

Linked Data for Production (LD4P)

TECHNICAL SERVICES WORKFLOW EVOLUTION THROUGH TRACER
BULLETS (STANFORD PROJECTS)

Arcadia Falcone
Josh Greben
Nancy Lorimer



Introduction:

LD4P, ITS GOALS & ITS CONTEXT IN THE CURRENT
LIBRARY TECHNICAL SERVICES PARADIGM

Linked Data for Production

- Overall focus:
 - Lay the groundwork for moving library technical services workflow into a linked data environment
- Subprojects within each institution:
 - ontology development
 - tools investigation
 - workflow analysis

Stanford Projects

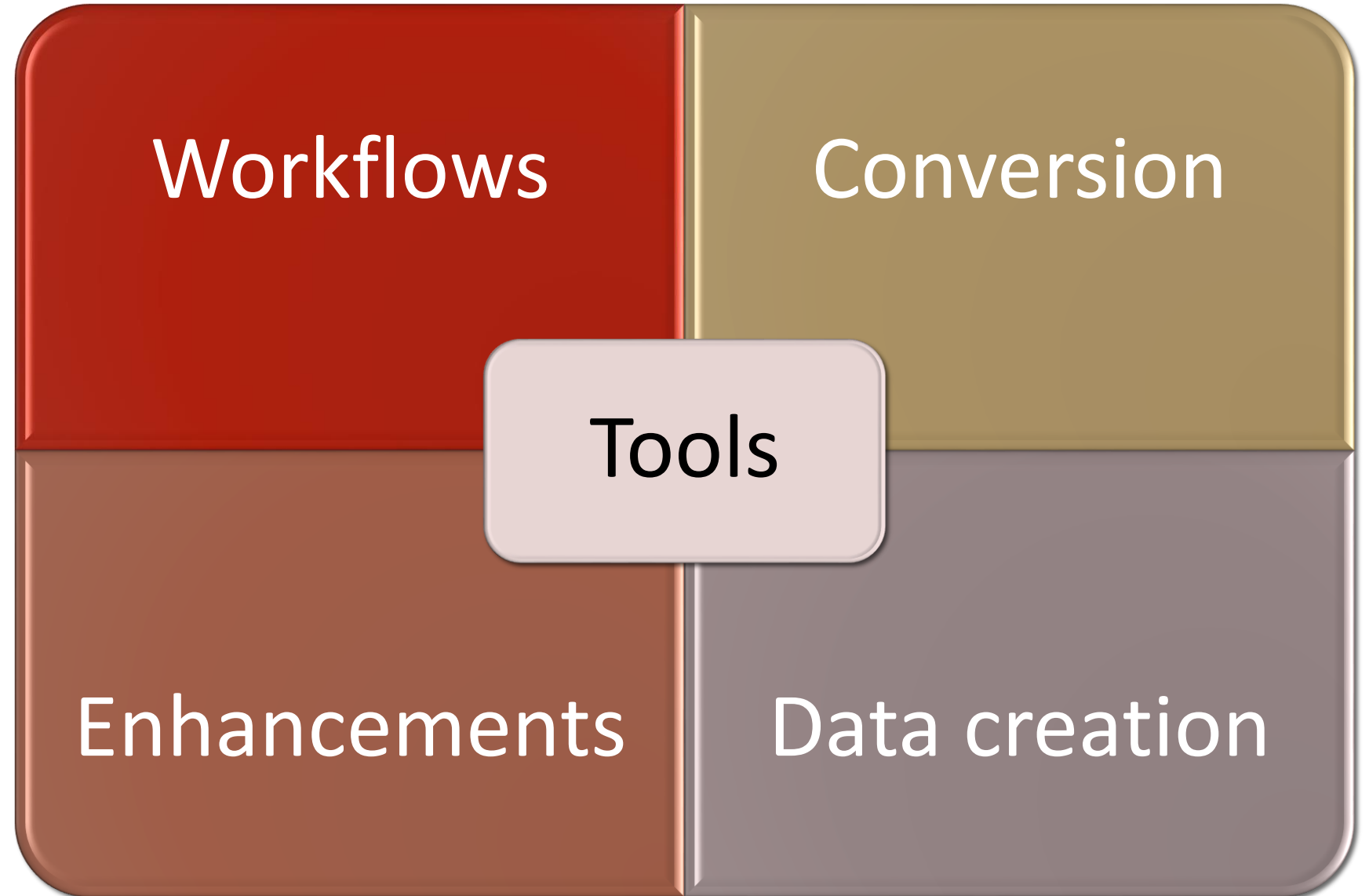
Performed Music Ontology (PMO)

- Extension to BIBFRAME 2.0

Workflows in Technical Services (“tracer bullets”)

- MARC-based workflows (vendor-supplied cataloging, original cataloging)
- digital repository workflows (individual & bulk deposit of metadata)

Themes



Workflows:

MODELING METADATA PROCESSES FOR A
HYBRID LINKED DATA ENVIRONMENT

Arcadia Falcone

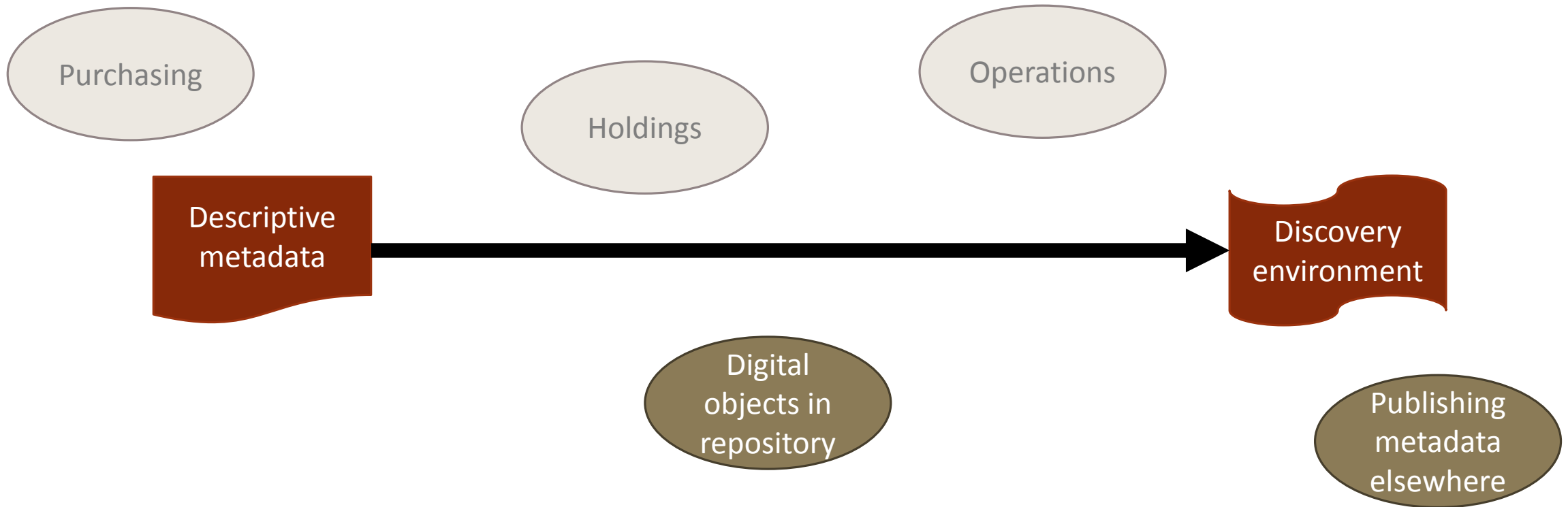


Goals

- To understand current technical services workflows both as specific tasks and generalized processes
- To model the processes of parallel linked data workflows, with their relationships to each other and to current workflows
- To begin identifying implementation specifications for systems, tools, and training

The “tracer bullet” paradigm

- Lightweight, end-to-end implementation with real data



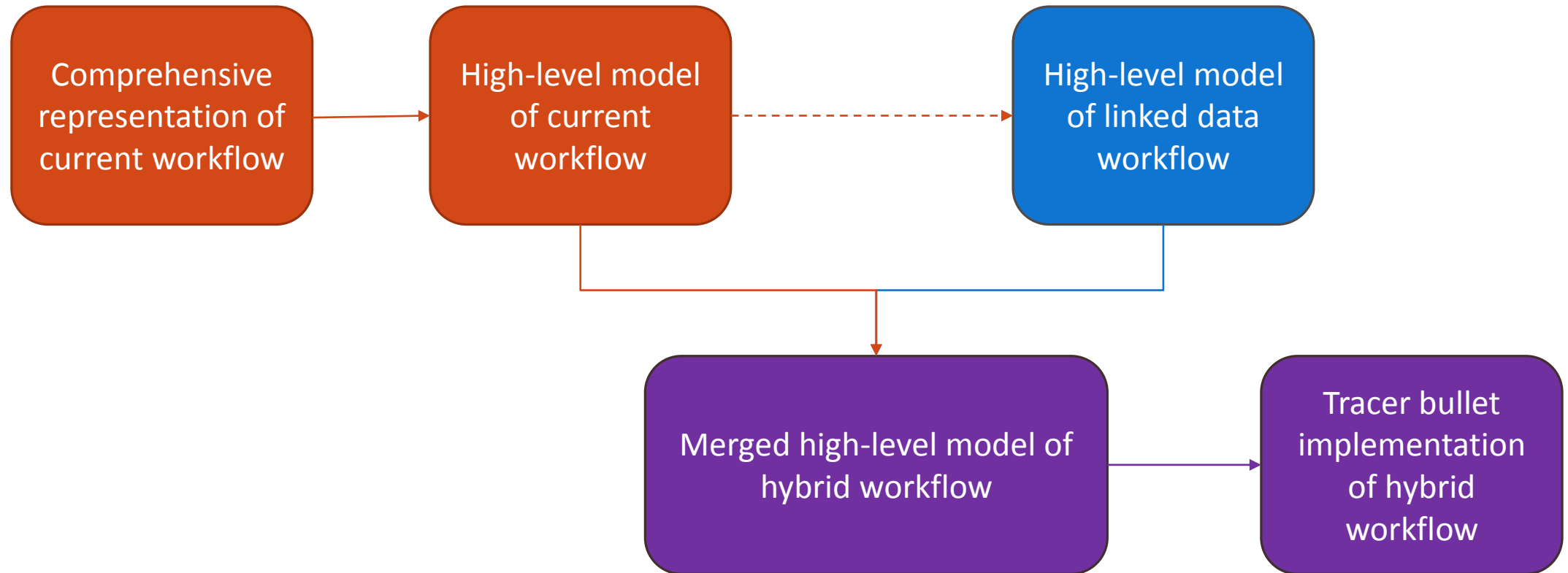
Parameters

- A hybrid environment involving MARC, MODS, BIBFRAME, and other standards will continue to exist both locally and globally
 - “Hybrid production” workflows
- The endpoint is a discovery layer that integrates MARC, MODS, and BIBFRAME data
- Processes should be scalable and require no additional human intervention beyond current workflows
- Processes should be defined so as to be modular and tool-agnostic

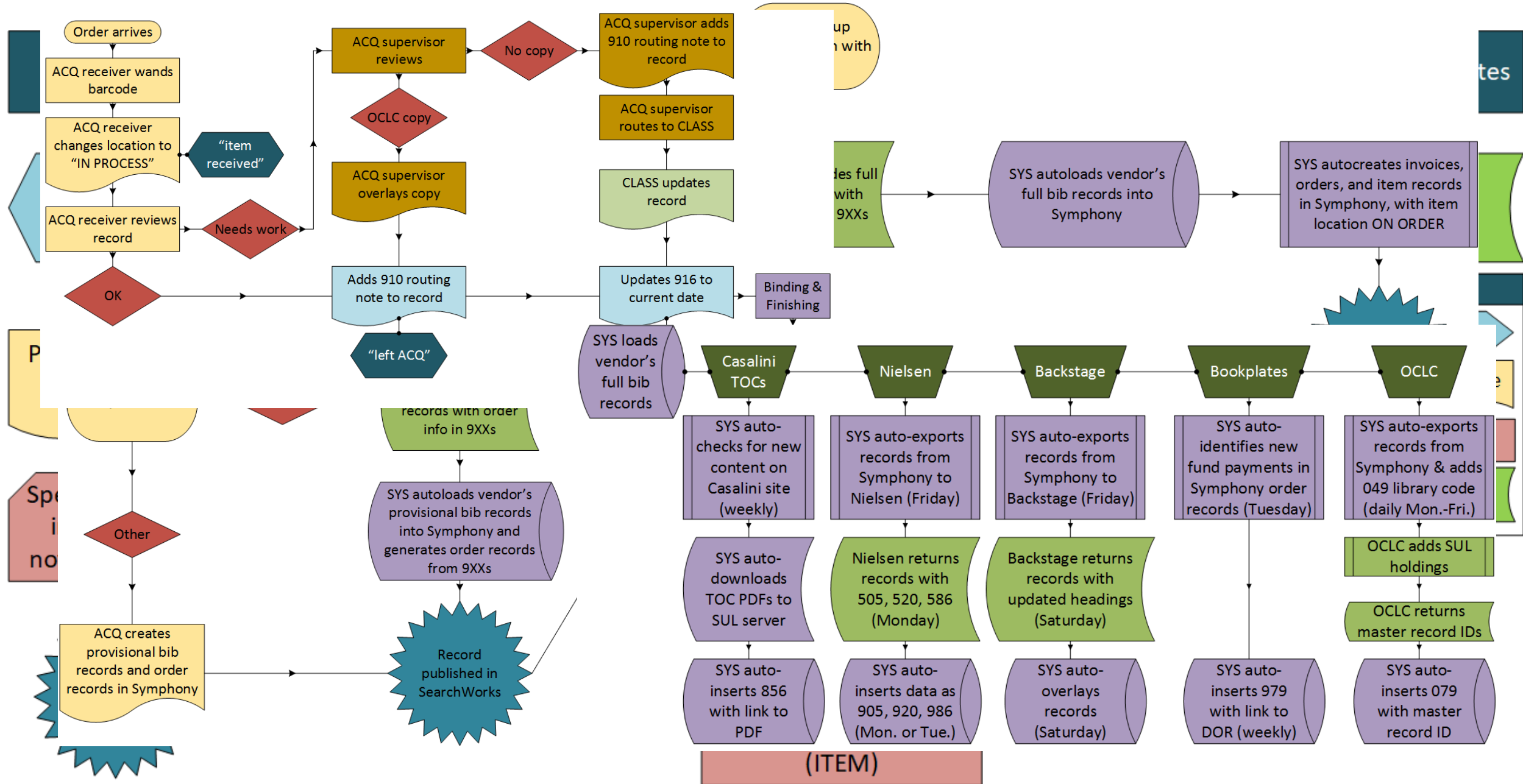
Four selected workflows

1. Vendors supply MARC records that an automated process loads into our ILS
...and into our triplestore as linked data
2. Metadata staff create original description of resources
...natively in a linked data editor
3. Users create description as part of digital object self-deposit in a web-based interface
...that is stored as linked data
4. A bulk process transforms structured metadata for a large collection of digital objects
...into linked data describing objects in our digital repository

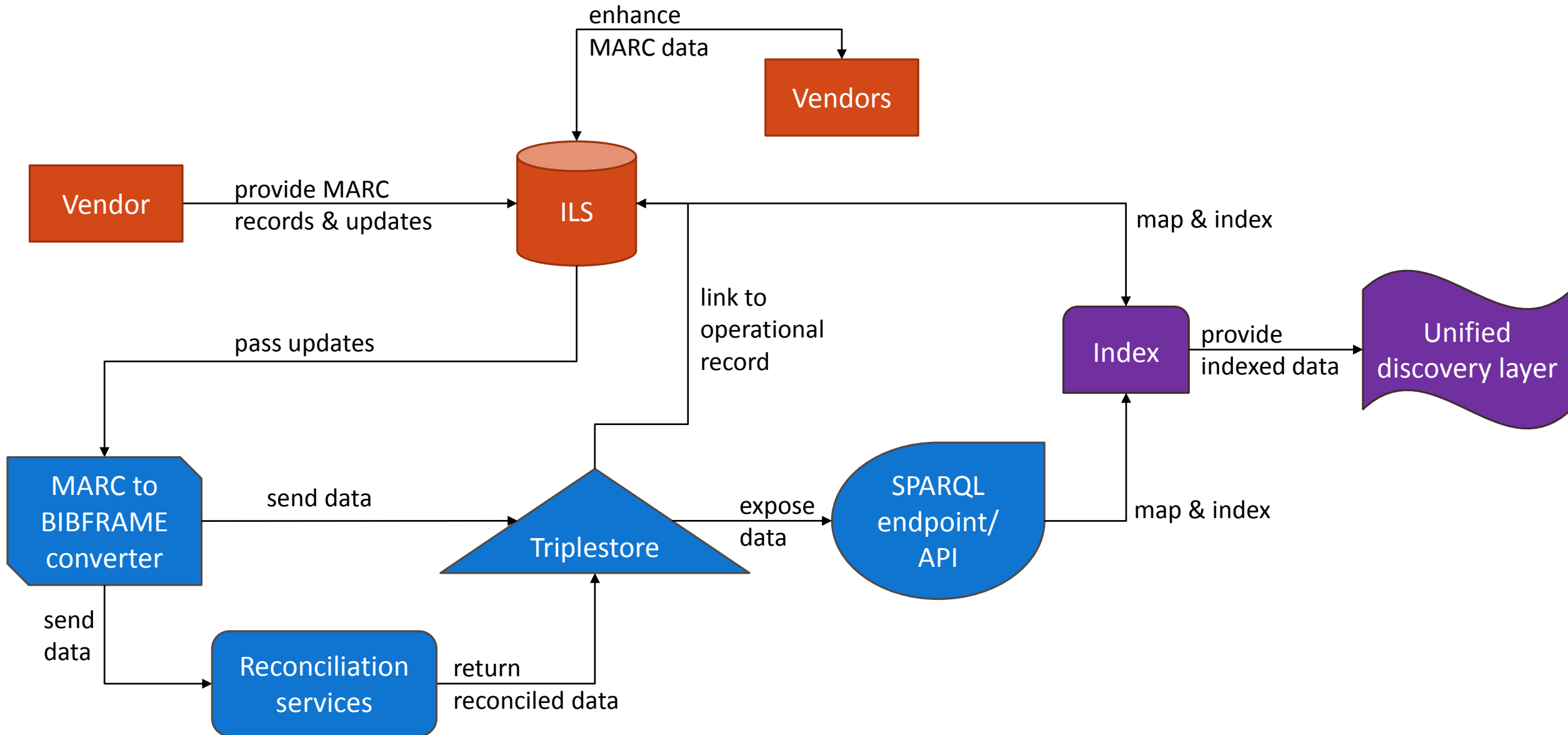
Roadmap for workflow analysis



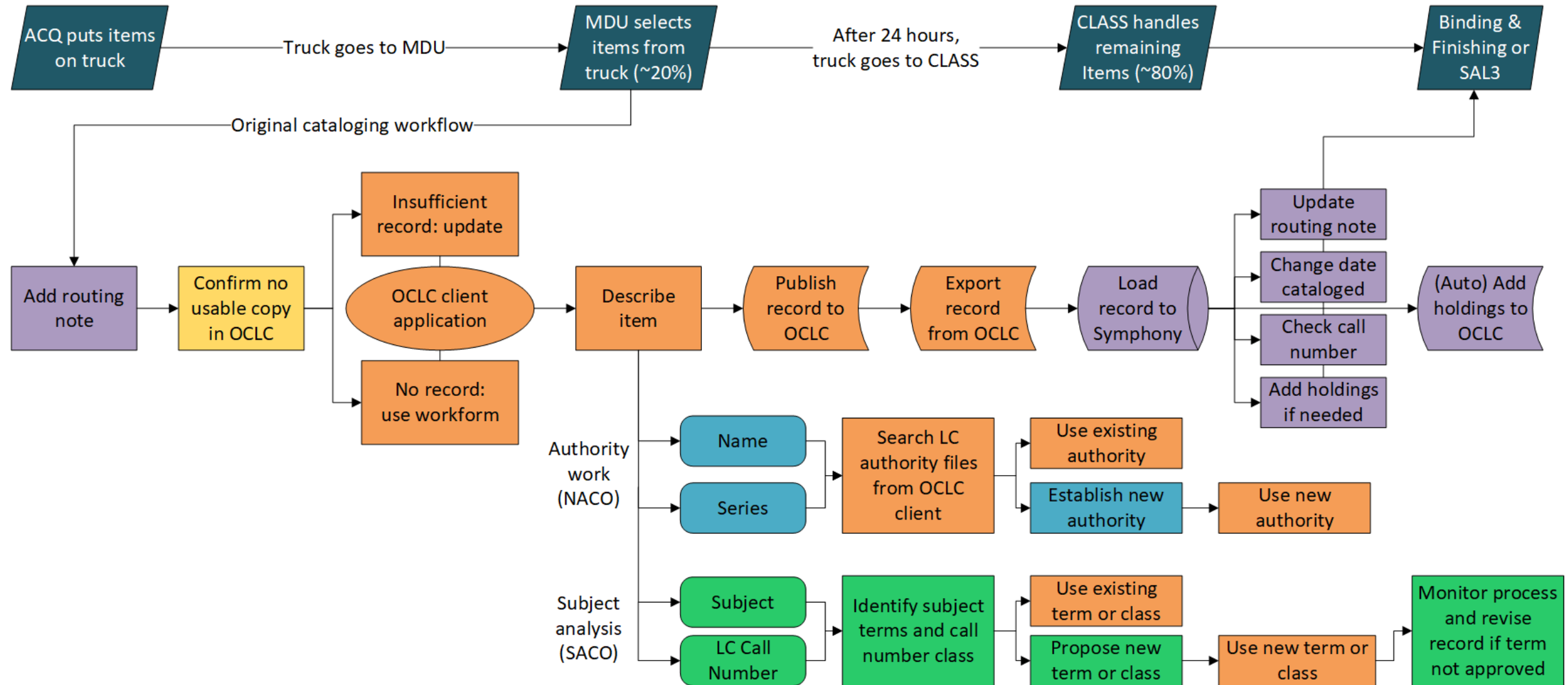
Workflow #1: task-based model



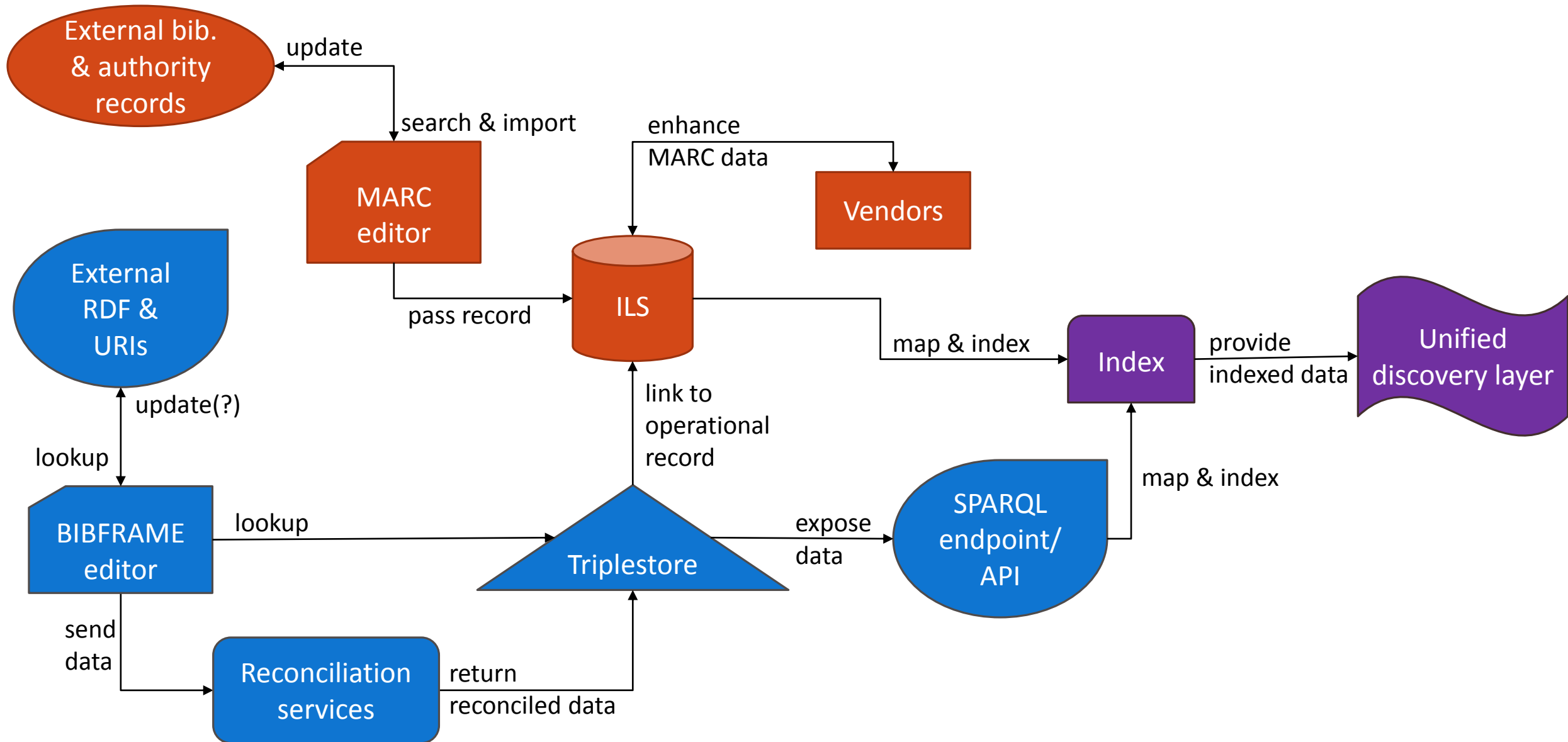
Workflow #1: process-based model



Workflow #2: task-based model



Workflow #2: process-based model



MARC TO BIBFRAME

EXPERIMENTS IN DATA ENHANCEMENT AND CONVERSION

Josh Greben
Nancy Lorimer





Ckey to Bibframe2 Conversion

Ckey for conversion:

Base URI for local namespace:

Do conversion

Main menu



Ckey to Bibframe2 Conversion

Do another conversion

MarcXML for ckey 123

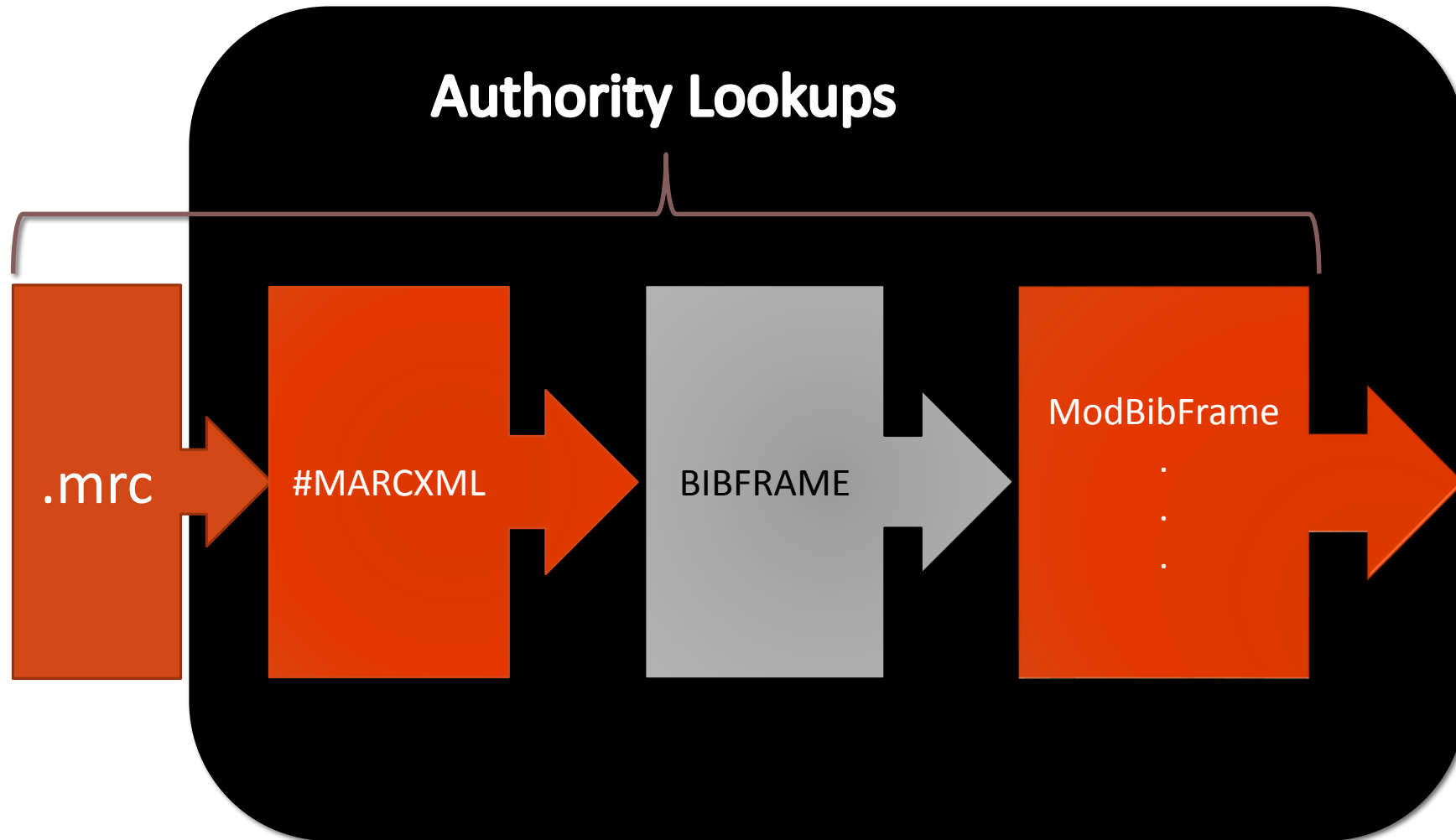
```
<?xml version="1.0" encoding="UTF-8"?>
<marcxml:collection xmlns:marcxml="http://www.loc.gov/MARC21/slim">
  <marcxml:record>
    <marcxml:leader>00924ccm a2200265 4500</marcxml:leader>
    <marcxml:controlfield tag="001">123</marcxml:controlfield>
    <marcxml:controlfield tag="003">SIRSI</marcxml:controlfield>
    <marcxml:controlfield tag="005">19900820141050.0</marcxml:controlfield>
    <marcxml:controlfield tag="008">731212|||||||fr ||| | fre </marcxml:controlfield>
    <marcxml:datafield tag="035" ind1=" " ind2=" ">
      <marcxml:subfield code="a">(OCoLC-M) 123010268</marcxml:subfield>
    </marcxml:datafield>
```

```
</marcxml:datafield>
</marcxml:record>
</marcxml:collection>
```

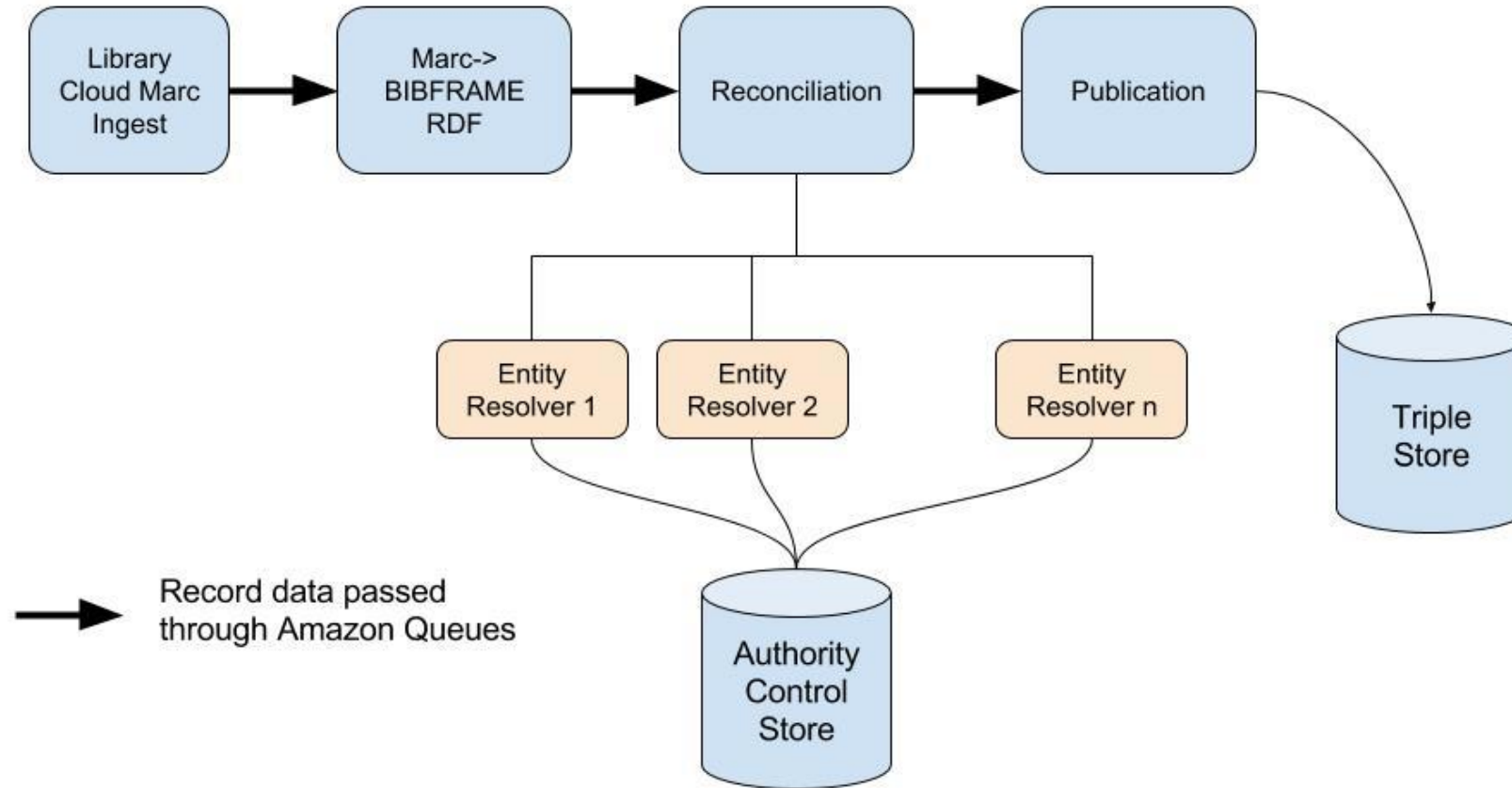
Bibframe2 for ckey 123

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:bf="http://id.loc.gov/ontologies/bibframe/" xmlns:bflc="http://id.loc.gov/ontologies/bflc/" xmlns:madsrdf="http://www.loc.gov/mads/rdf/v1#">
  <bf:Work rdf:about="http://ld4p.stanford.edu/123#Work">
    <bf:adminMetadata>
      <bf:AdminMetadata>
        <bf:generationProcess>
          <bf:GenerationProcess>
            <rdfs:label>DLC marc2bibframe2 v1.4.0-SNAPSHOT: 2017-10-16T21:05:21Z</rdfs:label>
          </bf:GenerationProcess>
        </bf:generationProcess>
      </bf:AdminMetadata>
    </bf:adminMetadata>
  </bf:Work>
</rdf:RDF>
```

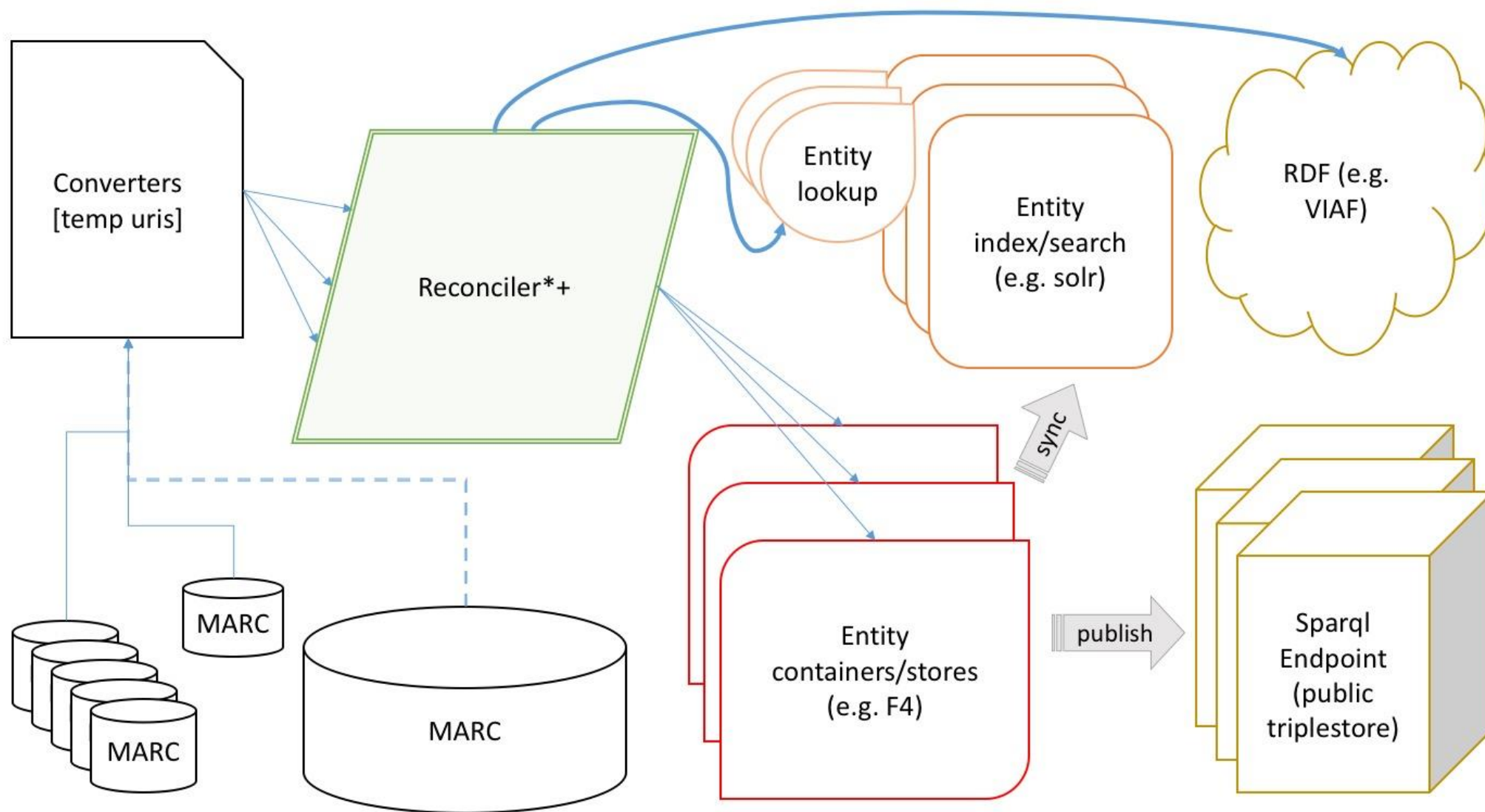
Bulk Conversion with URIs or Fingerprints



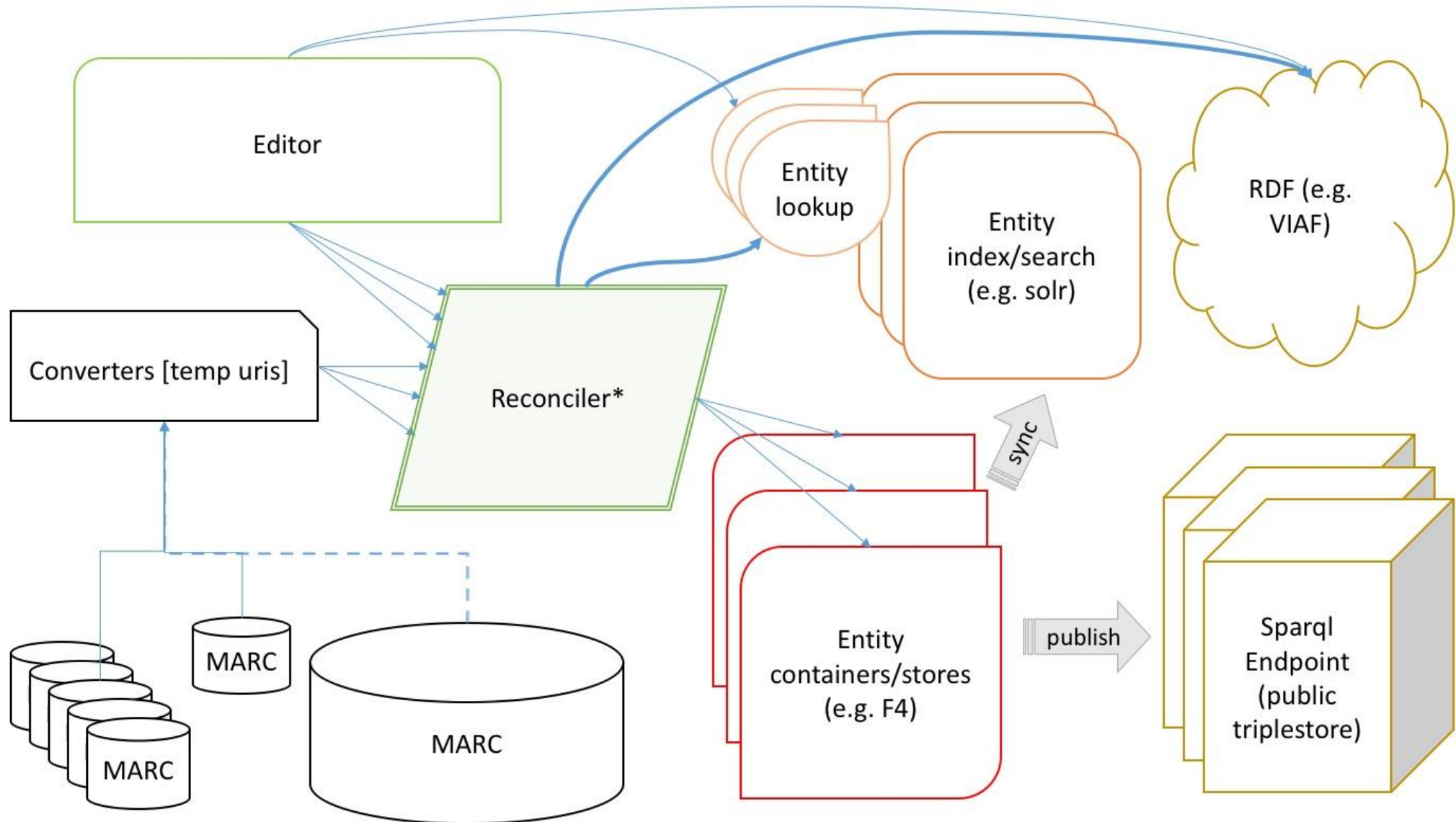
LD4L BIBFRAME Converter Pipeline



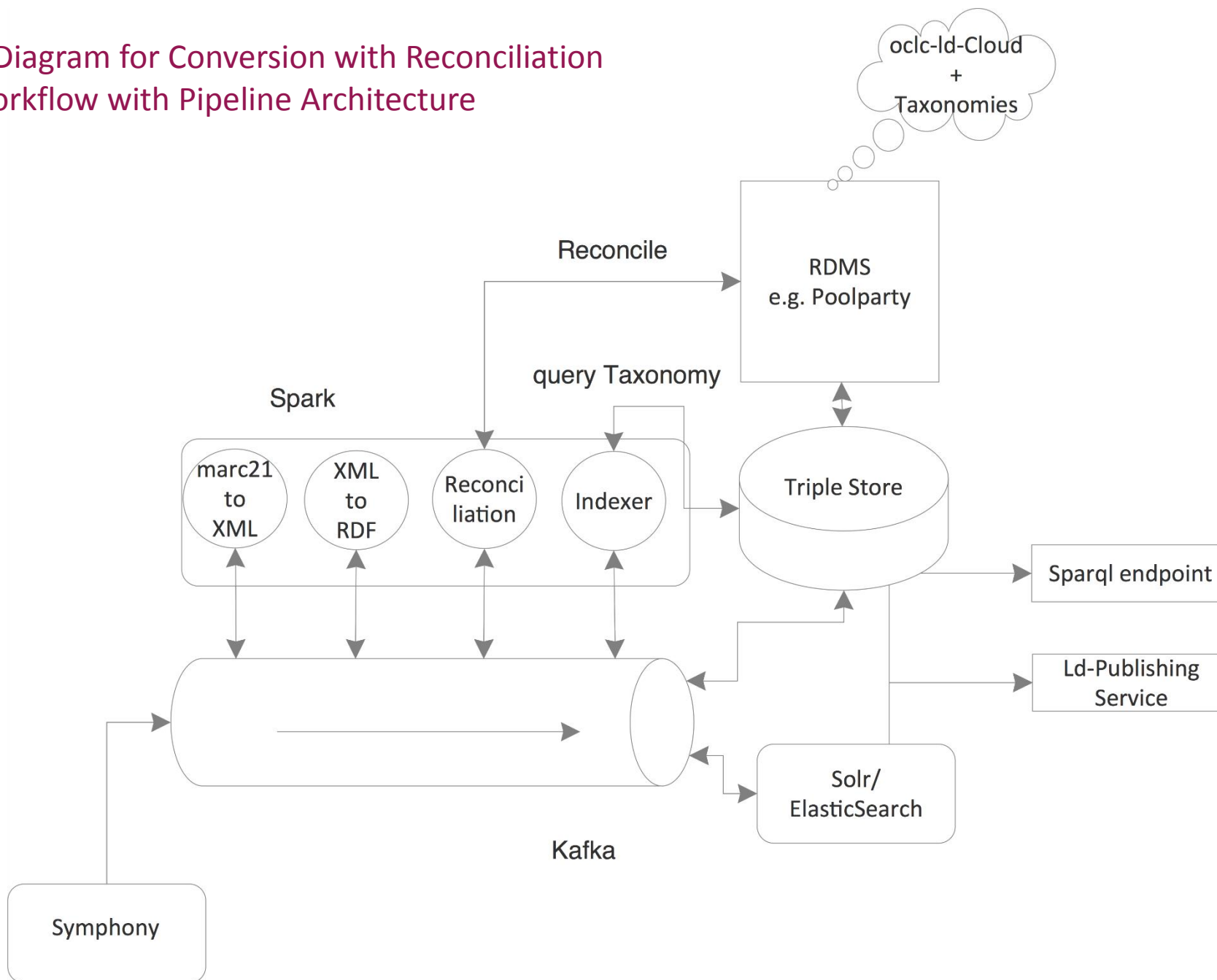
Component Straw-man for Conversion with Reconciliation Workflow



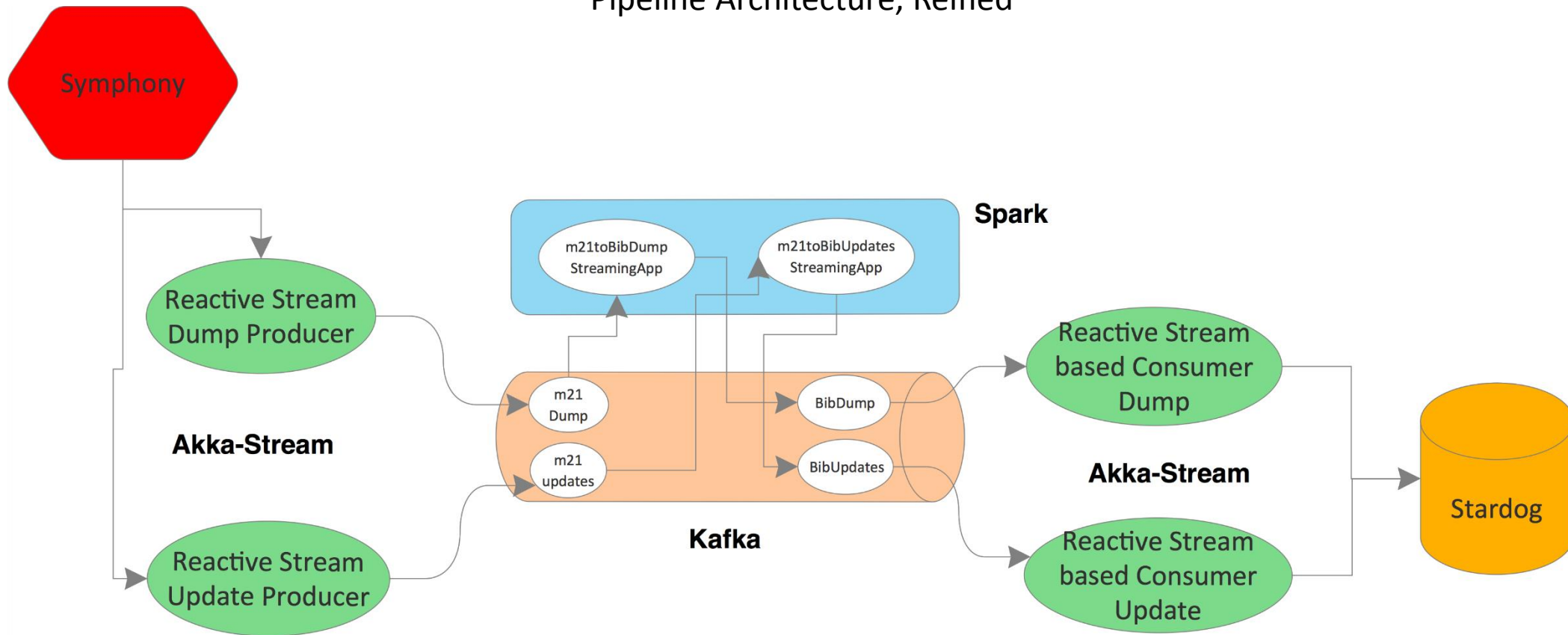
Component Straw-man for Conversion + Editor + Reconciliation Workflow



Component Diagram for Conversion with Reconciliation Workflow with Pipeline Architecture








Component Diagram for Conversion Workflow with Pipeline Architecture, Reified



-Reactive Stream based microservices build with **akka Stream** ensure a fast ingestion or digestion by performing an optimal use of the resources of the machines in term of concurrent. They implement the Reactive stream protocol that is based on a push and pull model (asynchronous non blocking and back-pressure)

BIBFRAME 2 to Solr Mapping

Solr Index Field	Description	LC MARC>BF2 Mapping Spec	Bibframe2
id 	ckey		derived from BF2:instance identifier?
all_search	all searchable text		value of all the labels associated with ..
collection 	constant: "bf2"		-
Title fields			
title_245a_search 	245a	mainTitle (I)	BF2:mainTitle?
title_245_search 	245abfgknps	approx: rdfs:label (I); does not include \$b approx: mainTitle (I) + subtitle (I) + originDate (W) + partNumber (I) + partName (I); both \$f and \$g map to originDate; \$k not mapped	
title_245a_display 	short title, without trailing punctuation \,/,: ??	mainTitle (I)	

BIBFRAME 2 SPARQL Queries

Gets the "main title" of an instance; title_245a_search

```
PREFIX bf: <http://id.loc.gov/ontologies/bibframe/>
SELECT ?o
WHERE {
  <http://ld4p-test.stanford.edu/6832810#Instance> bf:title ?t .
  ?t bf:mainTitle ?o .
  FILTER NOT EXISTS {
    ?t rdfs:subClassOf* bf:VariantTitle.
  }
}

select distinct ?s ?title ?titleType ?titleLabel
where {
  ?s a bf:Instance .
  ?s bf:title ?title .
  ?title a ?titleType .
  ?title bf:mainTitle ?titleLabel .
  Filter Not Exists {
    ?titleType rdfs:subClassOf* bf:VariantTitle.
  }
}
```

Conversion Questions

- URIs—where do you get them?
- are there other enhancements you can do?
- granularity of conversion
- adding local field conversions to a more generic converter
 - converter maintenance
- compatibility with other conversions and original metadata creation

Getting URIs

- BACKSTAGE LIBRARY WORKS
 - Providing LC-NAR, VIAF, ISNI URIs for a few years now in authority records
 - Recently began adding selected URIs directly in bib records
- SHARE-Virtual Discovery Environment
 - Has taken converted our entire bib file
 - Can convert MARC to BIBFRAME, and soon MODS to BIBFRAME
 - Has ability to reconcile at basic and enhanced levels

DATA CREATION

REQUIREMENTS & TOOLS

Josh Greben
Nancy Lorimer



The Bibframe Editor...

Needs prefabricated triples (i.e. profiles) and a way to apply them to your work

Needs a place to temporarily remember the data

- Memory store
- Loopback API Server (a la loopback.io)

Needs a way to fetch changes made to profiles

- Profile-edit server with http file endpoint
- Trigger file download

Needs a way to do lookups to id.loc.gov and other sources

- Cross-domain Scripting

Needs a way and a place to permanently store the triples data

- Reformat JSON to suit needs of posting to triplestore

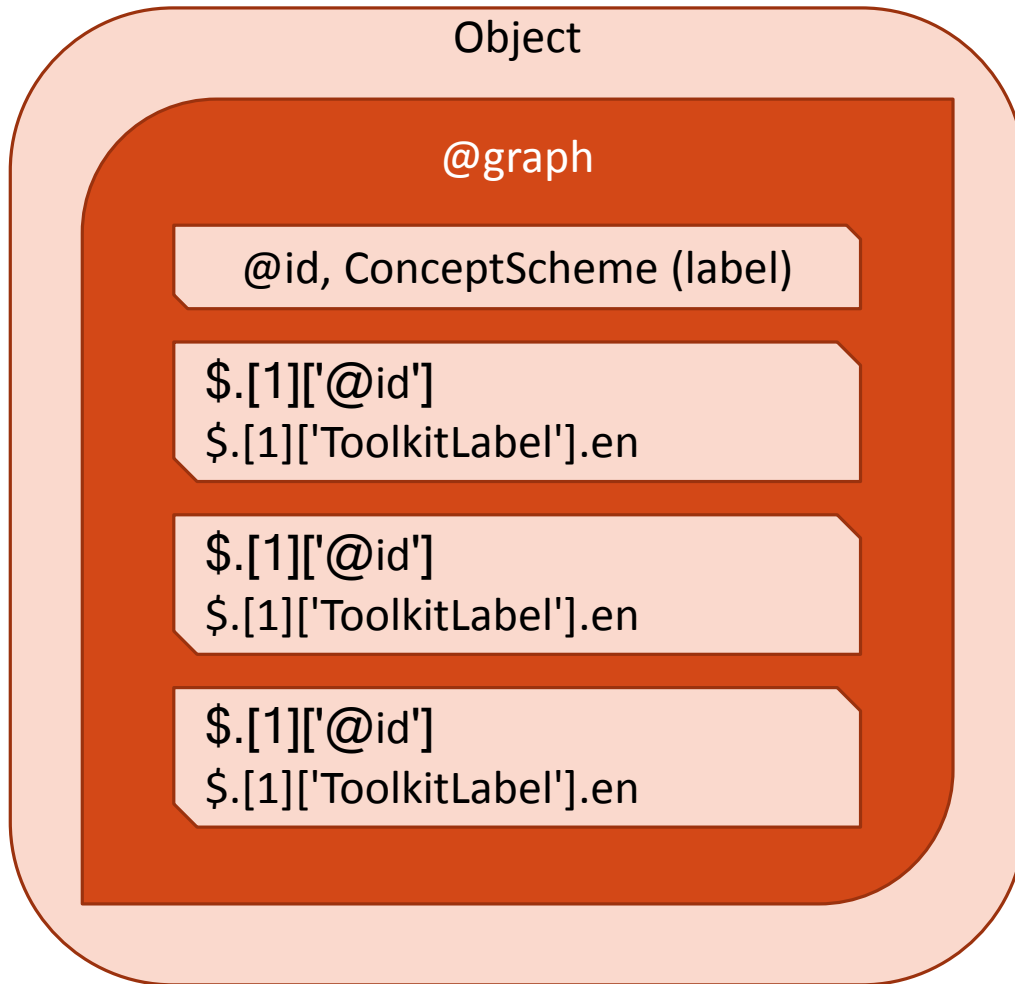
Needs a way to handle Reconciliation...

Lookups: LOC Suggest API

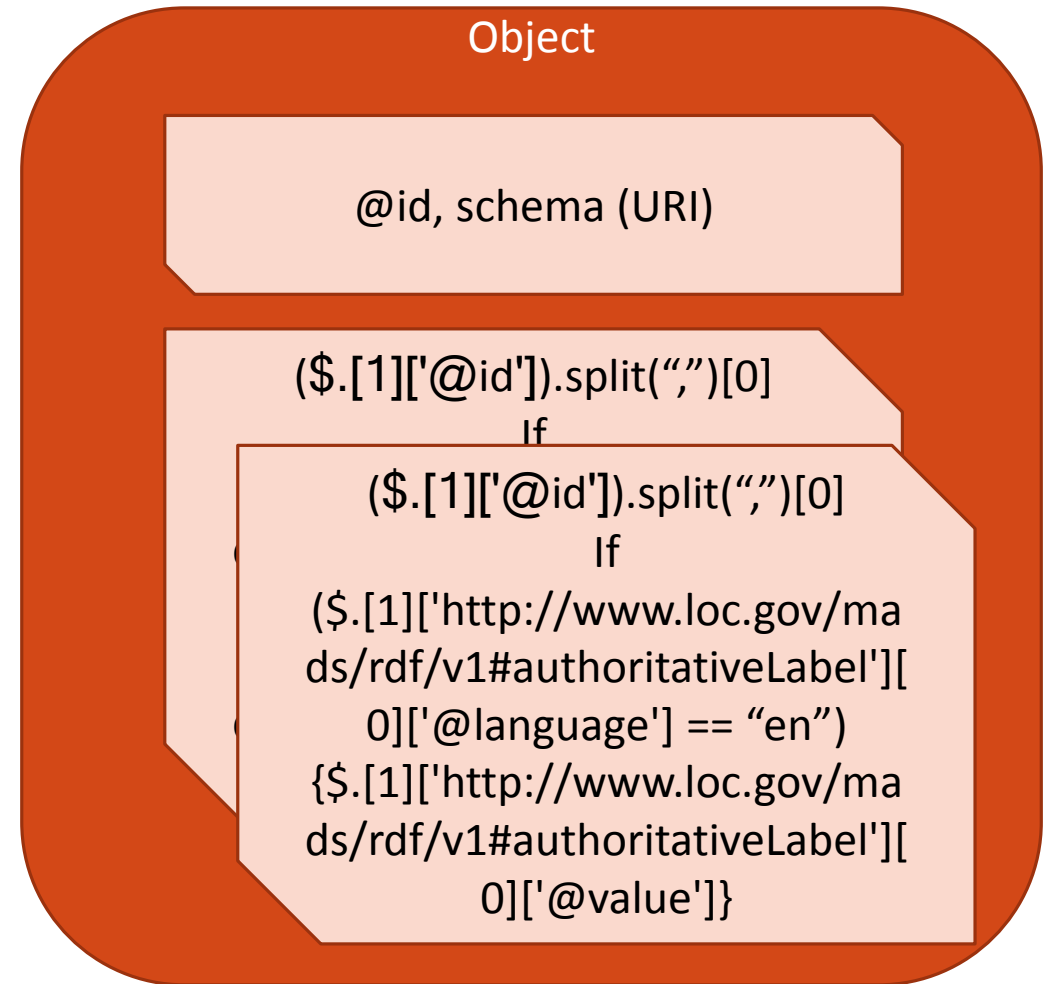
<http://id.loc.gov/authorities/performanceMediums/suggest/?ensemble>

```
[["Appalachian dulcimer","Baltic psaltery","Baroque lute","English guitar","English horn","Hardanger fiddle","Irish harp","Jew's harp","MIDI controller","Native American flute"],["1 result","1 result","1 result","1 result","1 result","1 result","1 result","1 result","1 result"],["http://id.loc.gov/authorities/performanceMediums/mp2013015022","http://id.loc.gov/authorities/performanceMediums/mp2013015373","http://id.loc.gov/authorities/performanceMediums/mp2013015059","http://id.loc.gov/authorities/performanceMediums/mp2013015250","http://id.loc.gov/authorities/performanceMediums/mp2013015251","http://id.loc.gov/authorities/performanceMediums/mp2013015321","http://id.loc.gov/authorities/performanceMediums/mp2013015356","http://id.loc.gov/authorities/performanceMediums/mp2013015360","http://id.loc.gov/authorities/performanceMediums/mp2013015474","http://id.loc.gov/authorities/performanceMediums/mp2013015495"]]
```

Lookups: rdaregistry.info & id.loc.gov getting ID and English label



<http://rdaregistry.info/termList/RDAproductionMethod.jsonld>



<http://id.loc.gov/authorities/performanceMediums.json>

RDF to TripleStore (BFE Produced)

```
{  
  "created": "2017-07-28T15:30:57.000Z",  
  "id": 4,  
  "modified": "2017-07-28T15:30:57.000Z",  
  "name": "HAr1501255857",  
  "profile": "http://localhost:8000/bf/static/profiles/bibframe/BIBFRAME 2.0 Serial.json",  
  "rdf": [  
    {}, {}, {}  
  ],  
  "url": "https://ld4p-loc-bfe-dev.stanford.edu/verso/api/bfs/HAr1501255857"  
}
```


RDF to TripleStore (JSON-LD)

```
{
  "@context": {
    "created": "2017-07-28T15:30:57.000Z",
    "id": 4,
    "modified": "2017-07-28T15:30:57.000Z",
    "name": "HAr1501255857",
    "profile": "http://localhost:8000/bf/static/profiles/bibframe/BIBFRAME 2.0 Serial.json",
    "url": "https://ld4p-loc-bfe-dev.stanford.edu/verso/api/bfs/HAr1501255857"
  },
  "@graph": [
    {}, {}, {}
  ]
}
```

PMO Sound Recording

[← Cancel](#)[Save](#)[Delete](#)

Profile

Performance

[≡ Change Resource](#)

BIBFRAME Work (RDA Work Elements)

[≡ Change Resource](#)

BIBFRAME Work (RDA Expression Elements)

[≡ Change Resource](#)

BIBFRAME Instance (RDA Manifestation)

[≡ Change Resource](#)

BIBFRAME Item (RDA Item)

[≡ Change Resource](#)

Performance name

[≡ Change Resource](#)[+ Add Resource Template](#)[Verbose Export](#)[Brief Export](#)

ID Resource URI Resource Label Contact Guiding statement for the use of this resource

▶ Lookup	 Delete Lookup	≡ Change Property
▶ Creator of Work (RDA 19.2)	 Delete Creator of Work (RDA 19.2)	≡ Change Property
▶ Title Information (Title Information (RDA 6.14.2, RDA 6.14.3))		≡ Change Property
	 Delete Title Information (Title Information (RDA 6.14.2, RDA 6.14.3))	
▶ Form of Work (RDA 6.3)	 Delete Form of Work (RDA 6.3)	≡ Change Property
▶ Date of Work (RDA 6.4)	 Delete Date of Work (RDA 6.4)	≡ Change Property
▶ Place of Origin of the Work (RDA 6.5)	 Delete Place of Origin of the Work (RDA 6.5)	≡ Change Property
▶ Other Distinguishing Characteristics of the Work (RDA 6.6)		≡ Change Property
	 Delete Other Distinguishing Characteristics of the Work (RDA 6.6)	
▶ Numerical Designation of a Musical Work (RDA 6.16)	 Delete Numerical Designation of a Musical Work (RDA 6.16)	≡ Change Property
▶ Medium of Performance	 Delete Medium of Performance	≡ Change Property
▶ Opus number statement (RDA 6.16.1.3.2)		≡ Change Property
▶ Thematic catalog statement	 Delete Opus number statement (RDA 6.16.1.3.2)	
▶ Key and mode (RDA 6.17)	 Delete Thematic catalog statement	≡ Change Property
▶ Music mode	 Delete Key and mode (RDA 6.17)	≡ Change Property
▶ Pitch center	≡ Change Property  Delete Music mode	≡ Change Property
▶ Nature of the Content (RDA 7.2)		≡ Change Property  Delete Pitch center
▶ (Geographic) Coverage of the Content (RDA 7.3)	≡ Change Property  Delete Nature of the Content (RDA 7.2)	
▶ (Temporal) Coverage of the Content (RDA 7.3)	 Delete (Geographic) Coverage of the Content (RDA 7.3)	
	 Delete (Temporal) Coverage of the Content (RDA 7.3)	≡ Change Property
▶ Intended Audience (RDA 7.7)	 Delete Intended Audience (RDA 7.7)	≡ Change Property
▶ Other Person, Family, or Corporate Body Associated With a Work (RDA 19.3)		≡ Change Property
	 Delete Other Person, Family, or Corporate Body Associated With a Work (RDA 19.3)	
▶ Subject of the Work (RDA Chapter 23)		≡ Change Property

BIBFRAME Instance (RDA Manifestation)

Instance of				
Title Information (RDA 2.3)	Instance Title			
Statement of Responsibility Relating to Title Proper (RDA 2.4.2)	Statement of Responsibility Relating to Title Proper (RDA 2.4.2)			+
Edition Statement (RDA 2.5)	Edition Statement (RDA 2.5)			+
Publication, Distribution, Manufacture Statements (RDA 2.8-2.10)	Publication Activity	Distribution Activity	Manufacture Activity	
Copyright Date (RDA 2.11)	Copyright Date (RDA 2.11)			+
Series Statement (RDA 2.12)	Series Statement (RDA 2.12)			+
Mode of Issuance (RDA 2.13)	Mode of issuance			
Issue, Distributor Number(s) (RDA 2.15)	Sound recording issue number	Music Distributor number		
Media type (RDA 3.2)	Media type			
Carrier (RDA 3.3)	Carrier type			
Extent (RDA 3.4)	Extent			
Dimensions (RDA 3.5)	Dimensions (RDA 3.5)			+
Base Material (RDA 3.6), Applied Material (RDA 3.7)	Base material (RDA 3.6)	Applied material (RDA 3.7)		
Type of Recording (RDA 3.16.2)	Type of recording (RDA 3.16.2)			
Playing Speed (RDA 3.16.4)	Playing speed (RDA 3.16.4)			
Recording Medium (RDA 3.16.3)	Recording medium (RDA 3.16.3)			
Disc characteristics (Groove, cutting) (RDA 3.16.5-3.16.6)	Groove characteristics (RDA 3.16.5)	Disc Cutting Technique		
Tape configuration (RDA 3.16.7)	Tape configuration (RDA 3.16.7)			
Configuration of Playback Channels (RDA 3.16.8)	Configuration of Playback Channels (RDA 3.16.8)			
Special Playback Characteristic (RDA 3.16.9)	Special Playback Characteristic (RDA 3.16.9)			
File type, encoding format, file size, bitrate (RDA 3.19)	File type (RDA 3.19.2)	Encoding format (RDA 3.19.3)	File size (RDA 3.19.4)	Encoded bitrate (RDA 3.19.7)

BIBFRAME Instance (RDA M

Title Information

Statement of Responsibility Relating to T

Edition Statement

Publication, Distribution, Manufacture Statements (RDA 2.8-2.10)

Copyright Date (RDA 2.11)

Series Statement (RDA 2.12)

Mode of Issuance (RDA 2.13)

Issue, Distributor Number(s) (RDA 2.15)

Media type (RDA 3.2)

Carrier (RDA 3.3)

Extent (RDA 3.4)

Dimensions (RDA 3.5)

Base Material (RDA 3.6), Applied Material (RDA 3.7)

Type of Recording, Recording Medium, Playing Speed (RDA 3.16.2-3.16.4)

Media type

Media type (RDA 3.2)

Media type (RDA 3.2)

RDA-Media-Types

audio
computer
microform
microscopic
other
projected
stereographic
unmediated
unspecified
video

Cancel

Save changes

Publication

Copyright D

Series Stat

Mode of iss

Sound recording issue number

Music Distributor number

Media type

Carrier type

Extent

Dimensions (RDA 3.5)

Base material (RDA 3.6)

Applied material (RDA 3.7)

Type of recording (RDA 3.16.2)

Recording medium (RDA 3.16.3)

Playing speed (RDA 3.16.4)

BioPortal/BiblioPortal

- repository of biomedical ontologies
- provides
 - ontology summaries & histories
 - viewing statistics
 - ontology details—classes & properties in hierarchies
 - mapping ability
- new “slice” called BiblioPortal
 - are working to make it a more independent portal

Submit New Ontology

Entry Type

- ☒ **Ontology** (9)
- ☐ **Ontology View** (0)
- ☐ **CIMI Model** (0)
- ☐ **NLM Value Set** (0)

Uploaded in the Last

Category

- ☐ **All Organisms** (0)
- ☐ **Anatomy** (0)
- ☐ **Animal Development** (0)
- ☐ **Animal Gross Anatomy** (0)
- ☐ **Arabidopsis** (0)
- ☐ **Biological Process** (0)
- ☐ **Biomedical Resources** (0)

Group

- ☒ **BIBLIO** (9)
- ☐ **BIS** (3)
- ☐ **CGIAR** (1)

BIBFRAME 2.0 (BIBFRAME)

Initiated by the Library of Congress, BIBFRAME provides a foundation for the future of bibliographic description, both on the web, and in the broader networked world

Uploaded: 6/8/17

projects
1

classes
188

schema.org (SCHEMA)

A collection of schemas that webmasters can use to markup HTML pages in ways recognized by major search providers, and that can also be used for structured data interoperability (e.g

Uploaded: 5/19/17

classes
780

Dublin Core (DC)

The Dublin Core Schema is a small set of vocabulary terms that can be used to describe several kinds of resources

Uploaded: 2/17/17

classes
11

CIDOC Conceptual Reference Model (CIDOC-CRM)

The CIDOC Conceptual Reference Model (CRM) provides definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation

Uploaded: 8/31/17

classes
85

Web Annotation Ontology (OA)

classes
25

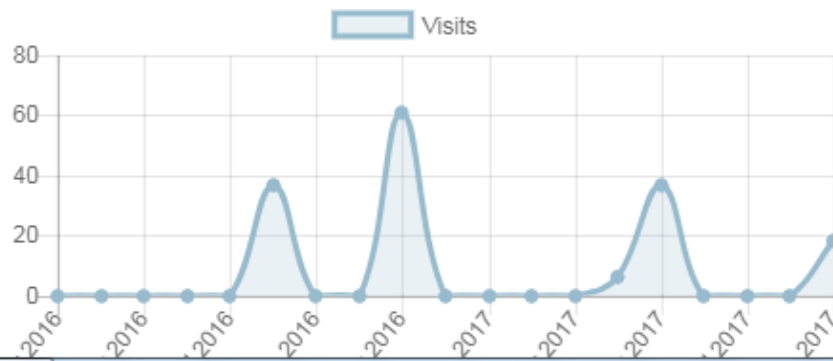
Details

ACRONYM	BIBFRAME
VISIBILITY	Public
BIOPORTAL PURL	http://purl.bioontology.org/ontology/BIBFRAME
DESCRIPTION	Initiated by the Library of Congress, BIBFRAME provides a foundation for the future of bibliographic description, both on the web, and in the broader networked world. In addition to being a replacement for MARC, BIBFRAME serves as a general model for expressing and connecting bibliographic data. A major focus of the initiative will be to determine a transition path for the MARC 21 formats while preserving a robust data exchange that has supported resource sharing and cataloging cost savings in recent decades.
STATUS	Production
FORMAT	OWL
CONTACT	John Graybeal, jgraybeal@stanford.edu
HOME PAGE	https://www.loc.gov/bibframe/
PUBLICATIONS PAGE	
DOCUMENTATION PAGE	https://www.loc.gov/bibframe/docs/index.html
CATEGORIES	Other, Upper Level Ontology
GROUPS	Bibliographic Materials Group

Metrics ?

NUMBER OF CLASSES:	188
NUMBER OF INDIVIDUALS:	0
NUMBER OF PROPERTIES:	195
MAXIMUM DEPTH:	2
MAXIMUM NUMBER OF CHILDREN:	75
AVERAGE NUMBER OF CHILDREN:	9
CLASSES WITH A SINGLE CHILD:	3
CLASSES WITH MORE THAN 25 CHILDREN:	2
CLASSES WITH NO DEFINITION:	2

Visits [Download as CSV](#)



BIBFRAME 2.0

Summary Classes Properties Notes Mappings Widgets

- Administrative metadata
- Applied material
- Arrangement of material
- Aspect ratio
- Assigner
- Associated agent
- Award note
- Base material
- Book format
- Capture of content
- Carrier type
- Cartographic ascension and declination
- Cartographic coordinates
- Cartographic data
- Cartographic equinox
- Cartographic G ring area excluded
- Cartographic outer G ring area covered
- Cartographic projection
- Classification
- Classification designation
- Classification item number
- Classification number
- Classification number span end
- Classification scheme edition
- Classification table identification
- Classification table sequence number
- Code
- Color content

Property Details

Labels

Book format

Definitions

Result of folding a printed sheet to form a gathering of leaves.

ID

<http://id.loc.gov/ontologies/bibframe/bookFormat>

CEDAR



CEDAR = The Center for Expanded Data Annotation and Retrieval

Mission: CEDAR will develop information technologies that make authoring complete metadata much more manageable, and that facilitate using the metadata in further research.

Elements:

- Interfaces and tools built and tested specifically for metadata creation
- Consistency in terminology
- Machine learning
- Editing capabilities
- Training and outreach
- Building on past work and leveraging ongoing collaborations

a

Type of resource

?

Contributor(s)

Contributor

Type of Contributor

?

Name

?

Contact

?

Affiliation

Type of entity

?

Affiliation

?

ISNI or ORCID

ISNI or ORCID

?

Identifier value

?

Role

Role

?

a

1

11

#

...

Q

The logo for CEDAR, featuring a stylized green tree icon to the left of the word "CEDAR" in a blue, serif, all-caps font.

Template

Find a property in BioPortal or [Enter Property URI](#)

[Start Over](#)

Search in BioPortal

Affiliation



10 results for the query 'Affiliation'. Click on a property below to select it

PROPERTY	DEFINITION	TYPE	SOURCE	ID
affiliation	An organization that this person is affiliated with. For example, a school/university, a club, or a team.	Annotation Property	SCHEMA	affiliation
affiliation	-	Datatype Property	CANCO	affiliation
affiliation	-	Annotation Property	DCAT	affiliation
Affiliation Ended	The date an individual ceased to be affiliated	Datatype Property	MADS-RDF	affiliationEnd

BF templates for RDA book cataloging

▼ RDA Book Work

- ▼ Work title
 - Work title
 - Main title
 - Title part
 - Title part number
- ▶ Content type
- ▶ Genre/Form
- ▶ Subject (LCSH)
- ▼ Identifier
 - Identifier type
 - Identifier value
- ▼ Contributor(s)
 - ▶ Agent
 - ▶ Role
 - ▶ Identifier
- Origin date

▼ RDA Book Instance

- ▶ Title Proper information
 - Statement of responsibility (RDA 2.4)
 - ▶ Parallel Title Information
 - ▶ Transliterated title information
 - ▶ Variant Title Information
- Edition statement (RDA 2.5)
- Publication statement
- Publication date
- ▶ Publication Information
- Copyright date
- Series statement
- Numbering within series
- ▶ Media type
- ▶ Carrier type
- ▶ Identifier
- ▶ Mode of issuance

LC Editor vs CEDAR: Similarities

- ability to do custom labelling that “hides” the ontology terms
- ability to do lookups to value vocabularies
- default values
- ability to repeat “fields”
- can use multiple ontologies
- primary output in JSON-LD
- neither deals well at the moment with multiple properties for the same class
- **the profile/template provides the primary definition of the application profile**

LC Editor vs CEDAR: Differences

LC

- properties & classes entered manually by profile creator
- individual elements are reusable, using the same profile; when the element changes in one place, it changes in every profile it is used in
- look ups restricted to full vocabularies (e.g. all LCGFT)
- no validation or extended application profile ability (e.g. date type) beyond basic profile

CEDAR

- properties and classes added through lookup & directly linked to ontology
- individual elements are reusable, but must be duplicated in each template; when the element changes in one place, it does not change in other places
- look ups can be restricted to individual children of a class or to hand-picked values
- validation of entries including text, date (provides xsd:date), URIs, numbers

Moving forward...

- Internal
 - working to complete workflow analysis
 - making current tracer bullets more robust & integrating SHARE-VDE & BSLW
 - further enhancement of CEDAR templates
- SHARE-VDE
 - more conversion (MODS to BF and MARC to BF extensions)
 - reconciliation of URIs
 - data enhancements
 - exploring potential for sharing data
- Broader Community
 - work with the PCC to host a linked data sandbox for community experimentation
 - filling out application profiles to include relationships from RDA Registry
 - working with the community to make BF a more community-based ontology