LIBER 2020 ONLINE

Tuesday 23rd June - Workshop

LIBER Linked Open Data Working Group – LOD Publication for Libraries

Semantic Interoperability between Bibliographic Models

Sofia Zapounidou[1,2], Michalis Sfakakis[1], Christos Papatheodorou[3]

(1) Ionian University, Dept. of Archives, Library Science & Museology, Corfu, Greece.
(2) Aristotle University of Thessaloniki, Library & Information Centre, Thessaloniki, Greece.
(3) National and Kapodistrian University of Athens, Dept. of History & Philosophy of Science, Athens, Greece.

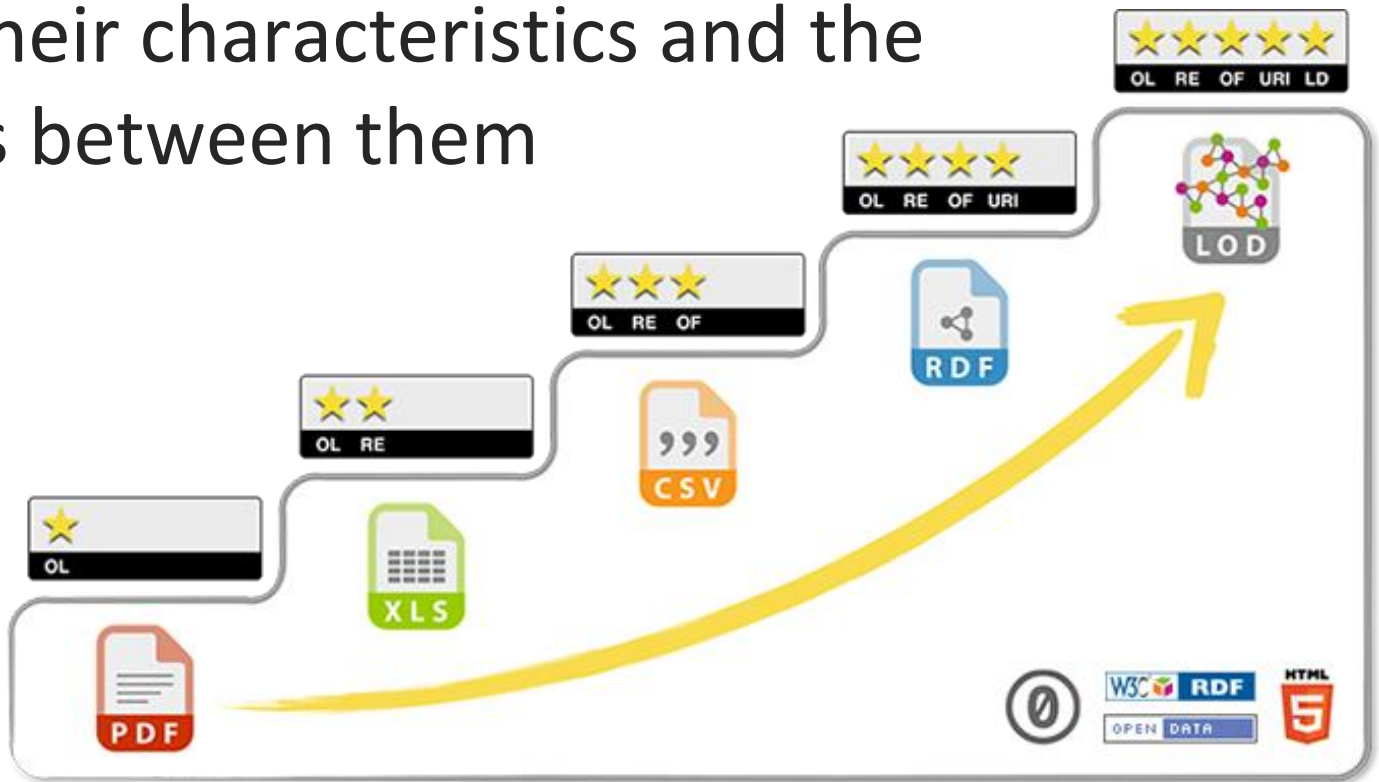
DBIS database & information systems group ionian university

Contents

- **Introduction**
  - SW – LOD
  - Models
  - Examples of datasets
  - The problem
  - The question
- **The Project**
  - Study of models
  - Mapping
  - Assessment
  - Findings
- **Good practices**
- **Final takeaway**

Semantic Web – Linked Open Data

- All things may be described
- Denote “things” with URIs
- Models define what are the “things” we want to talk about, their characteristics and the relationships between them



Models

- FRBR – milestone
- LRM – consolidates FRBR, FRAD, FRSAD
- FRBRoo – cooperation with museums
- RDA – rules for describing according to FRBR/LRM
- BIBFRAME – convert MARC records

!!! Same domain ≠ Differences regarding entities, entities' characteristics, & relationships

Datasets

- Spain – FRBR
- France – FRBR
- Germany – BIBFRAME
- LoC – BIBFRAME
- Sweden – BIBFRAME

!!! Differences regarding

- selected elements sets, &
- elements' values taken from controlled vocabularies.

The problem

- Linked data is created - technical interoperability
- The understanding of the created linked data is not ensured - semantic interoperability

The question

- Is semantic interoperability between bibliographic models feasible?
  - Is there some common ground?
  - Can there be mappings?
  - Are there any prerequisites / good practices ?

SI between Bibliographic Models Project

http://libdata.tab.ionio.gr/models/si-mapping/si_project.html

- Study of models
  - Similarities
  - Divergences

- Can we reconcile differences?

- Mapping
  - BIBFRAME-EDM application profile
  - FRBR-BIBFRAME
  - RDA-BIBFRAME
  - BIBFRAME-RDA

- Are there any prerequisites or good practices for better mappings?

- Assessment – Gold datasets
  - FRBR-BIBFRAME
  - RDA-BIBFRAME
  - BIBFRAME-RDA

- Are the mappings successful?
- Do we lose semantics?

Study of the models: Core entities/classes

Core entities/classes in each model clustered according to intellectual & material embodiment characteristics.

Level \ Models		FRBR	LRM	RDA	FRBRoo	BIBFRAME	EDM
Intellectual	Concepts	Work	LRM-E2 Work	C1001 Work	F1 Work & subclasses	Work & subclasses	Provided Cultural Heritage Object*
	Signs	Expression	LRM-E3 Expression	C1006 Expression	F2 Expression & subclasses		
Material embodiment		Manifestation	LRM-E4 Manifestation	C1007 Manifestation	F3 Manifestation Product Type / F4 Manifestation Singleton	Instance & subclasses	
		Item	LRM-E5 Item	C1003 Item	F5 Item	Item	Web Resource*

* Provided Cultural Heritage Object instances are described in EDM, only if there is at least one digital copy (born digital or digitized) of it. A Web Resource instance provides the URL pointing to the digital copy of a given Provided Cultural Heritage Object. Note that the *edm:ProvidedCHO* class, as equivalent to the union of the FRBR *Work-Expression-Manifestation* entities is expanded semantically to both intellectual and material embodiment levels.

Study of the models: relationships

Representation of relationships in each model.

Bibliographic Relationships	Models	FRBR	LRM	RDA	FRBRoo	BIBFRAME	EDM
Derivative	Adaptation	Work – has adaptation – Work	LRM-E2 Work – LRM-R22i was transformed into - LRM-E2 Work	rdac:C1001 Work - rdaw:P10155 is adapted as work - rdac:C1001 Work	F1 Work – R2i has derivative (type:adaptation) – F1 Work	bf:Work – bf:hasDerivative – bf:Work	edm:ProvidedCHO – dcterms:hasVersion - edm:ProvidedCHO
	Translation	Expression – has a translation – Expression	LRM-E3 Expression - LRM-R24i has derivation (type:translation) - LRM-E3 Expression	rdac:C1006 Expression- rdae:P20171 is translated as rdac:C1006 Expression -	F14 Individual Work – R2i has derivative (type:translation) – F14 Individual Work	bf:Work – bf:translation – bf:Work	edm:ProvidedCHO – edm:isDerivativeOf - edm:ProvidedCHO
Equivalence	Reproduction	Manifestation – has a reproduction – Manifestation	LRM-E4 Manifestation- LRM-E27 has reproduction - LRM-E4 Manifestation	rdac:C1007 Manifestation - rdam:P30039 is reproduced as manifestation - rdac:C1007 Manifestation	F3 Manifestation Product Type – P130i features are also found on (type of similarity:reproduction) – F3 Manifestation Product Type	bf:Instance – bf:hasReproduction- bf:Instance	edm:ProvidedCHO – dcterms:hasFormat - edm:ProvidedCHO*
Aggregates		Expression – is embodied in – Manifestation	LRM-E3 Expression – LRM-E3 is embodied in – LRM-E4 Manifestation	rdac:C1006 Expression- rdae:P20059 has manifestation of expression - rdac:C1007 Manifestation	F2 Expression – P165i is incorporated in – F24 Publication Expression	bf:Work – bf:hasInstance – bf:Instance	edm:ProvidedCHO – inverse of edm:incorporates - edm:ProvidedCHO

** In EDM, equivalence is represented in terms of different digital formats between the related edm:ProvidedCHO instances, e.g., a digitized publication in pdf format and daisy format.*

Study of the models: identifying & resolving differences

First, similarities & differences were identified.

Category	Type	Similarities	Heterogeneities
Semantic	Domain agreements / conflicts	Same or similar domain for bibliographic products	EDM cultural heritage domain.
		Capture same/similar info	Different conceptualizations of real-world bibliographic description cases e.g., core entities, types of bibliographic relationships, constraints
	Terminological (mis)matches	FRBR, FRBRoo, LRM, RDA – WEMI BIBFRAME Item – with WEMI Item entity	Work different in FRBR and BIBFRAME Common terms with different meaning, e.g. Work Different terms with same meaning, e.g., edition designation E-R versus Semantic Web/RDF terminology
		Many common terms, e.g. statement of responsibility	

Secondly, the project tried to reconcile each identified difference.

Category	Type	Heterogeneities	Project's approach
Semantic	Domain conflicts	EDM cultural heritage domain. Different conceptualizations of real-world bibliographic description cases e.g., core entities, types of bibliographic relationships, constraints	EDM application profile may add granularity with skos extension mechanism
	Terminological mismatches	Work different in FRBR and BIBFRAME Common terms with different meaning, e.g. Work Different terms with same meaning, e.g., edition designation E-R versus Semantic Web/RDF terminology	Study of each model's definitions Check LC conversions from MARC21 to BF BIBFRAME mailing list

!!! The Haslhofer & Klas categorization of structural and semantic metadata heterogeneities was used

SI between Bibliographic Models Project

http://libdata.tab.ionio.gr/models/si-mapping/si_project.html

- Study of models
  - Similarities
  - Divergences

- Can we reconcile differences?

- **Mappings**
  - BIBFRAME-EDM application profile
  - FRBR-BIBFRAME
  - RDA-BIBFRAME
  - BIBFRAME-RDA

- Are there any prerequisites or good practices for better mappings?

- Assessment – Gold datasets
  - FRBR-BIBFRAME
  - RDA-BIBFRAME
  - BIBFRAME-RDA

- Are the mappings successful?
- Do we lose semantics?

Mappings & Assessment: Core findings

- Representation approaches
- Controlled vocabularies
- Loss of semantics in certain cases

- Importance of cataloging policy

Examples follow …

Mappings: Representation approaches I

- Models enable different representation approaches without violating their primitives

!! Different approaches may trigger different mappings

1

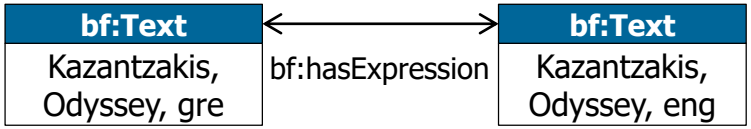
- Simple representation of two *Works* in BIBFRAME.
- No information regarding possible relationships between them.
- The *Works* remain unrelated.

Mappings: Representation approaches I

- Models enable different representation approaches without violating their primitives

!! Different approaches may trigger different mappings

2



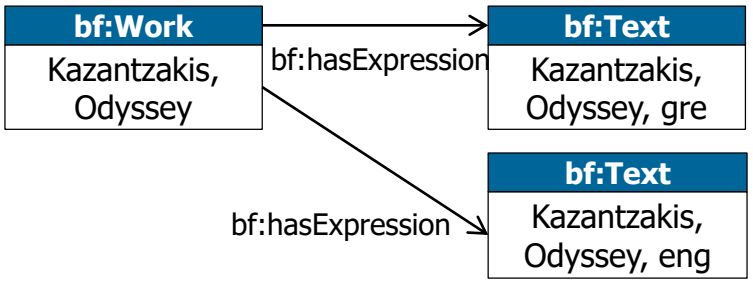
- The two *Works* are related with two *bf:hasExpression* relationships.
- Both *Works* share the same content (Kazantzakis, Odyssey) but in different signs/languages (Greek - English).

Mappings: Representation approaches I

- Models enable different representation approaches without violating their primitives

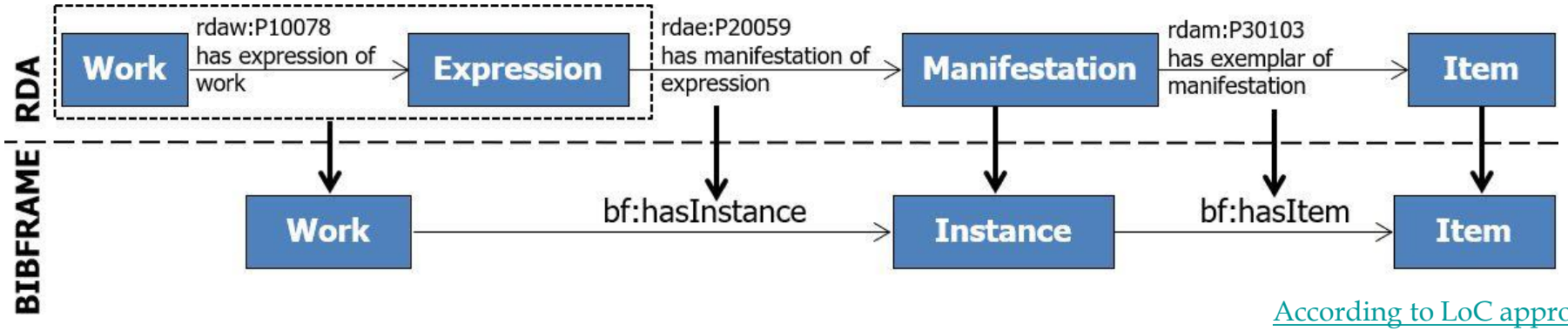
!! Different approaches may trigger different mappings

3



- A *Work* lacking signs-related information (Kazantzakis, Odyssey) is expressed in two *Works*.
- Not the same with FRBR/LRM/RDA
- The two *Works* still carry both ideas (Kazantzakis, Odyssey) and signs (Greek - English).

Mappings: Representation approaches III

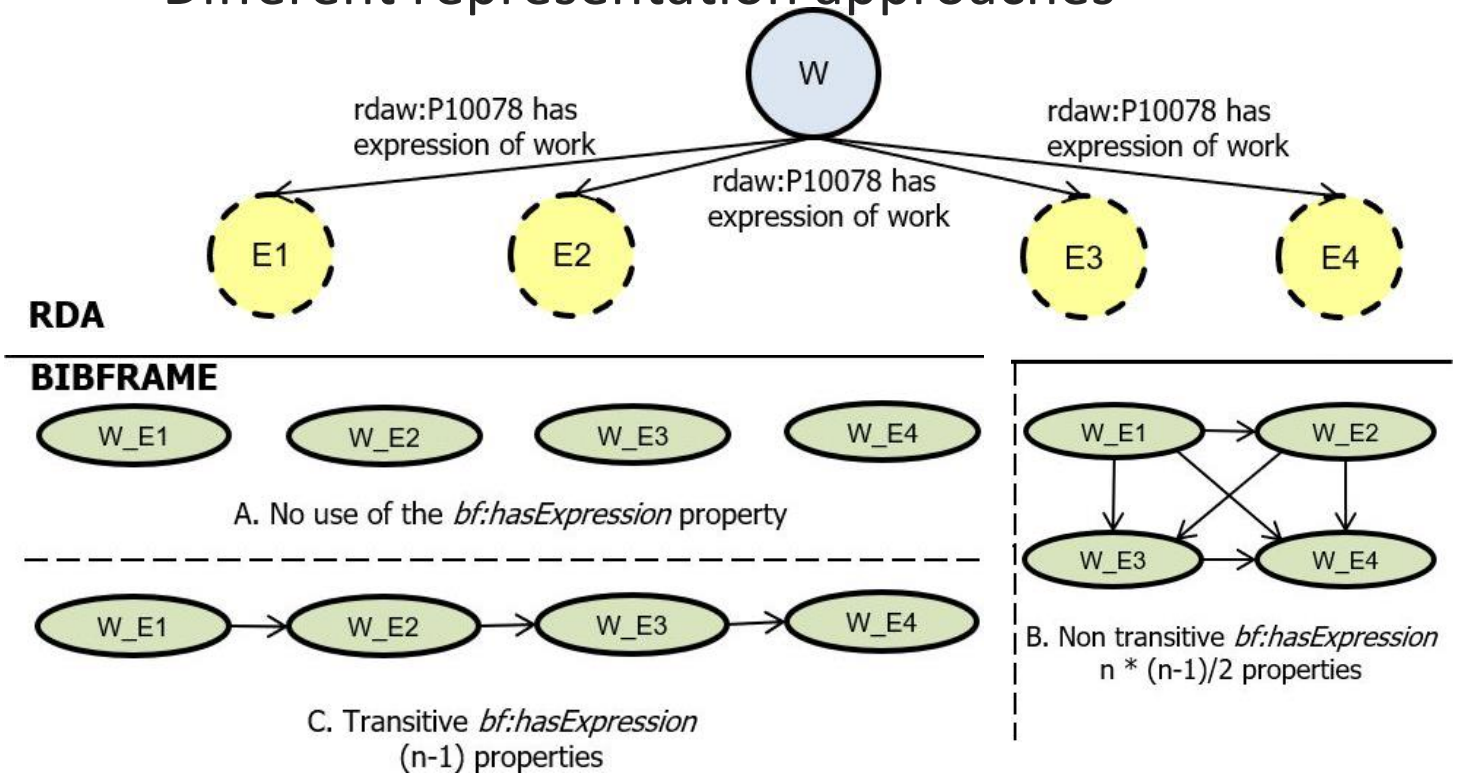


According to LoC approach

- Seems simple, right?
- Actually, no! It is not that simple.

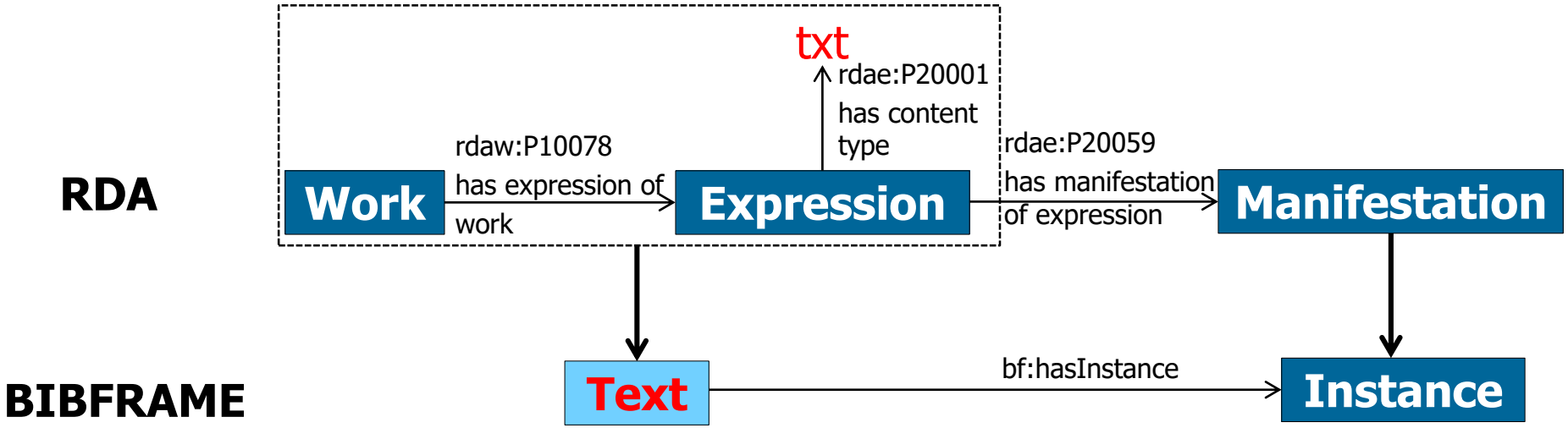
Mappings: Representation approaches IV

- How do we preserve the information that two *bf:Works* share the same ideational content?
- Different representation approaches



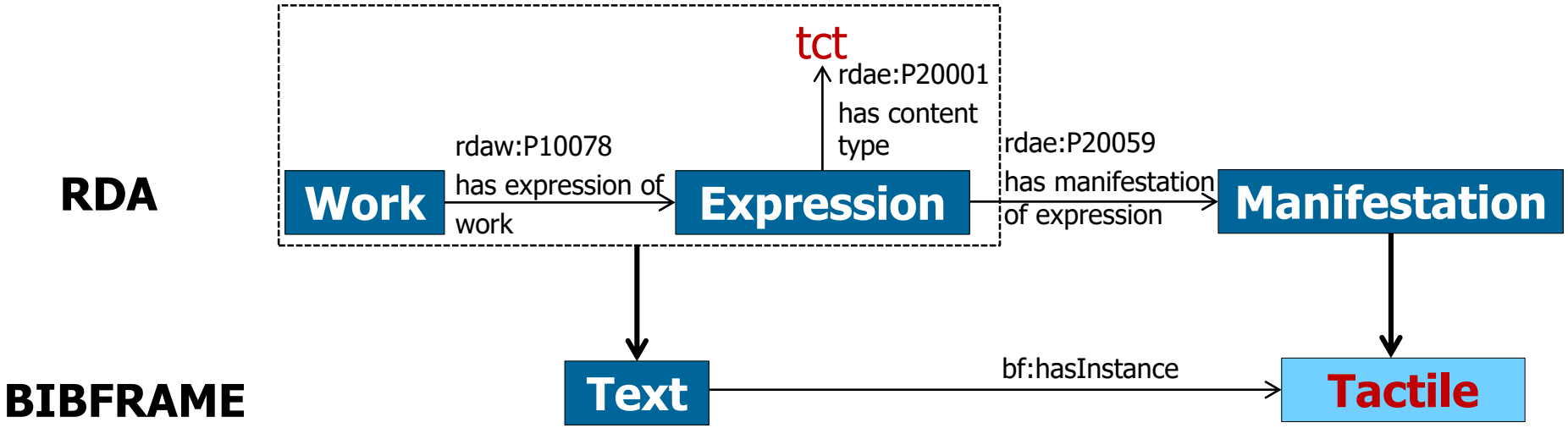
Mappings: Controlled vocabularies I

- They may trigger mappings

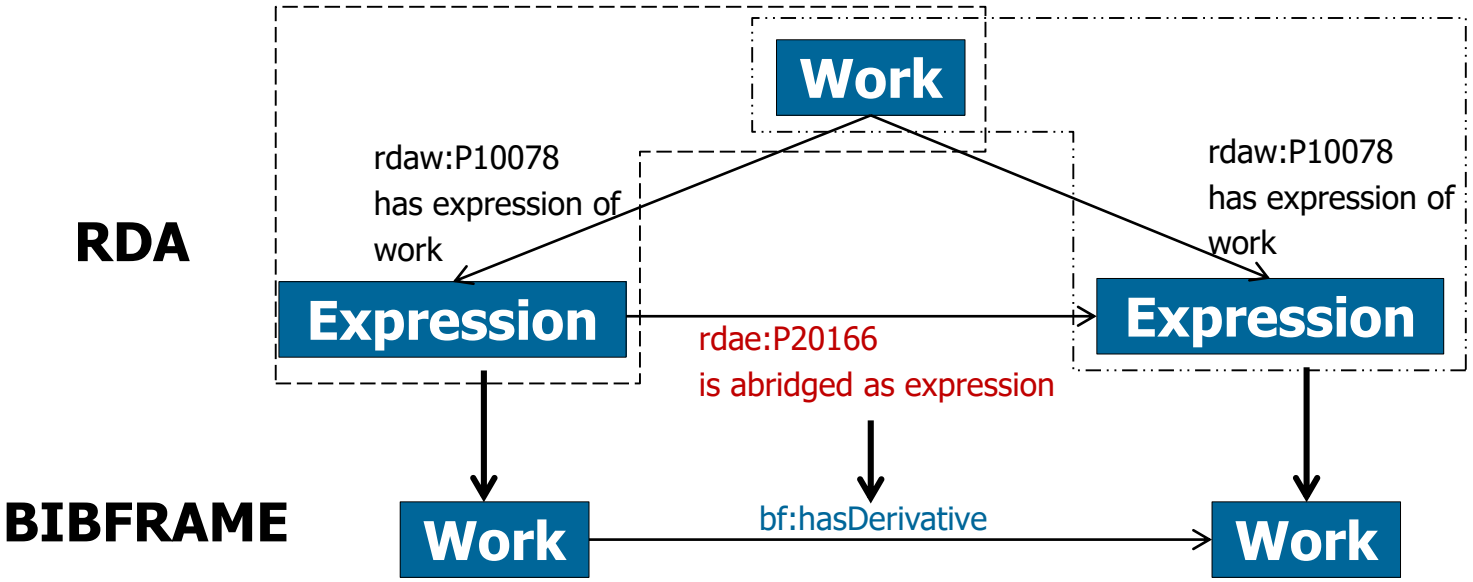


Mappings: Controlled vocabularies II

- They may trigger mappings

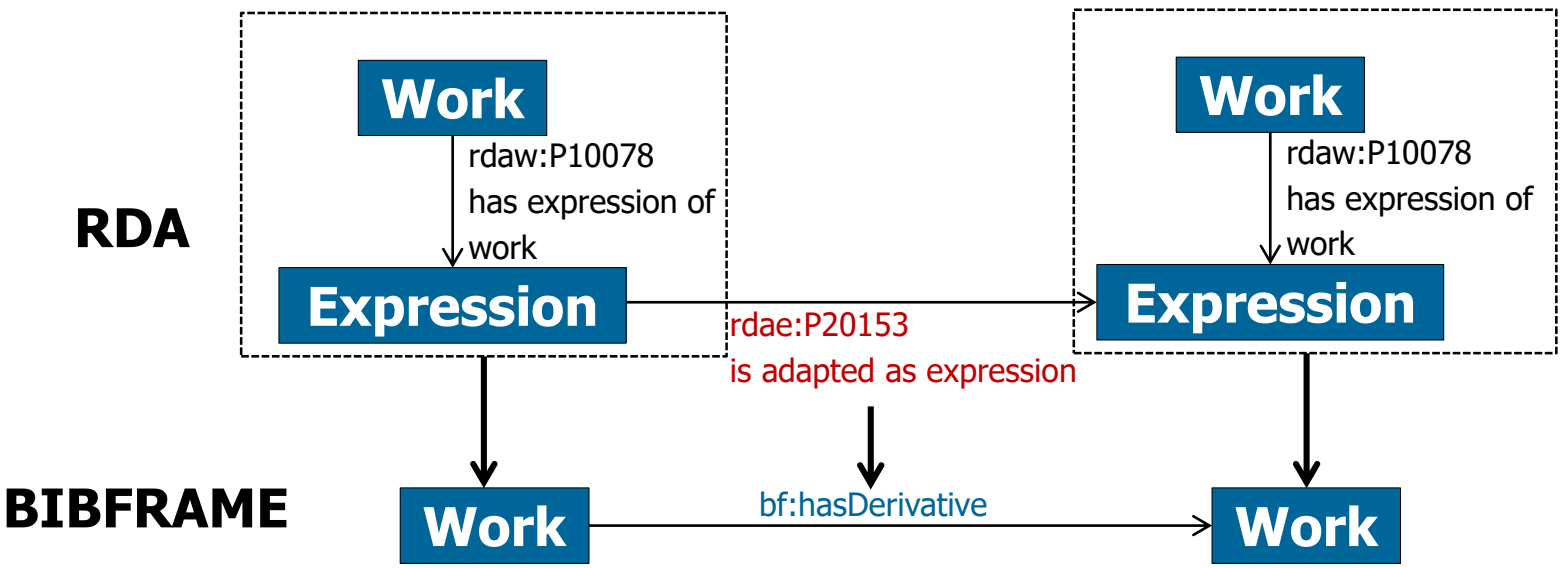


Mappings: Losses of semantics in relationships I



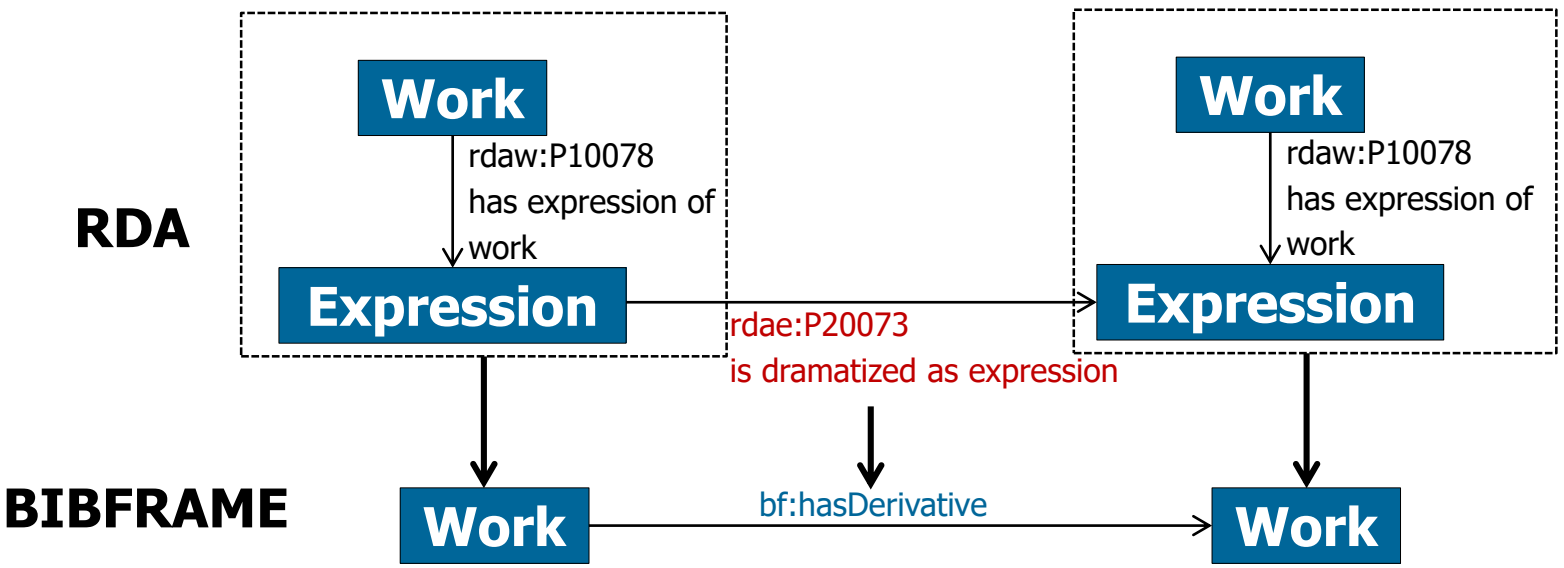
!!! The *Expression-Expression* relationship is preserved. Exact semantics is lost.

Mappings: Losses of semantics in relationships II



!!! The *Expression-Expression* relationship is preserved. Exact semantics is lost.

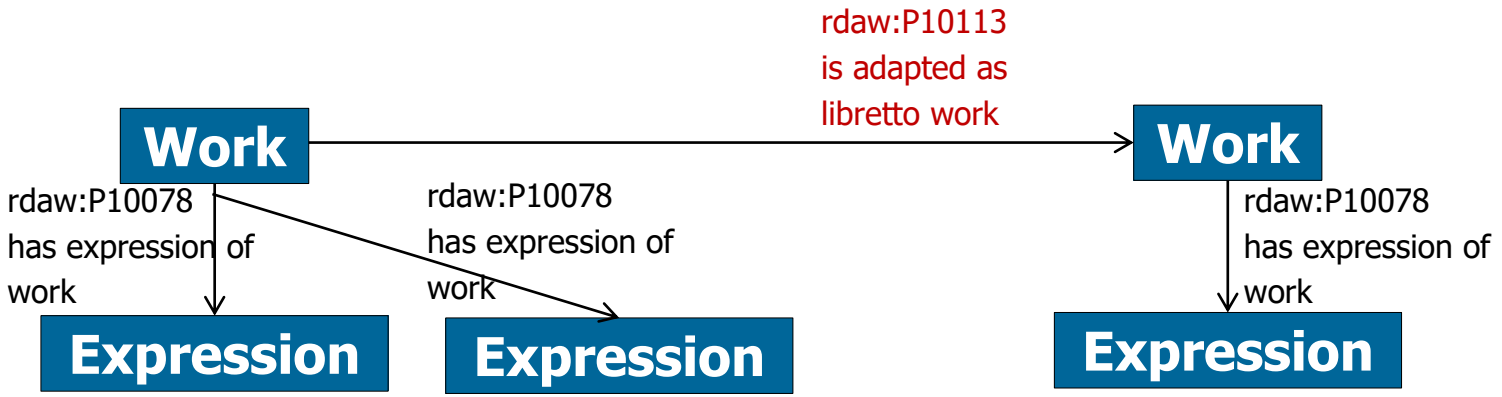
Mappings: Losses of semantics in relationships III



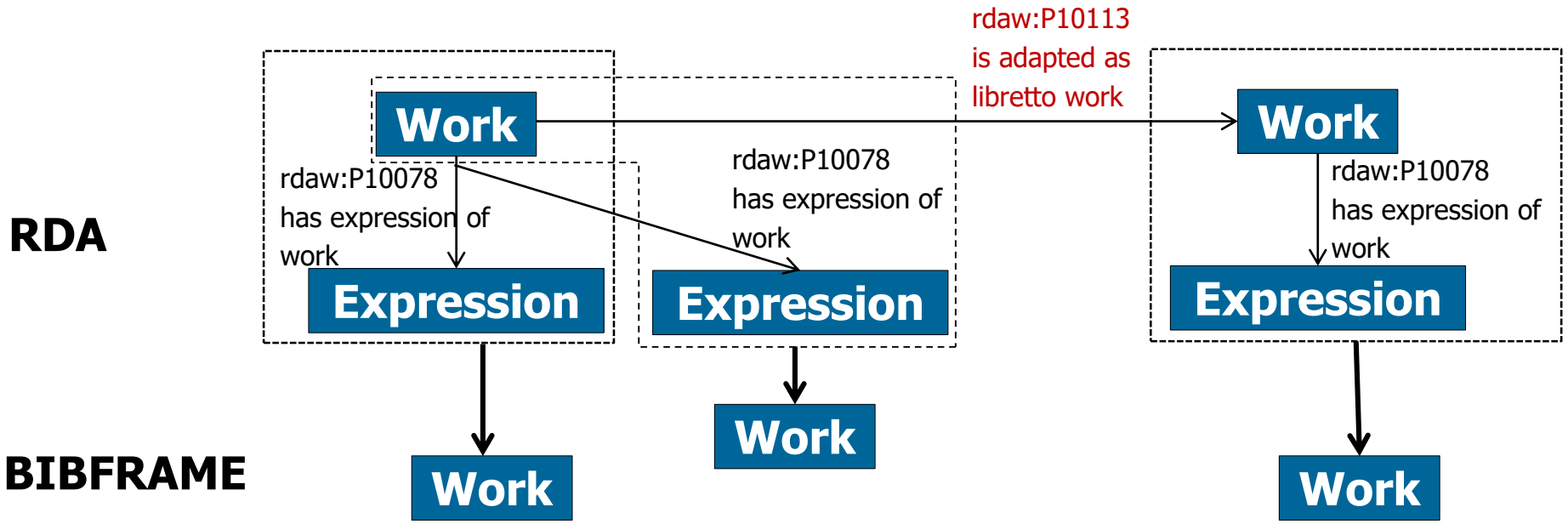
!!! The *Expression-Expression* relationship is preserved. Exact semantics is lost.

Mappings: Losses of semantics in relationships IV

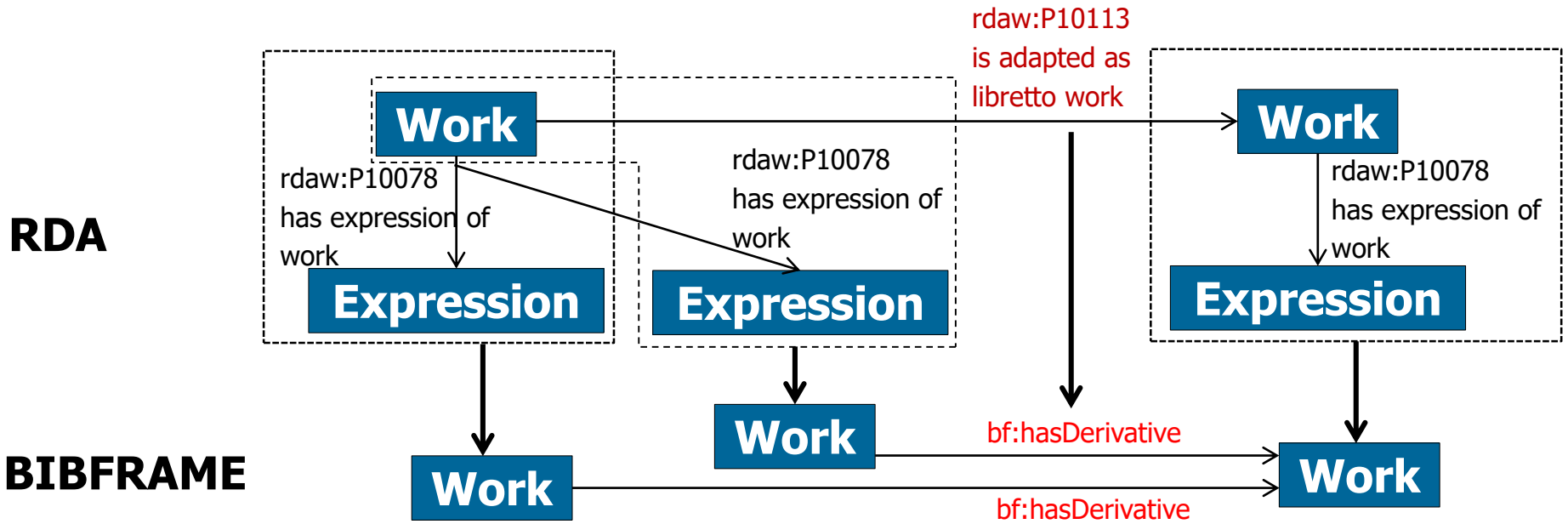
RDA



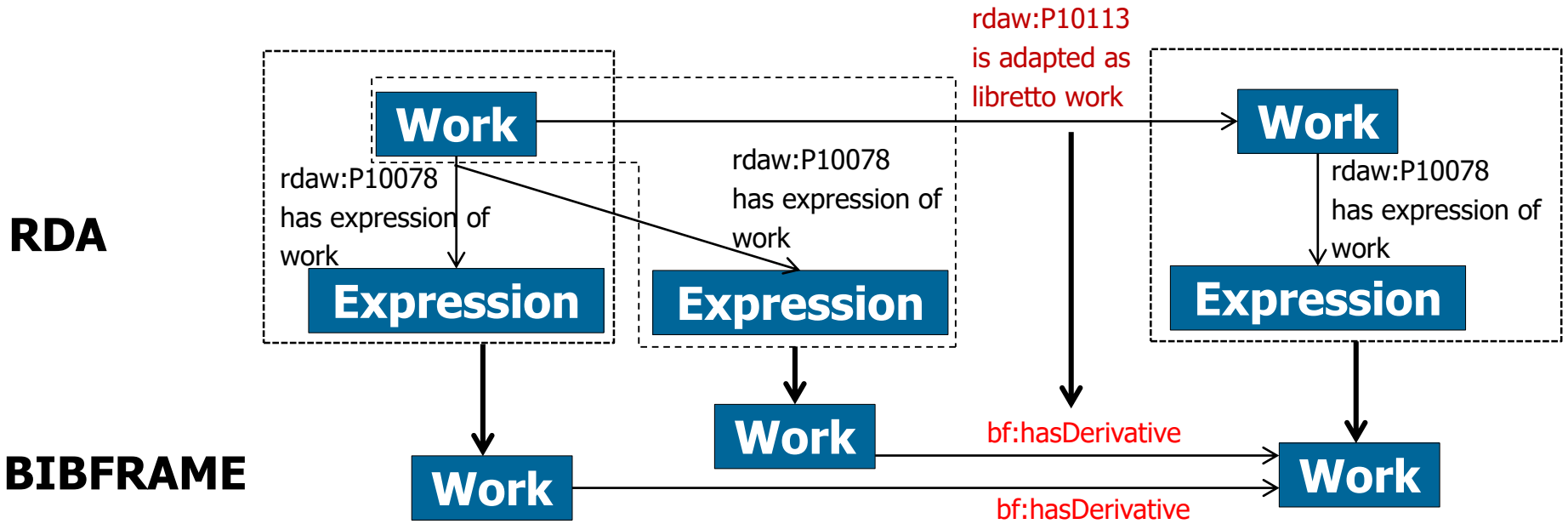
Mappings: Losses of semantics in relationships IV



Mappings: Losses of semantics in relationships IV

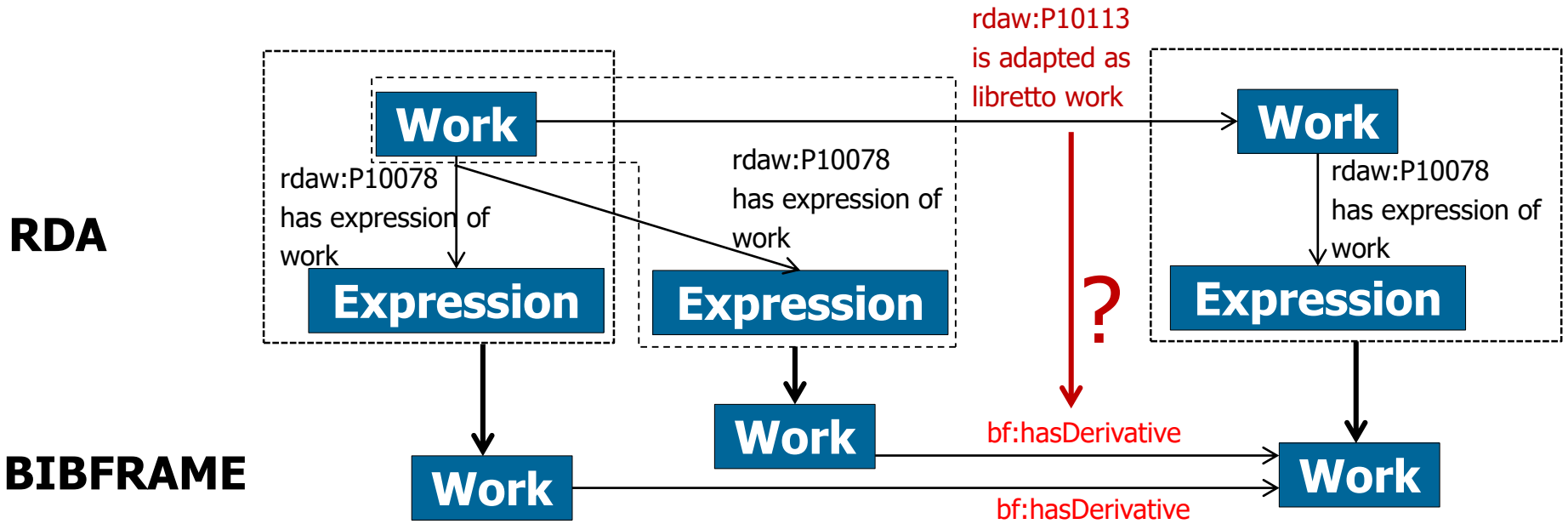


Mappings: Losses of semantics in relationships IV



!!! The mapping of *Work to Work* relationships from RDA to BIBFRAME generates many erroneous relationships

Mappings: Losses of semantics in relationships IV



!!! Ignore *Work to Work* relationships in RDA-BIBFRAME mappings to avoid generation of too many erroneous relationships

Assessment

- Real – world cases
  - 11 bibliographic families
  - Different languages
  - Many derivations
- Creation of Gold datasets
- Conversion & assessment
  - E.g. Gold RDA converted to BIBFRAME (BF2RDA) & BF2RDA compared to Gold BIBFRAME

http://libdata.tab.ionio.gr/models/si-mapping/si_project.html

DBIS database & information systems group ionian university

Semantic Interoperability between Bibliographic Conceptual Models

Introduction

Integration of library data into the semantic web (SW) demands a shift in conceptual data models and data format according to the SW principles and standards. New models have been developed for representing bibliographic information according to the new needs and formats. Well-known models are the Functional Requirements for Bibliographic Records (FRBR) and its consolidated version Library Reference Model (LRM), the FRBR Object-Oriented version (FRBRoo), the Resource Description and Access (RDA) which is based on the FRBR/LRM, and the Bibliographic Framework (BIBFRAME). There also exist cultural heritage models, such as the Europeana Data Model (EDM).

Semantic interoperability between bibliographic conceptual models is a prerequisite for seamless navigation into the bibliographic universe. Moreover, explicit representation of relationships between bibliographic entities may enable the navigability of bibliographic information even more. Taken into account that derivative relationships are common among the bibliographic universe, there is a focus on them in the framework of this project. This project seeks to contribute to semantic interoperability by:

- studying models' constructs to discover similarities and differences between the models.
- mappings in terms of core classes/entities, inherent relationships, and derivative relationships.
- assessing the mappings using a testbed.

Materials

Tools

- MARC to BIBFRAME 2 Convertor v0.0.1
- Virtuoso SPARQL Query Editor
- μέτα-Composer, meta-search work composer
- DBIS OPAC with a sample MARC21 dataset for experimentation

Data

MARC/XML

One MARC XML file containing a collection of 235 bibliographic records from the National Library of Spain and the Library of Congress. This file contains records for the following bibliographic families:

1. Cien años de soledad, 14 records
2. Преступление и наказание (Crime and Punishment), 24 records
3. Don Quijote, 11 records
4. Faust, 25 records
5. Ἰλιάς (Iliad), 25 records
6. Бра́тья Карама́зовы (Karamazov Brothers), 20 records
7. Madame Bovary, 29 records
8. Ὀδύσσεια (Odyssey), 19 records
9. The Scarlet letter, 19 records
10. Tom Sawyer, 31 records
11. Wuthering Heights, 18 records

Gold datasets

Two Gold Datasets have been developed to assess the mappings. They have been uploaded in the Virtuoso RDF server and SPARQL queries can be submitted.

Gold Datasets	SPARQL queries & Visualizations
Gold RDA dataset (.rdf)	Gold RDA dataset: graph IRIs, prefixes, SPARQL queries
Gold BIBFRAME dataset (.rdf)	Gold BIBFRAME dataset: graph IRIs, prefixes, SPARQL queries

Findings

- **Differences** exist but may be reconciled
- **Representation approaches** trigger different mappings
- **Property values** may be used to trigger mappings
- Selection of **value vocabularies** may play an important role in mappings.
- **Level** on which a bibliographic relationship is represented impacts interoperability, e.g. Work or Expression in RDA

Good practices

- Use commonly accepted representation approaches
  - E.g. translations are new *Expressions* of the same *Work* and NOT new *Works.*
- Represent relationships at the signs level, if possible
  - Between *Expressions* in RDA
  - Between *bf:Works* in BIBFRAME
- Use controlled vocabularies, well-known, if possible
- Use of *bf:hasExpression* for common ideational content

Final Takeaway

- Cataloging policies play a crucial role in minimizing the loss of semantics after conversions.
  - Understand the models' semantics
  - Formulate your cataloging policy accordingly

Thank you!
Questions?

Sofia Zapounidou

http://libdata.tab.ionio.gr/models/si-mapping/si_project.html