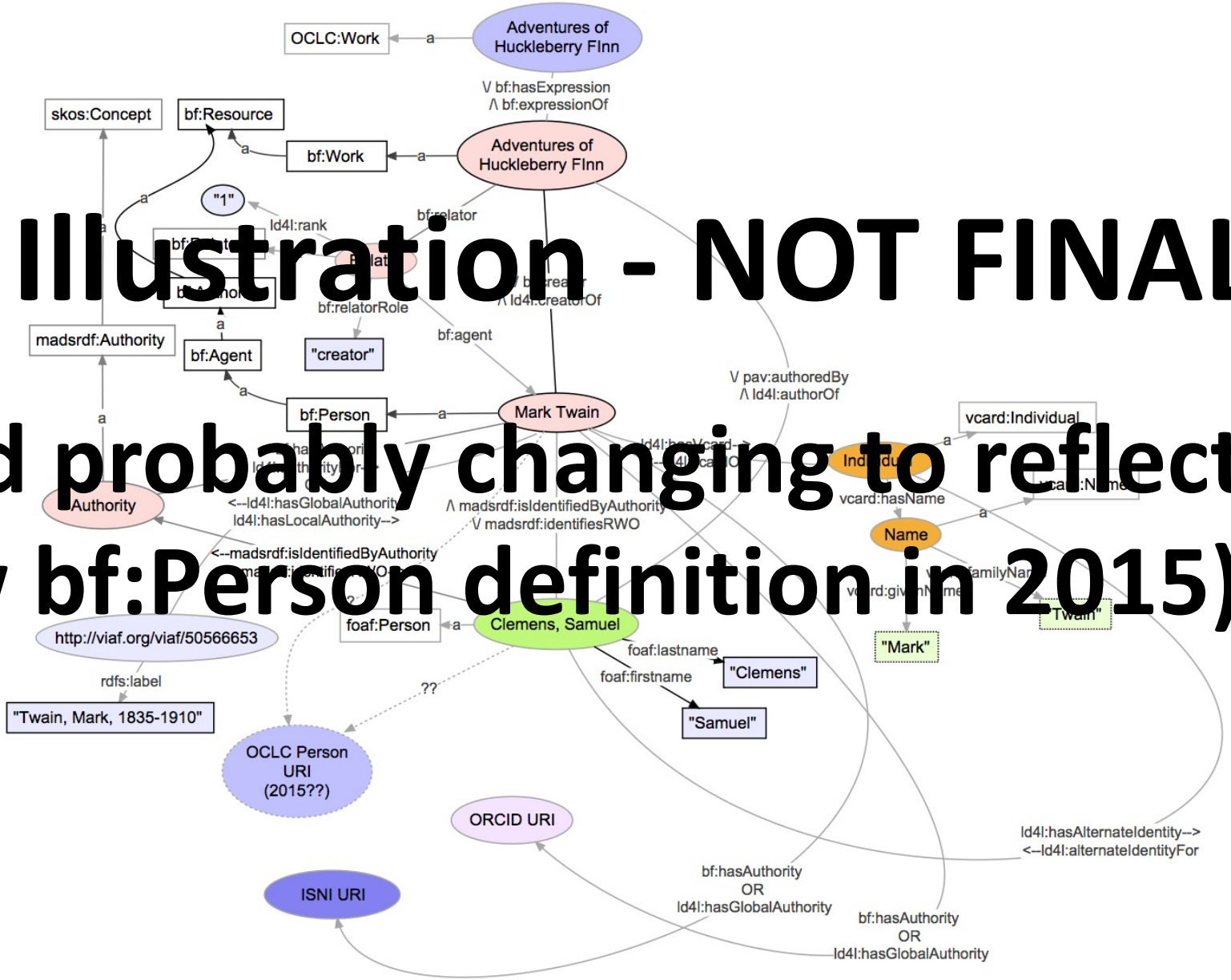
Leveraging OCLC work identifiers

- OCLC WorldCat functions as a union catalog of bibliographic identifiers shared across institutions
- Goal is to reference common OCLC Work URIs in bibliographic resources from Cornell, Harvard, and Stanford to support common search discovery and interlinking
  - Harvard: 82% of 13.6 million bib records can be matched to OCLC Work identifiers
  - Stanford and Cornell have ~2.6 million records in common out of ~5.8 million in each collection
  - Annotations and usage information can then be compared across 3 institutions
  - If it can work for 3, it can work for many

Likely components of the LD4L ontology

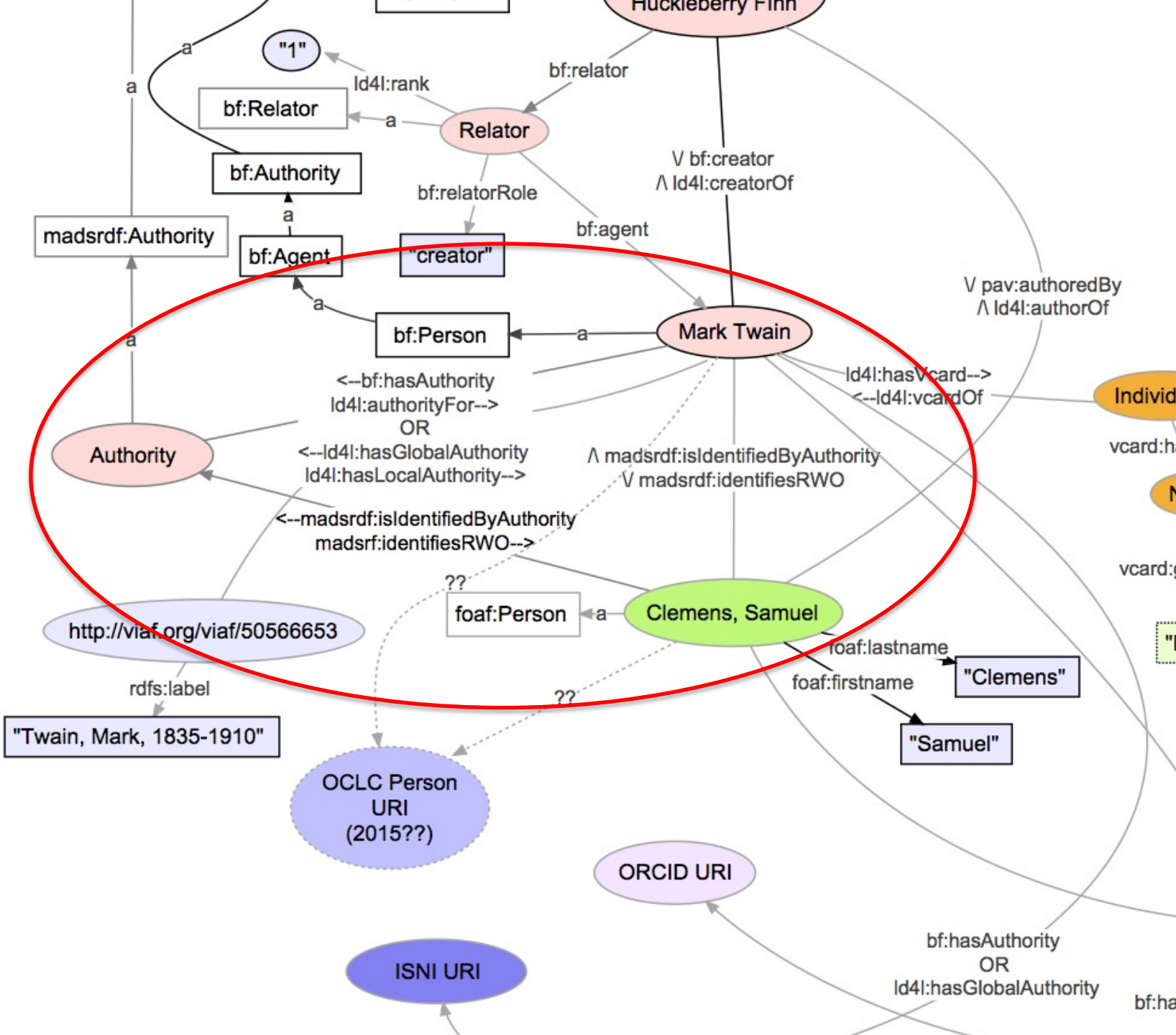
- Library resources: BIBFRAME
- Additional bibliographic types and partonomy relationships: FaBiO
- People/Organizations: VIVO-ISF (includes FOAF)
- Annotations: OpenAnnotation
- Provenance: PAV
- Virtual Collections and Structured Relationships: OAI-ORE
- Concepts: SKOS (or vocabularies such as Getty with stable URIs)
- Many identifiers: VIAF, ORCID, ISNI, OCLC Works

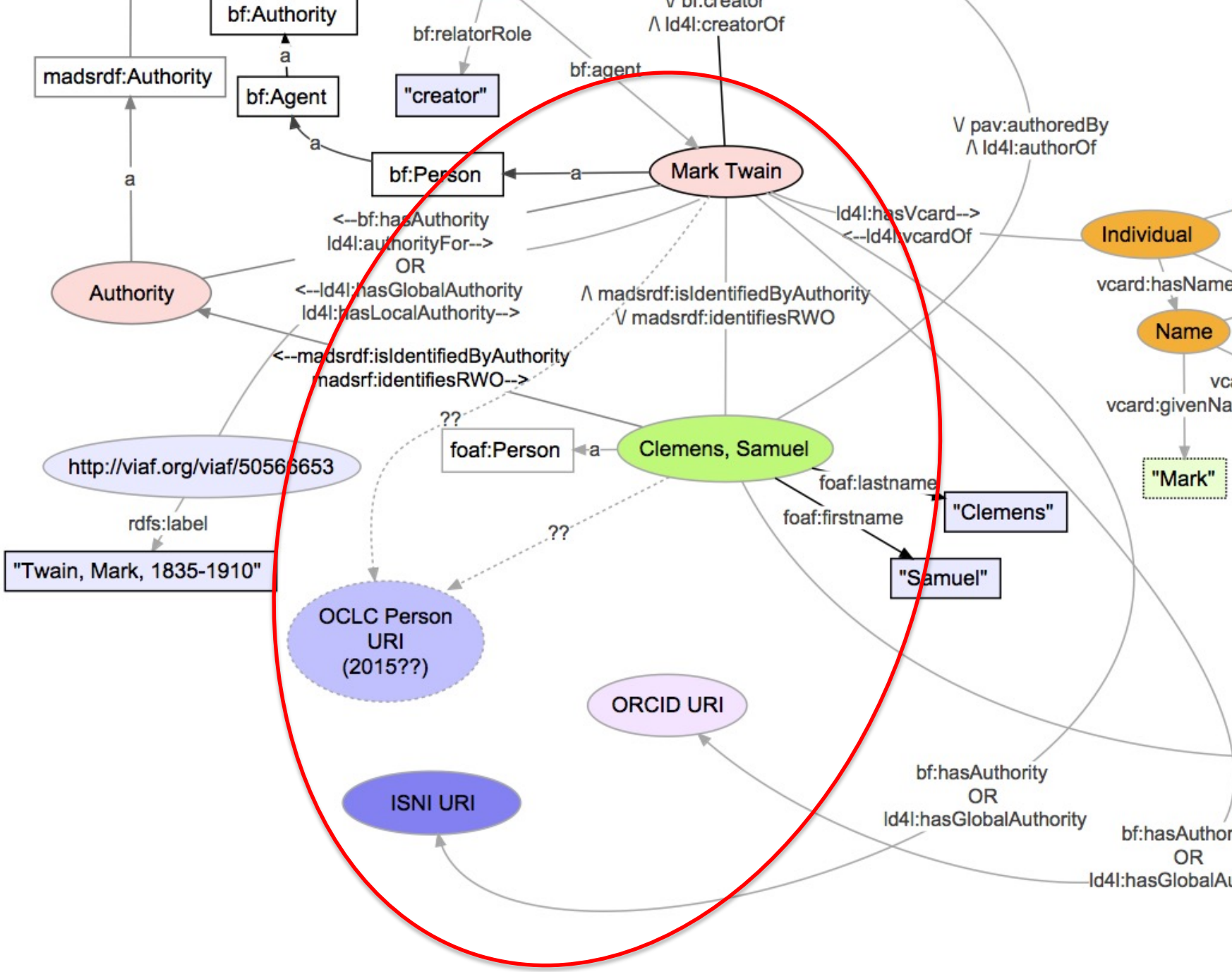
Snapshot of LD4L ontology discussion

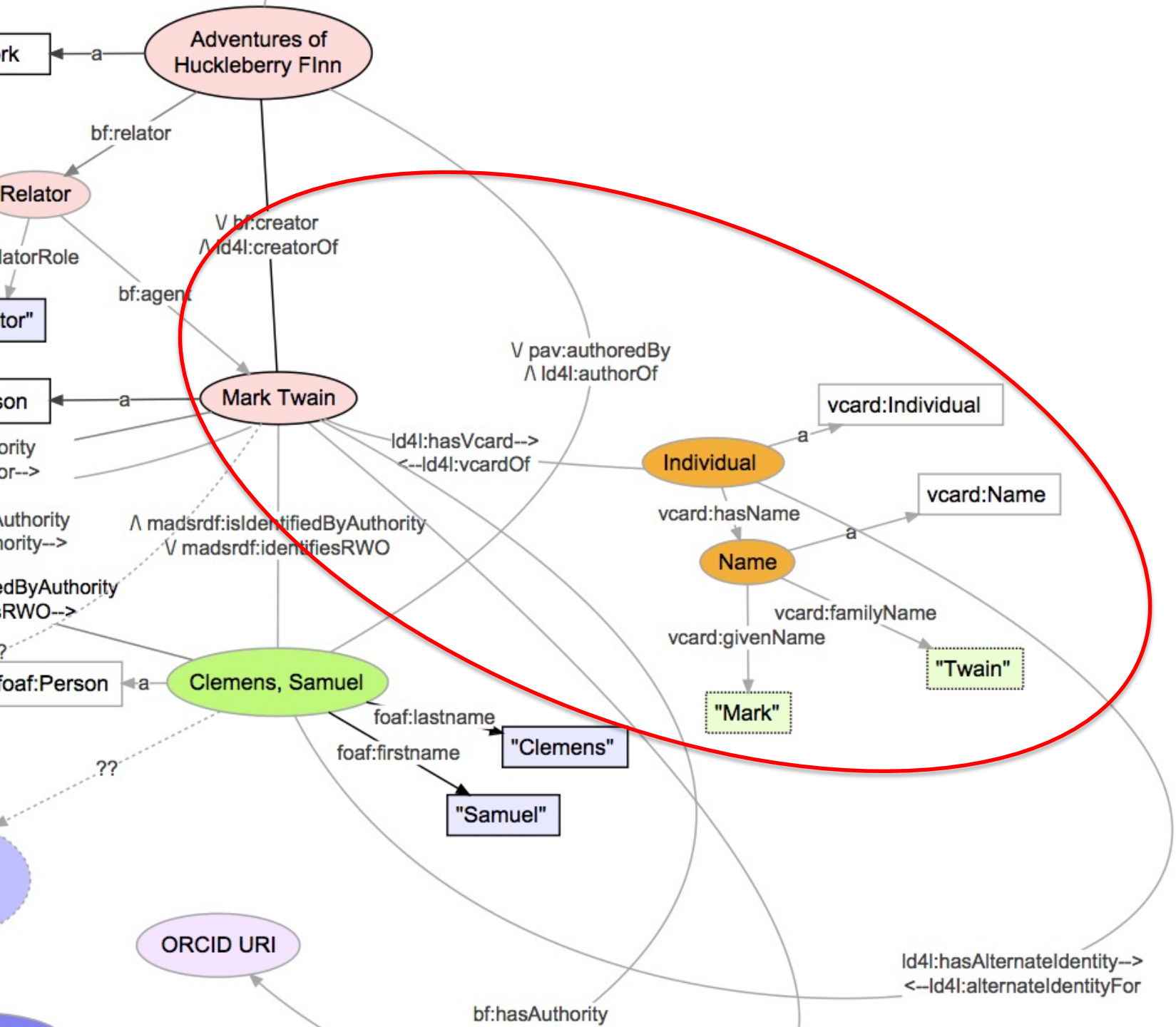


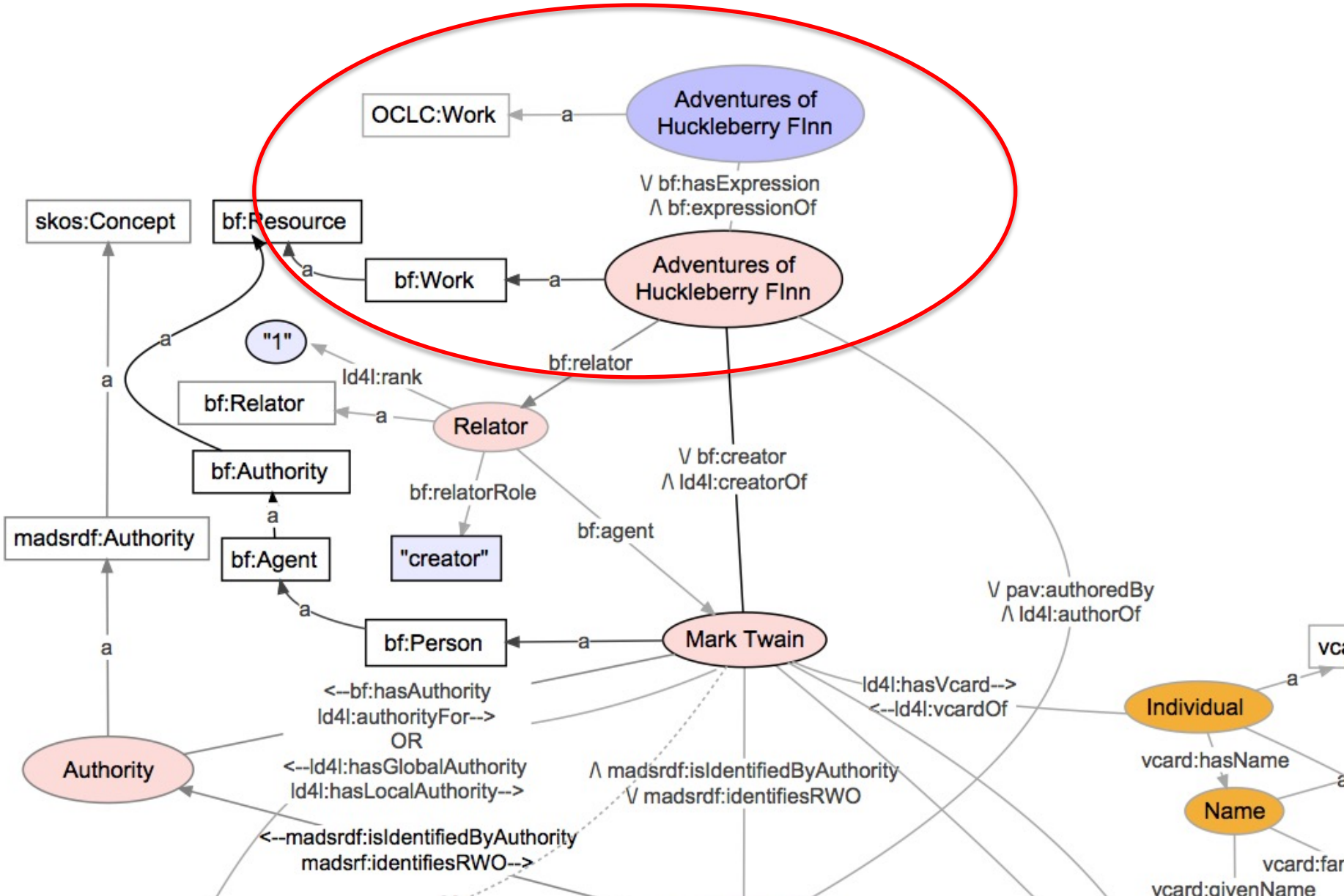
For Illustration - NOT FINAL

(And probably changing to reflect new bf:Person definition in 2015)


Huckleberry Finn
"1"
ld4l:rank
bf:relator
bf:Relator
a
Relator
bf:Authority
bf:relatorRole
∨ bf:creator
∧ ld4l:creatorOf
madsrdf:Authority
bf:Agent
"creator"
bf:agent
∨ pav:authoredBy
∧ ld4l:authorOf
bf:Person
Mark Twain
<--bf:hasAuthority
ld4l:authorityFor-->
OR
<--ld4l:hasGlobalAuthority
ld4l:hasLocalAuthority-->
ld4l:hasVcard-->
<--ld4l:vcardOf
Authority
∧ madsrdf:isIdentifiedByAuthority
∨ madsrdf:identifiesRWO
<--madsrdf:isIdentifiedByAuthority
madsrf:identifiesRWO-->
??
foaf:Person
Clemens, Samuel
http://viaf.org/viaf/50566653
foaf:lastname
"Clemens"
foaf:firstname
"Samuel"
rdfs:label
"Twain, Mark, 1835-1910"
OCLC Person
URI
(2015??)
ORCID URI
ISNI URI
bf:hasAuthority
OR
ld4l:hasGlobalAuthority


bf:Authority
madsrdf:Authority
a
bf:Agent
bf:relatorRole
"creator"
bf:agent
∨ bf:creator
∧ ld4l:creatorOf
∨ pav:authoredBy
∧ ld4l:authorOf
a
bf:Person
a
Mark Twain
a
<--bf:hasAuthority
ld4l:authorityFor-->
OR
<--ld4l:hasGlobalAuthority
ld4l:hasLocalAuthority-->
ld4l:hasVcard-->
<--ld4l:vcardOf
Individual
vcard:hasName
Name
vcard:givenNa
"Mark"
Authority
∧ madsrdf:isIdentifiedByAuthority
∨ madsrdf:identifiesRWO
<--madsrdf:isIdentifiedByAuthority
madsrf:identifiesRWO-->
??
foaf:Person
a
Clemens, Samuel
http://viaf.org/viaf/50566653
foaf:lastname
"Clemens"
foaf:firstname
rdfs:label
??
"Twain, Mark, 1835-1910"
"Samuel"
OCLC Person
URI
(2015??)
ORCID URI
ISNI URI
bf:hasAuthority
OR
ld4l:hasGlobalAuthority
bf:hasAuthor
OR
ld4l:hasGlobalAu


Adventures of Huckleberry FInn
a
bf:relator
Relator
V bf:creator
Λ ld4l:creatorOf
bf:agent
V pav:authoredBy
Λ ld4l:authorOf
Mark Twain
a
ld4l:hasVcard-->
<--ld4l:vcardOf
Individual
a
vcard:Individual
vcard:Name
vcard:hasName
a
Name
vcard:familyName
vcard:givenName
"Twain"
"Mark"
Λ madsrdf:isIdentifiedByAuthority
V madsrdf:identifiesRWO
Clemens, Samuel
a
foaf:Person
foaf:lastname
"Clemens"
foaf:firstname
"Samuel"
??
ORCID URI
ld4l:hasAlternateIdentity-->
<--ld4l:alternateIdentityFor
bf:hasAuthority


OCLC:Work
a
Adventures of Huckleberry FInn
∨ bf:hasExpression
∧ bf:expressionOf
skos:Concept
bf:Resource
a
bf:Work
a
Adventures of Huckleberry FInn
a
"1"
ld4l:rank
bf:relator
a
bf:Relator
a
Relator
bf:Authority
bf:relatorRole
∨ bf:creator
∧ ld4l:creatorOf
a
madsrdf:Authority
bf:Agent
"creator"
bf:agent
∨ pav:authoredBy
∧ ld4l:authorOf
a
bf:Person
a
Mark Twain
a
<--bf:hasAuthority
ld4l:authorityFor-->
OR
<--ld4l:hasGlobalAuthority
ld4l:hasLocalAuthority-->
ld4l:hasVcard-->
<--ld4l:vcardOf
Individual
a
Authority
∧ madsrdf:isIdentifiedByAuthority
∨ madsrdf:identifiesRWO
vcard:hasName
Name
<--madsrdf:isIdentifiedByAuthority
madsrf:identifiesRWO-->

Ontology Challenges

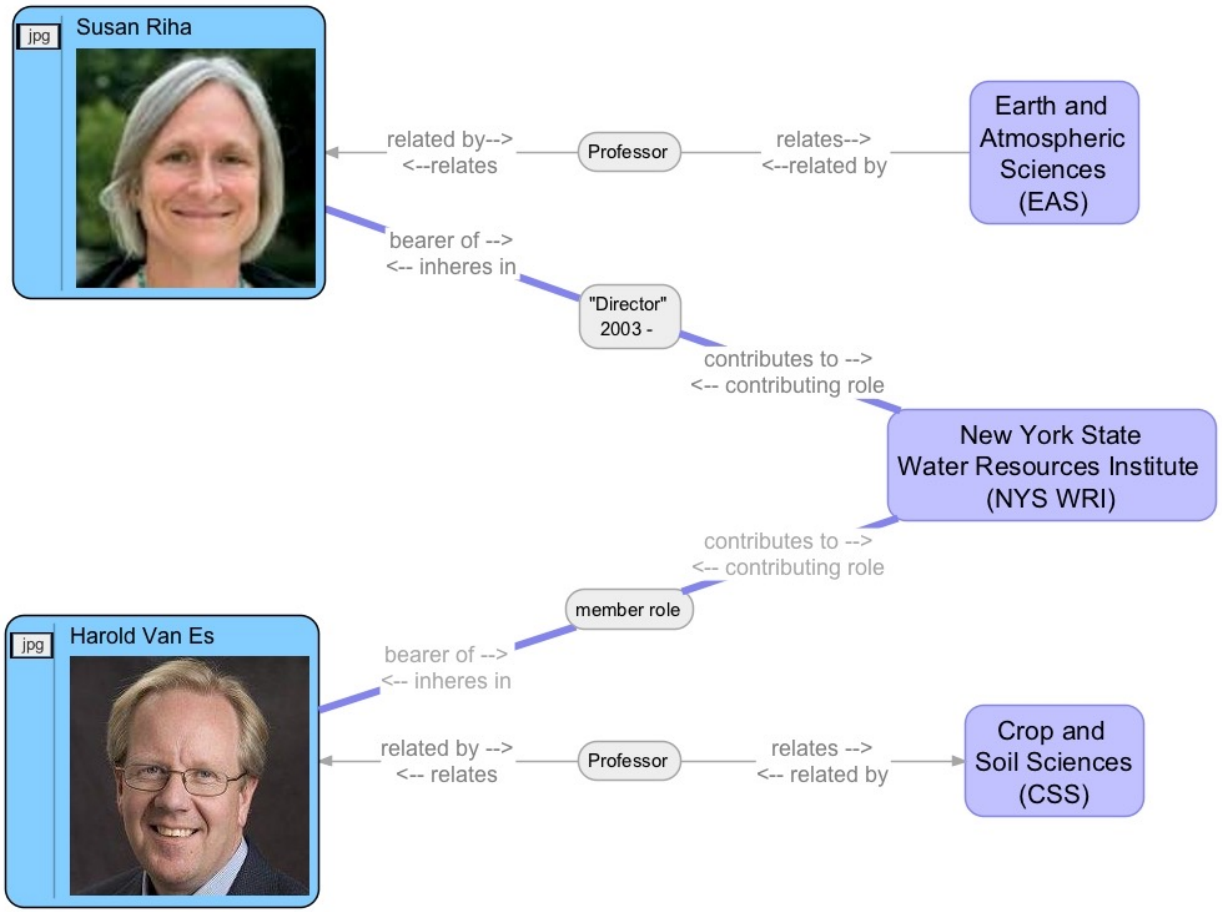
- We need to think about identifying people, and their relationships to other entities
- There are already identifiers for people and works, which need to be connected
- There are hard choices around the edges to be made, such as a single person with multiple identities, but that shouldn't stop progress being made for the 99.999%

Entity Reconciliation

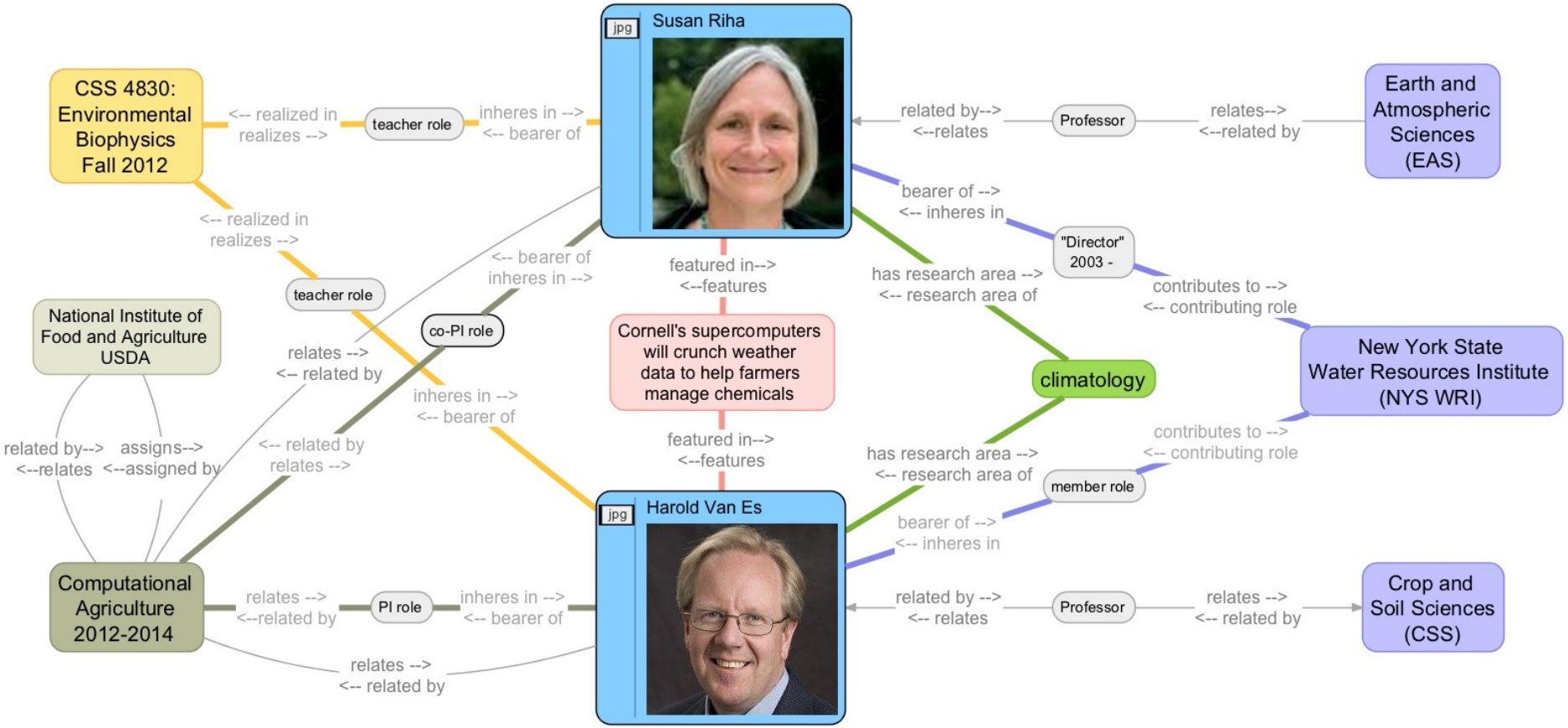
- Locally critical to link information across library system silos
- Essential to link across the three partners to support discovery, annotation, virtual collections: works, people, places, subjects, etc.
- Linking to web of LOD surfaces new relationships and networks
- Library role: expose our own unique entities and connect them to the rest of the world

The more we can link, the more we can discover

The More We Can Link: Loosely Connected Researchers

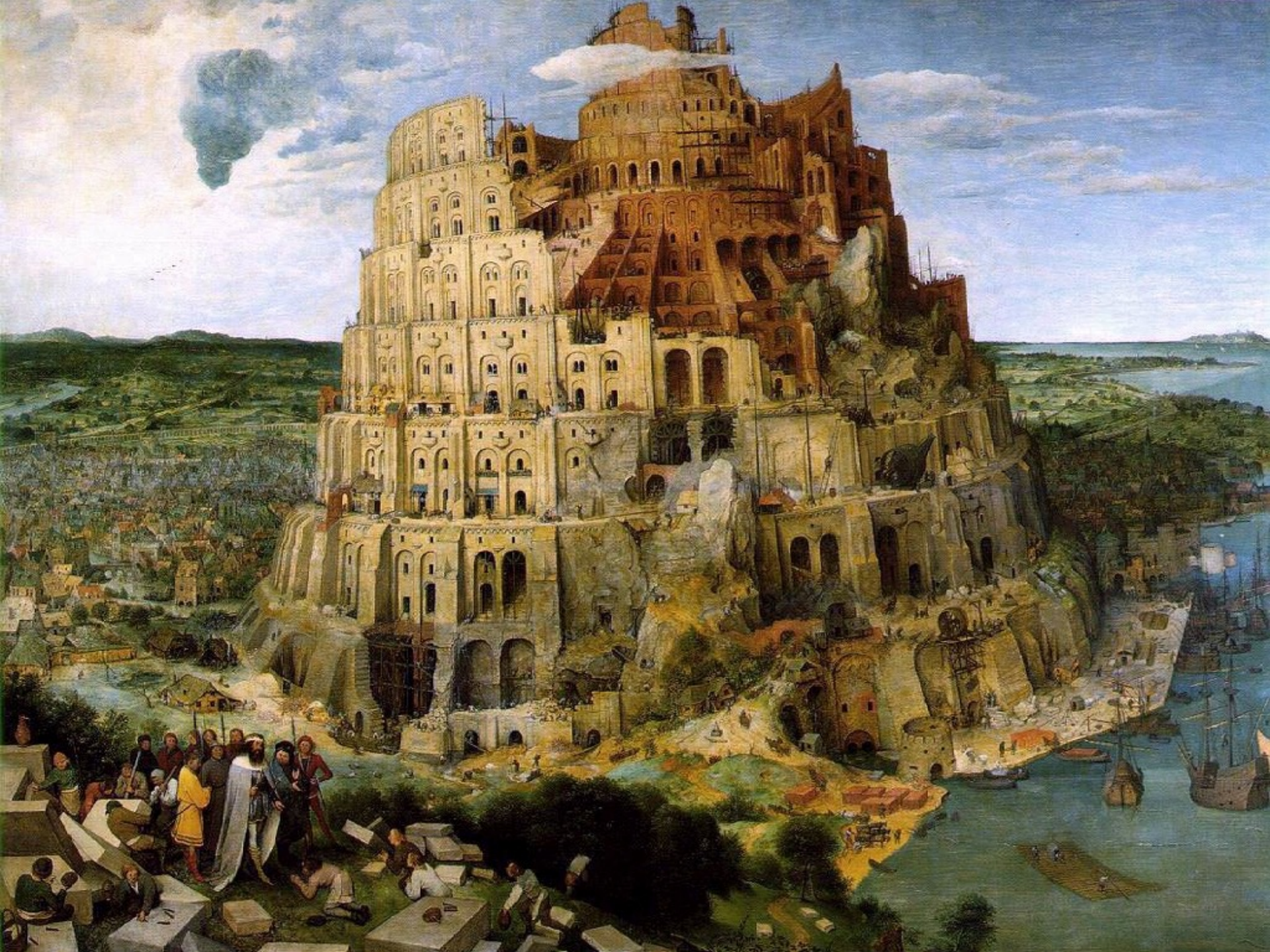


Strongly Connected Researchers

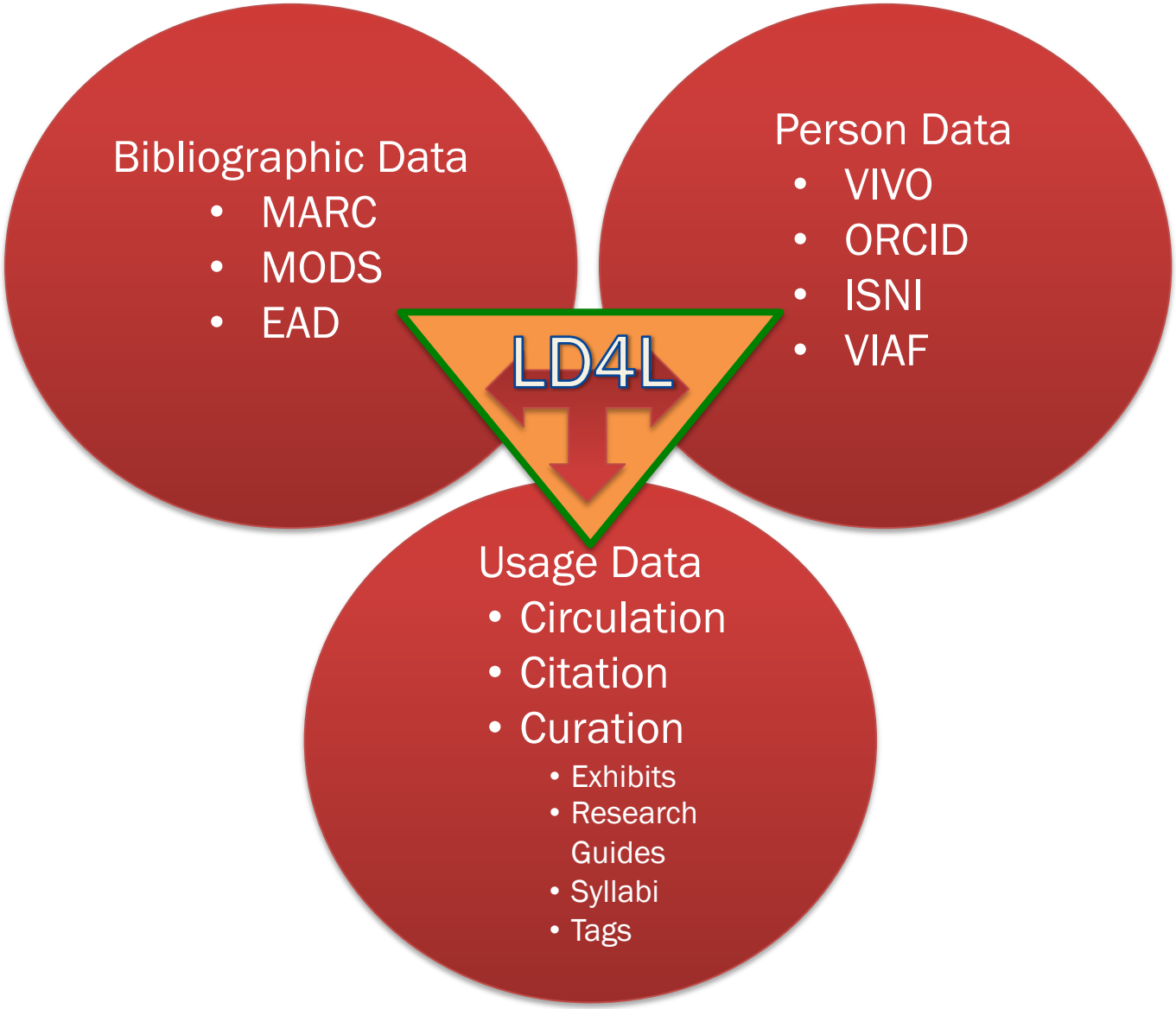


How will LD4L make these connections?

- By using ontologies commonly found in linked data
- By connecting with Cornell VIVO/Stanford CAP/Harvard Profiles information
- By using persistent, stable local identifiers (URIs)
- By linking stable local identifiers to global identifiers (ORCID, VIAF, ISNI)
- By supporting annotations with provenance
- By linking to external sources of networked relationships: Dbpedia, IMDB, OCLC



LD4L Data Sources



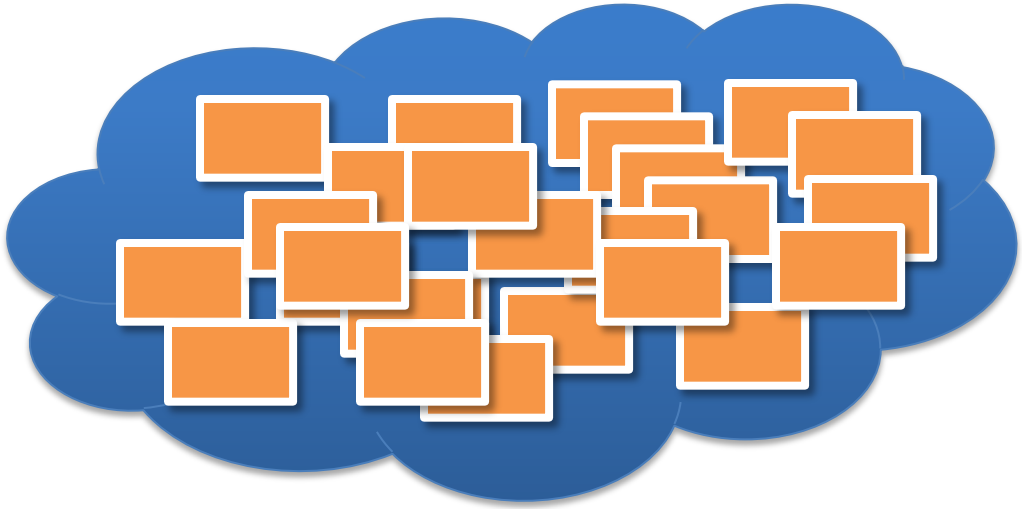
Stories as the Basis of Use Cases

As a _____, I want to _____, so that I can <realize this benefit>.

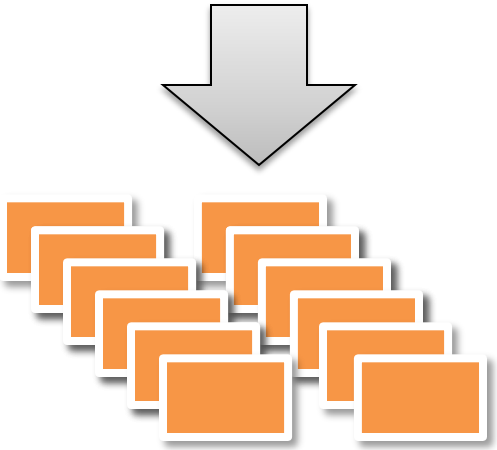
Potential Demonstrations:

A. Demo 1
B. Demo 2
C. Demo 3

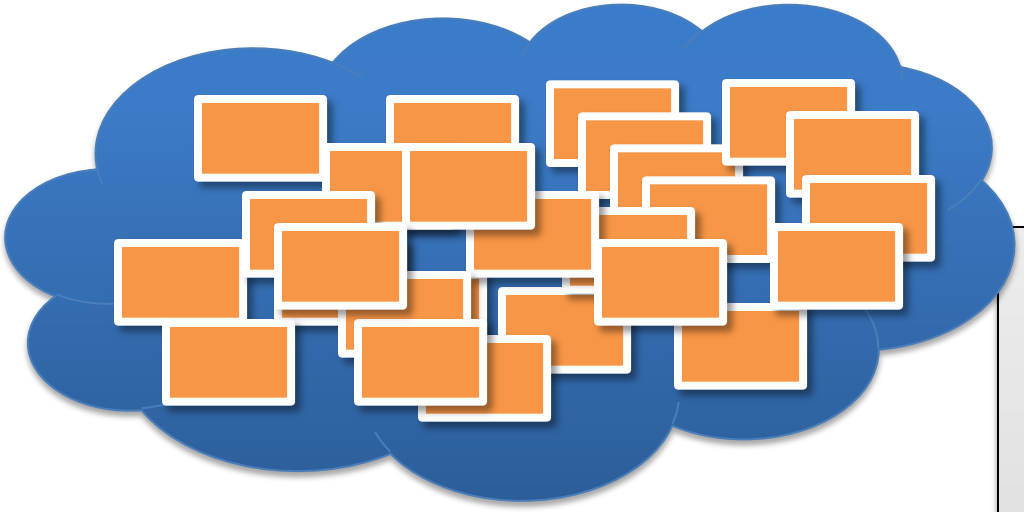
Data Sources Needed:
Ontology Requirements:
Engineering Work:



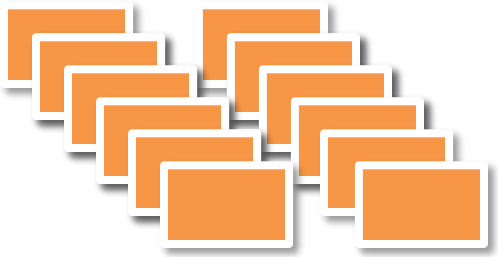
42 Raw Use Cases



12 Refined Use Cases
in 6 clusters…



42 Raw Use Cases



12 Refined Use Cases in 6 clusters…

LD4L Use Case Clusters

1. Bibliographic + curation data
2. Bibliographic + person data
3. Leveraging external data including authorities
4. Leveraging the deeper graph (via queries or patterns)
5. Leveraging usage data
6. Three-site services, e.g. cross-site search

Sample Refined Use Cases

1.1: Build a virtual collection

As a faculty member or librarian, I want to create a virtual collection or exhibit from multiple collections, so that I can share a focused collection with a <class, set of researchers, set of students in a disciplinary area>.

1.2: Tag scholarly information resources to support reuse

As a librarian, I would like to be able to tag scholarly information resources into curated lists, so that I can feed these these lists into subject guides, course reserves, or reference collections.

https://wiki.duraspace.org/display/ld4l/LD4L+Use+Cases

Engineering Work

Phase 1: Annotations

1.1 Build a Virtual Collection

1.2 Tag Scholarly Resources to Support Reuse

Phase 2: Authorities

2.1 Discover Works via People and their Relationships

3.1 Discover Works via Locations and their Relationships

3.2 Discover Works via Concepts and their Relationships

Phase 3: Linked Open Data

4.* Leverage the Deeper Graph

5.* Leverage Usage Data for <Research, Collection Building>

6.1 Cross-Institution Discovery

Engineering Work

Phase 1: Annotations

1.1 Build a Virtual Collection

1.2 Tag Scholarly Resources to Support Reuse

Engineering

Phase 2: Authorities

2.1 Discover Works via People and their Relationships

3.1 Discover Works via Locations and their Relationships

3.2 Discover Works via Concepts and their Relationships

Planning

Phase 3: Linked Open Data

4.* Leverage the Deeper Graph

5.* Leverage Usage Data

6.1 Cross-Institution Discovery

Research

http://projecthydra.org

What Is Hydra?

- A robust repository fronted by feature-rich, tailored applications and workflows (“heads”)
  - *One body, many heads*
- Collaboratively built “solution bundles” that can be adapted and modified to suit local needs.
- A community of developers and adopters extending and enhancing the core
  - *If you want to go fast, go alone. If you want to go far, go together.*

Technical Framework - Components

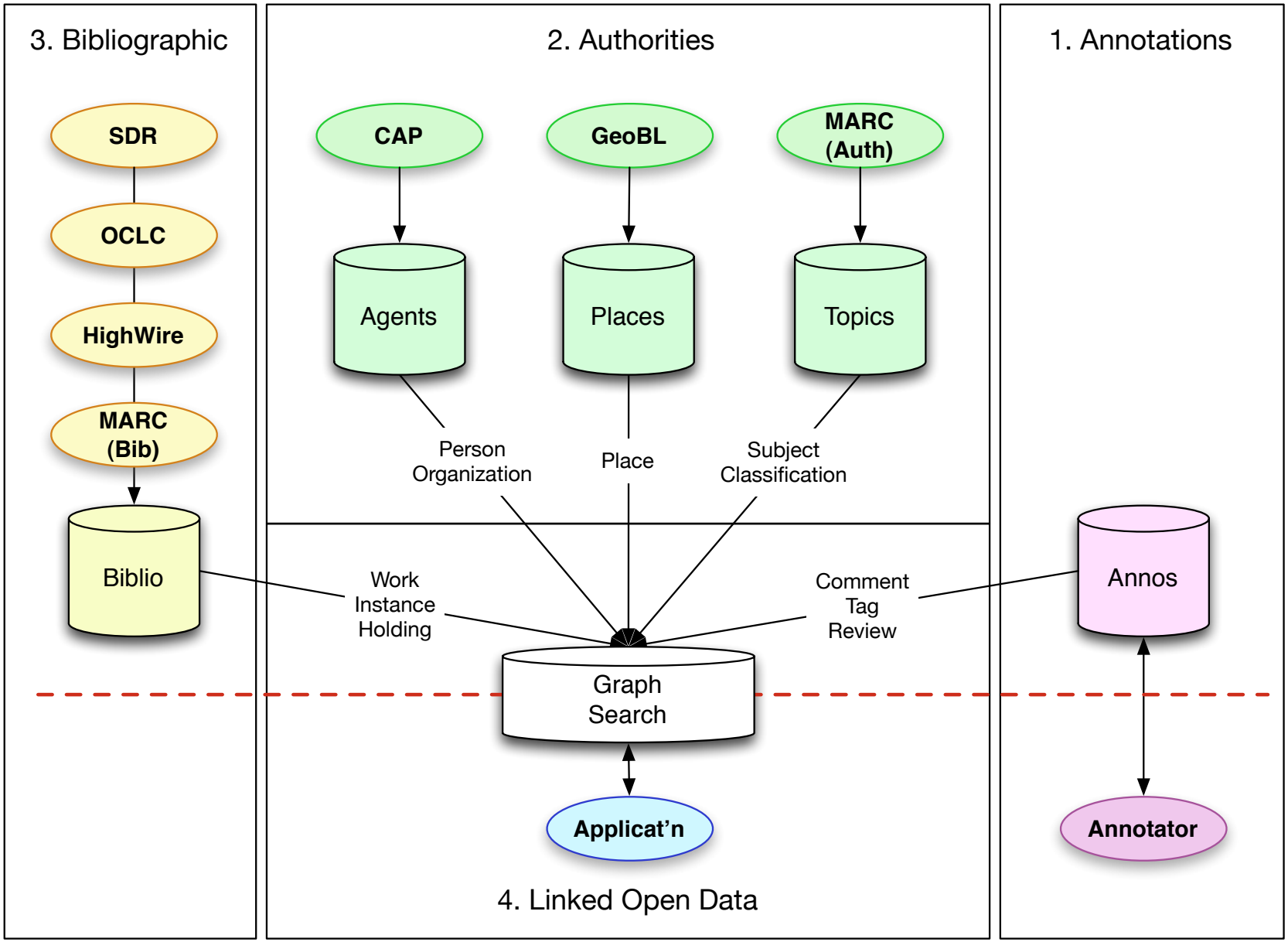
- Fedora provides a durable repository layer to support object management and persistence
- Solr, provides fast access to indexed information
- Blacklight, a Ruby on Rails plugin that sits atop solr and provides faceted search & tailored views on objects
- Hydra-Head, a Ruby on Rails plugin that provides create, update and delete actions against Fedora objects

How LD4L builds on Hydra

- We will augment the ActiveTriples gem to mimic ActiveFedora
- We will write code to store Open Annotations (OA) linked data in Fedora 4, natively
- We will use Blacklight as a UI for making/viewing annotations, and for searching data indexed from LD4L triple stores
- We will leverage the Questioning Authority Gem for *Use Case 3.4: LOD-based Data Entry*

- Provides model for access to library data
- Includes access to ShelfRank/StackScore for Harvard Library resources
- Provides concrete example for creating an ontology for usage
- Source data for Harvard LD4L instance

Assembling the Data



LD4L working assumptions

- Trying to do conversion and relation work at scale--with full sets of enterprise data
  - Harvard: 13.6 million bibliographic records
  - Stanford and Cornell: roughly 8 million bib records in each collection
- Trying to understand the pipeline / workflows that might be needed for this
- Looking to build useful, value-added services on top of the assembled triples

Challenges

- Perfection as the enemy of the good (e.g., in ontologies and reconciliation)
- Minting vs. finding identifiers
- Wider issues of entity reconciliation -- when do we mean sameAs vs. seeAlso?
- Scale
- Leveraging technologies developed for other purposes (e.g., LibGuides, DMS, CAP)
- Thinking outside the bibliographic box

Enhancing discovery and understanding with fully reconciled linked data

Ideally library descriptions of our information resources will:

- Refer to **identified** works, people, organizations, places, events, and subject headings
- Be discoverable in concert with other explicit and implicit library metadata of all types
- Join with local personal and organizational profile data to form a coherent, richer local authority picture
- Interoperate across libraries
- Interoperate with any other linked data via common, global identifiers and shared ontologies

Current state of the LD4L project

Courtesy National Gallery of Art, Washington

Project timeline 2014

- Jan-June 2014: Initial ontology design; identify data sources; identify external vocabularies; begin LD4L and Hydra ActiveTriples development
- July-Dec 2014: Complete initial ontology; complete initial ActiveTriples development; pilot initial data ingests into Vitro-based LD4L instance at Cornell

Workshop – February 2015

- Hold a two-day by invitation workshop for 25 attendees from 10-12 interested library, archive, and cultural memory institutions
- Demonstrate initial prototypes of LD4L and ontology
- Obtain feedback on initial ontology design
- Obtain feedback on overall design and approach
- Make connections to support participants in piloting this approach at their institutions
- Understand how institutions see this approach fitting in with their own multi-institutional collaborations and existing cross-institutional efforts such as the Digital Public Library of America, VIVO, and SHARE

Project timeline Jan-June 2015

- Pilot LD4L instances at Harvard and Stanford
- Populate Cornell LD4L instance from multiple data sources including MARC catalog records, EAD finding aids, VIVO data, CuLLR, and local digital collections
- Develop a test instance of the LD4L Search application harvesting RDF across the three partner institutions
- Integrate LD4L with ActiveTriples

Project timeline July-Dec 2015

- Implement fully functional LD4L instances at Cornell, Harvard, and Stanford
- Public release of open source LD4L code and ontology
- Public release of open source ActiveTriples Hydra Component
- Create public demonstration of LD4L Search-based discovery and access system across the three LD4L instances

LD4L Partnership

- Cornell, Harvard, and Stanford brought together by common interest in Linked Data
- Researchers, developers, and production metadata/cataloging all involved
- Will seek to expand LD4L community at workshop in February 2015
- Significant overlap between LD4L efforts and work of both VIVO and Hydra communities: code will go to those repositories

LD4L Team – June 2014

Project Outcomes

- Open source extensible LD4L ontology compatible with VIVO ontology, BIBFRAME, and other existing library LOD efforts
- Open source LD4L semantic editing, display, and discovery system
- Project Hydra compatible interface to LD4L, using ActiveTriples to support Blacklight search across multiple LD4L instances

Libraries must **evolve** and **collaborate** to meet the needs of tomorrow's scholars and researchers

J.W. Audubon - Ivory-billed Woodpecker

Questions?