WCCQ#1821856, VOL 0, ISS 0

Mapping Derivative Relationships from BIBFRAME 2.0 to RDA

Michalis Sfakakis, Sofia Zapounidou, and Christos Papatheodorou

QUERY SHEET

This page lists questions we have about your paper. The numbers displayed at left are hyperlinked to the location of the query in your paper.

The title and author names are listed on this sheet as they will be published, both on your paper and on the Table of Contents. Please review and ensure the information is correct and advise us if any changes need to be made. In addition, please review your paper as a whole for typographical and essential corrections.

Your PDF proof has been enabled so that you can comment on the proof directly using Adobe Acrobat. For further information on marking corrections using Acrobat, please visit http://journalauthors.tandf.co.uk/production/acrobat.asp; https://authorservices.taylorandfrancis.com/how-to-correct-proofs-with-adobe/

The CrossRef database (www.crossref.org/) has been used to validate the references.

AUTHOR QUERIES

Q1 Please check and resupply corresponding author email address if the given is inaccurate.

Q2 Please check that the heading levels have been correctly formatted throughout.

Q3 Please note that the ORCID section has been created from information supplied with your manuscript submission/CATS. Please correct if this is inaccurate.

CATALOGING & CLASSIFICATION QUARTERLY
https://doi.org/10.1080/01639374.2020.1821856

Mapping Derivative Relationships from BIBFRAME 2.0 to RDA

Q3 Michalis Sfakakis[a], Sofia Zapounidou[a], and Christos Papatheodorou[b]

[a]Department of Archives, Library Science and Museology, Ionian University, Corfu, Greece; [b]Department of History and Philosophy of Science, National and Kapodistrian University of Athens, Athens, Greece

ABSTRACT

The mapping from BIBFRAME 2.0 to Resource Description and Access (RDA) is studied focusing on core entities, inherent relationships, and derivative relationships. The proposed mapping rules are evaluated with two gold datasets. Findings indicate that 1) core entities, inherent and derivative relationships may be mapped to RDA, 2) the use of the *bf:hasExpression* property may cluster *bf:Works* with the same ideational content and enable their mapping to RDA Works with their Expressions, and 3) cataloging policies have a significant impact on the interoperability between RDA and BIBFRAME datasets. This work complements the investigation of semantic interoperability between the two models previously presented in this journal.

ARTICLE HISTORY
Received June 2020
Revised September 2020
Accepted September 2020

KEYWORDS
BIBFRAME; conceptual models; derivative bibliographic relationships; explore; interoperability; mappings; RDA

Introduction

Linked data technologies facilitate the publication of structured data in a machine-understandable way by providing links between the data that correlate them semantically. In this context, data models enable the understanding of data, while links provide the mechanism for further exploration. In the library domain, bibliographic models enable the definition and understanding of the entities of the bibliographic universe, while bibliographic relationships existing among bibliographic entities may provide further navigation opportunities. Today, there exist different bibliographic models identifying different entities and relationships among them.[1] Toward the vision of an integrated bibliographic universe where users may seamlessly explore bibliographic data, these different models need to interoperate. Mappings are going to be needed preserving information and semantics. Q2

CONTACT Sofia Zapounidou l12zapo@ionio.gr Department of Archives, Library Science and Museology, Ionian University, Corfu, Greece. Q1

This study aims to contribute to the interoperability of two well-known models, RDA and BIBFRAME 2.0. *RDA* is the de facto content standard used in libraries worldwide. The adoption of RDA enables the identification of bibliographic entities in legacy MARC records and prepares the ground for future conversions of legacy data to linked data. Although, the intention of RDA Toolkit[2] is not currently to cover linked data aspects, at the same time there is an RDA vocabulary[3] aligned with the FRBR/LRM models enabling the representation and publication of bibliographic data as linked data. The BIBFRAME model and conversion tools have been developed by the Library of Congress to convert MARC records to linked data. These two models will be, without doubt, paramount to the inclusion of the library data into the LOD context, therefore, they must be interoperable.

The study investigates and evaluates the structural and semantic interoperability between BIBFRAME and RDA by focusing on the core conceptualizations of each model, core entities and inherent relationships, and on derivative bibliographic relationships as linking mechanisms during the evolution of the intellectual or artistic content in time. The selection of derivative bibliographic relationships was made for the following reasons: (i) derivations are really common in the bibliographic universe, e.g., translations, revisions, abridgments, adaptations, dramatizations, etc.[4] (ii) *bibliographic families*, networks of related works somehow deriving from the same *progenitor* work, usually start with a derivation,[5] and (iii) representation of derivative bibliographic relationships supports the representation of works in the context of their families, as well as further exploration using the links relating the members of the family.

The current study, by developing and evaluating the BIBFRAME to RDA mapping, complements a previous work,[6] where the mapping of core entities, inherent relationships, and derivative relationships from RDA to BIBFRAME 2.0 was presented. The mapping from the granular model RDA to the less granular one BIBFRAME has not been straightforward due to the models' different conceptualizations. RDA core entities and inherent relationships were successfully mapped to BIBFRAME. However, the mapping of derivative relationships ended with the loss of specificity in most types of derivative relationships; the mapping of derivative relationships between RDA *Expressions* was successful, while the mapping of derivative relationships between RDA *Works* produced erroneous occurrences of the *bf:hasDerivative* relationship in BIBFRAME, which hinders models' interoperability and library data exploration.

The present work attempts a new complimentary step toward the structural and semantic interoperability between the two models. Developing and evaluating the BIBFRAME to RDA mapping for the same constructs (i.e., core entities, inherent relationships, and derivative relationships), the

following questions are considered: (a) Are there any conditions that must be met in both BIBFRAME semantics and implementation in order to make feasible and effective the mapping to RDA? (b) Can BIBFRAME data be mapped to RDA without any information loss?

In the next section, the methodology used in the RDA to BIBFRAME mapping is briefly presented. Then, the reverse mapping from BIBFRAME to RDA follows. The results of the mapping process are exhibited and further discussed in comparison with the RDA to BIBFRAME mapping results.

Mapping from RDA to BIBFRAME 2.0

RDA was selected for the mapping as the de facto content standard now used by the library community replacing the *Anglo-American Cataloguing Rules* (AACR). BIBFRAME is the model developed by the Library of Congress to integrate bibliographic descriptions on the web using linked data technologies. The described mapping focuses on core entities, inherent relationships and derivative relationships,[7] because the gradual development of a bibliographic family generally starts with a derivation, e.g., translation, revision.[8] Both RDA's and BIBFRAME's constructs, along with the description of bibliographic families and derivative relationships may be found in Zapounidou et al.[9] It is worth mentioning that the mapping from RDA to BIBFRAME, as well as the inverse mapping from BIBFRAME to RDA presented in this work, is based on the Library of Congress approach where the BIBFRAME *Work* class is considered as equal to the union of the RDA *Work* and *Expression* entities.[10]

The mapping was evaluated on the gold datasets, a Gold RDA and a Gold BIBFRAME dataset, already used to develop and evaluate the mapping rules for converting RDA data into BIBFRAME. The exact methodology for creating both gold datasets have been described in Zapounidou et al.[11] To assess the mapping, the Gold RDA dataset was converted to BIBFRAME and afterwards this new BIBFRAME dataset was compared to the Gold BIBFRAME one. The assessment of the mapping from RDA to BIBFRAME revealed five important findings, all summarized and considered in the discussion section.

Mapping from BIBFRAME 2.0 to RDA

The BIBFRAME to RDA mapping, similarly to the reverse one,[12] starts with the mapping of core entities and of their inherent relationships. It should be emphasized that the grouping of *bf:Work* instances is based on a specific set of *bf:Work* relationships and not on string matching of property

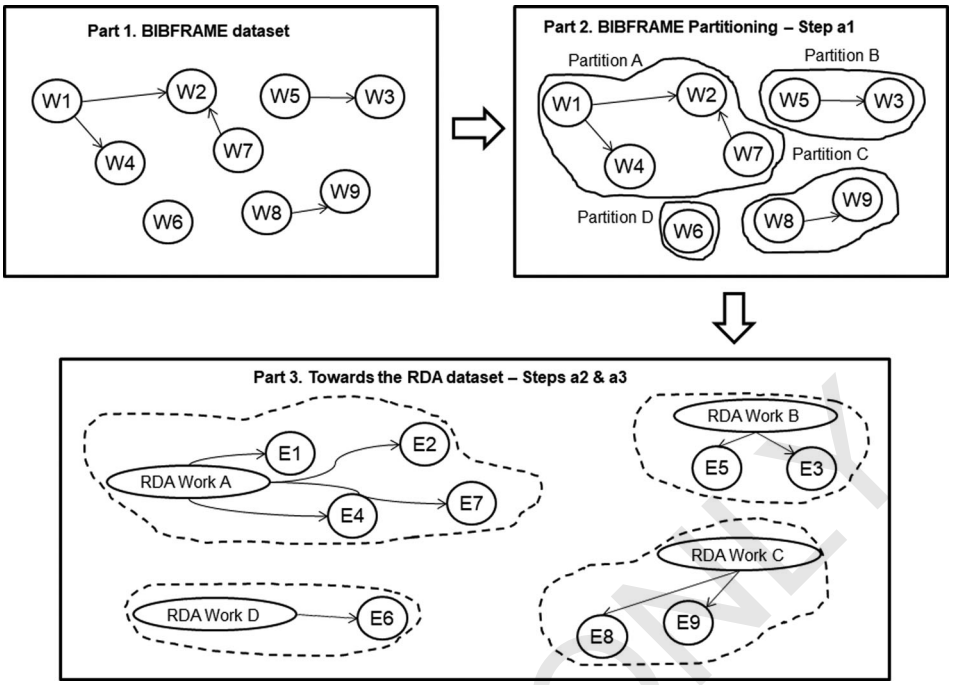

Figure 1. Partitioning and mapping a set of bf:Work instances to RDA Works and Expressions.

values. An intuitive description of the transformation of a BIBFRAME dataset to RDA is provided by the following outline:

Step (a): Map the set of *bf:Work* instances to the respective RDA entity instances, i.e., RDA *Works* along with their *Expression* instances.

Step (a1): Based on specific *bf:Work* relationships, partition the set of *bf:Work* instances to subsets, so that each subset contains all the different realizations of the same ideational content only (Figure 1 – Part 2).

Step (a2): For each subset create an instance of the RDA *Work* class (Figure 1 – Part 3).

Step (a3): For each *bf:Work* in the subset, create an instance of the RDA *Expression* class, and relate the aforementioned RDA *Work* instance to the generated RDA *Expression* instance with the property *rdaw:P10078 has expression of work* (Figure 1 – Part 3).

Step (b): Map derivative *bf:Work* relationships to their RDA *Expression* level relationships.

Step (c): Map *bf:Work* properties pertaining to the *bf:Work* class instances.

Step (d) Map *bf:Instance* and *bf:Item* classes along with their properties.

In the following sections all steps are justified and presented in detail. Specifically, the mapping of the *bf:Work* class to the RDA *Work* and *Expression* entities is presented first, and then the mapping of *bf:Work* derivative relationships. Thereafter, the mapping of properties pertaining to *bf:Work* class instances either to RDA *Work* or *Expression* properties is presented, and then the mapping of *bf:Instance* and *bf:Item* classes along with their properties.

Mapping of the bf:Work class

The dual nature of the *bf:Work* class, including both ideas and signs, suggests its mapping to a whole RDA path, '*Work - has expression of work – Expression*' one (Figure 2). This mapping adheres to the Library of Congress approach.[13] A critical point for transforming a set of *bf:Works* interlinked by particular relationships to RDA *Works* along with their *Expressions*, is to identify the proper subsets of *bf:Works*, so that each one of them contains all different realizations of the same ideational content. This approach reveals the RDA *Works* along with their *Expression* by exploiting the semantics of explicit relationships in the BIBFRAME dataset and avoids string matching of property values.

In BIBFRAME, *bf:hasExpression*, *bf:translation*, and *bf:otherEdition* are properties incorporating the aforementioned condition. All three properties imply that the *bf:Works* they relate share the same ideational content. More specifically, *bf:hasExpression* is used to cluster *bf:Works* containing different realizations of the same ideational content. The *bf:translation* property is used to represent literal translations, which are perceived in the library community as new realizations of the same ideational content. The *bf:otherEdition* property is used to represent other editions of the same content, e.g., edition of the same content in another media type.

It is worth noting that the subsets in *Step (a1)* must be disjoint due to the *one to many* cardinality (1:M) of the property *rdaw:P10078 has expression of work*. Moreover, the union of them has to be equal to the original set of *bf:Works* in order to assure that all *bf:Works* instances are mapped to RDA entities. Thereafter, the set of all resulting subsets forms a partition of the original set of *bf:Works*.

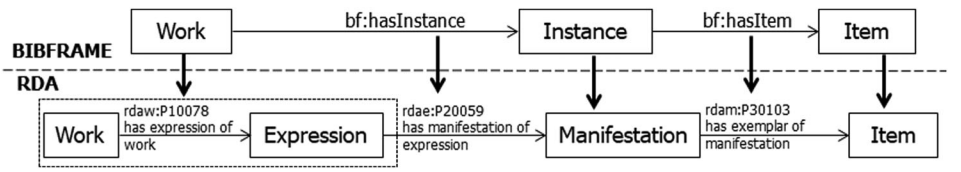

Figure 2. Core BIBFRAME entities and inherent relationships mapped to RDA.

The algorithm for the *Step (a)* mapping of *bf:Works* to RDA *Works* and *Expressions* has input a set of instances of the class *bf:Work*, outputs the corresponding RDA *Works* along with their *Expressions*, and consists of the following steps:

Step (a1): Separate the set of *bf:Work* instances to subsets such that each subset contains only all the different realizations of the same ideational content (Figure 1 – Part 2). Formally, the set of the *bf:Work* instances is represented by the graph *G(V, A)*, where the set of nodes *V* is the set of the *bf:Work* instances and the set of arcs *A* is the set of the *bf:hasExpression*, *bf:translation*, and *bf:otherEdition* properties that interlink the set of the *bf:Work* instances. This graph is ordinarily disconnected. In this stage the graph *G* is partitioned to its connected components (subgraphs), $\Gamma_i(V_i, A_i)$, where (i) $i = 1, 2, \ldots, N$ and N is the number of the components of the graph *G*; (ii) the set of the nodes V_i of a component Γ_i is subset of the set *V*, and the set of the arcs A_i of a component Γ_i is subset of *A*; (iii) the set of A_i are instances of the properties *bf:hasExpression*, *bf:translation*, and *bf:otherEdition* that connect the *bf:Work* instances that belong to the set V_i; (iv) the subgraphs Γ_i are disjoint to each other. Hence, all the sets V_i are disjoint to each other and all the sets A_i are disjoint to each other; (v) the subgraphs Γ_i are connected, i.e., either for any of the pairs (v_p, v_q) of its nodes V_i there is an (undirected) path connecting them or Γ_i consists of only one node.

Step (a2): For each subset create an instance of the RDA *Work* class (Figure 1 – Part 3). Actually, based on the construction of *Step (a1)*, for each component Γ_i an instance of the class RDA Work, *rdac:C10001 Work*, is generated in *Step (a2)*.

Step (a3): For each *bf:Work* in the subset, create an instance of the RDA *Expression* class, and relate the aforementioned RDA *Work* instance to the generated RDA *Expression* instance with the property *rdaw:P10078 has expression of work* (Figure 1 – Part 3). In *Step (a3)*, (i) the instances of the class *bf:Work* that are the members of the subset V_i of the component Γ_i are mapped to instances of the class RDA *Expression*, *rdac:C10006 Expression*, (ii) the generated instance of the *rdac:C10001 Work* class is connected to the generated instances of the class *rdac:C10006 Expression* with instances of the *rdaw:P10078 has expression of work* property. An example of applying the *Steps (a1)–(a3)* is presented in Figure 3.

Mapping derivative relationships

A derivative relationship between two *bf:Work* instances represents the relationship between two known realizations. Therefore, it is mapped to an RDA derivative relationship represented at the *Expression* level (*rdac:C10006*). In *Step (b)*, the *bf:Work* relationships are mapped to RDA

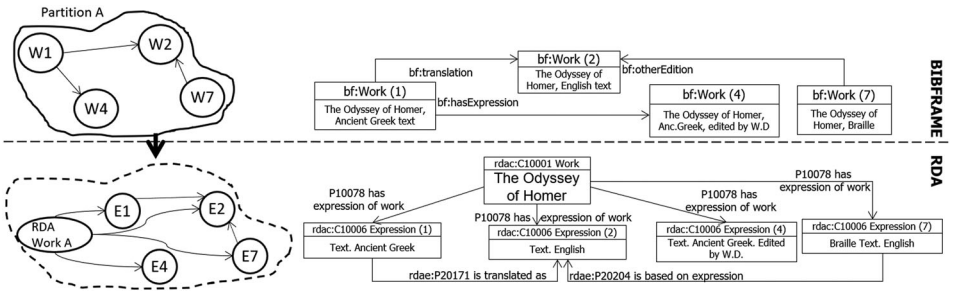

Figure 3. Example of mapping the Partition A that includes the bf:Works from the Odyssey family to RDA Work and Expressions following the Steps (a)–(b) of the mapping algorithm.

Expression-related relationships and not to RDA *Work*-related ones. It is also worth mentioning that a *bf:hasExpression* relationship which is mapped to *rdaw:P10078 has expression of work* relationship is not a derivative one. This approach was also followed for the reverse mapping from RDA to BIBFRAME as presented in our previous work.[14] The mappings of the BIBFRAME to RDA regarding the derivation relationships are presented in Appendix A - Table A1.

BIBFRAME provides three properties for derivate relationships between *bf:Works* (Figure 4): translation (*bf:translation/bf:translationOf*), 'other editions' (*bf:otherEdition* symmetric property), and derivation in general (*bf:hasDerivative/bf:derivative*).[15] "Other editions" refer to other available editions with different edition details, e.g., in another language or medium. The *bf:otherEdition* property may be used either between *bf:Works* (e.g., simultaneous edition in another language, edition of the same content in another media type), or between *bf:Instances* (e.g., reprints). All other types of derivative relationships, such as revision, abridgement, adaptation, etc., may be represented with the generic *bf:hasDerivative/bf:derivativeOf* properties due to the lack of other specialized properties for derivative relationships.

RDA is more analytic and provides relationships for derivations either between *Works*, or between *Expressions* (Figure 5). Representation of derivations among RDA *Works* denotes that there is a derivative relationship between the related RDA *Works,* but the exact sets of signs used in it remain unknown. Representation of derivations among RDA *Expressions* denotes that both sets of signs, source and derivative ones, involved in the derivation are known. RDA refines the derivative relationships identified in FRBR-FRAD models and provides a great number of specialized properties. For readability reasons, the subproperties of the *based on (work), derivative (work), based on (expression), and derivative (expression)* properties are not depicted in Figure 5.

Thereafter, the *bf:hasDerivative* property is mapped to the high-level *rdae:P20203 is derivative (expression)* property.[16] Due to the inexistence of

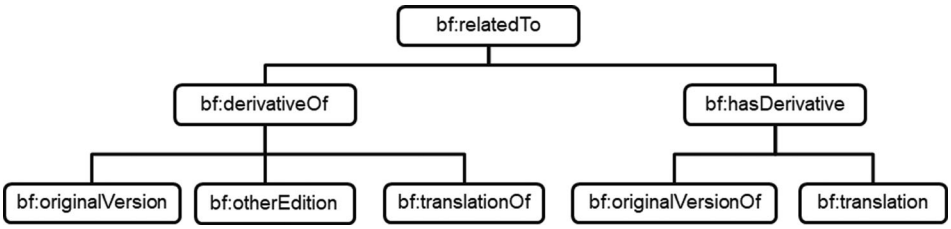

Figure 4. Hierarchy of properties for the representation of derivative relationships in BIBFRAME.

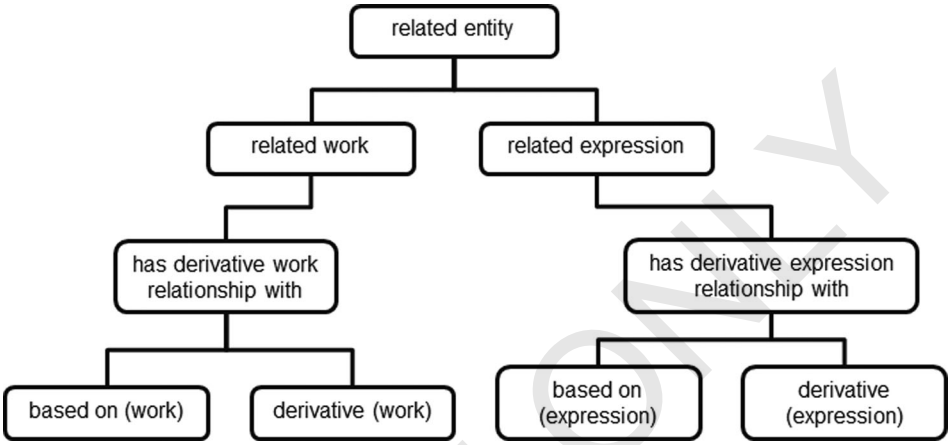

Figure 5. Part of the hierarchy of properties for the representation of derivative relationships in RDA.

specialized relationship for "other editions" in RDA, the *bf:otherEdition* property may be mapped either to the generic *rdae:P20203 is derivative (expression)* property, or to its inverse *rdae:P20204 is based on (expression)*. The *bf:otherEdition* property is symmetric, meaning that actually there is no discrimination regarding which instance of the class *bf:Work* is the domain of the property and which is the range. Therefore, its mapping depends on the cataloging policy implemented to create a BIBFRAME dataset. Furthermore, the *bf:translation* property is mapped to the *rdae:P20171 is translated as* property.

Mapping bf:Work properties

Regarding the mapping of the *bf:Work* properties to either RDA *Work* or *Expression*, *Step (c)*, the primary contribution information[17] and title of the *bf:Work* will be mapped to the properties of the instance of the RDA *Work*, while content type, non-primary contribution,[18] and language information will be mapped to RDA *Expression* properties. These mappings are further presented in Appendix A - Tables A2 and A3.

The dominant characteristics of works are usually defined by two key pieces of information: the author and the title.[19] The mappings for author

and title information represented with *bf:Work* properties to RDA is presented in Appendix A -Table A2. Title information is represented in BIBFRAME with the *bf:Title* class. In detail, it is represented with the following statement: *bf:Work – bf:title – bf:Title – bf:mainTitle – rdfs:literal*. This BIBFRAME statement is mapped to an RDA property, i.e., the *P10088 has title of work* data property, relating an RDA *Work* with its title represented by an instance of the class *rdfs:literal* (*rdac:C1001 – rdawd:P10088 – rdfs:literal*).

Authorship is represented with the *bf:Contribution* class. As an example, a person authoring a *bf:Work* is represented with the following two statements: 1) *bf:Work – bf:contribution – bf:Contribution – bf:agent – bf:Person,* and 2) *bf:Contribution – bf:role – bf:Role (relators:aut)*. The *bf:Contribution* class may be used for all types of contributions, primary and non-primary ones. For the explicit representation of primary contributions that are also used in name-title access points, another class may be used, namely the *bflc:PrimaryContribution* class, which for the time being has not been included in the BIBFRAME 2.0 ontology. This class has been defined in the BIBFRAME LC Extension Vocabulary in the context of Library of Congress' conversions from MARC to BIBFRAME. The utilization of this class is highly recommended by the authors because this class explicitly represents the primary contribution. By contrast, using the *bf:Contribution* class, primary contribution needs to be "discovered" exploiting the value of the *bf:Role* class in the following path *bf:Contribution - bf:role - bf:Role*. Furthermore, it must be noted that the *bf:Role* class may take several values. The roles used in this study are from the Library of Congress MARC Code List for Relators Scheme controlled vocabulary.[20] Other datasets may use other controlled vocabularies causing differentiations to the values used in the mapping rules.

RDA represents authorship differently; it uses different properties depending on the type of the creator and its contribution for relating an RDA *Work* with its author. Therefore, the mapping of primary contributions from BIBFRAME to RDA depends on (i) the type of the *bf:Agent* class, i.e., if the *bf:Agent* involved is typed as a *bf:Person, bf:Organization* or *bf:Family,* and (ii) the exact role of the *bf:Agent* (Appendix A - Table A2). In newer versions of RDA than the 2.7.3 used in this paper, the properties representing authorship were refined and more were added. Thus, the BIBFRAME mapping to newer versions of RDA needs to include more checks to determine the type of the agent and its exact role.

Step (c) proceeds with mapping rules for properties carrying pieces of information which are considered core to the identification of RDA *Expressions*, namely content type, language, and non-primary contribution (Appendix A - Table A3).[21] Content type is represented in BIBFRAME

with different *bf:Work* subclasses,[22] while in RDA the content type is represented as value of the *rdae:P20001 has content type* property. The values for this property are taken from a predetermined value vocabulary, the RDA Content Type Vocabulary.[23] The selection of a different value from the vocabulary may trigger different mappings.

Since primary contribution information refers to the ideational content included in a *bf:Work* instance, non-primary contribution consequently refers to the signs included in the same *bf:Work* instance. Therefore, the mapping of non-primary contributions to *bf:Work* instances will be made to RDA *Expression* properties. The mapping is similar to the mapping of primary contribution information. In detail, the mapping of the statement BIBFRAME *bf:Work - bf:contribution - bf: Contribution - bf:agent – bf:Agent* to the appropriate RDA *Expression* properties depends on the type of the *bf:Agent* (i.e., Person, Corporate Body, Family) and on the exact role this *bf:Agent* has (*bf:Contribution-bf:role-bf:Role*). Once again, the Library of Congress MARC Code List for Relators Scheme is used for determining the role values.[24]

Language information is represented in BIBFRAME by the *bf:Language* class and the following statement *bf:Work - bf:language – bf:Language (languages:value)*. The values for the *bf:Language* class are taken from the MARC List for Languages vocabulary.[25] This BIBFRAME triple will be mapped to the following RDA triple *rdac:C10006 Expression – rdaeo:P20006 has language of expression – languages:value*. The corresponding rules are presented analytically in Appendix A - Table A3.

Mapping bf:Instance and bf:Item classes

The *bf:Instance* class represents the 'material embodiments' of *bf:Works*, while the *bf:Item* class represents exemplars of *bf:Instances.* Therefore, the *bf:Instance* and *bf:Item* classes are mapped to the *rdac:C10007 Manifestation* and the *rdac:C10003 Item* classes respectively (Appendix A - Table A4). The *bf:hasInstance* property is mapped to the *rdae:P20059 has manifestation of expression* and the *bf:hasItem* property is mapped to the *rdam:P30103 has exemplar of manifestation* property (Figure 2).

The mapping of the BIBFRAME *Instance* class presents some challenges regarding the mapping of specific properties. BIBFRAME uses subclasses to represent particular attributes of *Instances*, while the respective representation is implemented by properties in RDA. For instance, information regarding extent is represented in BIBFRAME with the *bf:Extent* class, whereas RDA uses the *rdam:P30182 has extent of manifestation* property. The most challenging mapping involving *bf:Instance* properties has been the provision activity information. BIBFRAME uses the *bf:ProvisionActivity* class and its four subclasses (*bf:Publication, bf:Distribution, bf:Manufacture,*

bf:Production) to cluster individual pieces of information (e.g., agents, dates, places) all related to the described process. RDA – contrary to BIBFRAME's approach – represents the same information (agents, dates, places) with properties noting that these properties are different for publication, distribution, manufacture, and production. Wherever possible, we have used controlled vocabularies to preserve information and enable mappings (e.g., carrier types, media types). The mapping of *bf:Instance* properties is presented in Appendix A - Table A5.

Implementation of the mapping

The mapping has been implemented using Python and XSLT languages. This implementation uses the mapping rules presented earlier to transform BIBFRAME datasets encoded in RDFXML to RDA datasets encoded in RDFXML also. This implementation aims to assess the mappings of core entities, inherent relationships, and derivative relationships from BIBFRAME to RDA. Additionally, it aims to quantify the degree of interoperability between the two library data models.

We used the same two gold datasets, Gold BIBFRAME and Gold RDA, we had developed in our previous work.[26] Both datasets use cases belonging to eleven well-known bibliographic families, such as Don Quijote, Odyssey, Wuthering Heights, Madame Bovary, etc. The datasets were developed using the Protégé software, to describe derivative relationships (e.g., translation, revision, adaptation). The datasets have been uploaded in a Virtuoso RDF server and SPARQL queries can be submitted. The graph IRIs, along with the used prefixes and the SPARQL queries used for querying each dataset are openly available at our project's webpage.[27]

The Gold BIBFRAME dataset has been converted to a new RDA dataset, named BF2RDA. The BF2RDA dataset's graph IRIs and the SPARQL queries for querying it are also available through our project's webpage.[28]

There are three points regarding the implementation of the mapping that must be indicated. First, it is assumed that the *bf:hasExpression* property is implemented as a transitive property in the input data (Gold BIBFRAME dataset). According to the Web Ontology Language (OWL) Reference,[29] when a property P is defined as transitive, then for the pairs (A, B) and (B, C), which are instances of P, we can infer that the pair (A, C) is also an instance of P. Hence, a transitive property interlinks two instances A and C whenever it interlinks A with B and B with C. Therefore, if *bf:hasExpression* property is defined as OWL transitive, then in the case of three *bf:Work* instances *W1, W2, W3* with the relationships *W1* - *bf:hasExpression* – *W2* and *W2* - *bf:hasExpression* – *W3*, the *W1* - *bf:hasExpression* – *W3* relationship will be inferred. Regardless of whether

it is implemented as a transitive property or not, it affects only the first step (*Step (a1)*) for partitioning the initial set of *bf:Works*. Secondly, it is a matter of policy if the mapping will relate RDA classes' instances with the needed properties along with their reverse ones, or not. As an example, it is a matter of policy if the generated RDA *Works* will be related to their *Expressions* using the *rdaw:P10078 has expression of work or* using both the *rdaw:P10078 has expression of work* property and its reverse one, *rdae:P20231 has work expressed*. Our implementation of the mapping relates *bf:Works* in the generated BF2RDA dataset using both the needed properties and their reverse ones. Thirdly, in the Gold BIBFRAME dataset the symmetric property *bf:otherEdition* was instantiated relating the derivative *bf:Work* (domain) to the original *bf:Work* (range). Therefore, in transforming the Gold BIBFRAME dataset to RDA, the *bf:otherEdition* property was mapped to the *rdae:P20204 is based on (expression)* property. Nevertheless, a library may choose in its policy how the *bf:otherEdition* property will be used.

Results

The mapping of the Gold BIBFRAME dataset to RDA produced the BF2RDA dataset, which was later compared to the Gold RDA dataset to assess the qualitative characteristics of the mapping. Both Gold datasets and the BF2RDA dataset are presented analytically at the web page http://libdata.tab.ionio.gr/models/si-mapping/si_project.html.[30]

The first five columns of the Table 1 present the number of the instances of the core entities appeared in each of the bibliographic families selected to generate the Gold BIBFRAME and Gold RDA datasets respectively. The last three columns (in the right) present the number of the instances for the core RDA entities generated by the execution of the mapping (BF2RDA dataset). The results exhibit an accuracy of 100% regarding the mapping of

Table 1. Occurrences of core entities/classes in the three datasets.

	Gold BIBFRAME		Gold RDA			BF2RDA		
Family	Work	Inst.	Work	Expr.	Manif.	Work	Expr.	Manif.
Cien años	7	14	2	7	14	2	7	14
Crime&P	18	24	4	18	24	4	18	24
DonQuijote	12	11	4	12	11	4	12	11
Faust	25	25	7	25	25	7	25	25
Iliad	21	25	3	21	25	4	21	25
Karamazov	18	20	3	18	20	4	18	20
MmeBovary	22	29	3	22	29	5	22	29
Odyssey	15	19	2	15	19	3	15	19
ScarletLett	16	19	6	16	19	6	16	19
TSawyer	26	32	7	26	32	11	26	32
Wuthering	15	18	7	15	18	7	15	18
Total	**195**	**236**	**48**	**195**	**236**	**57**	**195**	**236**

the signs included in *bf:Works* to RDA *Expression* instances, and the mapping of *bf:Instances* to RDA *Manifestations*.

However, although in the Gold RDA dataset there are 48 instances of the RDA *rdac:C1001 Work* class, the mapping algorithm generated 57 such instances. This difference comes from the use of the generic *bf:hasDerivative* property in order to relate 9 abridgments to their original works. Actually, in the Gold datasets there exist 9 cases of abridgment. As already mentioned in BIBFRAME, there is no specific relationship for abridgments, and they are represented by the generic *bf:hasDerivative* property. Therefore, the differences on the realization aspects of the abridgments could not be exploited when the partitions are generated in *Step (a1)* of the mapping algorithm. Thereafter, the mapping generates nine more RDA *Work* instances reaching the total of 57 instances instead of the anticipated 48 ones. The difference of 9 *bf:Works* is fewer than expected. The Gold datasets include 10 abridgment cases and 6 revision ones. All of them are represented by the *bf:hasDerivative* property in BIBFRAME. Similarly to the representation of the abridgment relationship, the generic *bf:hasDerivative* property is used for revisions having the same result: loss of the needed information to include revisions in the proper subsets of the partition in *Step (a1)* of the mapping algorithm. Thus, a difference of 16 *bf:Works* between the Gold RDA and the BF2RDA was anticipated. Coincidentally, the 6 revision cases included in the Gold Datasets are also translations and 1 abridgment case is related to another *bf:Work* with the *bf:otherEdition* property. Thus, these 7 *bf:Works* have been included in the proper subset due to their linking with other *bf:Works* with a *bf:translation* instance (6 revision cases) or with a *bf:otherEdition* instance (1 abridgment case).

Regarding the mapping of derivative relationships, Table 2 presents the number of derivative relations per each bibliographic family in each dataset. Specifically, there are three columns in the Gold BIBFRAME dataset. The first one (Trl column) presents the instances of the *bf:translation* property, while the instances of the other two properties for derivative relationships (*bf:hasDerivative* and *bf:otherEdition*) are presented in the other two columns, hasDeriv and otherEd respectively. The three columns in the Gold RDA dataset present the derivative relationships represented between RDA *Works* (WorkDeriv), the translation relationships between RDA *Expressions* (Trl), and the rest of derivative relationships represented between RDA *Expressions* (ExprDeriv).

As mentioned, BIBFRAME *Work* relationships have been mapped to relationships relating RDA *Expressions*. Thirty-one instances of *Work-Work* relationships in the Gold RDA dataset (Table 2) cannot be represented in BIBFRAME because the exact related sets of signs are not known. These

Table 2. Occurrences of derivative relationships in the three datasets.

	Gold BIBFRAME			Gold RDA			BF2RDA		
Family	Trl	hasDeriv	otherEd	WorkDeriv	Trl	ExprDeriv	WorkDeriv	Trl	ExprDeriv
Cien años	4			1	4		0	4	
Crime&P	7		1	3	7	1	0	7	1
DonQuijote	2			1	2		0	2	
Faust	3			3	3		0	3	
Iliad	10	5	3	1	10	8	0	10	8
Karamazov	7	3		2	7	3	0	7	3
MmeBovary	3	3		2	3	3	0	3	3
Odyssey	8	4	2	1	8	6	0	8	6
ScarletLett	2			5	2		0	2	
TSawyer		4	3	6		7	0		7
Wuthering	3			6	3		0	3	
Total	**49**	**19**	**9**	**31**	**49**	**28**	**0**	**49**	**28**

cases in the Gold BIBFRAME dataset remain unrelated to other *bf:Works* of their bibliographic families. Therefore, the first column of the BF2RDA dataset presents zero occurrences of properties relating RDA *Works* (WorkDeriv). As a result, in the BF2RDA dataset there are no derivative relationships between RDA *Works* and possible exploration of data can be made only using the relationships at the *Expression* level.[31]

The second column of the BF2RDA dataset presents the instances of the *rdae:P20171 is translated as* property. Their number is the same to the number instances of the *rdae:P20171 is translated as* property in the Gold RDA dataset. Thus, the mapping of the *bf:translation* property has been successful without any loss of its semantics. The third column presents the total of other derivative relationships used to relate RDA *Expressions*. This total is the same to the sum of the mapped instances of the *bf:hasDerivative* and the *bf:otherEdition properties*. This total is also the same to the total of the ExprDeriv column in the Gold RDA dataset. Thus, the mapping of the *bf:hasDerivative* and the *bf:otherEdition properties* has also been successful. It must be noted that, even though the numbers of instances between the Gold RDA/ExprDeriv and BF2RDA/ExprDeriv are the same, different properties are instantiated in each column. This is due to the lack of specialized properties in BIBFRAME for the representation of derivative relationships. Translation is represented with the *bf:translation* property, while all the rest derivative relationships with the *bf:hasDerivative* property.

Moreover, there is no equivalent property for *bf:otherEdition* in RDA. Thus, apart from the *bf:translation* property, the rest of the BIBFRAME properties for derivative relationships are mapped to generic RDA ones, namely the *rdae:P20203 has derivative expression* and the *rdae:P20204 is based on expression* properties. It must be emphasized that the loss of semantics regarding the exact nature of derivation is due to the BIBFRAME semantics; it is not a result of the mapping process.

A significant result relating to the exploration of data must be reported. Due to the *bf:Work* semantics, all properties representing a derivative relationship have been mapped to RDA *Expression* properties. As a result, possible exploration of data can be made only using the relationships at the *Expression* level.

Discussion and conclusions

This study focuses on the semantic interoperability between BIBFRAME and RDA. A mapping algorithm was developed for the mapping of core entities, inherent relationships, and derivative relationships from BIBFRAME to RDA. As mentioned, BIBFRAME classes and properties have been used according to the Library of Congress interpretation regarding the *bf:Work* class being equivalent to the union of the RDA *Work* and *Expression* classes.[32]

RDA to BIBFRAME mapping core findings

It should be summarized that the previous paper[33] regarding the mapping from RDA to BIBFRAME revealed five important findings. First, RDA core entities and inherent relationships may be successfully mapped to BIBFRAME. Secondly, derivative relationships represented at the *Expression* level may be also mapped successfully to BIBFRAME but with loss of specificity due to BIBFRAME's fewer number of properties for the representation of derivative relationships.

Thirdly, the mapping of derivative relationships between RDA *Works* was ignored due to the generation of many false relationships (*bf:hasDerivative* property instances) between mapped *bf:Works*. When an RDA *Work* having a derivation relationship with another RDA *Work* and each of them having more than one *Expressions,* (e.g., there exist *n* Expressions for the first and *m* Expressions for the second RDA *Work*), then the mapping will generate a number of $n + m$ *bf:Work* instances and a number of *n* x *m* derivative relationships (*bf:hasDerivative* property) between them (actually a number of *m bf:hasDerivative* properties for each one of the *n bf:Work* instances). The generation of false relationships will hinder exploration of data and provide an unreliable context of relationships for each mapped *bf:Work*.

The fourth finding also associates with the exploration of a bibliographic family's members. Because the BIBFRAME *Work* class' semantics include both ideas and signs, derivative relationships are represented between known signs only. However, information regarding the exact source and derivative signs involved in a derivative relationship is not always available

to catalogers. As a result, the relationship may not be represented, *bf:Works* may be easily rendered 'orphan' and unrelated to other *bf:Works* with the same ideational content or members of the same *bibliographic family*.

Lastly, an interesting contribution in the previous study was that the *bf:hasExpression* property may be used to cluster different realizations of the same ideational content, therefore improving the semantic interoperability between the models, and the formulation of *bibliographic families*. In case the *bf:hasExpression* property is defined as transitive, it results in simpler clustering of the BIBFRAME *Works* realizing the same set of ideas and, hence, simpler representation of *bibliographic families*.

Findings regarding the mapping of the bf:Work class

This study revealed that the dual nature of the *bf:Work* class imposes more complicated mappings. The mapping of the *bf:Work* class to RDA *Works* and *Expressions* has been implemented successfully under the assumption that the relationships between *bf:Works* with the same intellectual content are expressed properly by the *bf:hasExpression*, *bf:translation*, and *bf:otherEdition* properties. The *bf:hasExpression* property may represent the existence of common intellectual content among two *bf:Works* when they have no other relationship, like *bf:translation*, and *bf:otherEdition*, to indicate their common intellectual content. Therefore, when the *bf:hasExpression* property is used, the BIBFRAME and RDA models could be considered aligned for the representation of both the intellectual content and signs using different modeling patterns.

The *bf:translation* property is used to represent translations, while the *bf:otherEdition* property is used to represent the other editions of the same content. Both properties indicate implicitly the existence of common intellectual content too. Contrary to RDA, the BIBRAME hierarchy of properties for the representation of derivative relationships (Figure 4) is limited. BIBFRAME does not provide specialized properties for the representation of other cases involving derivations of the same intellectual content, such as abridgment and revision and, therefore, it represents them using the generic property *bf:hasDerivative*. This property is also used for derivations involving different intellectual content, such as, adaptation, dramatization, etc. As a result, while the use of the *bf:hasDerivative* property for an abridgment or revision represents their derivative relationship, the property does not preserve the information about shared intellectual content between the related *bf:Works*.

This finding suggests that, in order to indicate that two related *bf:Works* where the one is the abridgment or revision of the other include the same ideas, two properties must be used to relate them: the *bf:hasDerivative*

property to represent the general derivative aspects of the abridgment/revision and the *bf:hasExpression* property to represent the existence of common ideational content. The implementation of this suggestion will improve the effectiveness of *Step (a)* of the mapping algorithm, so that all abridgments and revisions will be included in the proper subsets of the partition. This suggestion accords with an interesting observation in the datasets used in this study. As already presented in the results section, a difference of 16 *bf:Works* between the Gold RDA and the BF2RDA was anticipated taking under consideration that the Gold BIBFRAME dataset includes 16 abridgment and revision cases represented with *bf:hasDerivative* property instances. Yet, a difference of only 9 *bf:Works* was observed between Gold RDA and BF2RDA datasets. Seven *bf:Works* were included in the proper subset due to existence of additional relationship properties indicating the shared intellectual content, namely all 6 revision cases in the Gold BIBFRAME dataset are also translations and are related with a *bf:translation* instance, and 1 abridgment case is also another edition of the same content in another media type (audio) and is related with a *bf:otherEdition* instance.

The selection of which pieces of information are going to be preserved and how they will be represented using the BIBFRAME model's primitives is a matter of cataloging policy. For the preservation of shared ideational content, the use of the *bf:hasExpression* property is suggested. As demonstrated in this study, the representation approach of using the *bf:hasExpression* property for denoting the existence of shared ideational content among the related *bf:Works* enabled (i) the clustering of *bf:Works* in BIBFRAME, and (ii) the high improvement of the interoperability of BIBFRAME and RDA by identifying the common ideational content in the BIBFRAME and mapping them properly to RDA *Works* with their *Expressions*. The clustering of *bf:Works* may be observed in the Gold BIBFRAME dataset visualizations,[34] while the proper mapping of *bf:Works* to RDA *Works* with their *Expressions* is demonstrated in Table 1. The decision of whether the *bf:hasExpression* property is going to be used as a transitive or non-transitive one is another cataloging policy decision, unless the model is extended to make the property transitive. In the Gold BIBFRAME dataset, the *bf:hasExpression* property was used as a transitive one resulting in simpler representation patterns.

Regarding the clustering of *bf:Works* two comments must made. First, in this study, the clustering of *bf:Works* does not depend on string matching of title/primary contribution information but on explicit relationships only. Thus, possible inconsistencies between title and/or primary contribution information describing *bf:Works* realizing the same content are avoided. It should be noted that Title/Primary contribution inconsistencies do not exist

in the Gold Datasets. However, in the real world, inconsistencies exist, and literals are likely to change. Related research has proven that string matching for identifying Works (in RDA terms) is challenging and error prone demanding prior corrections and enrichment of data.[35] The mapping of this study focuses on the mapping of core constructs. Possible inconsistencies regarding title and/or primary contribution do not affect the mapping of *bf:Work* class to RDA *Work* and *Expression* entities, but they may have an impact on merging and mapping certain properties, e.g., merging and mapping title-related properties from *bf:Work* instances to its RDA *Work* instance title-related properties.

Secondly, the authors are aware of the *bflc:Hub* class that is presented as a collocation mechanism by the Library of Congress.[36] Its use was not considered in this study for the following reasons: (i) the *bflc:Hub* class is not part of the BIBFRAME model, (ii) up to now, there is no official definition regarding its use in the BIBFRAME ontology, nor in the BIBFRAME LC Extension (bflc vocabulary), and (iii) the Library of Congress still experiments with the possible uses of this class to "collect like or related Things."[37] Once the definition and use of the *bflc:Hub* class are finalized, this class needs to be further studied with regard to the interoperability between BIBFRAME and RDA.

Findings regarding the mapping of the properties of bf:Work and bf:Instance classes

Due to the dual nature of the *bf:Work* class, some of its properties are mapped to RDA *Work* ones, while others to RDA *Expression* ones. In order to generate proper RDA *Work* and *Expression* instances containing the information defined by their classes' semantics, certain pieces of information must be present in the *bf:Work* instances. The RDA *Work*-related information that must be present in *bf:Work* instances is primary contribution and title. The RDA *Expression*-related information that must be present in *bf:Work* instances is content type, language, and other contributions.

These pieces of information are represented with different primitives in BIBFRAME and RDA; BIBFRAME uses classes, while RDA uses properties. As an example, in BIBFRAME contributions are represented with the *bf:Contribution* class which groups together instances of the *bf:Agent* and the *bf:Role* classes. The *bf:Contribution* class may be used to represent both primary and non-primary contributions. Therefore, it needs to be specified if the contribution will be considered as an RDA *Work* or an *Expression* property. The information if a contribution is primary or not may be extracted by the value of the *bf:Role* class instance, e.g., author. The *bf:Role* values may be either taken from a controlled vocabulary or be a plain literal value. This cannot be characterized as a sound construct for the

representation of primary contributions. The *bflc:PrimaryContribution* class, defined in the BIBFRAME LC Extension List, may represent primary contribution information explicitly. Using the *bflc:PrimaryContribution* class, the primary contribution is explicitly represented by the class itself and there is no need for further checking the *bf:Role* instances' values. The inclusion of this class in the official BIBFRAME may be proven useful for better representation of primary contributions using the BIBFRAME constructs.

Contrary to the use of the *bf:Contribution* class in BIBFRAME, RDA provides different properties to represent the contribution of each agent type (person, organization, family) with a specific role. The role of author may better illustrate the RDA modeling approach. RDA provides three properties, one per agent type; namely, *rdaw:P10061 has author agent* when the agent involved is a person, *rdaw:P10530 has author corporate body* when the agent involved is a corporate body, and *rdaw:P10577 has author family* when the agent involved is a family.[38] To successfully map the authorship information to the specialized RDA property, controlled vocabularies must be used. Similarly, a controlled vocabulary is needed for the language information, while the proper *bf:Work* subclass, e.g., *bf:Text, bf:Audio,* must be used to map content type information to the corresponding RDA *Expression* property (*rdae:P20001 has content type*).

The issue of different primitives in BIBFRAME and RDA for the representation of the same information is also evident in the mapping of *bf:Instances* properties to RDA. Appropriate *bf:Instance* subclasses may reveal the carrier type, while the *bf:ProvisionActivity* class gathers all information regarding the agent, the date, and the place of a certain embodiment process. In RDA, there are different agent, date, place-related properties for each type of provision activity. Therefore, in BIBFRAME, *bf:ProvisionActivity* subclasses must be used (i) to represent the exact process of the *bf:Instance's* embodiment, as well as the type and role of the agent(s) involved in the process, and (ii) to enable the mapping to specific RDA *Manifestation* properties. The observations made during the mapping of *bf:Work* and *bf:Instance* properties advocate for the use of specialized classes and controlled vocabularies in BIBFRAME to provide successful mappings to RDA without loss of information. The importance of controlled vocabularies in mappings was stated in one of our previous studies[39] and is in accordance with other studies advocating for fewer literals, more structured information, and controlled vocabularies.[40]

Findings regarding the mapping of derivative relationships

Regarding derivative relationships, the mapping has been successful with only one case of specificity loss (*bf:otherEdition*). It is important to bear in

mind two differences between BIBFRAME and RDA. First, BIBFRAME defines a small number of derivative relationships in comparison to RDA (Figure 4). Secondly, these relationships involve derivation between known signs; *bf:Work* instances include both ideas and signs, and the representation of a relationship between two *bf:Work* instances indicates that both original and derivative signs are known. Contrary to RDA that permits the representation of derivation at the abstract *Work* level (Figure 5), a derivative relationship when either original or derivative signs remain unknown cannot be represented in BIBFRAME. Unawareness of the exact signs involved in a derivation is really common during cataloging. This means that a *bf:Work* may easily be rendered unrelated to others owing to the semantics of the *bf:Work* class. This may have an impact of the formulation of bibliographic families and on the explorability of data. This has been reported in the results of this study with the unrelated *bf:Works* in the Gold BIBFRAME dataset and the exploration of families through the *Expression* properties only in BF2RDA. Based on this finding, cataloging policies need to take under consideration that the representation of relationships must be made at the most specific level. Given the amount of time and expertise needed for this task, it may be decided to implement this policy in collections of great interest or in collaboration with experts.[41] Again, this is a cataloging policy issue.

Conclusions

In general, BIBFRAME and RDA use different primitives and modeling approaches. The mappings have revealed that information may be preserved, even though there are cases where specificity is lost. High-levels of interoperability were achieved by using BIBFRAME primitives in specific representation patterns to capture certain semantics: the proper use of relationship properties denoting the existence of common ideational content among related *bf:Works* (*bf:hasExpression*), the presence of certain pieces of information regarding the description of ideas and signs contained in *bf:Work* instances (use of *bf:Work* subclasses, contributions, etc), the selection of the most specialized class to enable the mapping to the semantically corresponding RDA property (e.g., selecting the *bf:ProvisionActivity* subclass to map its information to RDA using the proper agent, data, and place related properties), and the use of controlled vocabularies. These patterns may be considered as a cataloging policy described in a BIBFRAME profile. The study has provided evidence that modeling decisions have an impact on the interoperability between the models. Interoperability between BIBFRAME and RDA is not solely a matter of both models' semantics and

structures; it also depends on the cataloging policy and the modeling approach used in the implementations of the models.

With regard to the limitations of our study, it must be noted that both mappings use the 2.7.3. version of the RDA vocabularies. Since then, newer versions have been published. Yet, as referred to in related endnotes, the findings of both mappings remain relevant in the newer RDA versions. Both mappings were developed within the scope of studying the interoperability between BIBFRAME and RDA. The focus has been on mapping core constructs (core classes and inherent relationships) and derivative relationships. Thus, the mapping of properties pertaining to other BIBFRAME classes, such as *bf:Content, bf:Title, bf:Contribution, bf:ProvisionActivity,* needs to be revisited. The Gold Datasets were implemented according to specific cataloging policies assuming that there are no divergences regarding titles and name authorities. Thus, real-world problems such as inconsistencies regarding title/primary contributor or other cataloging errors are not addressed in the mapping. The proposed algorithm performs more effectively when running on the whole dataset, where all relationships exists, rather than to a single entity or incomplete subset. Despite these limitations, the mappings offer valuable insights regarding the role of cataloging policies in the interoperability of bibliographic data and advocate for the policies implemented in the two mappings to be considered good practices.

Derivative relationships are one type of bibliographic relationships. A study for other bibliographic relationships, such as whole-part, aggregates, and equivalence, is needed to further examine the semantic interoperability between the models, and to identify good practices for interoperability and for the preservation of other bibliographic relationships after mappings.

ORCID

Michalis Sfakakis http://orcid.org/0000-0003-2973-7455
Sofia Zapounidou http://orcid.org/0000-0001-8784-9581
Christos Papatheodorou http://orcid.org/0000-0002-9025-6469

Notes

1. Kim Tallerås, "From Many Records to One Graph: Heterogeneity Conflicts in the Linked Data Restructuring Cycle," *Information Research* 18, no. 3 (2013), http://InformationR.net/ir/18-3/colis/paperC18.html; Trond Aalberg, Kim Tallerås, and David Massey, "The Impact of New Bibliographic Models on the Search Experience," in *Proceedings of CoLIS, the Tenth International Conference on Conceptions of Library and Information Science* (Ljubljana, Slovenia, June 16–19, 2019). *Information Research* 24, no. 4 (2019), http://InformationR.net/ir/24-4/colis/colis1915.html; Hyoungjoo Park, and Margaret Kipp, "Library Linked Data Models: Library Data in the Semantic

Web," *Cataloging & Classification Quarterly* 57, no. 5 (2019): 261–277, https://doi.org/10.1080/01639374.2019.1641171; Michele Seikel, and Thomas Steele, "Comparison of Key Entities Within Bibliographic Conceptual Models and Implementations: Definitions, Evolution, and Relationships," *Library Resources & Technical Services* 64, no. 2 (2020): 62–71, https://doi.org/10.5860/lrts.64n2.62.

2. http://access.rdatoolkit.org/
3. http://www.rdaregistry.info/
4. Barbara B. Tillett, "Bibliographic Relationships: Toward a Conceptual Structure of Bibliographic Information used in Cataloging" (Ph.D. Dissertation, University of California, Los Angeles, 1987), https://search.proquest.com/docview/303545088 (accessed June 4, 2020); Richard P. Smiraglia, "Authority Control and the Extent of Derivative Bibliographic Relationships" (Ph.D. Dissertation, University of Chicago, 1992), https://search.proquest.com/docview/304038939 (accessed June 4, 2020); Richard P. Smiraglia, and Gregory H. Leazer, "Derivative Bibliographic Relationships: The Work Relationship in a Global Bibliographic Database," *Journal of the Association for Information Science and Technology* 50, no. 6 (1999): 493–504, doi: 10.1002/(SICI)1097-4571(1999)50:6 < 493::AID-ASI4 > 3.0.CO;2-U; Richard P. Smiraglia, "Derivative Bibliographic Relationships among Theological Works," *Proceedings of the ASIS Annual Meeting* 36 (1999): 497–506; Marija Petek, "Derivative Bibliographic Relationships in the Slovenian Online Catalogue COBIB," *Journal of Documentation* 63, no. 3 (2007): 398–423, doi: 10.1108/00220410710743315.
5. Smiraglia and Leazer, "Derivative Bibliographic Relationships;" Smiraglia, Richard P. "Instantiation: Toward a Theory," in *Data, Information and Knowledge in a Networked World, Annual Conference of the Canadian Association for Information Science*, ed. Liwen Vaughan (London, Ontario, June 2–4 2005), 1–8, http://www.cais-acsi.ca/ojs/index.php/cais/article/view/226.
6. Sofia Zapounidou, Michalis Sfakakis, and Christos Papatheodorou, "Mapping Derivative Relationships from RDA to BIBFRAME 2," *Cataloging & Classification Quarterly* 57, no. 5 (2019): 278–308, https://doi.org/10.1080/01639374.2019.1650152. Both mappings, 'RDA to BIBFRAME' presented in our earlier paper and the 'RDA to BIBFRAME' described in the present one, use the version 2.7.3 of the RDA vocabularies.
7. Ibid.
8. Smiraglia and Leazer, "Derivative Bibliographic Relationships."
9. Zapounidou, Sfakakis, and Papatheodorou, "Mapping Derivative Relationships."
10. BIBFRAME - Bibliographic Framework Initiative, "BIBFRAME Profiles: Introduction and Specification (Draft - 5 May 2014)," 2014, http://www.loc.gov/bibframe/docs/bibframe-profiles.html; Sally H. McCallum, "BIBFRAME Development," *JLIS.It* 8, no. 3 (2017): 72–85, https://doi.org/10.4403/jlis.it-12415; Sally McCallum, "BIBFRAME Introduction," in *European Bibframe Workshop* (Florence, 2018), http://www.casalini.it/EBW2018/web_content/2018/presentations/McCallum_1.pdf.
11. Zapounidou, Sfakakis, and Papatheodorou, "Mapping Derivative Relationships."
12. Ibid.
13. BIBFRAME - Bibliographic Framework Initiative, "BIBFRAME Profiles;" McCallum, "BIBFRAME Development;" McCallum, "BIBFRAME Introduction."
14. Zapounidou, Sfakakis, and Papatheodorou, "Mapping Derivative Relationships."
15. There is another derivative relationship in the BIBFRAME hierarchy of derivative relationships/properties represented with the *bf:originalVersion* / *bf:originalVersionOf* properties. As already discussed in Zapounidou, Sfakakis, and Papatheodorou,

"Mapping Derivative Relationships," we consider that these two properties may be used between *bf:Instances* for the representation of reproduction. Therefore, their use has been excluded from the study.

16. It must be noted that in this paper we have downloaded RDA vocabularies (version 2.7.3) to create the Gold RDA dataset in a Protégé installation. Therefore, in our text and the webpages developed for this study, we use the RDA element sets' labels for the description of classes and properties. There are cases where there are slight differences between a class'/property's label in the vocabulary and on the RDA Registry website. As an example, the *rdae:P20203* property has the label 'is derivative (expression)' in the RDA element set, while on the RDA registry its label is 'has derivative expression'.
17. Primary contribution information pertains to the agent that is chiefly responsible for the creation of a *bf:Work*. Other contributions that are not considered as definitive of a *bf:Work's* ideational content are characterized in this paper as "non-primary contributions." A typical example of a non-primary contribution is the contribution of a translator.
18. Please check the previous endnote for the use of the following terms in this paper: primary contribution, and non-primary contribution.
19. Trond Aalberg and Maja Žumer, "Looking for Entities in Bibliographic Records," Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 5362 LNCS (2008): 327–30, https://doi.org/10.1007/978-3-540-89533-6-36; Thomas B. Hickey and Edward T. O'Neill, "FRBRizing OCLC's WorldCat," Cataloging & Classification Quarterly 39, no. 3–4 (2009): 239–51, https://doi.org/10.1300/J104v39n03_15; Manolis Peponakis, Michalis Sfakakis, and Sarantos Kapidakis, "FRBRization: Using UNIMARC Link Fields to Identify Works," World Library and Information Congress : 77th IFLA General Conference and Assembly (2011), 1–12, https://www.ifla.org/past-wlic/2011/187-peponakis-en.pdf; Naimdjon Takhirov, "Extracting Knowledge for Cultural Heritage Knowledge Base Population" (PhD thesis, Norwegian University of Science and Technology, 2013), http://hdl.handle.net/11250/253516.
20. "LC Linked Data Service: MARC Code List for Relators Scheme," Library of Congress, http://id.loc.gov/vocabulary/relators.html (accessed May 26, 2020).
21. E. T. O'Neill, "FRBR: Functional Requirements for Bibliographic Records. Application of the Entity-Relationship Model to Humphry Clinker," Library Resources and Technical Services 46, no. 4 (2002): 150–59, http://dx.doi.org/10.5860/lrts.46n4.150; Martha M. Yee, "FRBRization: A Method for Turning Online Public Finding Lists into Online Public Catalogs," Information Technology and Libraries 24, no. 2 (2005): 77–95, https://doi.org/10.6017/ital.v24i2.3368; Aalberg, and Žumer. "Looking for Entities in Bibliographic Records."
22. BIBFRAME provides two ways for representing the fundamental form of communication in which a *bf:Work*'s content is expressed. The first one is the use of *bf:Work* subclasses and the second one is the use of the *bf:Work-bf:content-bf:Content*. These two ways can also be used simultaneously. For the creation of the Gold Datasets in this study, the authors used *bf:Work* subclasses as a more stable way for representing content comparing to the *bf:content* property. Selecting the *bf:content* property for representing content demands the use of controlled vocabulary further proving the authors' argument for the role of cataloging decisions in bibliographic data interoperability.

23. "Term and Code List for RDA Content Types," Library of Congress, 2014, https://www.loc.gov/standards/valuelist/rdacontent.html (accessed May 26, 2020).
24. "LC Linked Data Service: MARC Code List for Relators Scheme."
25. "LC Linked Data Service: MARC List for Languages," Library of Congress, 2011, http://id.loc.gov/vocabulary/languages.html (accessed May 26, 2020).
26. Zapounidou, Sfakakis, and Papatheodorou, "Mapping Derivative Relationships."
27. Sofia Zapounidou, Michalis Sfakakis, and Christos Papatheodorou, "Semantic Interoperability between Bibliographic Conceptual Models," Database & Information Systems Group, Ionian University, 2020. http://libdata.tab.ionio.gr/models/si-mapping/si_project.html (accessed May 26, 2020).
28. Ibid.
29. Sean Bechhofer, Frank van Harmelen, Jim Hendler, Ian Horrocks, Deborah L. McGuinness, Peter F. Patel-Schneider, and Lynn Andrea Stein, "OWL Web Ontology Language Reference: W3C Recommendation 10 February 2004," last modified February 10, 2004, accessed August 3, 2020, https://www.w3.org/TR/owl-ref/.
30. Ibid.
31. Zapounidou, Sfakakis, and Papatheodorou, "Semantic Interoperability between Bibliographic Conceptual Models." In the webpage presenting the BF2RDA dataset, there are two sections proving that the exploration of each bibliographic family in the dataset is impossible without the inclusion of the RDA Expressions in the SPARQL queries. These sections are "Visualize all Works in a family" and "Visualize Works and Expressions in families." In the first one, no visualizations are generated, while in the second one proper visualizations of bibliographic families are generated.
32. BIBFRAME - Bibliographic Framework Initiative, "BIBFRAME Profiles;" McCallum, "BIBFRAME Development;" McCallum, "BIBFRAME Introduction."
33. Zapounidou, Sfakakis, and Papatheodorou, "Mapping Derivative Relationships."
34. Zapounidou, Sfakakis, and Papatheodorou, "Semantic Interoperability between Bibliographic Conceptual Models." In the webpage presenting the Gold BIBFRAME dataset, the use of the *bf:hasExpression* property for clustering *bf:Works* sharing the same ideational content may be observed in the section "Visualize Works in a family."
35. Richard P. Smiraglia, "The 'Works' Phenomenon and Best Selling Books," *Cataloging & Classification Quarterly* 44, no. 3–4 (2007): 179–95, https://doi.org/10.1300/j104v44n03_02; Aalberg and Žumer, "Looking for Entities in Bibliographic Records;" Trond Aalberg, Tanja Merčun, and Maja Žumer, "Coding FRBR-Structured Bibliographic Information in MARC," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 7008 LNCS (2011): 128–37, https://doi.org/10.1007/978-3-642-24826-9_18.
36. Kevin Ford, "Concerning Identities: For Things, but not the Easy Things [Contains Hubs Part I]," in *3rd European BIBFRAME Workshop* (Stockholm: National Library of Sweden, 2019), https://www.kb.se/download/18.d0e4d5b16cd18f600eaea/1569322333140/2019-kefo-identity.pdf; Kevin Ford, "Concerning Relationships : Hubs Part II," in *3rd European BIBFRAME Workshop* (Stockholm: National Library of Sweden, 2019), https://www.kb.se/download/18.d0e4d5b16cd18f600eafb/1569324367132/2019-kefo-relationships.pdf; Sally McCallum, "BIBFRAME Expansion and Access," in *3rd European BIBFRAME Workshop* (Stockholm: National Library of Sweden, 2019), https://www.kb.se/download/18.d0e4d5b16cd18f600eabd/1569242059929/stockholm2019-EWS-BFdevelopment-mccallum.pdf.
37. Ford, "Concerning Identities;" Ford, "Concerning Relationships."

38. At this point it must be highlighted that in the 2.7.3 version of the RDA vocabularies, there is only one property for authors (P10061) that can be used for person and collective agents. In the following RDA versions, the property was refined and new ones have been added to represent person, corporate body, or family as author. These properties (P10061, P10530, and P10577) are the ones referred in the text. These refinements demand that checks are made before the mapping to specific RDA properties. The checks need to consider the type of agent and its role.
39. Sofia Zapounidou, Michalis Sfakakis, and Christos Papatheodorou, "Preserving Bibliographic Relationships in Mappings from FRBR to BIBFRAME 2.0," in *Research and Advanced Technology for Digital Libraries: 21st International Conference on Theory and Practice of Digital Libraries, TPDL 2017, Thessaloniki, Greece, September 18–21, 2017, Proceedings*, ed. Jaap Kamps, Giannis Tsakonas, Yannis Manolopoulos, Lazaros Iliadis, and Ioannis Karydis, LNCS 10450:15–26 (Springer International Publishing AG, 2017), https://doi.org/10.1007/978-3-319-67008-9_2.
40. Aidan Hogan et al., "An Empirical Survey of Linked Data Conformance," *Journal of Web Semantics* 14 (2012): 14–44, https://doi.org/10.1016/j.websem.2012.02.001; Gordon Dunsire, Diane Hillmann, and Jon Phipps, "Reconsidering Universal Bibliographic Control in Light of the Semantic Web," *Journal of Library Metadata* 12, no. 2–3 (2012): 164–76, https://doi.org/10.1080/19386389.2012.699831; Marcia Lei Zeng, "Interoperability," *Knowledge Organization* 46, no. 2 (2019): 122–46, https://doi.org/10.5771/0943-7444-2019-2-122.
41. Laurence S. Creider, "Cataloging, Reception, and the Boundaries of a 'Work,'" *Cataloging & Classification Quarterly* 42, no. 2 (2006): 3–19, https://doi.org/http://dx.doi.org/10.1300/J104v42n02_02; Pauline Rafferty, "FRBR, Information, and Intertextuality," *Library Trends* 63, no. 3 (2015): 487–511, http://doi.org/10.1353/lib.2015.0008; Henrik Wallheim, "From Complex Reality to Formal Description: Bibliographic Relationships and Problems of Operationalization in RDA," *Cataloging & Classification Quarterly* 54, no. 7 (2016): 483–503, https://doi.org/10.1080/01639374.2016.1200169.

Appendix A

Table A1. Step (b) – Mapping derivative relationships.

BIBFRAME	RDA
bf:translation	*rdae:P20171 is translated as*
bf:otherEdition	*rdae:P20203 is derivative (expression)* *rdae:P20204 is based on expression*
bf:hasDerivative	*rdae:P20203 has derivative expression*

Table A2. Step (c)- Mapping BIBFRAME title and primary contribution information from BIBFRAME Work to RDA Work properties.

Information	BIBFRAME	RDA
Title	bf:Work - bf:title - bf:Title - bf:mainTitle - rdfs:literal	rdac:C1001-rdawd:P10088-rdfs:literal
Primary Contribution	bf:Work - bf:contribution - bflc:PrimaryContribution - bf:agent - bf:Agent	rdac:C1001-rdawo:P10061-rdac:C1002
	If bf:Agent - rdf:type - bf:Person	rdac:C1001-rdawo:P10445-rdac:C1002
	then if	rdac:C1001-rdawo:P10442-rdac:C1002
	bflc:PrimaryContribution-bf:role-bf:Role(relators:aut)	rdac:C1001-rdawo:P10065-rdac:C1002
	else if	rdac:C1001-rdawo:P10444-rdac:C1002
	bflc:PrimaryContribution-bf:role-bf:Role(relators:pht)	rdac:C1001-rdawo:P10455-rdac:C1002
	else if	rdac:C1001–rdawo:P10530–rdac:C10005
	bflc:PrimaryContribution-bf:role-bf:Role(relators:cmp)	rdac:C1001-rdawo:P10539-rdac:C10005
	else if	rdac:C1001-rdawo:P10536-rdac:C10005
	bflc:PrimaryContribution-bf:role-bf:Role(relators:spk)	rdac:C1001–rdawo:P10577–rdac:C10008
	else if	rdac:C1001-rdawo:P10586-rdac:C10008
	bflc:PrimaryContribution-bf:role-bf:Role(relators:edc)	rdac:C1001-rdawo:P10583-rdac:C10008
	else if	
	bflc:PrimaryContribution-bf:role-bf:Role(relators:drt)	
	if bf:Agent-rdf:type-bf:Organization	
	then if	
	bflc:PrimaryContribution-bf:role-bf:Role(relators:aut)	
	else if	
	bflc:PrimaryContribution-bf:role-bf:Role(relators:pht)	
	else if	
	bflc:PrimaryContribution-bf:role-bf:Role(relators:cmp)	
	if bf:Agent-rdf:type-bf:Family	
	then if	
	bflc:PrimaryContribution-bf:role-bf:Role(relators:aut)	
	else if	
	bflc:PrimaryContribution-bf:role-bf:Role(relators:pht)	
	else if	
	bflc:PrimaryContribution-bf:role-bf:Role(relators:cmp)	

Table A3. Step (c) - Mapping BIBFRAME content type, language, and non-primary contribution information from BIBFRAME Work to RDA Expression properties.

Information	BIBFRAME	RDA
Content type	bf:Work - rdf:type – (bf:Work subclass)	rdac:C10006- rdaeo:P20001-
	or (bf:Work subclass)	rdaco:1020
	if bf:Text	rdaco:1012
	if bf:Audio	rdaco:1010
	if bf:NotatedMusic	rdaco:1002
	if bf:Cartography	rdaco:1009
	if bf:NotatedMovement	rdaco:1007
	if bf:Dataset	rdaco:1014
	if bf:StillImage	rdaco:1023
	if bf:MovingImage	rdaco:1021
	if bf:Object	rdaco:1008
	if bf:Multimedia	
Non-primary contributions	bf:Work - bf:contribution-bf:Contribution-bf:agent – bf:Agent	rdac:C10006-rdaeo: P20346 - rdac:C1002
	If bf:Agent- rdf:type –bf:Person	rdac:C10006-rdaeo:P20338 - rdac:C1002
	then	rdac:C10006-rdaeo:P20378 - rdac:C1002
	if	rdac:C10006-rdaeo:P20351 - rdac:C1002
	bf:Contribution-bf:role-bf:Role(relators:trl)	rdac:C10006-rdaeo:P20389 - rdac:C1002
	if	rdac:C10006-rdaeo:P20380 - rdac:C1002
	bf:Contribution-bf:role-bf:Role(relators:edt)	rdac:C10006-rdaeo:P20338 - rdac:C1002
	if	rdac:C10006-rdaeo:P20357 - rdac:C1002
	bf:Contribution-bf:role-bf:Role(relators:nrt)	rdac:C10006-rdaeo:P20053 - rdac:C1002
	if	
	bf:Contribution-bf:role-bf:Role(relators:prf)	
	if	
	bf:Contribution-bf:role-bf:Role(relators:ctb)	
	if	
	bf:Contribution-bf:role-bf:Role(relators:spk)	
	if	
	bf:Contribution-bf:role-bf:Role(relators:edc)	
	if	
	bf:Contribution-bf:role-bf:Role(relators:abr)	
	if	
	bf:Contribution-bf:role-bf:Role(relators:lbt)	
language	bf:Work - bf:language-bf:Language (languages:value)	rdac:C10006-rdaeo:P20006-languages:value

Table A4. Step (d) – Mapping the bf:Instance and bf:Item classes.

bf:Instance	*rdac:C10007 Manifestation*
bf:hasInstance	rdae:P20059 has manifestation of expression
bf:Item	rdac:C10003 Item
bf:hasItem	rdam:P30103 has exemplar of manifestation

Table A5. Step (d) – Mapping bf:Instance properties.

Information	BIBFRAME	RDA
	bf:Instance - rdf:type – if bf:Print if bf:Manuscript if bf:Archival if bf:Tactile if bf:Electronic	rdac:C10007 - rdam:P30001 "has carrier type"- carriers:nc rdac:C10007 - rdam:P30001 "has carrier type"- carriers:nz rdac:C10007 - rdam:P30001 "has carrier type"- carriers:zu rdac:C10006- rdaeo:P20001- tct AND rdac:C10007 - rdam:P30001 "has carrier type"- carriers:nc rdac:C10007 - rdam:P30001 "has carrier type"- carriers:cr
Carrier	bf:Instance-bf:carrier-carriers:value	rdac:C10007 - rdam:P30001 "has carrier type"- carriers:value
Issuance	bf:Instance-bf:issuance - issuance:value	rdac:C10007 - rdam:P30003 "has mode of issuance"- issuance:value
Media	bf:Instance-bf:media- mediaTypes:value	rdac:C10007 - rdam:P30002 "has media type" - mediaTypes:value
Dimensions	bf:Instance-bf:dimensions-rdfs:literal	rdac:C10007 - rdam:P30169 "has dimensions"-rdfs:literal
Extent	bf:Instance-bf:extent-bf:Extent - rdfs:label	rdac:C10007 - rdam:P30182 "has extent of manifestation" -rdfs:literal
Title	bf:Instance - bf:title-bf:Title- bf:mainTitle-rdfs:literal	rdac:C10007 - rdam:P30134 "has title of manifestation" - rdfs:literal
Provision Activity Statement	If there is one bf:Instance - bf:provisionActivityStatement- rdfs:literal statement else Check the following Provision Activity	rdac:C10007-rdam:P30292 "has manifestation statement"- rdf:literal
Provision Activity	IF bf:Instance-bf:provisionActivity-bf:ProvisionActivity-rdf:type-bf:Publication then bf:Instance - bf:provisionActivityStatement-rdfs:literal bf:Instance-bf:provisionActivity-bf:ProvisionActivity-bf:agent-bf:Agent bf:Instance-bf:provisionActivity-bf:ProvisionActivity-bf:date-xsd:integer bf:Instance-bf:provisionActivity-bf:ProvisionActivity-bf:place-bf:Place	rdac:C10007- rdam:P30289"has manifestation publication statement" - rdfs:literal rdac:C10007 - rdam:P30083 "has publisher agent" - rdac:C10002 "agent" rdac:C10007-rdamd:P30011 "has date of publication" - xsd:integer rdac:C10007 - rdam:P30088 "has place of publication" - rdac:C10009 "place"

(*continued*)

Table A5. Continued.

Information	BIBFRAME	RDA
	IF bf:Instance-bf:provisionActivity- bf:ProvisionActivity-rdf:type- bf:Distribution then bf:Instance - bf:provisionActivityStatement- rdfs:literal bf:Instance-bf:provisionActivity- bf:ProvisionActivity-bf:agent- bf:Agent bf:Instance-bf:provisionActivity- bf:ProvisionActivity-bf:date- xsd:integer bf:Instance-bf:provisionActivity- bf:ProvisionActivity-bf:place-bf:Place	rdac:C10007- rdam:P30283"has manifestation distribution statement" - rdfs:literal rdac:C10007 - rdam:P30080 "has distributor agent" - rdac:C10002 "agent" rdac:C10007 –rdam:P30008 "has date of distribution" - xsd:integer rdac:C10007 - rdam:P30085 "has place of distribution" - rdac:C10009 "place"
	IF bf:Instance-bf:provisionActivity- bf:ProvisionActivity-rdf:type- bf:Manufacture then bf:Instance - bf:provisionActivityStatement- rdfs:literal bf:Instance-bf:provisionActivity- bf:ProvisionActivity-bf:agent- bf:Agent bf:Instance-bf:provisionActivity- bf:ProvisionActivity-bf:date- xsd:integer bf:Instance-bf:provisionActivity- bf:ProvisionActivity-bf:place-bf:Place	rdac:C10007- rdam:P30287"has manifestation manufacture statement" - rdfs:literal rdac:C10007 - rdam:P30082 "has manufacturer agent" - rdac:C10002 "agent" rdac:C10007-rdam:P30010 "has date of manufacture" - xsd:integer rdac:C10007 - rdam:P30087 "has place of manufacture" - rdac:C10009 "place"
	IF bf:Instance-bf:provisionActivity- bf:ProvisionActivity-rdf:type- bf:Production then bf:Instance - bf:provisionActivityStatement- rdfs:literal bf:Instance-bf:provisionActivity- bf:ProvisionActivity-bf:agent- bf:Agent bf:Instance-bf:provisionActivity- bf:ProvisionActivity-bf:date- xsd:integer bf:Instance-bf:provisionActivity- bf:ProvisionActivity-bf:place-bf:Place	rdac:C10007- rdam:P30288"has manifestation production statement" - rdfs:literal rdac:C10007 - rdam:P30081 "has producer agent of unpublished manifestation" - rdac:C10002 "agent" rdac:C10007-rdam:P30009 "has date of production" - xsd:integer rdac:C10007 - rdam:P30086 "has place of production" - rdac:C10009 "place"
Relationships	bf:Instance - bf:accompaniedBy - bf:Instance bf:Instance - bf:hasPart - bf:Instance bf:Instance - bf:hasReproduction - bf:Instance	rdac:C10007 - rdam:P30027 "is accompanied by manifestation"- rdac:C10007 rdac:C10007 - rdam:P30033 "has part manifestation" - rdac:C10007 rdac:C10007 - rdam:P30039 "is reproduced as manifestation" - rdac:C10007