

Opus Ex Machina

Modelling SuperWork and Work Entities in BIBFRAME



Ian Bigelow, Cataloguing Coordinator
University of Alberta Libraries, Cataloguing and Metadata Strategies

The Story Begins: RDA Implementation

The importance of description of work elements and the identifier for work

0.6.6 Section 2: Recording Attributes of Work and Expression

When recording data identifying a work, include as a minimum the following elements that are applicable and readily ascertainable.

Preferred title for work

Identifier for work

A preferred title for work is a basis for an authorized access point representing the work. When constructing the access point, combine an authorized access point for an agent, if appropriate, and a preferred title for work.

If a preferred title for work is the same as or similar to a title for a different work, or to a name for an agent, differentiate them by recording as many of the additional identifying elements in the following list as necessary. Record these elements as separate elements, as parts of the authorized access point representing the work, or as both.



Initial attempts: OCLS Cataloguing Workflow

Where do we put the work identifiers in MARC and what do we use?

Identifier for the work	024 1# \$a7822183031 024 7# \$a http://experiment.worldcat.org/entity/work/data/1840249565 \$2uri See http://www.loc.gov/standards/sourcelist/standard-identifier.html for standard identifier source codes for subfield 2.
Preferred title of the work	130 0# \$aGreat Gatsby (Motion picture : 2013)\$0 http://viaf.org/viaf/300876477
Preferred title of the work	240 10 \$aSchneewittchen\$0 http://d-nb.info/gnd/4116406-4
Title of a work	710 22 \$aNational Geographic Society (U.S.).\$bCartographic Division.\$tEarth's fractured surface.\$0 http://worldcat.org/entity/work/id/1090078494

PCC Task Group on URI in MARC: 758 Reasons why work identifiers are important

MARC PROPOSAL NO. 2017-09

DATE: May 16, 2017

REVISED:

NAME: Defining Field 758 (Resource Identifier) in the MARC 21 Bibliographic Format

SOURCE: PCC Task Group on URIs in MARC, Program for Cooperative Cataloging (PCC)

SUMMARY: This paper proposes establishing a new field 758 in the MARC 21 Bibliographic format to identify related resources.

KEYWORDS: Field 758 (BD); Resource Identifier (BD)

RELATED: [2017-DP02](#)

STATUS/COMMENTS:

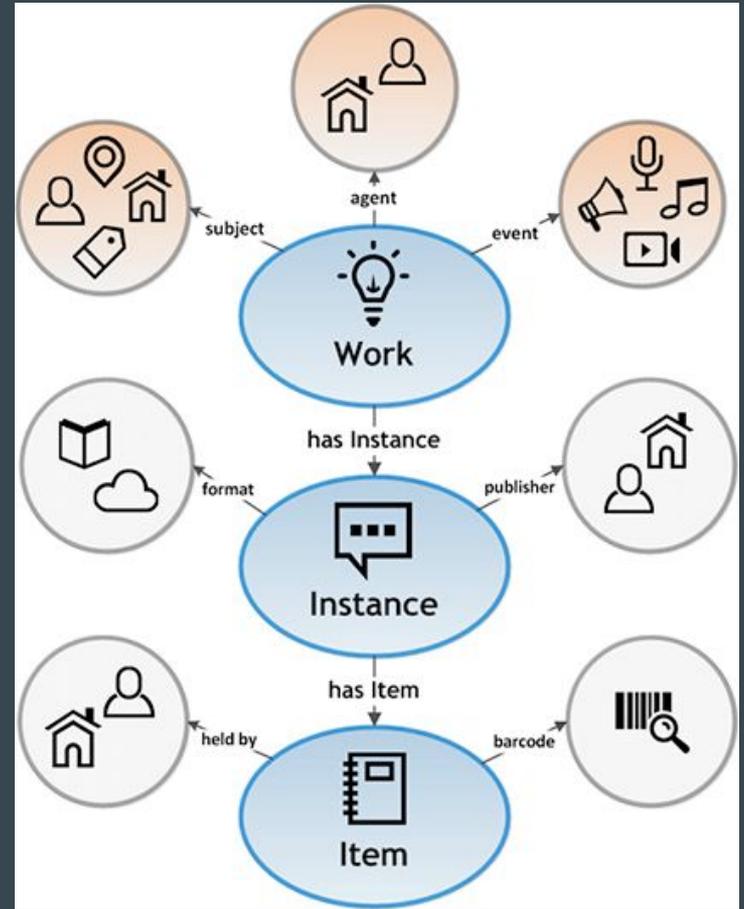
05/16/17 – Made available to the MARC community for discussion.

06/24/17 – [Results of MARC Advisory Committee discussion](#): Approved, with the amendment that subfield \$1 will be added to field 758; the first sentence of the field definition will be amended to read “An identifier for a resource that is either the resource described in the bibliographic record or a resource to which it is related.” It was noted that, while the inclusion of an explicit relationship in field 758 may be desirable, the MARC formats do not mandate the use of specific subfields in field definitions; the application of field 758 would need to be developed as a matter of best practice by the community.

08/07/17 - [Results of MARC Steering Group review](#) - Agreed with the MAC decision.

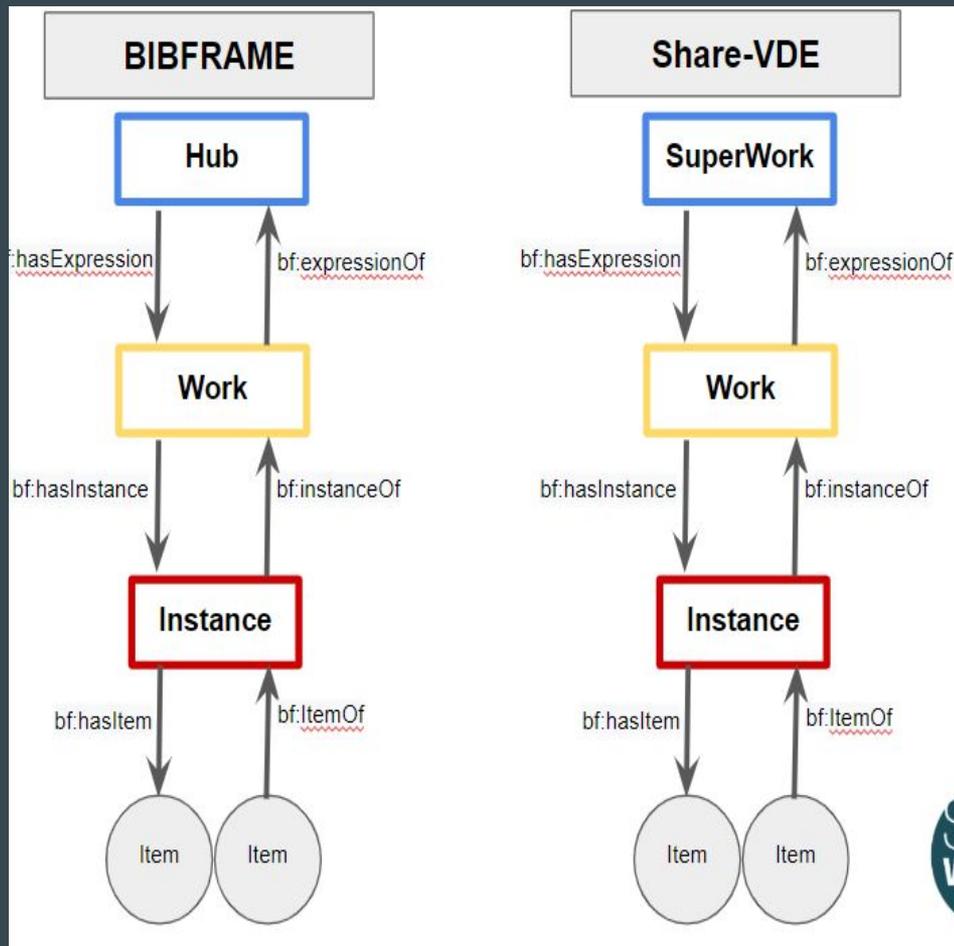
BIBFRAME (as you know it)

- Three core levels of abstraction
 - Work
 - Instance
 - Item
- Additional key concepts
 - Agents
 - Subjects
 - Events
- Consists of RDF classes and properties
 - members of a class share certain characteristics and may have subclasses
 - properties describe characteristics of resources as well as relationships among resources

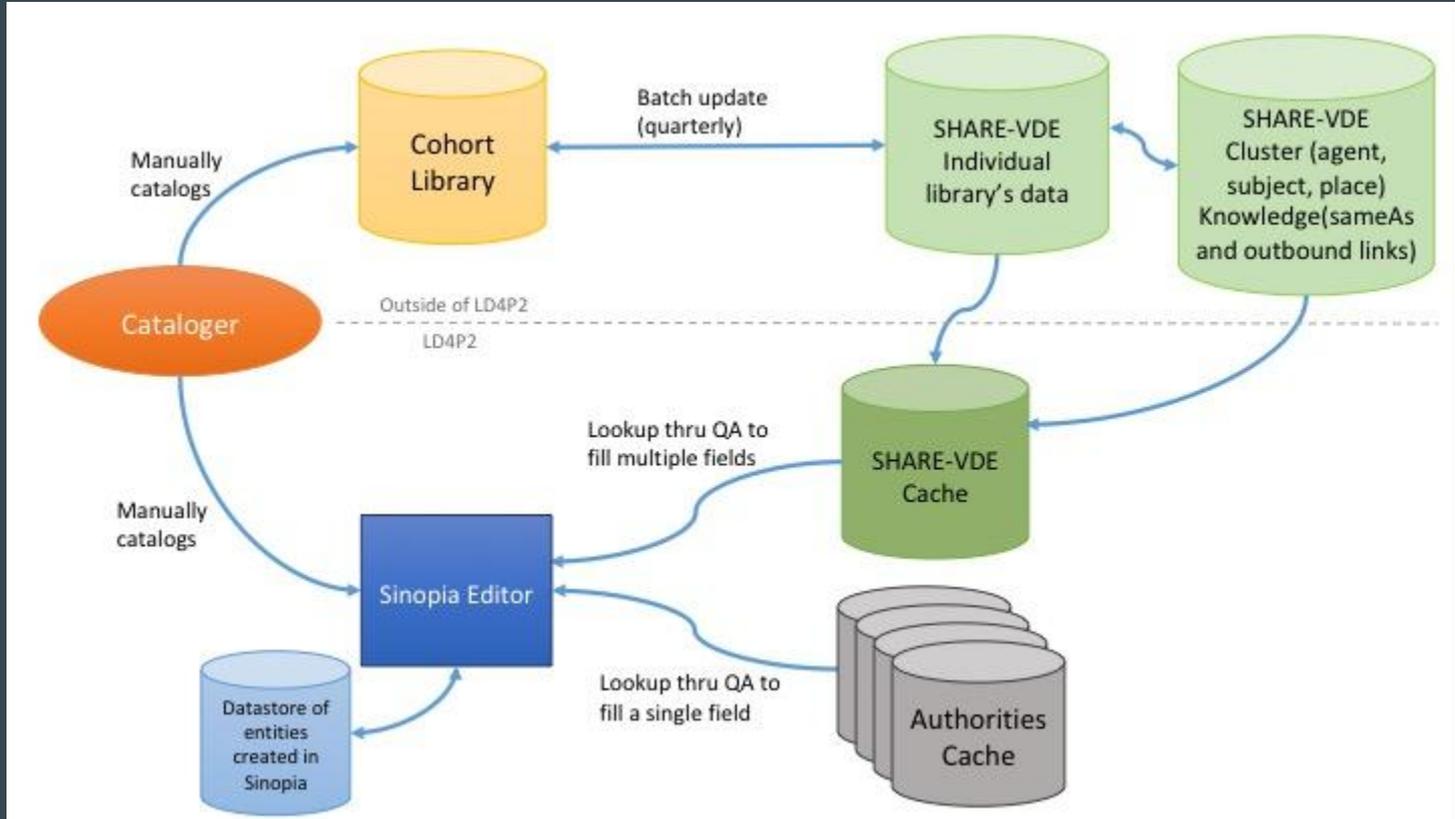


BIBFRAME today

- Four core levels of abstraction
 - *Opus (Hub, SuperWork)*
 - Work
 - *Master* Instance
 - Item
- Additional key concepts
 - Agents
 - Subjects
 - Events
- Consists of RDF classes and properties
 - members of a class share certain characteristics and may have subclasses
 - properties describe characteristics of resources as well as relationships among resources



LD4P2 Data Flow: For this to work we need identifiers



Seemingly Insoluble Difficulties

1. What URI vocabularies are available for primary resource types that:
 - a. Can be clearly identified with a predicate
 - i. For RDA, what identifiers can be used that match range definitions?
 - ii. For BIBFRAME, what identifiers can be defined as
<http://id.loc.gov/ontologies/bibframe/instanceOf>
 - b. Have universal, or at least wide application / work in an international context
 - c. Be able to create them for use in MARC, original BIBFRAME description, and through conversion from MARC to BIBFRAME



The Share-VDE Work ID Working Group

At the LD4 Workshop at Stanford University in May of 2018 conversations identified the need for a group to work on refining the creation of work identifiers for the Share Virtual Discovery Environment (SVDE). In the months following members of the Work ID Working Group (WIDWG) were identified from within stakeholder institutions and based on relevant expertise. Initial representation included Casalini Libri, @Cult, George Washington University, Library of Congress, NLM, Ohio State University, PCC, Stanford University, UCDavis, University of Alberta Libraries, and University of Chicago.

Charge:

1. Review the Share work clustering outline and submit feedback on potential improvements or optimizations
2. Review the use of primary resource identifiers in the Share-VDE data set and provide feedback as appropriate
3. Engage with the PCC to identify and/or develop best practices for use of these identifiers in BIBFRAME and MARC data



Existing Vocabularies?

OCLC Work ID

- Challenges with technical methods of inclusion for non OCLC members
- schema:exampleOfWork, but no open outline of how the identifiers are created or defined
- Until recently OCLC works were still considered experimental
 - Could be a next step for enrichment of SVDE data
 - It would be interesting to compare approaches and come to a standard for creation of primary resource identifiers through algorithmic processes

Library of Congress Work ID

- Based on LCCN
 - Example: [<http://id.loc.gov/resources/works/c010579972>]
- Or making use of nametitle authorities
 - Example: [<http://id.loc.gov/resources/works/no98044787>]
- Not of universal applicability, and at the time was not ready for use in Share-VDE
 - Could be a next step for enrichment in SVDE depending on approach

The Challenge: Lots of Work in a Short Time -- LC bf: example

```
<?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://example.org/6020396#Work">
    <bf:adminMetadata>
      <bf:AdminMetadata>
        <bf:generationProcess>
          <bf:GenerationProcess>
            <rdfs:label>DLC marc2bibframe2 v1.3.0-SNAPSHOT</rdfs:label>
          </bf:GenerationProcess>
        </bf:generationProcess>
        <bf:status>
          <bf:Status>
            <bf:code>p</bf:code>
          </bf:Status>
        </bf:status>
      </bf:AdminMetadata>
    </bf:adminMetadata>
  </bf:Work>
</rdf:RDF>
```

Defining Works

LRM

Definition: “The intellectual or artistic content of a distinct creation”²

Scope note: “A work is an abstract entity that permits the grouping of expressions that are considered functional equivalents or near equivalents. A work is a conceptual object, no single material object can be identified as the work. The essence of the work is the constellation of concepts and ideas that form the shared content of what we define to be expressions of the same work. A work is perceived through the identification of the commonality of content between and among various expressions.”²

BIBFRAME

“The highest level of abstraction, a Work, in the BIBFRAME context, reflects the conceptual essence of the cataloged resource: authors, languages, and what it is about (subjects).”³

There are differences in definition, and this has been a subject of discussion for many years, but the WIDWG needed a practical solution in a short period of time. We wanted to keep an open mind about the definition of “Work”, and how work identifiers are created in Share-VDE.

2. Riva, P., Le Boeuf, P., Žumer, M., & IFLA FRBR Review Group. (2017). IFLA Library Reference Model: A conceptual model for bibliographic information. Den Haag, Netherlands : IFLA.

3. Library of Congress. (2017). Overview of the BIBFRAME 2.0 model. Retrieved from <https://www.loc.gov/bibframe/docs/bibframe2-model.html>



SuperWork!

Are Work to Work Relationships Sufficient?

1. While Work -> Expression relationships can currently be expressed in BIBFRAME, these are ultimately Work-Work relationships, and determining the initial or primary work, or hierarchical relationships between works may prove difficult with this structure.
2. Through conversion from MARC to BIBFRAME, or automatic work ID generation based on BIBFRAME elements, unless we can define a difference (a fingerprint for each cluster or constellation) between Work and SuperWork elements then these relationships (work-expression) cannot be captured through conversion or automated processing. With the scale of data conversion underway, not doing this would seem like a missed opportunity. Once a separate fingerprint is defined for this primary work, it needs a name, thus the creation of SuperWork.

Defining Share VDE Work and SuperWork

Share VDE Work

- Is equivalent to a BIBFRAME Work, but is no longer the highest level of abstraction
- Identifiers for Share VDE Work are created algorithmically based on unique constellations of elements for BIBFRAME Works (including RDA work and expression level elements)
- The types of Share VDE Work and the definitions for which elements are used in its creation are outlined in the Work ID Cluster Mapping

Share VDE SuperWork

- The highest level of abstraction in Share VDE, the new SuperWork class (subclass of bf:Work) is meant to aggregate or group functional or near equivalent bf:Work clusters
- Identifiers for Share VDE SuperWork are created algorithmically based on unique constellations of elements for BIBFRAME Works, minus RDA expression level elements

Refined by:

Hubs x

shakespeare, william

GO

RESET

Results: 1-20 of 1,434

< 1 2 3 4 5 6 ... >

Refine your results

Scheme

Hubs

Type

Hub

1,434

Work

1,434

Collection

No values

Created Date

2010s

1,434

Modified Date

	Label	Vocabulary	Concept	Subdivision	Identifier
1.	Shakespeare, William, 1564-1616 Plays. Selections	Hubs	Work		254f0f53...
	Shakespeare, William, 1564-1616 Hollywood immortals perform Shakespeare ; Shakespeare, William, 1564-1616 Seven comedies ; Shakespeare, William, 1564-1616 Four comedies ; Shakespeare, William, 1564-1616 Pericles, Prince of Tyre ; Cymbeline ; The two noble kinsmen ; Shakespeare, William, 1564-1616 Hi...				
2.	Shakespeare, William, 1564-1616 Plays	Hubs	Work		c4c931d7...
	Shakespeare, William, 1564-1616 Plays of William Shakespeare ; Shakespeare, William, 1564-1616 Dramatic works of William Shakespeare ; Shakespeare, William, 1564-1616 Complete plays of William Shakespeare ; Shakespeare, William, 1564-1616 Contemporary Shakespeare series (University Press of America ...				
3.	Shakespeare, William, 1564-1616 Sonnets. Selections	Hubs	Work		19cd1943...

Opus: Hub vs SuperWork vs IFLA LRM Work

In January 2019 a new SuperWork class was introduced in Share VDE data. Shortly after, just prior to ALA Annual 2019 LC introduced the Hub to their data. While further analysis and refinement of practice for these parallel processes is needed, ultimately they both seem to serve the same function in BIBFRAME and are hereafter referred to as the **Opus** in this discussion.

The distinction between types of Opus (Share VDE SuperWork and LC Hub) and the relationship between the updated model to IFLA-LRM warrant further discussion within the **International BIBFRAME Community** and RDA. The BIBFRAME Workshop in Europe may well be an excellent starting point to clarify definitions of primary resources in BIBFRAME.



Finishing the Opera - Opus Ex Machina!

Opus (bf: SuperWork, Hub):

The highest level of abstraction, an opus is an entity that permits the grouping of works that are considered functional or near equivalents. The opus is a subclass of work, defined by a constellation of elements that form the shared content of aggregated works.

Work:

A work is a resource reflecting the conceptual essence of a cataloguing resource. A work is defined by a constellation of elements representing the specific intellectual or artistic form that an opus takes each time it is “realized”, paired with related elements required to define the conceptual essence of said realization.

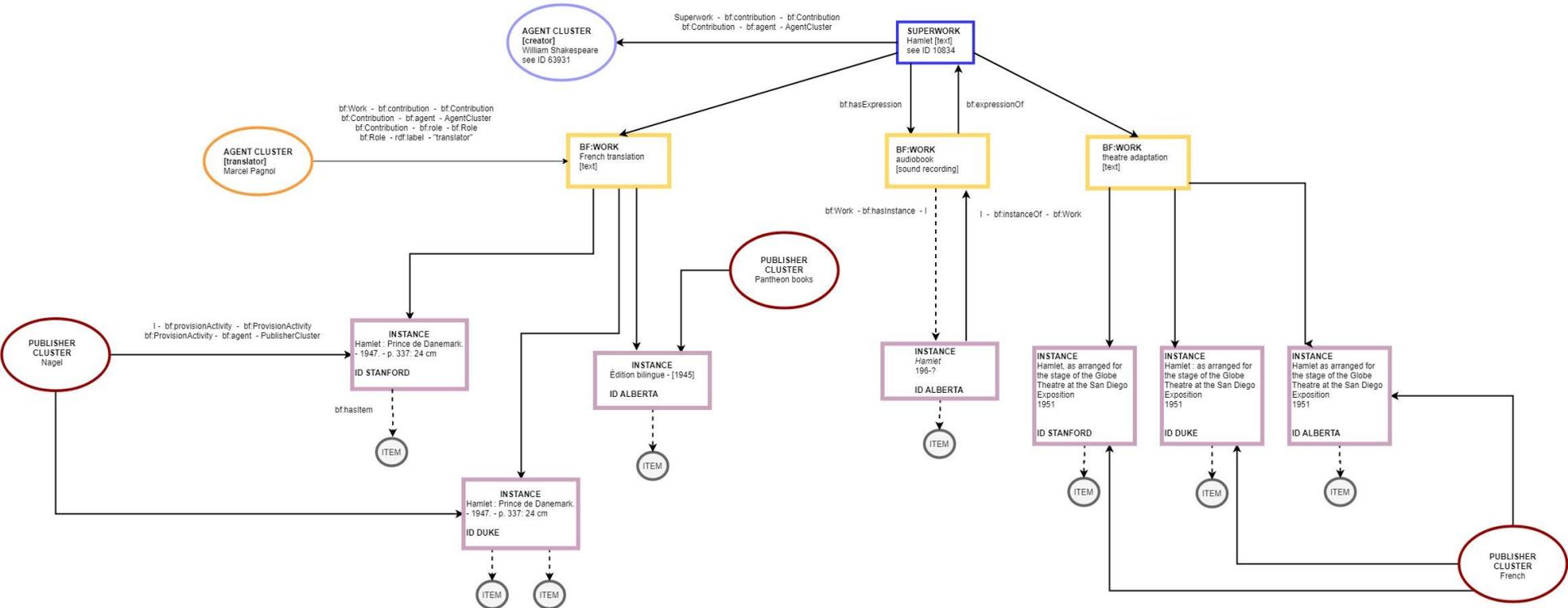


Work from the Machine: Work ID Cluster Mapping

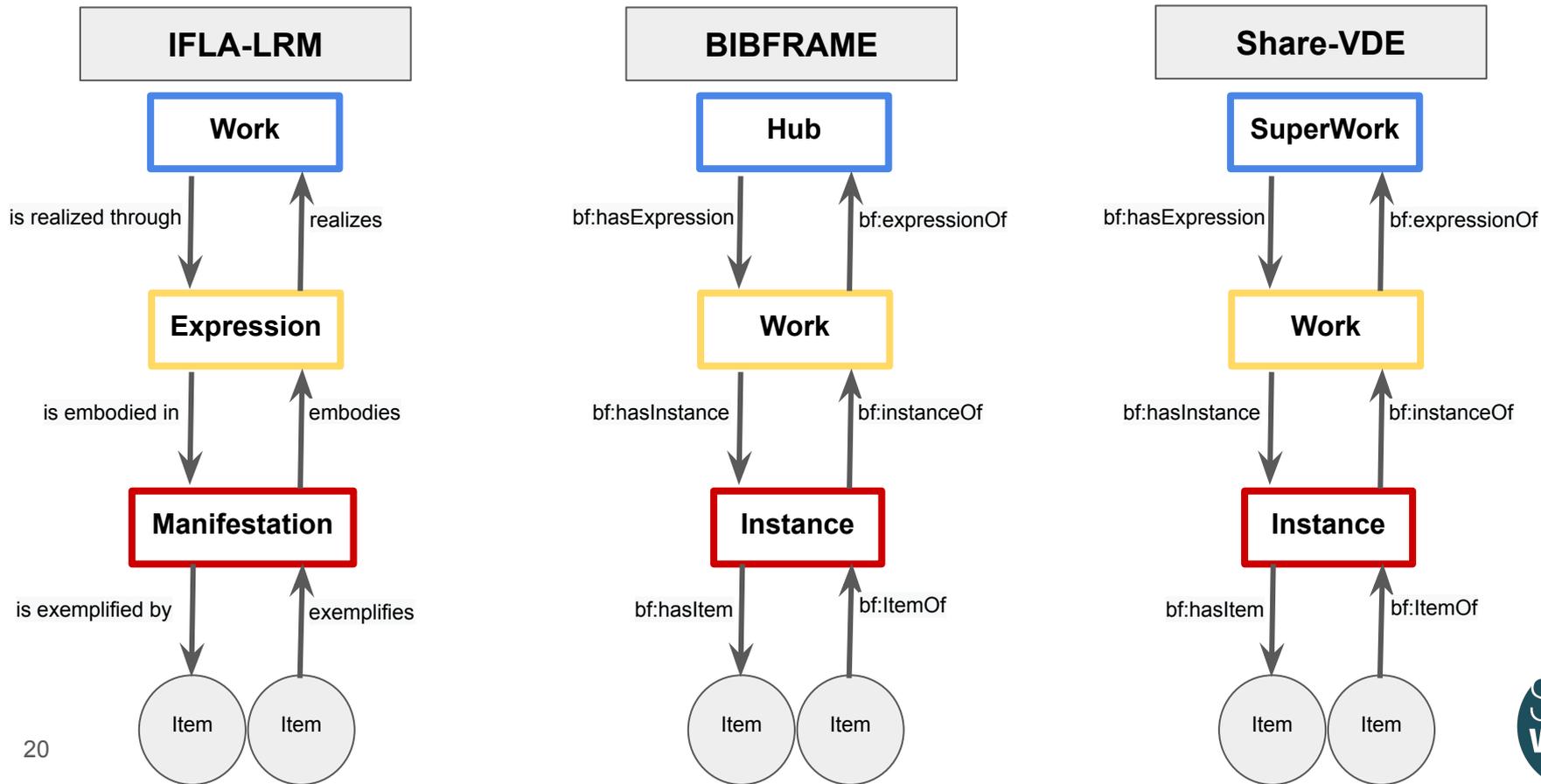
Conversion specifications for SVDE MARC to BIBFRAME processes can be found here				
	Field	Indicators	Subfields	
SuperWork				
W1		130	\$a\$d\$f\$g\$k\$m\$n\$r\$t\$p	\$n may or may not indicate a part; \$p is a part (portion of a work)
W2		240 1XX and 7XX	\$a\$d\$f\$g\$k\$m\$n\$r\$t\$p	\$n may or may not indicate a part; \$p is a part (portion of a work) Used for clustering - Outline here
W3		245 1XX and 7XX	\$a\$n\$p	Only if no W1 or W2 identified* Used for clustering - Outline here
W4		700/710/711	\$a\$d\$f\$g\$k\$m\$n\$r\$t\$p	Note differentiation in relationships based on 2nd indicator 2 (analytic) or "blank" (related work)
W5		730	\$a\$d\$f\$g\$k\$m\$n\$r\$t\$p	Note differentiation in relationships based on 2nd indicator 2 (analytic) or "blank" (related work)
W6		758	\$a\$4\$0\$1	As a separate work or clustered in conjunction with other work types. Use identifiers and relationships in SVDE data
W7		780	\$a\$g\$s\$t\$w	Note 2nd indicator values for relationship assignment (0-7) ; also only use \$w if the identifier is for a work ID (LCCN, Work URI ...)
W8		785	\$a\$g\$s\$t\$w	Note 2nd indicator values for relationship assignment (0-8) ; also only use \$w if the identifier is for a work ID (LCCN, Work URI ...)

The current SHARE-VDE entity model

Share-VDE Super Work graph (simplified for UI/UX purposes) - draft 21st February 2019



Ongoing BIBFRAME Developments



Sapientia is Key

What is a Cluster Knowledge Base?

Why is it important?



Seemingly Insoluble Difficulties

All resolved suddenly and in unexpected ways (not really)

1. What URI vocabularies are available for primary resource types that:
 - a. Can be clearly identified with a predicate
 - i. For RDA, what identifiers can be used that match range definitions?
 - ii. For BIBFRAME, what identifiers can be defined as
<http://id.loc.gov/ontologies/bibframe/instanceOf>
 - b. Have universal, or at least wide application / work in an international context
 - c. Be able to create them for use in MARC, original BIBFRAME description, and through conversion from MARC to BIBFRAME



Reality - More work ahead

1. What URI vocabularies are available for primary resource types that:
 - a. Can be clearly identified with a predicate
 - i. For RDA, what identifiers can be used that match range definitions?
 1. Can Opus type works fit this need
 2. If not, what does the RDA community intend to use?
 - ii. For BIBFRAME, what identifiers can be defined as <http://id.loc.gov/ontologies/bibframe/instanceOf>
 1. For Opus and Work, definitions (conceptual and entity based) need to be refined and confirmed
 - b. Have universal, or at least wide application / work in an international context
 - i. Currently in id.loc.gov and Share VDE, and Sapiaentia has potential for universal application
 - c. Be able to create them for use in MARC, original BIBFRAME description, and through conversion from MARC to BIBFRAME
 - i. This work based on BIBFRAME elements needs to be developed



Thank You