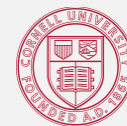# Linked Data for Libraries (LD4L): an update

Nancy Lorimer

Stanford University

ALA Midwinter 2015

Cornell University Library
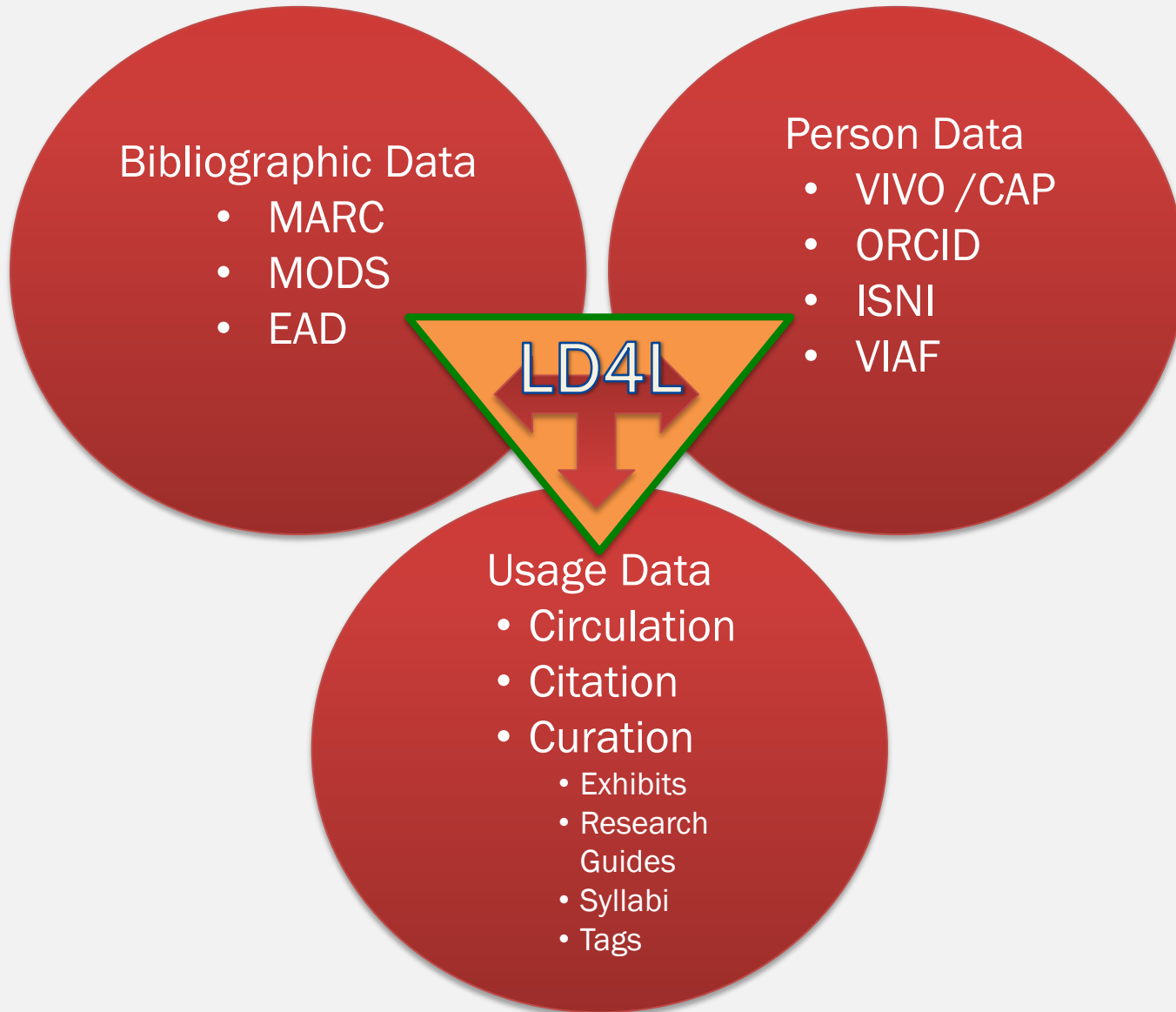
HARVARD LIBRARY

Linked Data for Libraries (LD4L)

- Two-year grant to Cornell, Harvard, and Stanford starting Jan '14 from the Mellon Foundation

- 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, the Hydra Partnership.

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

LD4L Data Sources

Bibliographic Data
- MARC
- MODS
- EAD

Person Data
- VIVO /CAP
- ORCID
- ISNI
- VIAF

LD4L

Usage Data
- Circulation
- Citation
- Curation
  - Exhibits
  - Research Guides
  - Syllabi
  - Tags

THE ONTOLOGY…

BIBFRAME

- The Library of Congress (LoC) has developed the BIBFRAME ontology as an (eventual) replacement for MARC

- Bibframe is the leading candidate to replace MARC and provides the most complete capture of MARC data when transformed

- Converters of MARC to Bibframe through MARCxml in development

- MARC data is by far our largest pool of data to be transformed
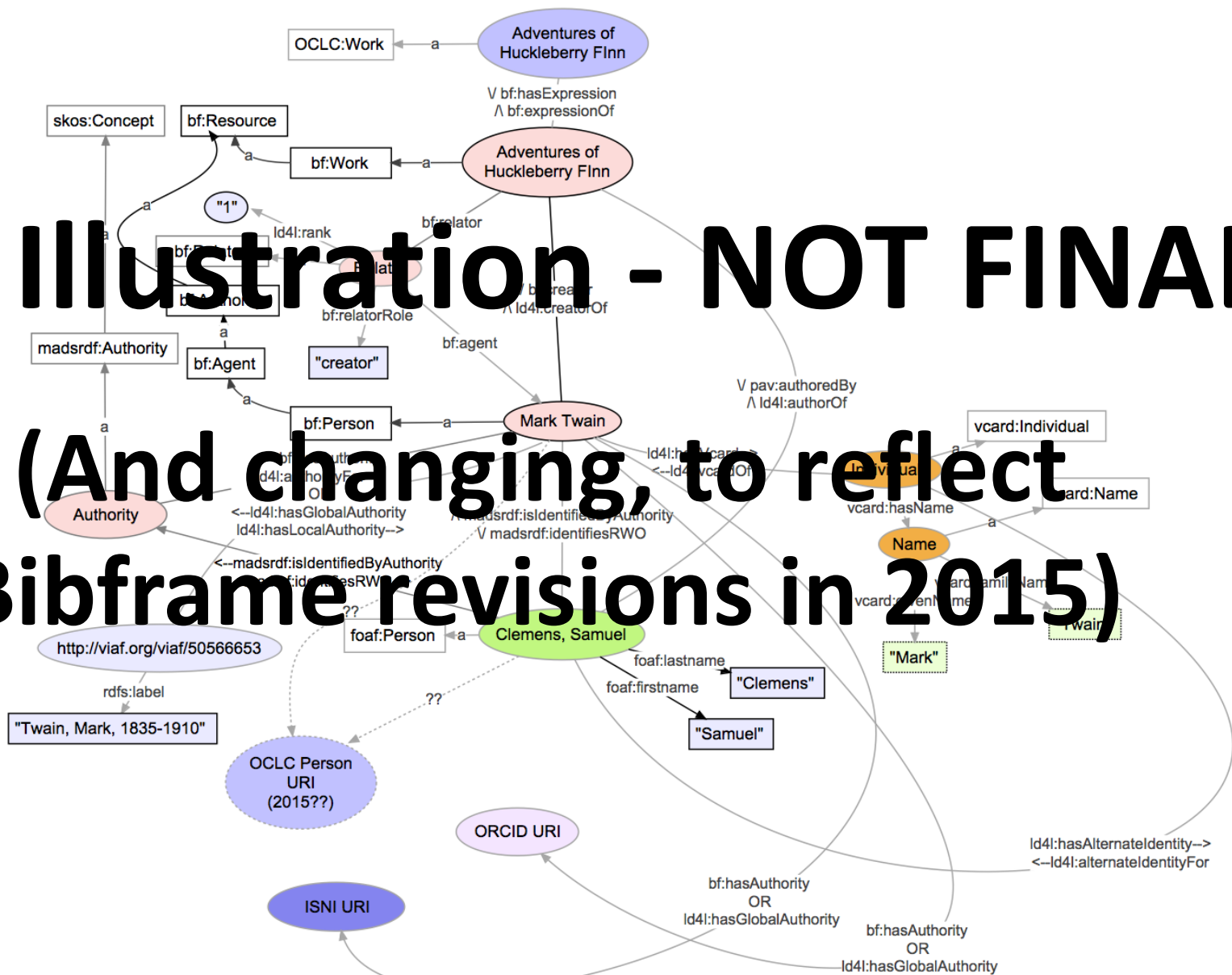
http://bibframe.org/

Likely components of the LD4L ontology

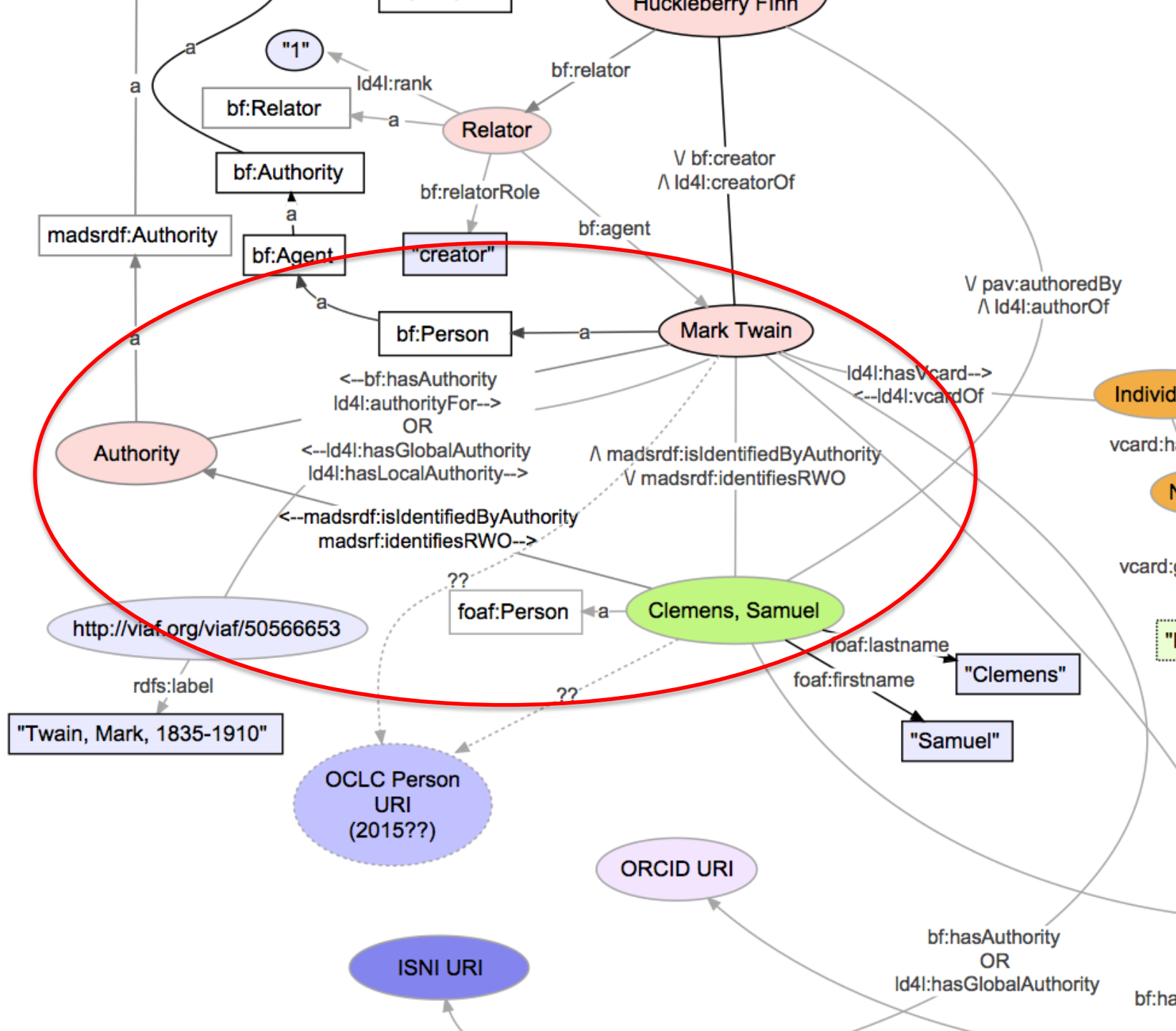- Library resources: BIBFRAME
- Additional bibliographic types and partonomy relationships: FaBiO, Music Ontology (?), Schema
- People/Organizations: VIVO-ISF (includes FOAF)
- Annotations: OpenAnnotation
- Provenance: PAV
- Virtual Collections and Structured Relationships: OAI-ORE
- Concepts: SKOS (or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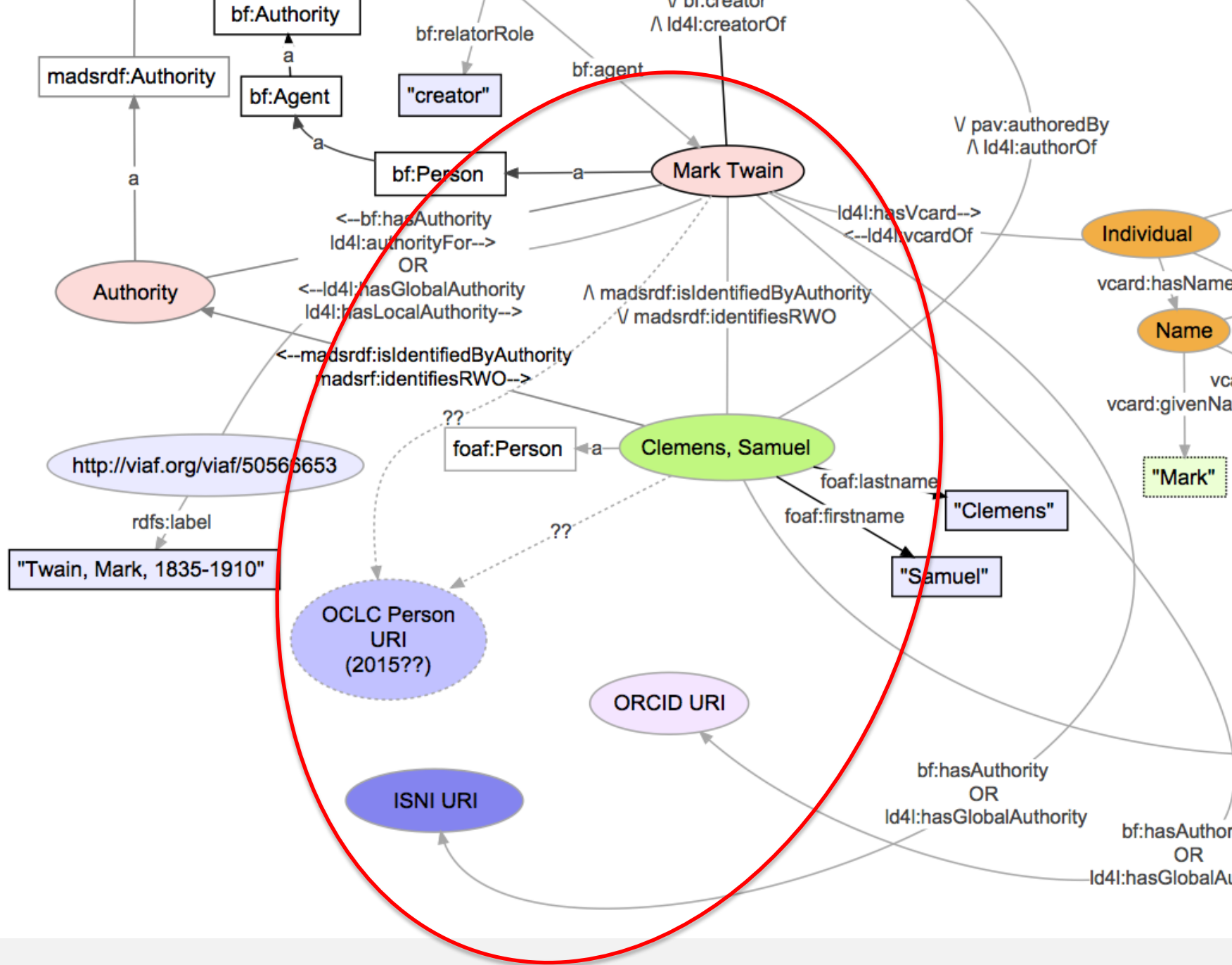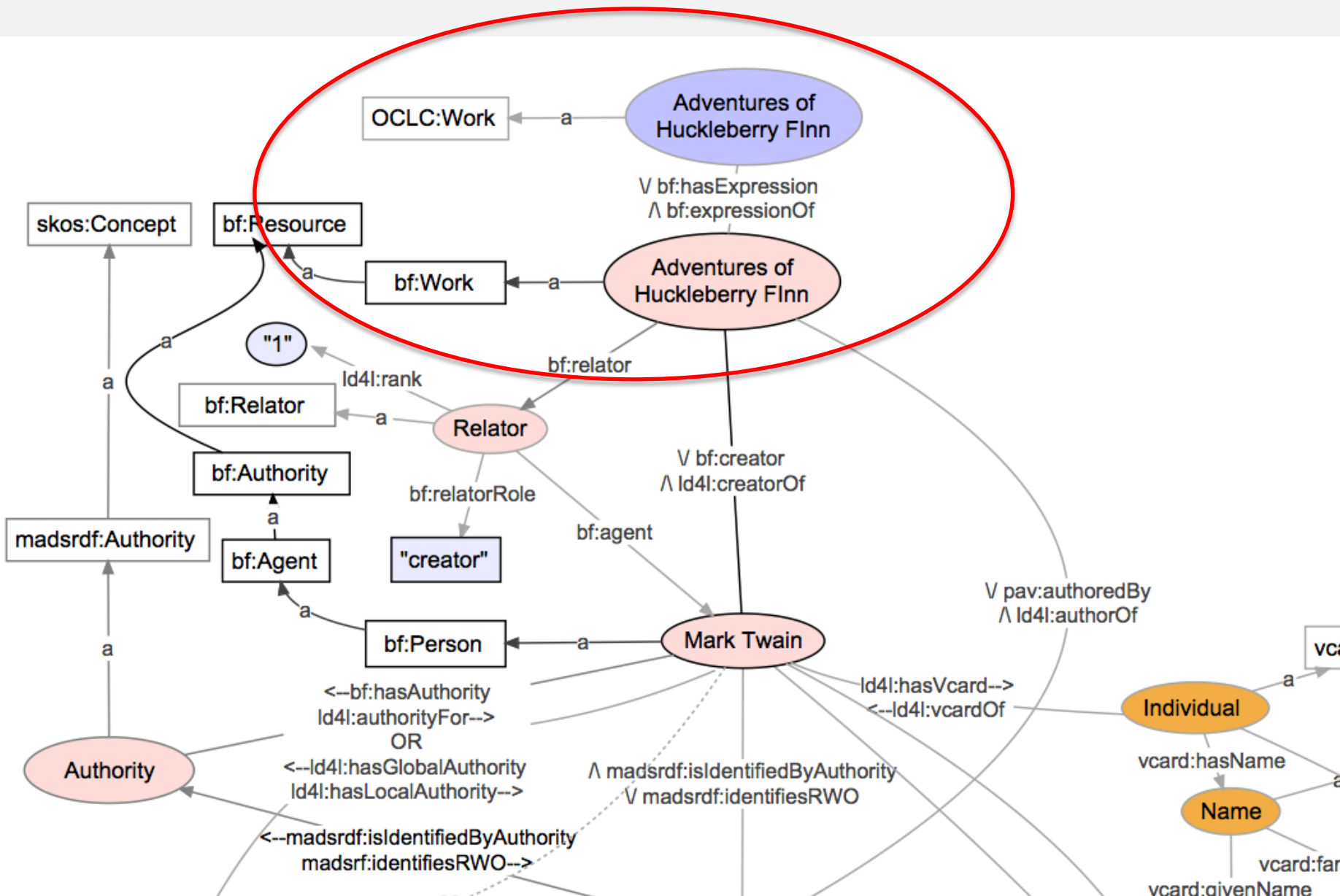- Many identifiers: VIAF, ORCID, ISNI, OCLC Works

For Illustration - NOT FINAL

(And changing, to reflect Bibframe revisions in 2015)

bf:Authority

madsrdf:Authority

bf:Agent

a

bf:Person

bf:relatorRole

"creator"

bf:agent

∨ bf:creator
∧ ld4l:creatorOf

Mark Twain

a

∨ pav:authoredBy
∧ ld4l:authorOf

ld4l:hasVcard-->
<--ld4l:vcardOf

Individual

vcard:hasName

Name

vcard:givenNa

<--bf:hasAuthority
ld4l:authorityFor-->
OR
<--ld4l:hasGlobalAuthority
ld4l:hasLocalAuthority-->

∧ madsrdf:isIdentifiedByAuthority
∨ madsrdf:identifiesRWO

Authority

<--madsrdf:isIdentifiedByAuthority
madsrf:identifiesRWO-->

??

foaf:Person

a

Clemens, Samuel

foaf:lastname

"Clemens"

"Mark"

http://viaf.org/viaf/50566653

rdfs:label

"Twain, Mark, 1835-1910"

??

foaf:firstname

"Samuel"

OCLC Person
URI
(2015??)

ORCID URI

bf:hasAuthority
OR
ld4l:hasGlobalAuthority

ISNI URI

bf:hasAuthor
OR
ld4l:hasGlobalA

Ontology challenges

- Bibframe is a moving target
  - Interpretation of a bf:Person
  - The modelling of the bf:Event
- Entity reconciliation and its processes
  - Integrating existing identifiers for people and works

THE USE CASES...

Use Cases   Status

Phase 1: Annotations (Bibliographic + curation data)

    1.1   Build a Virtual Collection

    1.1   Tag Scholarly Resources to Support Reuse

Phase 2: Authorities (Bibliographic + person data)

    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 (Leveraging external data)

    4.   Leverage the Deeper Graph

    5.   Leverage Usage Data

    6.1   Cross-Institution Discovery

Use Cases   Status

Phase 1: Annotations   Engineering

    1.1   Build a Virtual Collection

    1.1   Tag Scholarly Resources to Support Reuse

Standards

    Digital Publishing Interest Group

    EPUB standards agency has adopted Open Annotation

    Web Annotation Working Group formed

    Engagement with Linked Data Platform (LDP) Working Group

Engineering

    Stanford currently implementing full prototype system, with goal to share

    Implements Open Annotation over LDP

    Compatible with Annotator and other existing products

    Building out valuable infrastructure services (e.g. IIIF)

Use Cases   Status

Phase 1: Annotations   Engineering

    1.1   Build a Virtual Collection

    1.1   Tag Scholarly Resources to Support Reuse

Phase 2 Authorities   Planning/ Engineering

    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   Planning/ Engineering

    4.*   Leverage the Deeper Graph

    5.*   Leverage Usage Data

    6.1   Cross-Institution Discovery

Project timeline Jan-June 2015

- Pilot LD4L instances at Harvard and Stanford

- Populate Cornell LD4L instance from multiple data sources

- Develop a test instance of the LD4L Search application harvesting RDF across the three partner institutions

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
  - DEMOS!

- Obtain feedback

- 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

- Seek to expand the LD4L community & support participants in piloting this approach at their institutions

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 Hydra Component

- Create public demonstration of LD4L Search-based discovery and access system across the three LD4L instances

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, that supports Blacklight search across multiple LD4L instances

LD4L Partnership

- Cornell, Harvard, and Stanford brought together by common interest in Linked Data

- Researchers, developers, and production metadata/cataloger librarians all involved

- Will seek to expand LD4L community at workshop in February 2015