



Virtual
Discovery
Environment

Data modeling in and beyond BIBFRAME

Tiziana Possemato, @Cult - Casalini Libri

SWIB19
Semantic Web in Libraries

25 - 27 November 2019
Hamburg, Germany

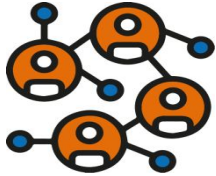
Share-VDE initiative in SWIB

- **SWIB 2017: Will you be my bf: forever? Analysing Techniques for Conversion to BIBFRAME at the University of Alberta**
Ian Bigelow / Sharon Farnel -- *University of Alberta, Canada*
- **SWIB 2018: Share virtual discovery environment in Linked Data (SHARE-VDE)**
Michele Casalini [Lightning talks]
- **SWIB 2019: Data modeling in and beyond BIBFRAME**
Tiziana Possemato

Share-VDE initiative and its goals



What is Share-VDE?



Share Virtual Discovery Environment in Linked Data is a library-driven initiative to establish an **effective working environment** for the use of linked data by libraries within a global context.

Library data are **enriched** with additional information and relationships, and bibliographic and authority data are converted into linked data.



A **virtual discovery platform** with the structure of the **BIBFRAME** data model is created to simplify the way in which that data is consumed.

The network of resources created is the basis for the **Share-VDE Sapiientia Cluster Knowledge Base**, the common authoritative source of clusters accessible in RDF, open to the entire Share-VDE community.

Who is responsible for it?

Share-VDE is a collaborative endeavour based on the needs of libraries, developed by:



the joint effort of the **Share-VDE Advisory Council** and of the **Working Groups**;



Casalini Libri, provider of bibliographic and authority data as member of the Program for Cooperative Cataloguing;



@Cult, provider of ILS, Discovery tools and Semantic web solutions for the cultural heritage sector;



influenced by the vision of the **LD4P initiative**;



with input and active participation from an **international group of research libraries**.



Share-VDE overall goals

Enrichment of MARC records with URIs

Conversion from MARC to RDF using the BIBFRAME vocabulary (and other ontologies)

Data publication according to the **BIBFRAME** data model

Batch/automated **data updating** procedures

Batch/automated **data dissemination** to libraries

Progressive implementation of **use cases**, with priorities defined by the Share-VDE community



Share-VDE phases

Phase 1

R&D: 2016 – 2017

1985 and 2015 imprint titles; 2,249,397 bib-records and 3,601,327 auth-records.

Phase 2

R&D: 2017 – 2018

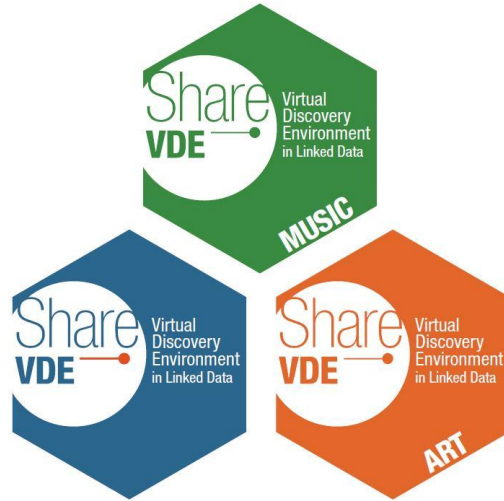
Entire catalogues for all resource types; 94,378,728 bib-records and 24,150,238 auth-records.

Phase 3

Production environment: 2019 -

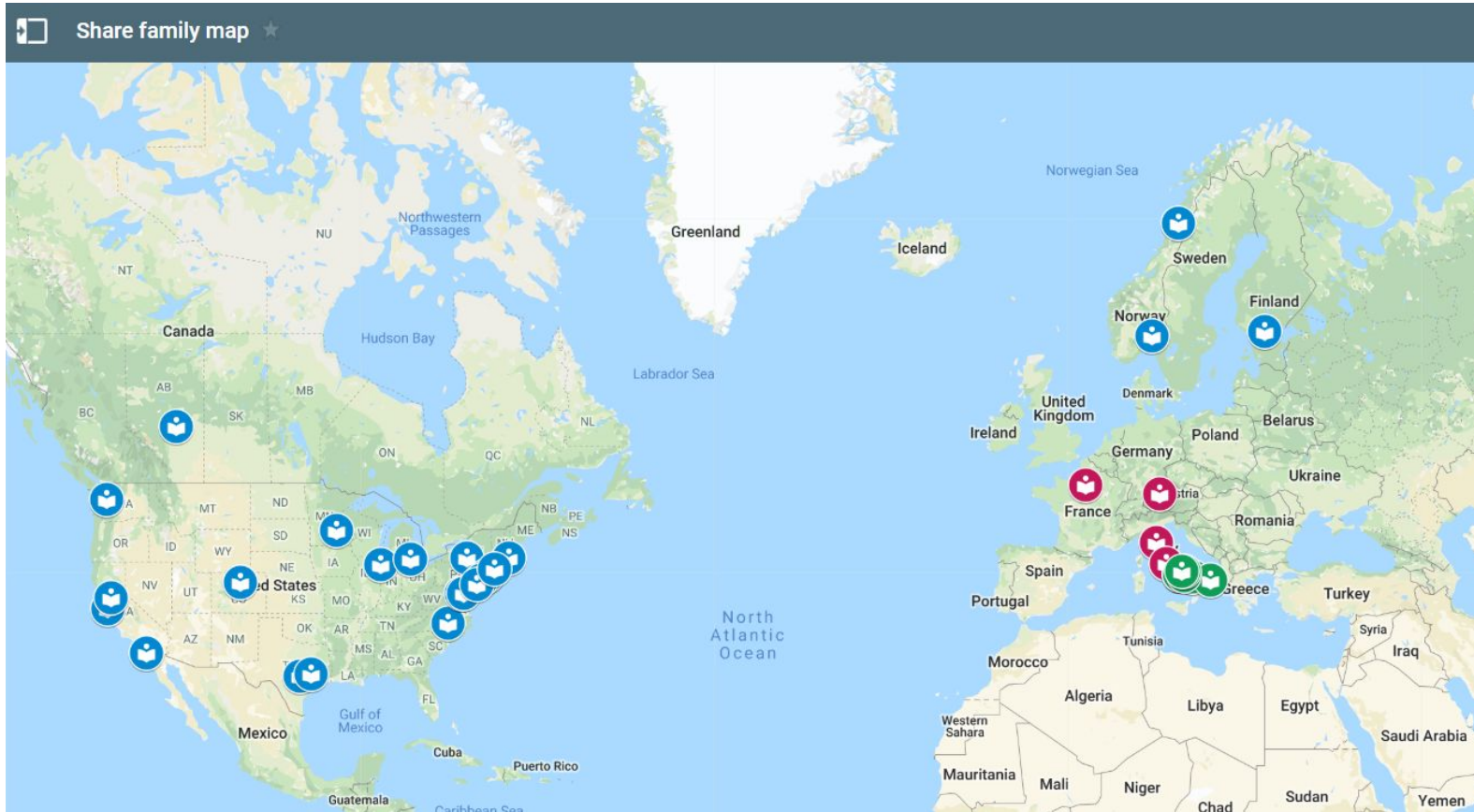
In progress.

The Share family



The [Share family](#) of initiatives based on linked data comprises **Share-VDE**, **Share-Catalogue** (the Italian network of university libraries applying the Share principles), **Share-ART** (the [Kubikat-LOD](#) project including the Art History libraries of the Max Planck Institut), and **Share-MUSIC** (a pilot in the music domain). The different characteristics of each field are a useful asset that can be used to the advantage not only of the Share family as a whole, but for each single discipline.

The Share family map around the world



The Share family participating institutions

Share VDE Full members

Duke University
New York University
Stanford University
University of Alberta – NEOS consortium
University of Chicago
University of Michigan at Ann Arbor
University of Pennsylvania
Yale University

National Libraries

National Library of Norway
National Library of Finland

With the cooperation of

Library of Congress

LD4P Cohort members

Cornell University
Frick Art Reference Library
Harry Ransom Center Texas A&M
Harvard University
National Library of Medicine
Northwestern University
Princeton University
UC Davis
UC San Diego
University Colorado at Boulder
University of Minnesota
University of Texas A&M
University of Washington

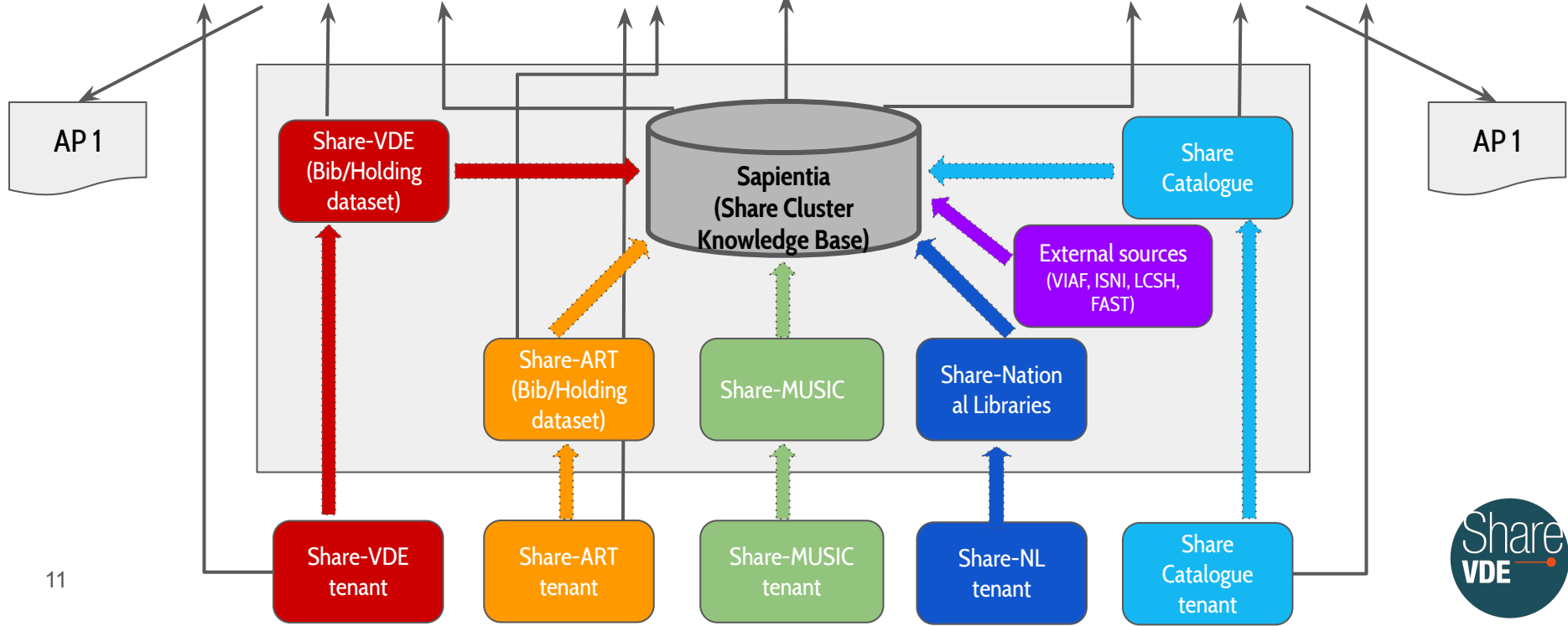
Share-Catalogue Institutions

Università Degli Studi di Napoli "Federico II"
Università degli Studi della Basilicata
Università Degli Studi di Napoli L'Orientale
Università degli Studi di Napoli Parthenope
Università del Salento
Università degli Studi di Salerno
Università degli Studi del Sannio RCost
Università degli Studi della Campania "Luigi Vanvitelli"

Share-Art (Kubikat-LOD) project

Max-Planck-Institut
Kunsthistorisches Institut in Florenz
Biblioteca Hertziana Rome
Central Institute of Art History Munich
Deutsches Forum für Kunstgeschichte Paris
Centre allemand d'histoire de l'art Paris

Common Share User Interface



Share-VDE Advisory Council & Working Groups

The Share-VDE Advisory Council's role is to provide insight and analysis of the MARC to BIBFRAME transformation to make recommendations for improvements based on member library data analysis, and project documentation. The AC also provides overall guidance to the activities of Share-VDE initiative.

There are different sub-committees focusing on specific areas:

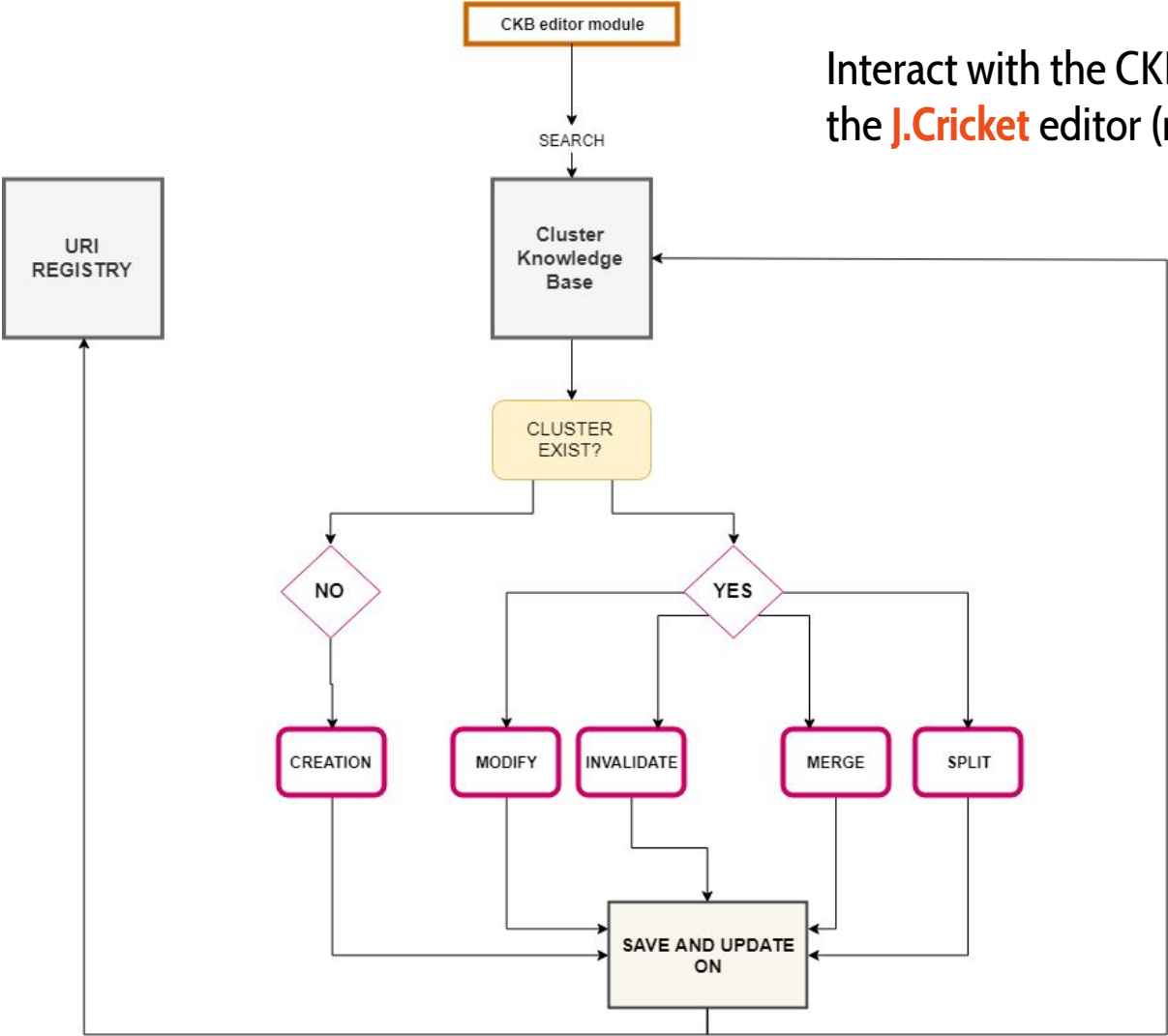
- Entity Identification Working Group
- Authority/Identifier Management Services Working Group
- Cluster Knowledge Base Editor Working Group
- User experience/User Interface Working Group
- Automatic Update processes Task Group

Cluster Knowledge Base Maintenance Working Group

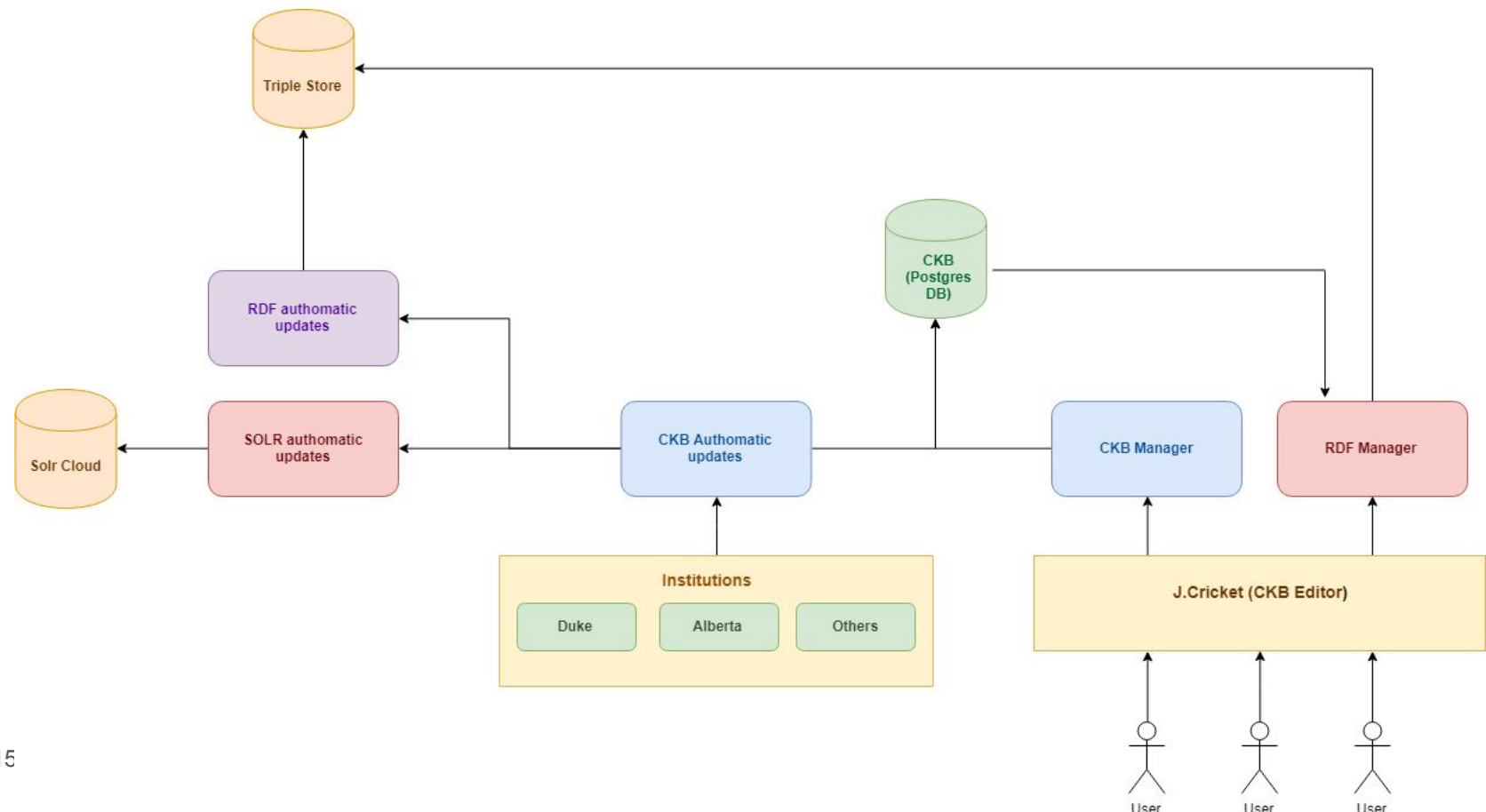
The role of **J. Cricket** (the Share CKB editor) on update processes is defined by the Share Cluster Knowledge Base Maintenance Working Group:

- an essential part of the conversion process from MARC to RDF is the maintenance of metadata that have been produced and registered on the Share CKB (Sapientia);
- the group analysis how participant libraries interact with the Sapientia CKB and how they use the tool to interact (create/modify/delete) the data;
- the same approach will be applied to the data originally created in BIBFRAME (using Sinopia and other LD editors).

Interact with the CKB **Sapientia** using the **J.Cricket** editor (manual process)



Automatic and manual data updates: primary/replica relationship



All changes need to be ‘registered’

The role of the URI Registry in the Share-VDE datasets

“Within this changed context, the management of URIs (Uniform Resource Identifiers) must be carefully evaluated. URIs play the role of universal unique identifiers in the technological environment of linked open data: as the issue typical of the “Web of documents” of locating resources or web pages is becoming less relevant, in the semantic Web URIs identify a specific object (thing) or, using proper terminology, an entity. In addition to having to respond to the characteristics of dereferencing, simplicity, stability and manageability, a well-structured URI must be persistent, i.e. it must not undergo changes over time in order to guarantee the correct recovery of the identified entity and the information connected to it. This aspect of persistence over time is more and more urgent, especially in the context of Linked Open Data, which opens up scenarios of use and re-use of the data much wider than the traditional context.”

URI Registry to record changes

PROCESS I: changes resulting from DELTA

UC A1 - Records created

UC A1a - Authority records

UC A1b - Bibliographic records

UC A2 - Modified records

UC A2a - Minor changes to the data

UC A2b - Substantial changes to the data

UC A3 - Deleted records

UC A3a - Authority record

UC A3b - Bibliographic record

UC A4 - Mash-up/merged records

UC A4a - Authority record

UC A4b - Bibliographic record

UC A5 - Split records

PROCESS II: changes resulting from the CKB Editor

UC B1 - Creation

UC B1a - Cluster creation

UC B1b - Creation of the URI

UC B2 - Modification

UC B3 - Invalidation

UC B3a- cluster Super Work invalidation

UC B3b- cluster Agent invalidation

UC B3c- cluster Instance invalidation

UC B3d- cluster Publisher invalidation

UC B4 - Merge

UC B5 - Split

Share-VDE data modeling

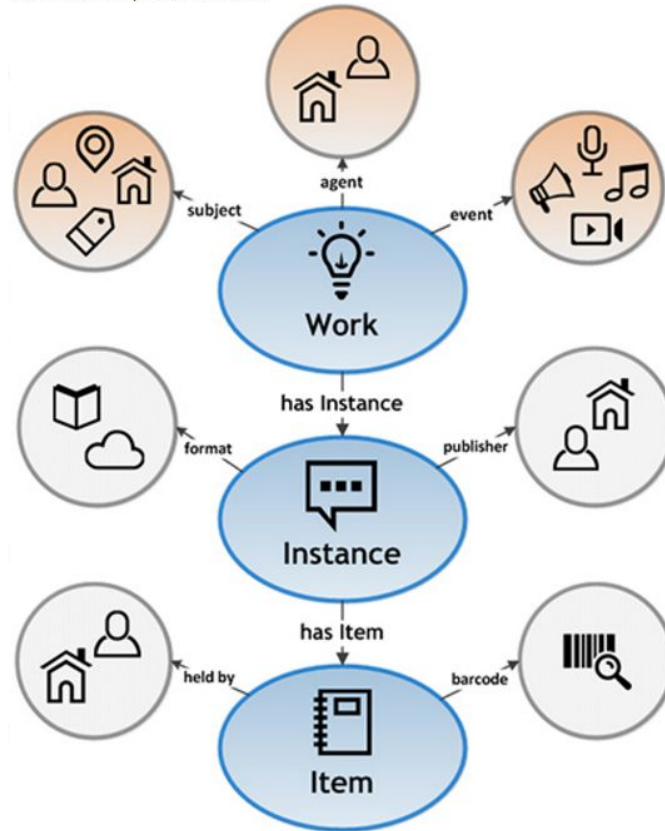


Data modeling

Ongoing discussions with Share family members and external parties around the **evolution of the entity models**:

- the Share-VDE **SuperWork** entity level has been **related with** the very recent **Library of Congress Hub** property. Analysis of similarities and possible interoperability layers are ongoing in the Entity Identification Working Group;
- after analysis and discussions among the Share-VDE community, one of the future enhancements of the data model will include the **MasterInstance** in order to help the relationship between the shared data elements and the local ones for the Instance layer.

The BIBFRAME 2.0 data model



Entity definitions: BIBFRAME

Hub: it's still under analysis and testing

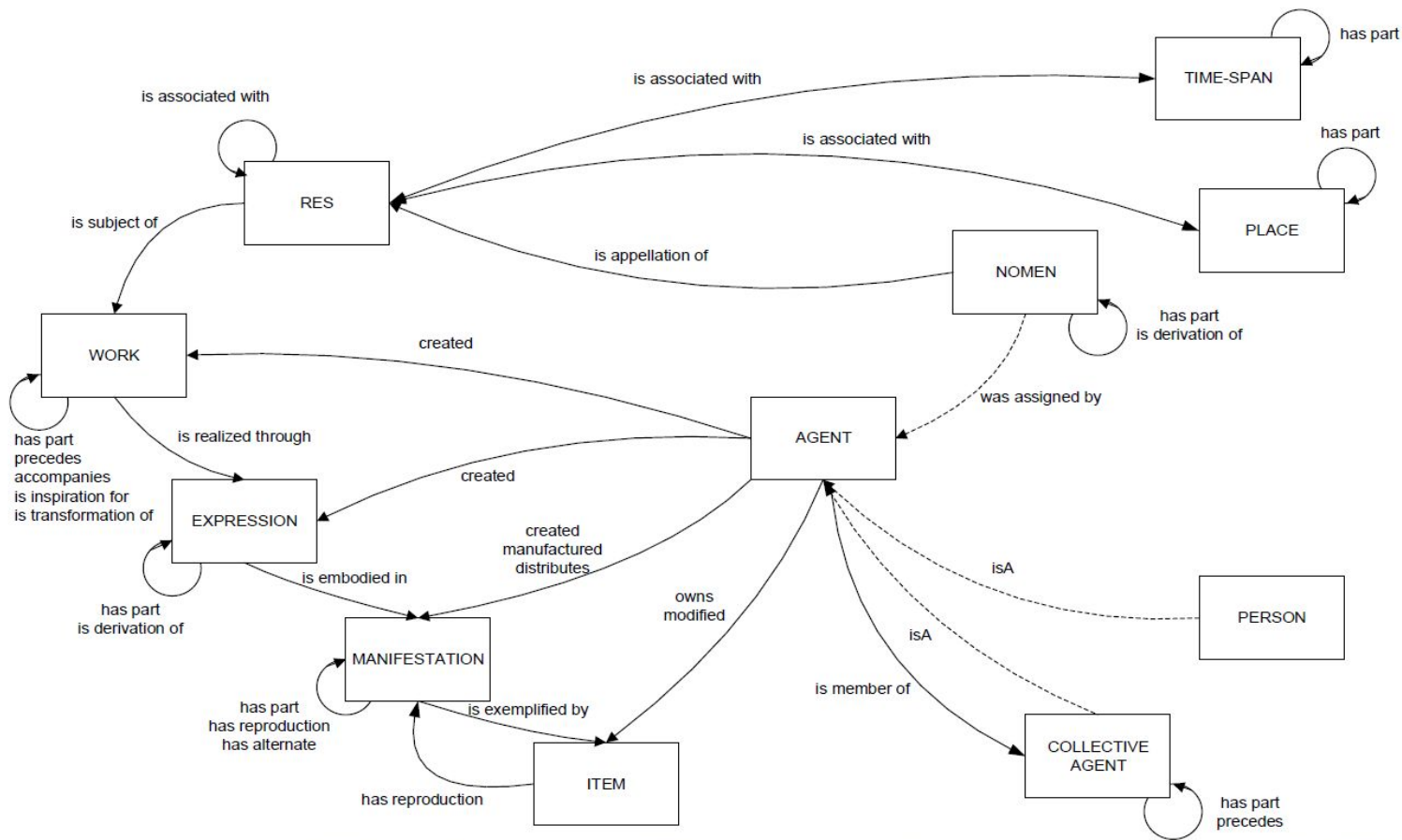
Work <http://id.loc.gov/ontologies/bibframe.html#c> **Work:** resource reflecting a conceptual essence of a cataloging resource.

Instance <http://id.loc.gov/ontologies/bibframe.html#c> **Instance:** resource reflecting an individual, material embodiment of a Work.

Item <http://id.loc.gov/ontologies/bibframe.html#c> **Item:** single example of an Instance.

Source: <http://id.loc.gov/ontologies/bibframe.html>

The LRM data model



Entity definitions: IFLA-LRM

Work: the intellectual or artistic content of a distinct creation.

Expression: a distinct combination of signs conveying intellectual or artistic content.

Manifestation: a set of all carriers that are assumed to share the same characteristics as to intellectual or artistic content and aspects of physical form. That set is defined by both the overall content and the production plan for its carrier or carriers.

Item: an object or objects carrying signs intended to convey intellectual or artistic content.

Source: https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017_rev201712.pdf

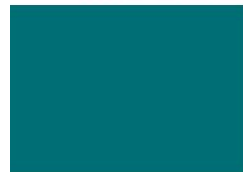
BIBFRAME vs LRM



Work, Instance, Item (BIBFRAME)

vs

Work, Expression, Manifestation, Item (LRM)



IFLA Library Reference Model

**A Conceptual Model for
Bibliographic Information**

SuperWork Plain Language Description*

A new class is being tested for implementation in the Share-VDE and Linked Data for Production (LD4P) Cohort: the **SuperWork** entity

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.

*Work Identification Working Group, SuperWork Plain Language Description

SuperWork Plain Language Description*

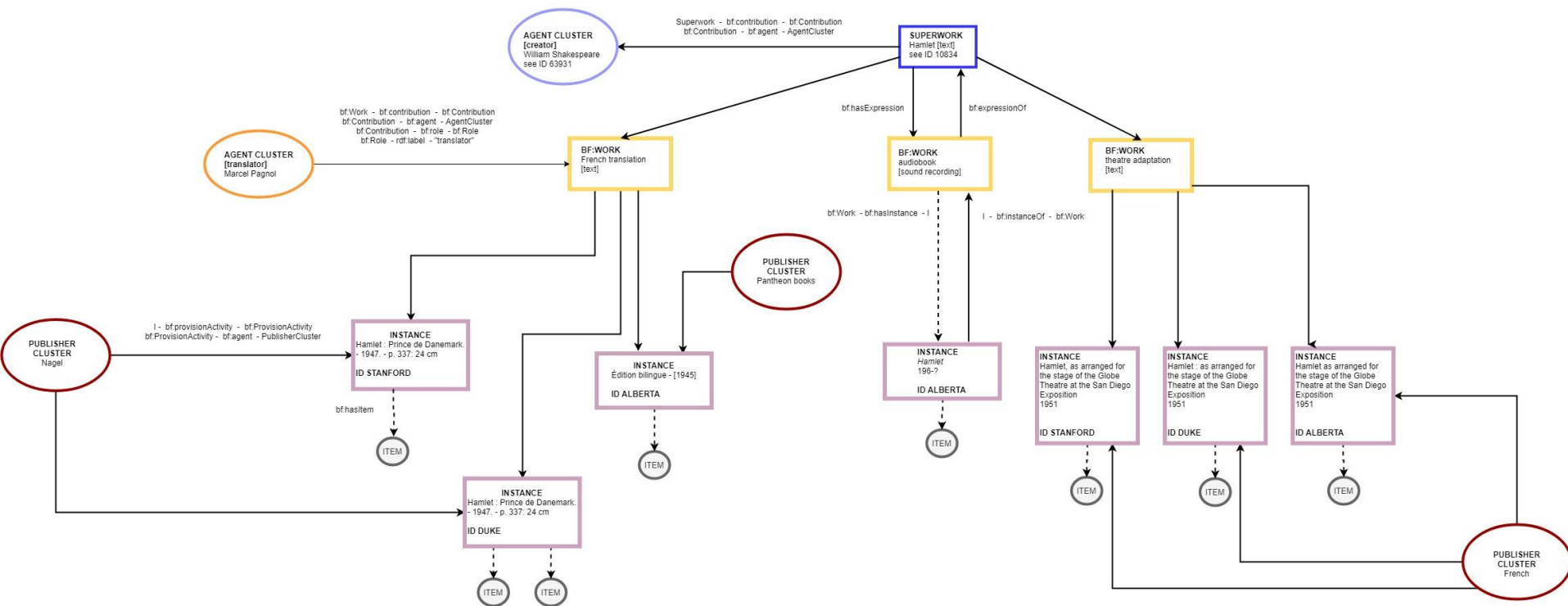
Share-VDE SuperWork:

- the highest level of abstraction in Share-VDE data model, the new SuperWork class 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.

*Work Identification Working Group, SuperWork Plain Language Description

The current Share-VDE entity model

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



How to manage Instances in a shared environment?



Instance vs Manifestation



Instance (in BIBFRAME): a Work may have one or more individual, material embodiments, for example, a particular published form. These are Instances of the Work. An Instance reflects information such as its publisher, place and date of publication, and format.

IFLA Library Reference Model

A Conceptual Model for
Bibliographic Information

Manifestation (in LRM): a set of all carriers that are assumed to share the same characteristics as to intellectual or artistic content and aspects of physical form. That set is defined by both the overall content and the production plan for its carrier or carriers.

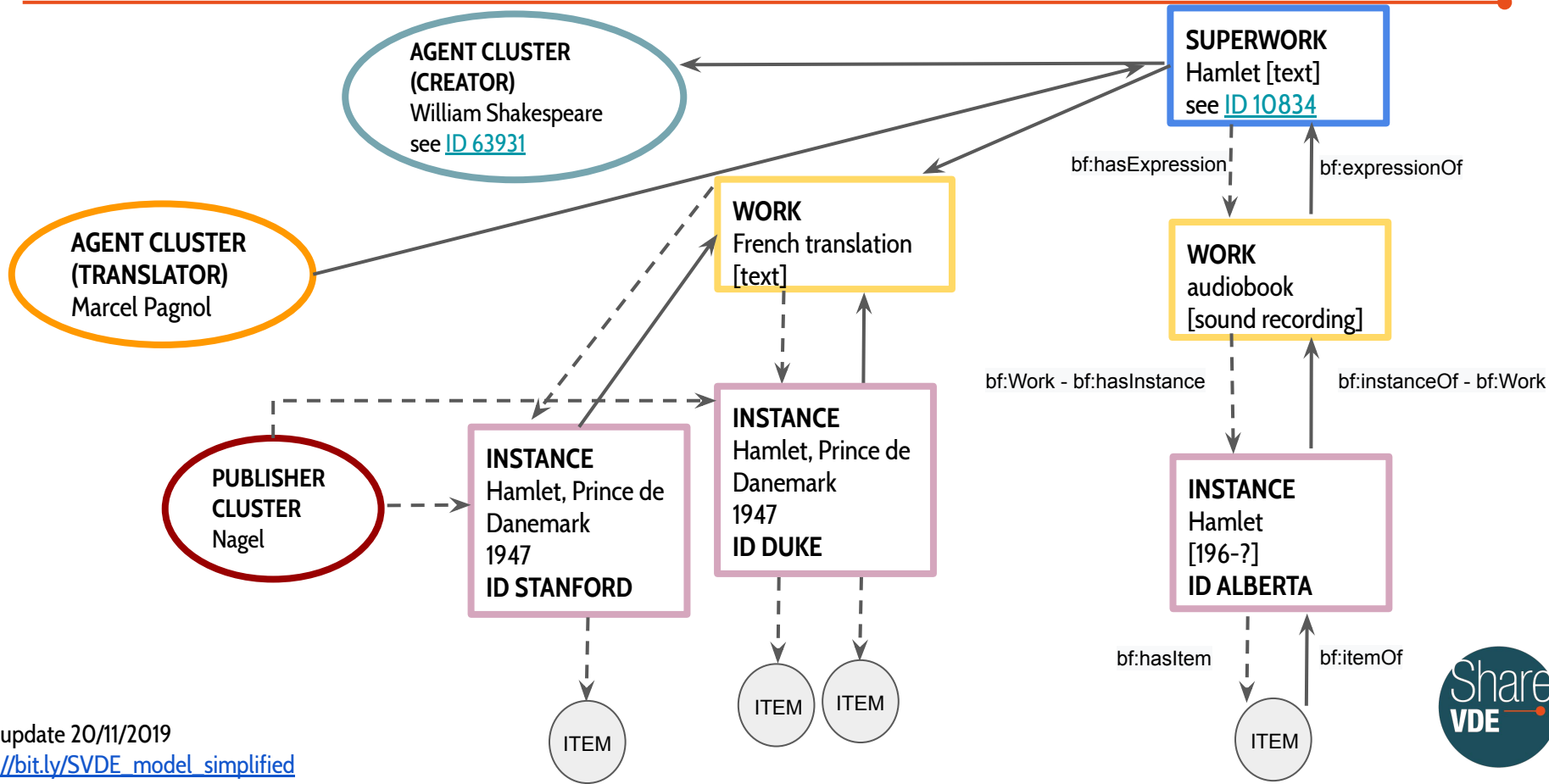
The Share-VDE future entities model

In the current Share-VDE entity model, an Instance is not really identified as an Entity, but as a description of an entity made by a particular Institution. The first proof of this is the instance URI: it is built using the Share-VDE + type of entity + source (the institution that created the original record) + the ID of the original record:

<http://share-vde.org/sharevde/rdfBibframe/Instance/UALBERTA6947549>

The Share-VDE Advisory Council with its subcommittees is discussing the evolution of the Share-VDE instance from a "description of" to an "entity".

Current Share-VDE model simplified



The Instance as a *Master Instance*



Instance (as logical *Master Instance*)

A new entity for the bibliographic management of the Union Catalog

Draft under revision, last update [11/09/2019](#)

Introduction

The recent development phases of Share-VDE have involved an overall adjustment of the management of data flows, of the ontological model of entities and of the structuring of the user interface, with the aim of further enhancing the information potential on the platform and solving the most critical issues regarding data inconsistencies and discrepancies.

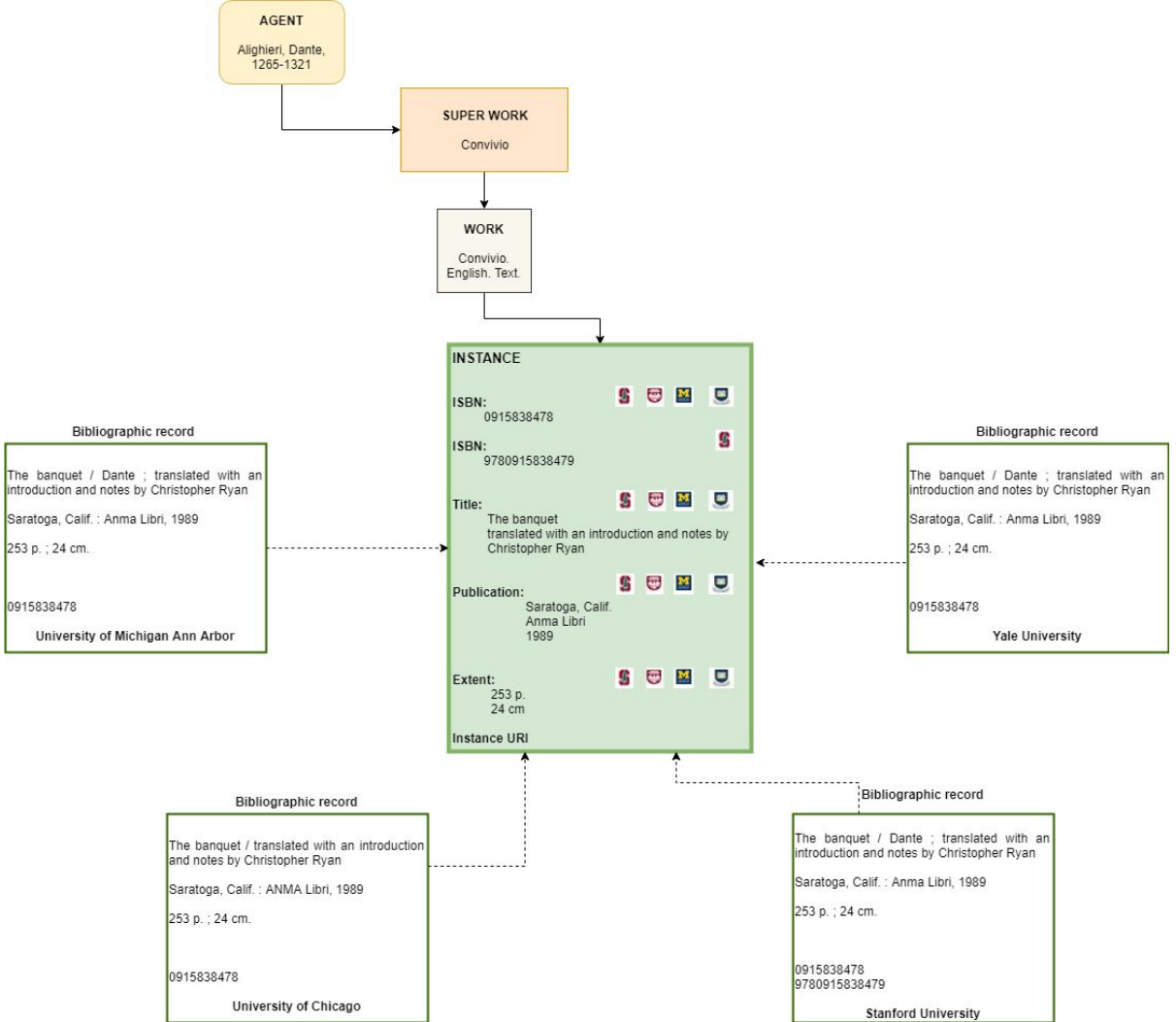
On this, it must be noted that Share-VDE currently includes over 20 research libraries, each one equipped with its own internal OPAC and linked to different cataloging practices, although to a large extent referable to the same RDA standard. The loading of millions of bibliographic and authority data coming from such different situations has consequently implied the detection of a series of problems at data processing level, which can be summarized in two main aspects:

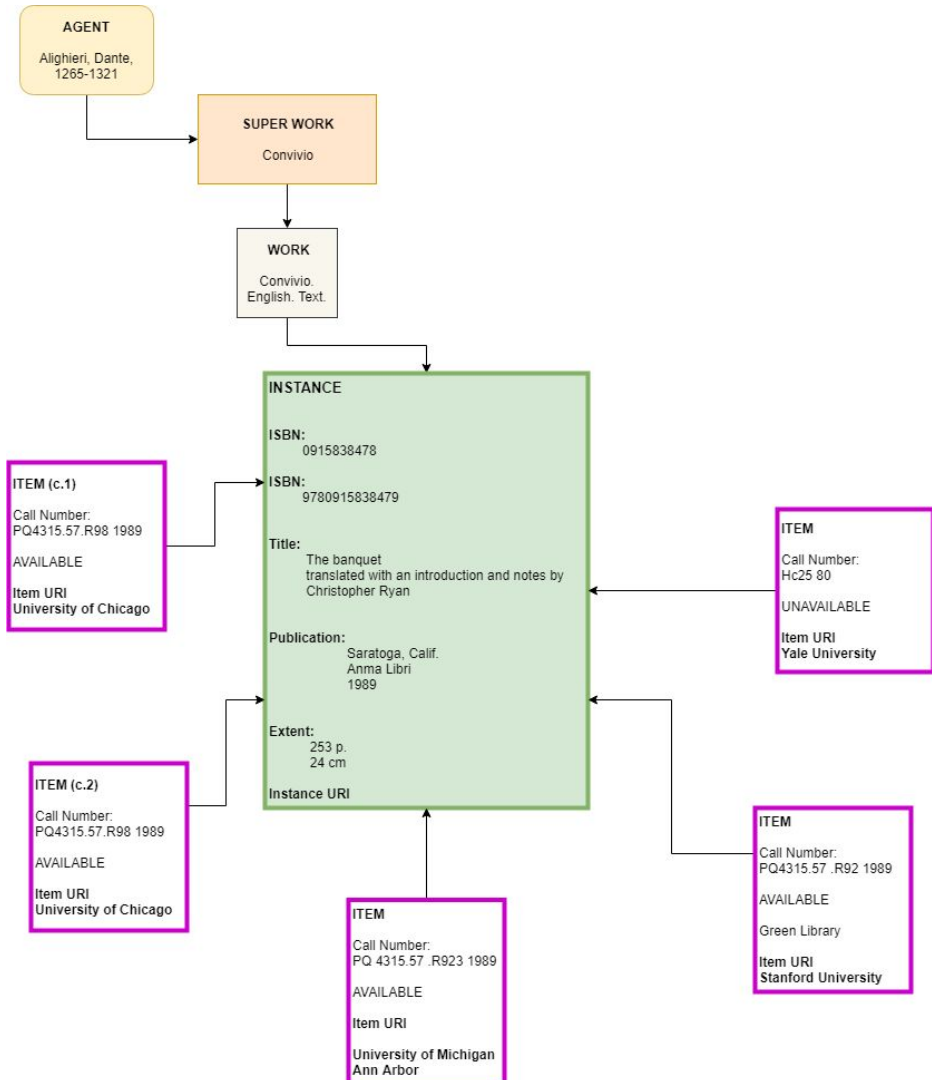
1. multiplication of entities during clustering processes;
2. multiplication of Instances.]

**Draft version of the
Master Instance analysis**



Instance as a *MasterInstance*

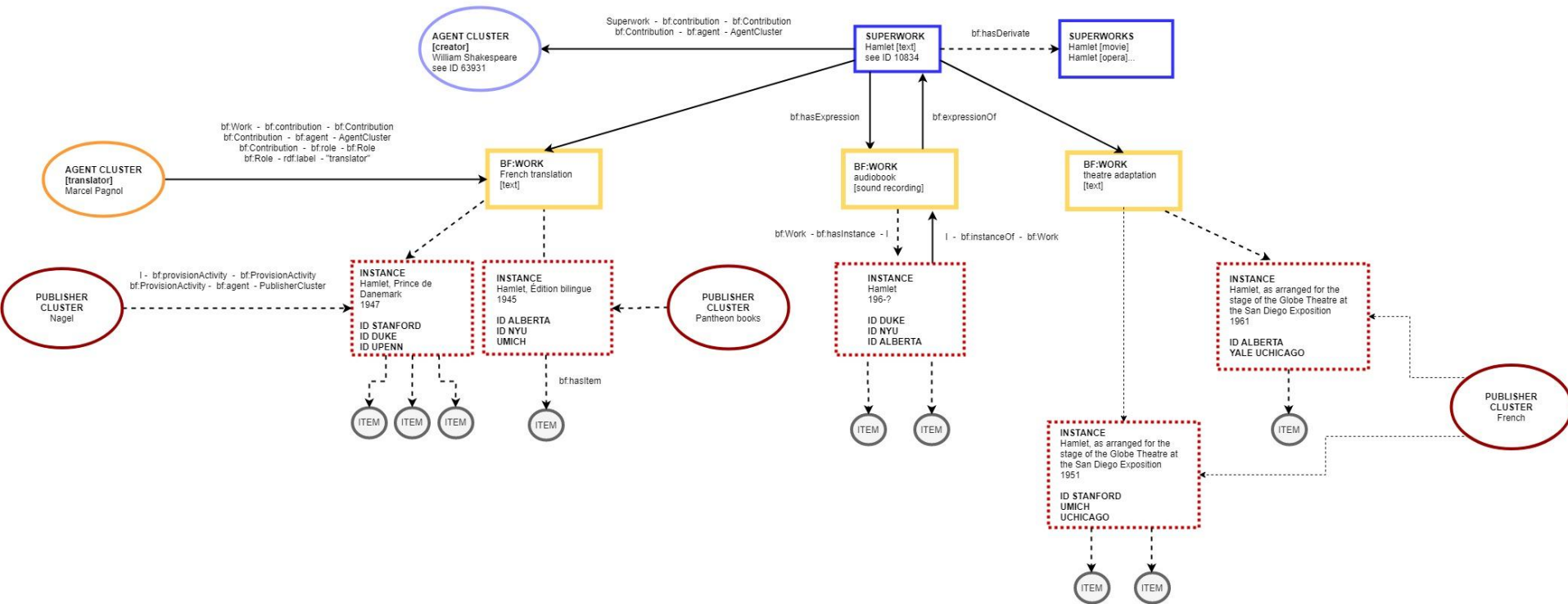




Instance (as a *MasterInstance*) and the related Items

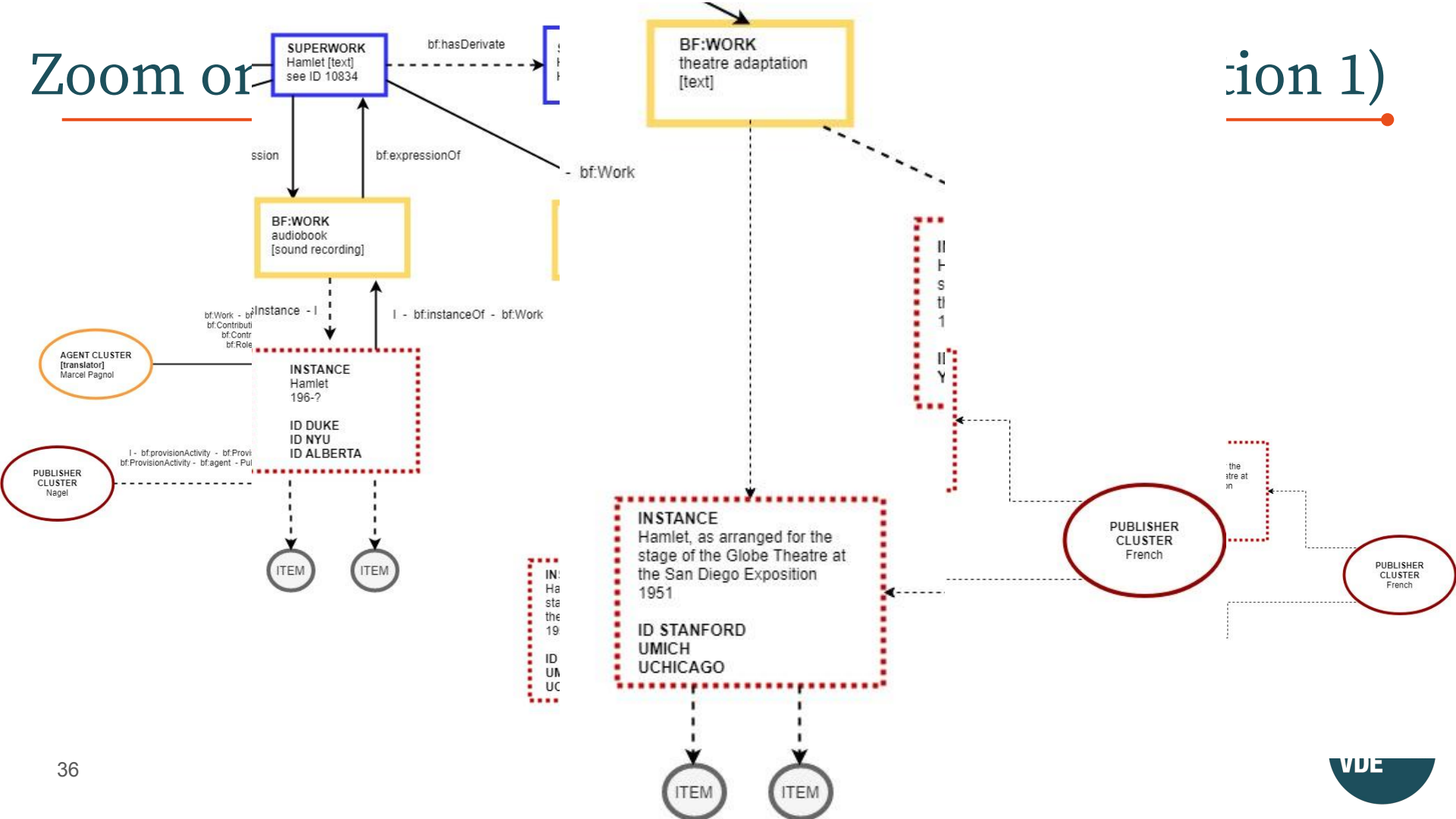
The Share-VDE future entities model (option 1)

SHARE-VDE Entity graph (simplified for UI/UX purposes) - Future version (1)



Zoom on

tion 1)



The Share-VDE future entities model (option 1)

Key concepts of this model:

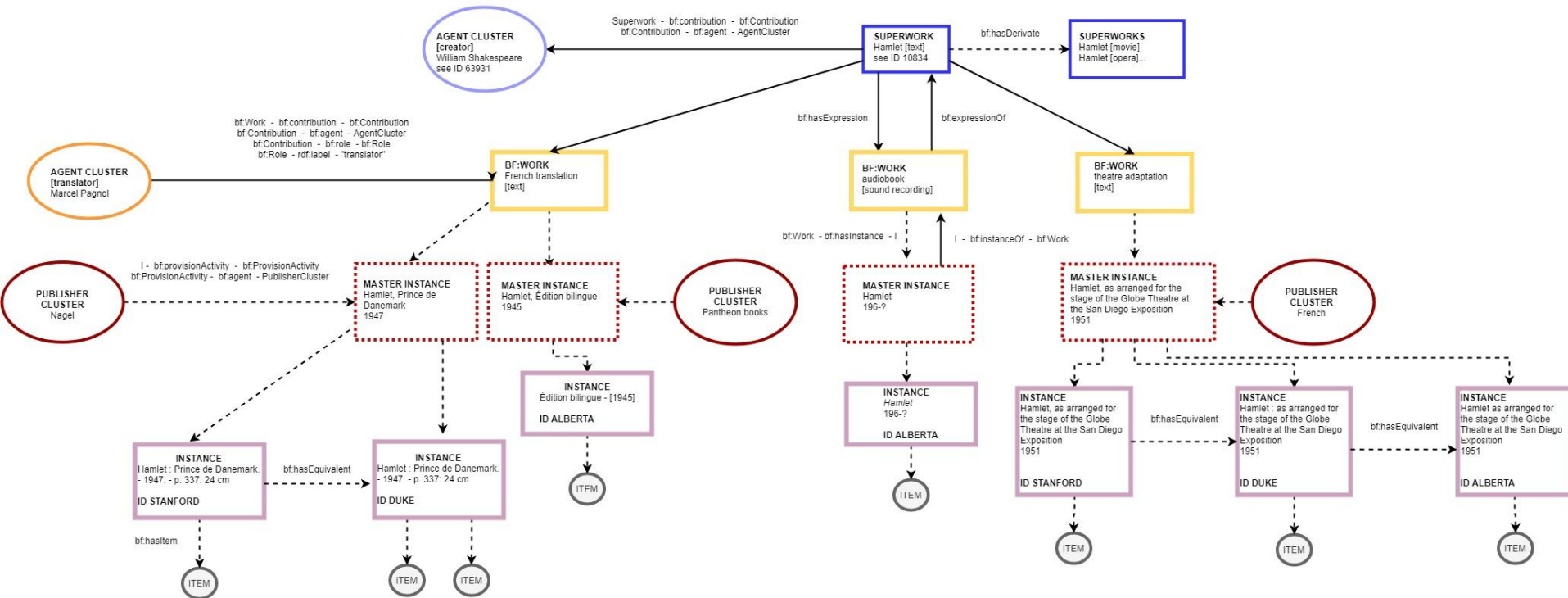
In this scenario the Instance assumes a Share-VDE ID (URI), which does not reflect the "owner" (= the original ID of the library) but an "ideal" Instance representing the "real" instance of BIBFRAME.

To link each one of these instances to each library, we have (at least) two options (or perhaps both together):

- moving local data and (library) information to the Item level;
- including the Provenance to each triple to identify local description of the same Instance (in case the institutions were interested in preserving some specific attributes)

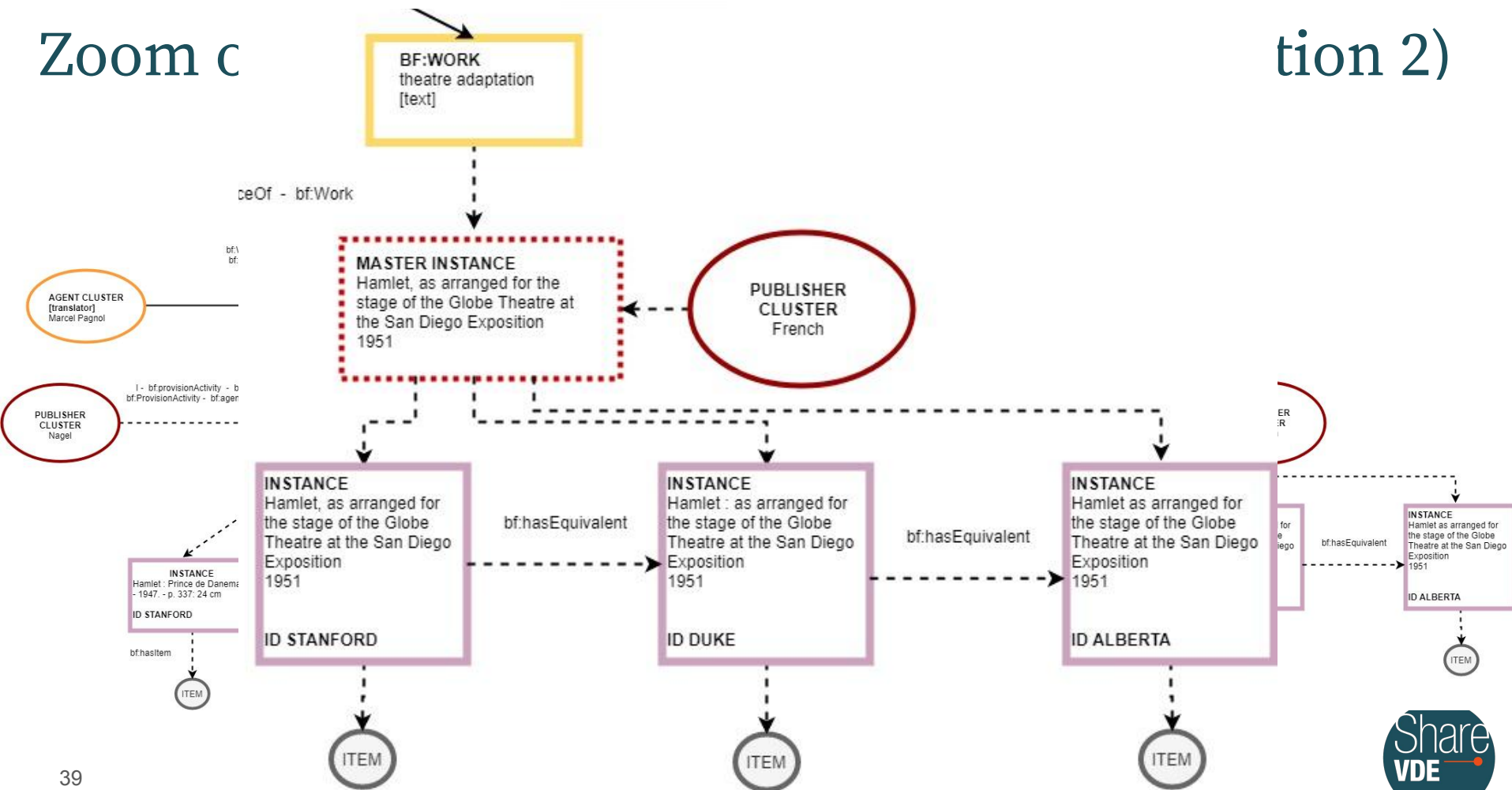
The Share-VDE future entities model (option 2)

SHARE-VDE Entity graph (simplified for UI/UX purposes) - Future version (2)



Zoom c

tion 2)



The Share-VDE future entities model (option 2)

Key concepts of this model:

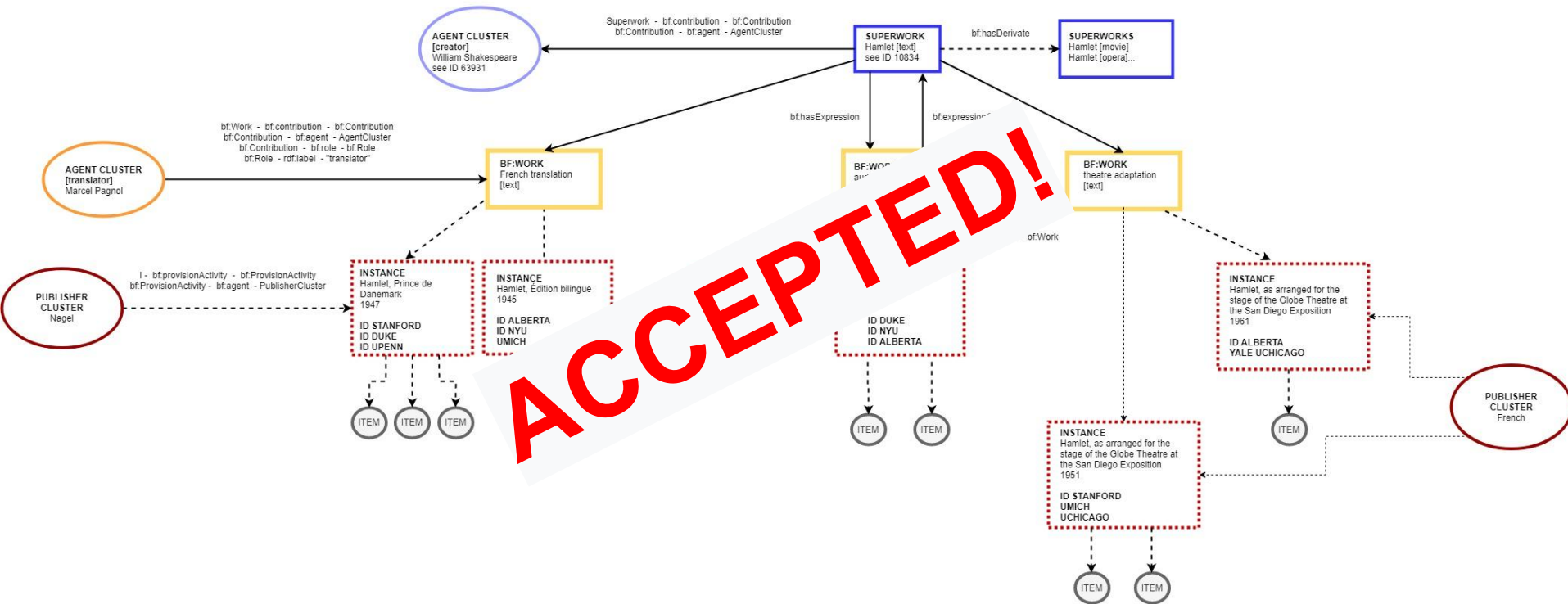
In this scenario a new level is introduced: the Master Instance, that corresponds completely to the BIBFRAME Instance. It assumes a Share-VDE ID (URI), which does not reflect the "owner" (= the original ID of the library) but an "ideal" Instance representing the "real" instance of BIBFRAME.

Under the Master Instance, this scenario proposes the Instances coming from each library, identified by a library ID (URI).

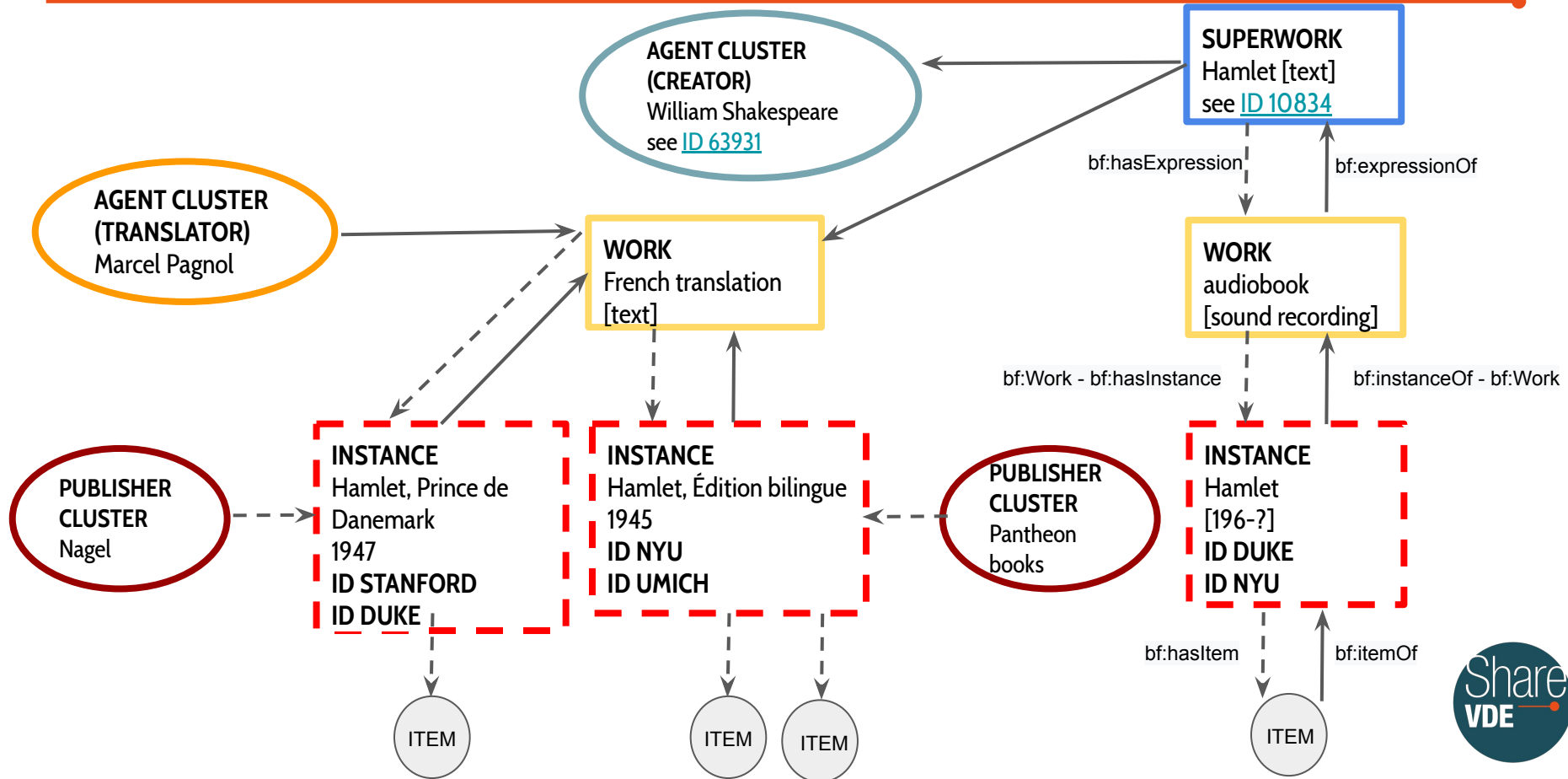
To link the Master Instance with the Instances we need to design a specific predicate (something like "has description") to express a possible "variant" form of the instance description coming from different libraries.

The Share-VDE future entities model (option 1)

SHARE-VDE Entity graph (simplified for UI/UX purposes) - Future version (1)



Future Share-VDE model simplified



How to redesign a model that could be accepted by a wider community



Comparison IFLA-LRM BIBFRAME Share-VDE

IFLA-LRM

Work

Expression

Manifestation

Item

Item

is realized through

realizes

is embodied in

embodies

is exemplified by

exemplifies

BIBFRAME

Hub

Work

Instance

Item

Item

bf:hasExpression

bf:expressionOf

bf:hasInstance

bf:instanceOf

bf:hasItem

bf:itemOf

Share-VDE

SuperWork

Work

Instance

Item

Item

bf:hasExpression

bf:expressionOf

bf:hasInstance

bf:instanceOf

bf:hasItem

bf:itemOf

Entity definitions in Share-VDE

The Work Identification Working Group is starting an interesting conversation around the topic, that is reported, to share opinions and feedback from participants, on an in progress document:

Introducing the OPUS:

A paper to discuss updated entity and model definitions for BIBFRAME and the relationship to IFLA-LRM

“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 serve the same function in BIBFRAME and are hereafter referred to as the Opus in this discussion [...]”.

We all are participating and waiting for results to evaluate how much has to be maintained and how much has to be changed in the model, and in the related data!



Entity definitions in Share-VDE – First step

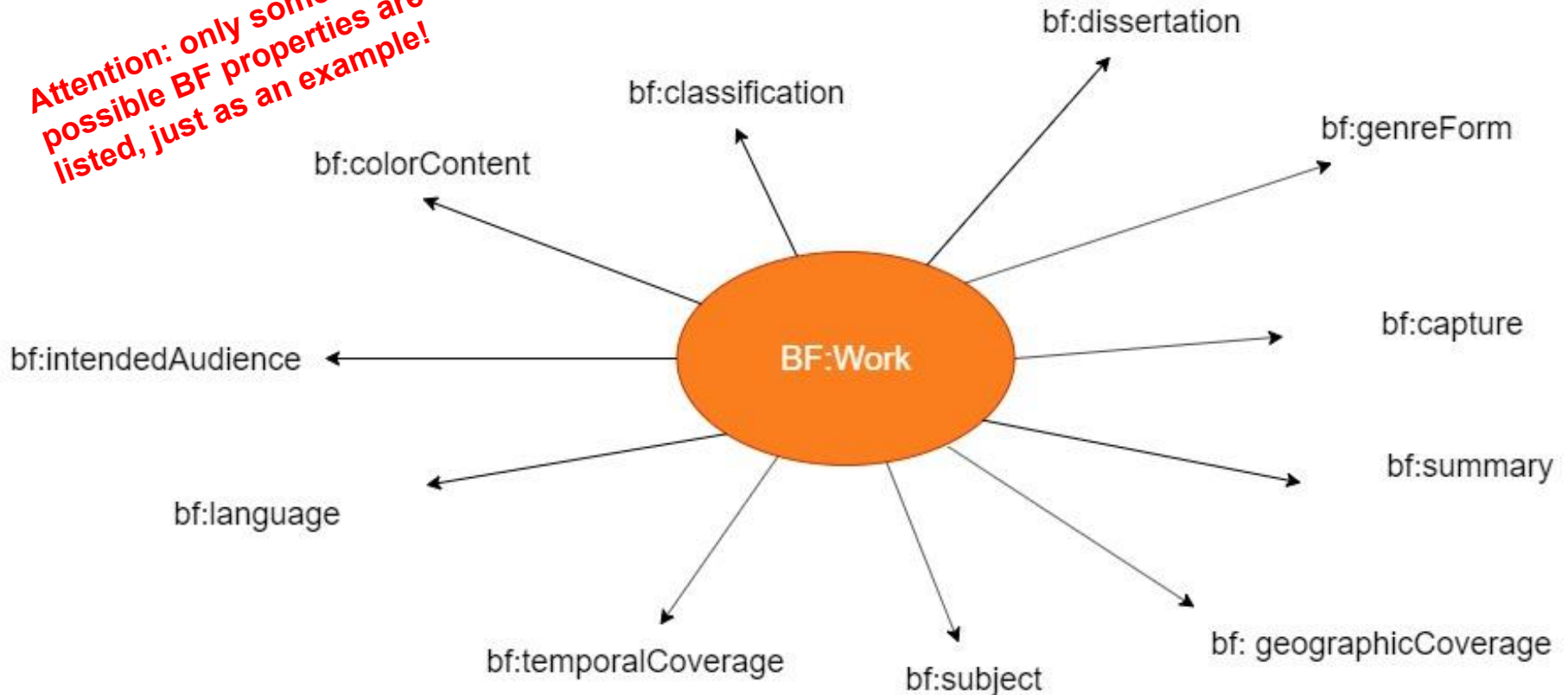
Defining a new entity in a semantic world is not something that concerns a "word" (how to assign a label to a description) but something that concerns a "meaning"



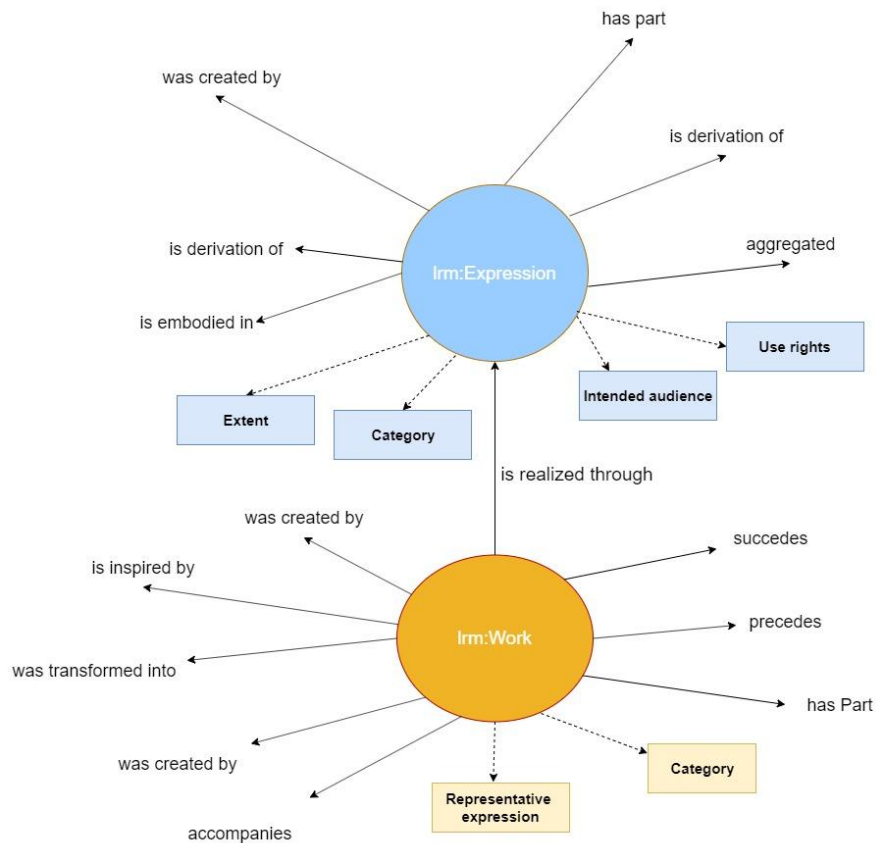
Think having in mind the starting point (MARC 21)
but trying to forget it and going to the meaning of an Entity

Work as an Entity in BIBFRAME

Attention: only some of possible BF properties are listed, just as an example!

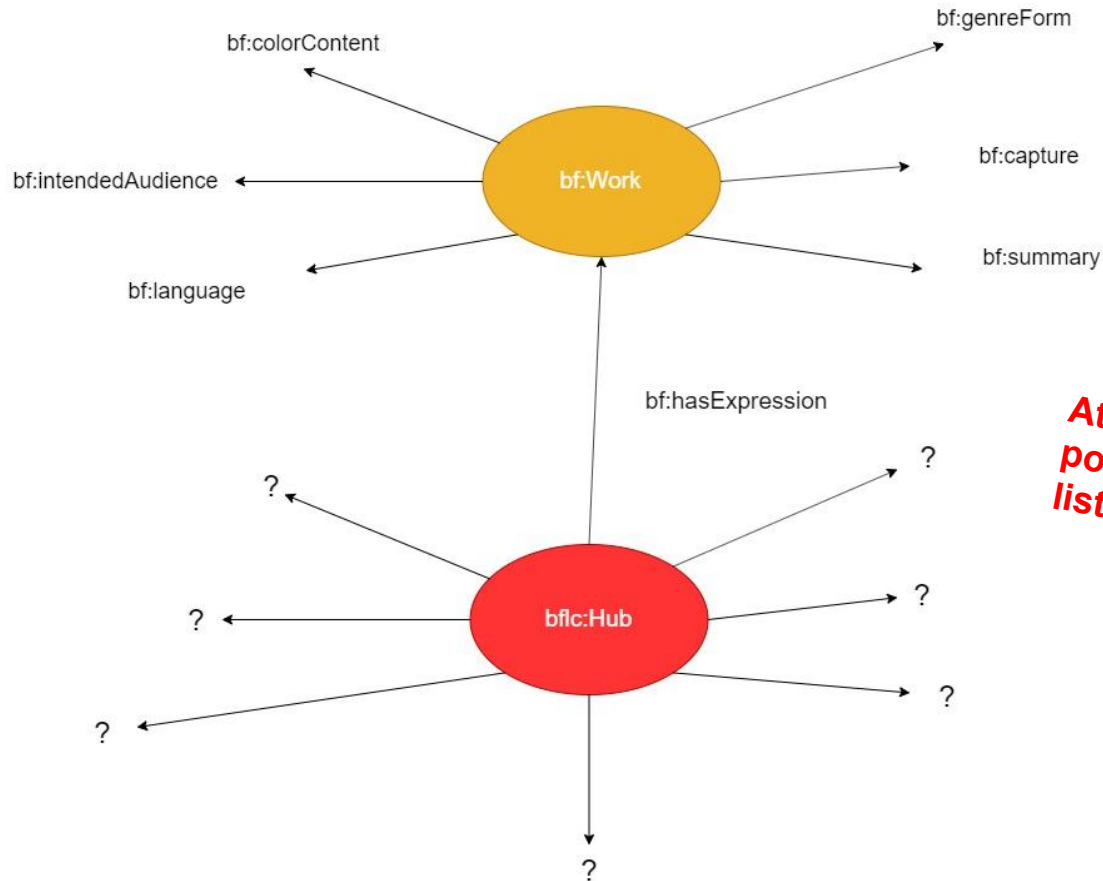


Work and Expression as entities in LRM



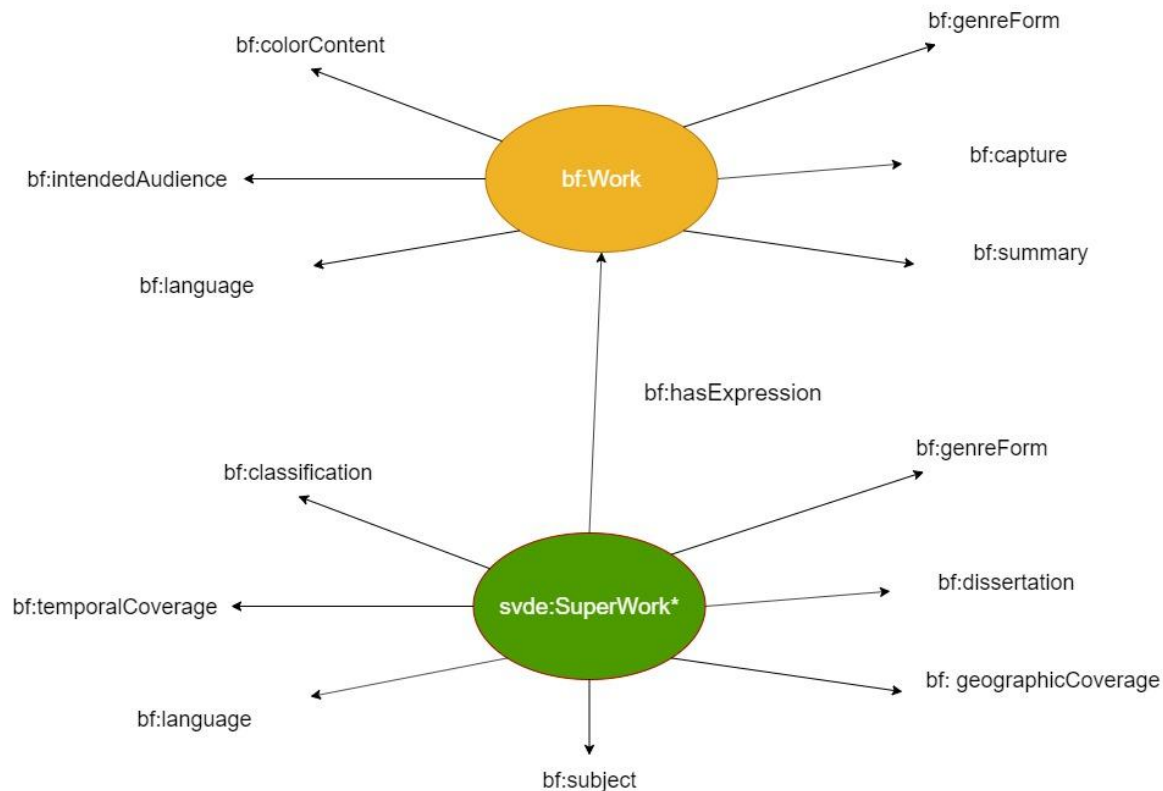
Attention: only some of possible LRM properties and relationships are listed, just as an example!

How to manage Hub as an Entity?



Attention: only some of possible BF properties are listed, just as an example!

SuperWork as an Entity – An option



Entity detection – A natural language analysis



Entity_detection



File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive



Share

100%

€ % .0 .00 123

Times New...

12

B

I

A



fx

	A	B	C	D	E
1	URI Relator term	Relator term	Relator code	English	Italian
2	http://id.loc.gov/vocabulary/relators/abr	Abridger	abr		
3				abridged by	abbreviato da
4				abridged by	abbreviata da
5				abridged by	abbreviati da
6				abridged by	abbreviate da
7				quantitative synthesis by	sintesi quantitativa di
8				quantitative syntheses by	sintesi quantitative di
9				theatrical reduction of	riduzione teatrale di
10				theatrical reductions of	riduzioni teatrali di
11	http://id.loc.gov/vocabulary/relators/act	Actor	act		
12				as arranged by	come disposto da
13				as it is acted by	come è recitato da
14				as it is now acted by	come è recitato adesso da
15				by [name], comedian	di [nome], comico
16				by [name], comedian	di [nome], attore comico
17				by [name], comedian	di [nome], attrice comica
18				by [number] players with the [Company]	con [numero] attori della [Compagnia]



fx

	A	B	C	D	E
1	URI Relator term	Relator term	Relator code	English	Italian
60				with epilogues by	con epiloghi di
61	http://id.loc.gov/vocabulary/relators/aud	Author of dialog	aud		
62	http://id.loc.gov/vocabulary/relators/auj	Author of introduction, etc.	auj		
63				a foreword by	una prefazione di
64				a general introduction, and prefaces and notes to each of the plays by	un'introduzione generale, e prefazioni e note per ciascuna delle scene di
65				an introduction to each play by	un'introduzione per ogni opera teatrale di
66				an introd. to each play by	un'introd. per ogni opera teatrale di
67				an introductory pref. by	una pref. introduttiva di
68				ed. with an introductory memoir by	ed. con un ricordo introduttivo di
69					
70				edited with an introduction by	pubblicato con un'introduzione di
71				edited with an introduction and notes by	pubblicato con un'introduzione e note di
72				edited with an intr. by	pubblicato con un'intr. di
73				edited with an introd. by	pubblicato con un'introd. di
74				edited with forewords by	pubblicato con prefazioni di

A long list of phrases from Marc 21 Tag 245 \$c The list goes on



fx

	A	B	C	D	E
1	URI Relator term	Relator term	Relator code	English	Italian
75				edited with intr. by	pubblicato con intr. di
76				edited with introd. by	pubblicato con introd. di
77				edited with introd., running commentary, glosses, and notes, by	pubblicato con introd., cronaca, glosse, e note, di di
78				edited with introductions by	pubblicato con introduzioni di
79				edited with introductions and notes by	pubblicato con introduzioni e note di
80				edited with introduction, notes, and material for discussion by	modificato con introduzione, note e materiale di discussione di
81				foreword	prefazione
82				forewords	prefazioni
83				foreword by	prefazione di
84				forewords by	prefazioni di
85				foreword of the bibliographer	prefazione del bibliografo
86				forewords of the bibliographer	prefazioni del bibliografo
87				foreword and introductions by	prefazione e introduzioni di
88				forewords and introduction by	prefazioni e introduzione di
89				foreword and additional material by	prefazione e materiale aggiuntivo di

...and it continues...

SuperWork vs Work vs Hub – A conversation



svde:SuperWork - lrm:Work - bf:Hub: some thoughts around the semantics expressed in each entity

Discussion document

The following text has been extracted from the initial conversations on Slack

Initial discussion started on Slack

We assumed to include the Hub entity following the mapping specifications coming from the LoC (I refer to the spreadsheet sent by @kefo), and to manage the SuperWork entity independently: in this scenario, some Works will be exactly the same (both a Hub and a SuperWork) others not. I think this is absolutely correct and reasonable. If we decide to go in





Virtual
Discovery
Environment

Thank you!

tiziana.possemato@atcult.it
tiziana.possemato@casalini.it