IONIAN UNIVERSITY

DEPARTMENT OF ARCHIVES, LIBRARY SCIENCE AND MUSEOLOGY

Sofia Zapounidou

PhD Thesis

Study of library data models in the Semantic Web environment

Corfu, 2020

Zapounidou, S. (2020). Study of library data models in the Semantic Web environment (Doctoral dissertation). Ionian University, Department of Archives, Library Science and Museology, Corfu, Greece. doi:10.5281/zenodo.4018523

ORCID ID: 0000-0001-8784-9581

Study of library data models in the Semantic Web environment

PhD Thesis

Sofia Zapounidou

Advisory Committee

Christos PAPATHEODOROU (Supervisor), Professor, Dept. of History and Philosophy of Science, National and Kapodistrian University of Athens.

Manolis GERGATSOULIS, Professor, Dept. of Archives, Library Science and Museology, Ionian University.

Theodore DALAMAGAS, Research Director, Vice Director of Information Management Systems Institute at ATHENA Research Center.

Members of the PhD Defense Committee

Panos CONSTANTOPOULOS, Professor, Dept. of Informatics, Athens University of Economics and Business.

Michalis SFAKAKIS, Associate Professor, Dept. of Archives, Library Science and Museology, Ionian University.

Manolis DRAKAKIS, Assistant Professor, Dept. of Archives, Library Science and Museology, Ionian University.

Emmanouel GAROUFALLOU, Assistant Professor, Dept. of Library Science, Archives and Information Systems, International Hellenic University.

Μελέτη μοντέλων δεδομένων βιβλιοθηκών στο περιβάλλον του Σημασιολογικού Ιστού

Διδακτορική διατριβή

Σοφία Ζαπουνίδου

Συμβουλευτική Επιτροπή

Χρήστος ΠΑΠΑΘΕΟΔΩΡΟΥ (Επιβλέπων), Καθηγητής, Τμ. Ιστορίας και Φιλοσοφίας της Επιστήμης, Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών.

Μανόλης ΓΕΡΓΑΤΣΟΥΛΗΣ, Καθηγητής,Τμ. Αρχειονομίας, Βιβλιοθηκονομίας & Μουσειολογίας, Ιόνιο Πανεπιστήμιο.

Θοδωρής ΔΑΛΑΜΑΓΚΑΣ, ερευνητής Α' του Ερευνητικού Κέντρου “ΑΘΗΝΑ”.

Μέλη της εξεταστικής επιτροπής

Πάνος ΚΩΝΣΤΑΝΤΟΠΟΥΛΟΣ, Καθηγητής, Τμ. Πληροφορικής, Οικονομικό Πανεπιστήμιο Αθηνών.

Μιχάλης ΣΦΑΚΑΚΗΣ, Αναπληρωτής Καθηγητής, Τμ. Αρχειονομίας, Βιβλιοθηκονομίας & Μουσειολογίας, Ιόνιο Πανεπιστήμιο.

Εμμανουήλ ΔΡΑΚΑΚΗΣ, Επίκουρος Καθηγητής, Τμ. Αρχειονομίας, Βιβλιοθηκονομίας & Μουσειολογίας, Ιόνιο Πανεπιστήμιο.

Εμμανουήλ ΓΑΡΟΥΦΑΛΛΟΥ, Επίκουρος Καθηγητής, Τμ. Βιβλιοθηκονομίας, Αρχειονομίας και Συστημάτων Πληροφόρησης, Διεθνές Πανεπιστήμιο Ελλάδας.

Abstract

To service users' information needs is the mission of libraries. Inevitably, the Library Catalog as a tool and its objectives evolve based on users' needs and seeking behavior. During the 20th century, the evolution of the catalog's objectives provoked the identification of bibliographic entities, their attributes and their relationships. Thanks to the entity-relationship modeling, these conceptualizations were first expressed by the FRBR model (Functional Requirements for Bibliographic Records). Nowadays, bibliographic entities need to be remodeled exploiting current technologies, known as the Semantic Web, that render data machine-understandable, and provide structure, meaning and trust to the existing World Wide Web. Linked data is a step further enabling the linking of these machine-understandable representations. In this context, bibliographic relationships and families may serve the linking of bibliographic entities and exploration within and beyond the Library Catalog. Thus, library data will be linked to other data to serve new user tasks out of the library environment and in a wide variety of domains.

The overview of the current bibliographic conceptual models presents an abundance of them with differences in terms of the numbers of bibliographic entities and relationships they define. Existing library linked datasets that have exploited these models are very different to one another in terms of modeling and selection of vocabularies. Thus, even though linked data technologies are used, the understanding of the data in the datasets is not ensured. This is a semantic interoperability issue that needs to be resolved to avoid the development of library linked datasets that end up isolated and unused. There have been taken some related initiatives; two mappings between non-library models (schema.org and European Data Model) and the FRBR have been attempted, and studies mostly with regard to the interoperability between models' core entities. There are no mappings between library models and almost no study exists on the preservation of bibliographic relationships as linking mechanisms in the linked data environment. Toward the goal of semantic interoperability and mappings, bibliographic conceptual models need to be compared to discover similarities and divergences in terms of modeling, granularity, constructs, and linking mechanisms.

The main research question of the thesis is: "Is semantic interoperability between conceptual bibliographic data models feasible?" To answer this question, the thesis poses four objectives: 1) to study and to compare bibliographic models identifying similarities and differences, 2) to develop mappings between the models, 3) to assess the mappings using a testbed, and 4) to identify any possible prerequisites or good cataloging practices for better mappings. The study of the models focuses on 5 models of the library domain, FRBR and its consolidation LRM, FRBRoo, RDA, BIBFRAME, and the EDM, a cultural heritage domain model. The inspection uses real-world cases to discover how core bibliographic entities, common bibliographic relationships (derivative, equivalence, and aggregates), and bibliographic families are represented by each model. This study reveals similarities that may enable semantic interoperability, as well as important differences that may impede it. The results have been organized using the Haslhofer and Klas categorization of metadata heterogeneities. A BIBFRAME-EDM application profile and three mappings (FRBR-BIBFRAME, RDA-BIBFRAME, and BIBFRAME-RDA) have been developed attempting to reconcile the heterogeneities identified between the models. The mappings are assessed using Gold Datasets to ultimately exhibit the success of the mappings. There are cases that semantics is lost after the conversions, but these losses are due to the models' conceptualizations and not due to the mappings.

The results of the thesis confirm that semantic interoperability may be achieved under specific conditions. All the conditions, prerequisites and good practices identified during the study of the models, the development of the mappings and their assessment using the approach of the Gold Datasets, involve cataloging policy decisions. Thus, the final thesis statement advocates for better cooperation between stakeholders and the adoption of a common mindset and practices to resolve heterogeneities of the past and to prevent new ones from happening.

Keywords: Cataloging; Conceptual models; Library linked data; Mappings; Semantic Interoperability; Semantic Web;

Μελέτη μοντέλων δεδομένων βιβλιοθηκών στο περιβάλλον του Σημασιολογικού Ιστού

Περίληψη

Κύριο αντικείμενο των βιβλιοθηκών είναι η εξυπηρέτηση των πληροφοριακών αναγκών των χρηστών τους. Οι κατάλογοι βιβλιοθηκών και οι στόχοι τους αναπόφευκτα εξελίσσονται με βάση τις ανάγκες των χρηστών και τον τρόπο που οι τελευταίοι αναζητούν. Κατά τη διάρκεια του 20ου αιώνα, η εξέλιξη των στόχων των καταλόγων οδήγησε στον καθορισμό των βιβλιογραφικών οντοτήτων, των χαρακτηριστικών τους και των σχέσεών τους. Αυτές οι έννοιες για πρώτη φορά εκφράστηκαν όλες μαζί σε ένα εννοιολογικό μοντέλο οντοτήτων-σχέσεων, το FRBR (Functional Requirements for Bibliographic Records). Σήμερα απαιτείται η εκ νέου μοντελοποίηση των βιβλιογραφικών οντοτήτων βάσει των νέων τεχνολογιών που προσφέρει ο Σημασιολογικός Ιστός, οι οποίες επιτρέπουν την κατανόηση των δεδομένων από μηχανικά συστήματα και μπορούν να παράσχουν στον υπάρχοντα Παγκόσμιο Ιστό δομή, νόημα και αξιοπιστία δεδομένων. Τα Συνδεδεμένα Δεδομένα αποτελούν ένα βήμα πέρα από τον Σημασιολογικό Ιστό επιτρέποντας τη διασύνδεση των μηχανικά κατανοητών δεδομένων. Σε αυτό το πλαίσιο, οι βιβλιογραφικές σχέσεις και οι βιβλιογραφικές οικογένειες μπορούν να υποστηρίξουν τη διασύνδεση των βιβλιογραφικών οντοτήτων και την εξερεύνηση των βιβλιογραφικών δεδομένων εντός και πέρα του Καταλόγου Βιβλιοθήκης. ‘Ετσι τα βιβλιοθηκονομικά δεδομένα θα μπορούν να διασυνδεθούν με άλλα δεδομένα και να υποστηρίξουν νέες ανάγκες των χρηστών βιβλιοθηκών εκτός του περιβάλλοντος της βιβλιοθήκης και σε μια ευρεία ποικιλία θεματικών πεδίων.

Η επισκόπηση των υπαρχόντων βιβλιογραφικών εννοιολογικών μοντέλων αναδεικνύει την ύπαρξη μίας πληθώρας μοντέλων με σημαντικές διαφορές μεταξύ τους όσον αφορά στον αριθμό των οντοτήτων και των βιβλιογραφικών σχέσεων που ορίζουν. Υπάρχοντα σύνολα συνδεδεμένων δεδομένων στο χώρο των βιβλιοθηκών που έχουν χρησιμοποιήσει αυτά τα μοντέλα παρουσιάζουν επίσης σημαντικές διαφοροποιήσεις όσον αφορά το μοντέλο και τα λεξιλόγια που επιλέχθηκαν για την υλοποίησή τους. Έτσι, αν και χρησιμοποιούνται τεχνολογίες συνδεδεμένων δεδομένων, η κατανόηση των δεδομένων δεν εξασφαλίζεται. Αυτό αποτελεί ζήτημα σημασιολογικής διαλειτουργικότητας η επίλυση του οποίου είναι απαραίτητη για την αποφυγή ύπαρξης συνόλων δεδομένων στον Σημασιολογικό Ιστό που καταλήγουν απομονωμένα και αχρησιμοποίητα. Ήδη έχουν αναληφθεί κάποιες πρωτοβουλίες σχετικά με τη σημασιολογική διαλειτουργικότητα. Συγκεκριμένα έχουν επιχειρηθεί δύο αντιστοιχίσεις (mappings) μεταξύ δυο μη-βιβλιοθηκονομικών μοντέλων (schema.org και Europeana Data Model) με το βιβλιοθηκονομικό FRBR, όπως και σχετικές έρευνες με εστίαση στη διαλειτουργικότητα μεταξύ των κύριων βιβλιογραφικών οντοτήτων των μοντέλων. Αντιστοιχίσεις μεταξύ βιβλιοθηκονομικών μοντέλων δεν υπάρχουν και σχεδόν καμία έρευνα δεν έχει διεξαχθεί που να μελετά τη διατήρηση των βιβλιογραφικών σχέσεων ως μηχανισμών διασύνδεσης στο περιβάλλον των συνδεδεμένων δεδομένων.

Το κύριο ερώτημα της διατριβής είναι: “Είναι η σημασιολογική διαλειτουργικότητα μεταξύ των εννοιολογικών μοντέλων των δεδομένων των βιβλιοθηκών εφικτή;” Για την απάντηση του ερωτήματος, η διατριβή θέτει 4 στόχους: 1) τη μελέτη και σύγκριση εννοιολογικών μοντέλων δεδομένων βιβλιοθηκών με στόχο την ανίχνευση ομοιοτήτων και ετερογενειών, 2) την ανάπτυξη αντιστοιχίσεων μεταξύ των μοντέλων, 3) την αξιολόγηση των αντιστοιχίσεων χρησιμοποιώντας δεδομένα, και 4) τον προσδιορισμό προϋποθέσεων ή καλών πρακτικών καταλογογράφησης για καλύτερες αντιστοιχίσεις. Η μελέτη των μοντέλων εστιάζει σε 5 βιβλιοθηκονομικά μοντέλα (FRBR και την αναθεώρησή του LRM, FRBRoo, RDA, BIBFRAME) και σε ένα μοντέλο του χώρου της πολιτιστικής κληρονομιάς, το EDM. Στη μελέτη αυτή αξιοποιούνται πραγματικές περιπτώσεις με στόχο να ανιχνευθεί πώς το κάθε μοντέλο αναπαριστά βασικές βιβλιογραφικές οντότητες, κοινές

βιβλιογραφικές σχέσεις (σχέση παραγωγής, σχέση ισοδυναμίας και έργα που συναθροίζουν – συσσωματώνουν άλλα έργα) και βιβλιογραφικές οικογένειες. Η μελέτη αυτή αποκαλύπτει ομοιότητες μεταξύ των μοντέλων οι οποίες διευκολύνουν την σημασιολογική διαλειτουργικότητα και ετερογένειες που μπορεί να την δυσκολέψουν. Τα αποτελέσματα αυτής της μελέτης έχουν οργανωθεί σύμφωνα με την κατηγοριοποίηση για ετερογένειες μεταξύ μεταδεδομένων των Haslhofer και Klas. Στο πλαίσιο της διατριβής έχουν αναπτυχθεί ένα προφίλ εφαρμογής από το BIBFRAME στο EDM και τρεις αντιστοιχίσεις (FRBR-BIBFRAME, RDA-BIBFRAME, BIBFRAME-RDA) με στόχο την εξισορρόπηση των ετερογενειών μεταξύ των μοντέλων. Η επιτυχία των αντιστοιχίσεων αξιολογείται με τη χρήση πρότυπων συνόλων δεδομένων (Gold datasets) που αναπτύχθηκαν για αυτό τον σκοπό. Υπάρχουν περιπτώσεις όπου η σημασιολογία των δεδομένων δεν διατηρήθηκε, αλλά αυτές οι απώλειες οφείλονται σε ενδογενή σημασιολογικά χαρακτηριστικά (conceptualizations) των μοντέλων και όχι στις αντιστοιχίσεις.

Τα αποτελέσματα της διατριβής επιβεβαιώνουν ότι η σημασιολογική διαλειτουργικότητα μπορεί να επιτευχθεί κάτω από ορισμένες συνθήκες. Όλες οι συνθήκες, τα προαπαιτούμενα και οι καλές πρακτικές που ανιχνεύθηκαν κατά τη μελέτη των μοντέλων, την ανάπτυξη των αντιστοιχίσεων, των πρότυπων συνόλων δεδομένων και της αξιολόγησης των αντιστοιχίσεων, σχετίζονται με τους στόχους και τις πολιτικές ανάπτυξης καταλόγων που εφαρμόζονται. Η τελική δήλωση της διατριβής, λοιπόν, τάσσεται υπέρ της καλύτερης συνεργασίας μεταξύ των εμπλεκομένων φορέων και της υιοθέτησης κοινού τρόπου σκέψης και πρακτικών για την επίλυση ετερογενειών του παρελθόντος και την αποφυγή δημιουργίας νέων.

Λέξεις-κλειδιά: Αντιστοιχίσεις, Εννοιολογικά μοντέλα, Καταλογογράφηση, Σημασιολογική διαλειτουργικότητα, Σημασιολογικός Ιστός, Συνδεδεμένα δεδομένα βιβλιοθηκών.

Στους τρεις δασκάλους της ζωής μου!

Στον Γρηγόρη που εγώ ονειρευόμουν να του μάθω απλά να διαβάζει και αυτός μου δίδαξε άλλα πράγματα μεγαλύτερα και ουσιαστικά: αγάπη, δύναμη, πίστη. Δεν έχεις πάψει να μου λείπεις.

Στον Άγγελο-Γρηγόρη που με τις διπλές του χάρες μού έμαθε ξανά από την αρχή να ζω. Μου έμαθε να χαμογελάω ξανά.

Στη Σμαράγδα, στο μοναδικό μου πετράδι, που μου έμαθε ξανά να γελώ δυνατά και να χορεύω.

Acknowledgements

I would like to thank my supervisor, Professor Christos Papatheodorou, for his guidance, support and encouragement through each stage of my PhD. I would also like to thank Associate Professor Michalis Sfakakis for collaborating with Professor Papatheodorou and me in many articles helping me evolve as a researcher. The frequent online meetings and conversations with both of them were inspiring and challenging. Yet, I would not change a thing.

Ευχαριστίες

Στα ελληνικά μπορώ να εκφράσω τις ευχαριστίες μου από καρδιάς.

Η διαδρομή προς τη συγγραφή του διδακτορικού ήταν πολυετής και χωρίς την κατανόηση και καθοδήγηση του επιβλέποντα καθηγητή μου, Χρήστου Παπαθεοδώρου, δεν υπήρχε περίπτωση να τα καταφέρω. Δίπλα του έμαθα πολλά για τον άγνωστο και συνάμα γοητευτικό κόσμο της έρευνας. Και αν αυτά που έμαθα φαίνονται αρχικά ως μαθήματα προς μία νέα ερευνήτρια, αν το καλοσκεφτείς είναι και μαθήματα ζωής: η αποτυχία και πώς να τη χειριστείς, η επιμονή, η επίγνωση του λάθους και πώς να το διορθώσεις, η συνεχής βελτίωση, η απορία και πώς να την εκφράσεις, το εύρημα και πώς να το παρουσιάσεις με τον πιο απλό τρόπο αλλιώς ούτε κι εσύ δεν το έχεις καταλάβει. Σε αυτό το συναρπαστικό ταξίδι πέρα από την πετυχημένη επιλογή του κ.Παπαθεοδώρου στη θέση του επιβλέποντα είχα και την τύχη να συνεργαστώ με τον Μιχάλη Σφακάκη, άοκνος ερευνητής και πολύτιμος συνεργάτης. Ο συνδυασμός των προσωπικοτήτων και της θετικής αύρας των δύο αυτών ανθρώπων αποτελεί τη συνταγή του «δεν έτυχε-πέτυχε» της διατριβής μου. Τους ευχαριστώ πολύ και τους δυο.

Θα ήθελα να ευχαριστήσω τη διευθύντριά μου στη Βιβλιοθήκη & Κέντρο Πληροφόρησης ΑΠΘ Κατερίνα Νάστα για την κατανόηση και την στήριξή της. Επίσης, την συναδέλφισσα Ελευθερία Κοσέογλου για την εξαιρετική συνεργασία που έχω μαζί της όλα αυτά τα χρόνια και που ποτέ δε σταμάτησε να πιστεύει σε εμένα. Θα ήθελα να ευχαριστήσω τις κυρίες Αφροδίτη Μάλλιαρη και Ασπασία Τόγια, καθηγήτριες και οι δύο στο Τμήμα Βιβλιοθηκονομίας, Αρχειονομίας και Συστημάτων Πληροφόρησης του Διεθνούς Πανεπιστημίου Ελλάδας, για τη φιλία τους, τις όμορφες συνεργασίες και την εμπιστοσύνη τους στο πρόσωπό μου. Τέλος, ευχαριστώ τη συναδέλφισσα Ελένη Δρεπανίδου για τη βοήθειά της σε θέματα καταλογογράφησης και τον συνάδελφο Ηλία Κυριαζή που με διευκόλυνε με την αποστολή άρθρων στα οποία δεν είχα πρόσβαση.

Προς τις αγαπημένες μου φίλες Ειρήνη, Γιώτα, Χρηστίνα, Αντιγόνη, Αλεξάνδρα, Βένη, Μαρία, Γιάννα, Βάσω και Ρέα που με υπομονή δέχτηκαν αγόγγυστα την απουσία μου από τη ζωή τους για μεγάλα χρονικά διαστήματα λόγω των υποχρεώσεων του διδακτορικού, θέλω να πω «Κορίτσια μού λείψατε και έρχομαι για καφέ!».

Τα μέλη της οικογένειάς μου επωμίστηκαν αρκετά φορτία εξαιτίας μου. Θα μου επιτρέψετε να τους ευχαριστήσω έναν-έναν. Ο σύζυγός μου Δημήτρης έχει αποδείξει πώς είναι πάντα μαζί μου, στα δύσκολα και στα εύκολα. Μπορεί να φτάσει στην άλλη άκρη της γης για να κάνει πραγματικότητα τα όνειρα της οικογένειας. Αυτή η σιγουριά είναι και η δύναμη που μου χάρισε. Τον ευχαριστώ. Η αδερφή μου Ελένη είναι πρότυπο για μένα: ήρεμη δύναμη και έμπρακτη υποστηρίκτρια των ανθρώπων που αγαπάει. Αν και είμαι η μεγαλύτερη αδερφή, μεγαλώνοντας θέλω να της μοιάσω. Ένα ευχαριστώ είναι τόσο λίγο προς τους γονείς μου, Γιάννη και Σμαρώ, που ακούραστα από την πρώτη στιγμή που γεννήθηκα αγωνίζονται για μένα, θυσιάζονται, λύνουν προβλήματα, διαλύουν όποια απογοήτευση κάνοντας αυτό που ξέρουν οι γονείς να κάνουν καλύτερα: να δίνουν φτερά στα παιδιά τους. Είμαι ευγνώμων.

List of contents

List of figures

List of tables

List of abbreviations

AACR	Anglo-American Cataloguing Rules
BIBFRAME	Bibliographic Framework
BIBO	Bibliographic Ontology
BnE	Biblioteca Nacional de España
BnF	Bibliothèque nationale de France
CIDOC CRM	CIDOC Conceptual Reference Model (see also ICOM CIDOC)
DNB	Deutsche Nationalbibliothek
DPLA	Digital Public Library of America
DPLA MAP	Digital Public Library of America Metadata Application Profile
EDM	Europeana Data Model
E-R	Entity-Relationship
FaBiO	FRBR-aligned Bibliographic Ontology
FOAF	Friend of a Friend
FRAD	Functional Requirements for Authority Data
FRBR	Functional Requirements for Bibliographic Records
FRBROO	Functional Requirements for Bibliographic Records object-oriented version
FRSAD	Functional Requirements for Subject Authority Data
HTML	Hypertext Markup Language
ICOM	International Council of Museums
ICOM CIDOC	ICOM International Committee for Documentation
ICP	International Cataloguing Principles
IFLA	International Federation of Library Associations and Institutions
ILS	Integrated Library System
ISBD	International Standard Bibliographic Description
LRM	Library Reference Model
MADS	Metadata Authority Description Schema
MARC	Machine-Readable Cataloging
MODS	Metadata Object Description Schema
RDA	Resource Description and Access
SKOS	Simple Knowledge Organization System
W3C	World Wide Web Consortium
WEMI	Work, Expression, Manifestation, Item
XML	Extensible Markup Language

1. Introduction

This introduction briefly presents the context of the thesis and the definition of the problems arisen by the relevant state of the art, the aim of the thesis and its contribution. Lastly, the structure of the thesis is presented.

1.1. Establishing the context: Library data in the linked data universe

Semantic Web (SW) enables the publication of any type of data in a structured way so that applications can query data and draw inferences from it (W3C, 2015). Most important to develop a *Web of Data* using Semantic Web technologies, the published structured data needs to be linked to other data. This network of interrelated datasets is known as Linked Data. In the linked data universe, there has already been published library data by different libraries or institutions. The publication of these library datasets on the Semantic Web is based on different vocabularies that carry different semantics. These vocabularies are structured, standardized and encoded according to the existing technologies, and formulate conceptual models. Obviously, the conceptual models encompass structural (syntactic) and of course semantic dissimilarities. Due to these dissimilarities, the datasets implementing the models cannot be linked to one another, and thus different semantics inevitably challenge the understandability, linking and the reusability of published data by third-party applications.

Hence, a semantic interoperability issue arises threatening the datasets to be isolated and hindering their interlinking. Toward the creation of a Library Linked Open Data Cloud, library datasets need to be linked to one another and share common semantics that applications can use to query and inference. Concerns have already been raised about the semantic interoperability between the models used in library linked datasets (H. Park & Kipp, 2019; Patrício, Cordeiro, & Ramos, 2020; Suominen & Hyvönen, 2017; Tallerås, 2018). The question that emerges is if semantic differences between the models may be overcome, and if so, how.

1.2. Basic concepts – Relevant academic literature

This paragraph tries to provide all basic concepts used in this thesis by shortly describing the historical context in which they were formulated.

Library catalogs have been developed to serve user needs. Starting from the late 19th century, scholars tried to identify users' exact needs. These needs were first expressed by librarian C.A.Cutter as the "objects of the catalog" (Cutter, 1904). The three objectives of the catalog were *finding, collocating,* and *selecting.* The first objective involves finding a document when certain of its characteristics are known, i.e., author, title, subject. The second objective involves finding a document that is unknown to the user through the collocation of the document with other documents exhibiting a known shared characteristic, i.e., author, subject, kind of literature. The third objective involves selection. The library catalog must provide further information regarding the documents' edition or genre to help the user decide if the book in question may serve his/her information needs. Cutter's "object of the catalog" were later updated and expanded. Nowadays, they are known as "**user tasks**".

To serve user tasks, scholars identified the **entities** of the publication world in which users are mostly interested. Typical examples are author, title or subject of a book. In this context, the cataloging theory literature of the 19th and 20th centuries is considered: Panizzi's 91 rules for the British Museum Library Catalog (Panizzi, 1841), Cutter's definitions on the objectives of the Library Catalog (Cutter, 1904), the revision of these objectives by Lubetzky (Lubetzky, 1969, 1986) and the adoption of the revised objectives at the international level (International Federation of Library Associations, 1974; Spalding et al., 1967; *Statement of Principles adopted by The International Conference on Cataloguing Principles Paris, October 1961*, 1961), the identification of bibliographic entities that Library Catalogs should describe by prominent scholars, such as Cutter (Cutter, 1904), Lubetzky (Lubetzky, 1969, 1986), Petee (Pettee, 1936), Ranganathan (Ranganathan, 1955), and Verona (Verona, 1959).

After the "entities of interest" were defined, rules and content **standards** tried to provide guidelines regarding what information must be recorded and how it must be recorded to describe each entity in a standardized way. During almost the whole 20th century, bibliographic data was recorded on library catalog cards using the International Standard Book Description (ISBD) (International Federation of Library Associations, 1974) and the Anglo-American Cataloguing Rules (AACR) (Gorman, Winkler, & American Library Association, 1978; Spalding et al., 1967) standards. Each **library catalog card** gathered in one description different pieces of bibliographic information about different entities. The need for recording relationships between the entities was recorded in authority files about persons, organizations, events, places, and titles. Yet, to find these recorded relationships, one needed to further check the authority files, which were recorded on individual cards alphabetically, and located in a distinct part of the card catalog.

As technology progressed during the 20th century, libraries tried to take advantage of the current technological developments to better serve the identified user tasks and the bibliographic entities' descriptive needs. In late 1960s, libraries took advantage of the magnetic tapes storage technology of the time and developed the **Machine Readable Cataloging Standard – MARC** to store and exchange bibliographic data (Avram, 1975). The current version of MARC is MARC21. The information recorded in library catalog cards was copied to MARC records. Similarly to the library catalog cards, MARC **bibliographic records** in online catalogs accumulated different pieces of bibliographic information recorded serially in predefined fixed fields or value fields (Knapp, 1968). Automation enabled easier searching of the information stored in bibliographic records. Similarly to the card catalogs, online catalogs using the MARC structure enabled the recording of relationships in separate authority records that could be browsed alphabetically.

Later, database technology improved library catalogs furnishing more effective searching options. Entity-relationship modeling principles, on which the database technology was based, offered libraries the potential to better record the relationships existing between bibliographic entities. As a result, the library community developed a new model signaling the transition of future library catalogs from records-based to entity-based. The new model "**Functional Requirements for Bibliographic Records – FRBR**" (IFLA Study Group on the Functional Requirements for Bibliographic Records, 1998) was developed by the International Federation of Library Associations and Institutions (IFLA). FRBR redefined user tasks based on the potential of the then current technologies. Relying on the redefined user tasks and the cataloging theory of the early 20th century regarding the bibliographic entities of interest, the FRBR determined the entities in its context, the entities' attributes, and the relationships between the defined entities. The **core entities** of the model are *Work*, *Expression*, *Manifestation*, and *Item*, and they are also known as *WEMI.* They correspondingly represent the ideas of an author (*Work*), the set of signs used to realize the ideas (Expression), the embodiment of the signs in a palpable object such as a publication (*Manifestation*), and an exemplar of the embodiment kept by a library (*Item*). FRBR is considered a milestone in the history of cataloging theory (Denton, 2007).

The publication of FRBR enabled for the first time the systematic record of relationships between the bibliographic entities differentiating between **inherent** relationships and **bibliographic** relationships. Inherent relationships are the ones provided purposely by the model and correlate its core entities, e.g., a *Work is realized through an Expression.* Bibliographic relationships exist between the *WEMI* entities revealing either a content relationship (e.g., has a translation), or a structural relationship (e.g., has part). The exact nature of bibliographic relationships which were of interest to users and were somehow recorded by the library community in library catalogs was discovered by Tillett in her thesis (Tillett, 1987). Tillett studied both cataloging rules starting from Panizzi's 91 rules (Panizzi, 1841), and MARC records in library catalogs to identify the types of relationships that librarians tried to record. The result of her work was a taxonomy of bibliographic relationships:

1. Equivalence relationship to represent exact copies in cases of reproduction, reprinting, etc.
2. Derivative relationship to represent modifications such as translations, adaptations, dramatizations, etc.
3. Descriptive relationship to represent reviews, commentaries, annotated editions, etc.
4. Whole-part relationship to represent the structure between parts and their whole, e.g., book-chapter.
5. Accompanying relationship to represent supplements, complements, etc.
6. Sequential relationship to represent prequels, sequels, changes in the titles of serials, etc.
7. Shared characteristic relationship to represent the sharing of a mutual attribute, e.g., two books published by the same publisher, or translated by the same translator.

Later, Smiraglia expanded these categories enriching the derivative bibliographic relationship (Smiraglia, 1992, 2005). The representation of bibliographic relationships enables the formulation of **bibliographic families**, that is a set of related *Works* all originating from a common earlier *Work* known as the **progenitor**. The term "bibliographic family" was first used by Wilson (Wilson, 1968); Smiraglia further worked on the bibliographic families concept discovering that bibliographic families most frequently start expanding through derivation (Smiraglia, 2005; Smiraglia & Leazer, 1999).

Technology progressed rapidly after the first publication of the FRBR in 1998. Internet and the World Wide Web provided users with new types of material, new search tools, and new navigation possibilities. Nowadays, **Semantic Web** (Berners-Lee, 1998) and **Linked data** (Berners-Lee, 2009) provide the potential of representing entities of interest in a machine understandable format, linking these representations, and enabling applications to inference. Library data mostly remain in closed systems, out of the web, while users daily browse the internet consuming online services. Once again, technology developments have instigated the reformulation of user tasks and the addition of a new one: to **navigate and explore** (Galeffi, Bertolini, Bothmann, Rodríguez, & McGarry, 2017). With regard to the newly-added *explore* user task, bibliographic relationships and bibliographic families are considered key enablers. Especially, the derivative bibliographic relationship, with which most often bibliographic families start expanding, may really serve the explorability of bibliographic data in a library linked data universe.

Toward the creation of a library linked data universe and the goal of integrating library data into the web, libraries have undertaken related projects. Library linked data projects use various vocabularies and conceptualizations, meaning they have defined **new bibliographic conceptual models** for the representation of their data. Even though many of the used models are inspired by the FRBR model, (i) they represent the same semantics using totally different constructs, e.g., one model uses a property for representing the publisher of a book, while another model uses a whole path with classes and properties to represent the same piece of information, and most importantly (ii) they define different entities, attributes, and relationships representing different semantics. These differences have an impact on semantic interoperability.

Semantic interoperability relates to the common understanding of meaning, and may involve varying issues, such as, modeling, standards, schemas, value vocabularies, etc. (Zeng, 2019). Studies about library linked data projects have already identified important differences between the projects with regard to semantic interoperability issues, meaning the selected models, schemas and value vocabularies (Smith-Yoshimura, 2016, 2018; Suominen & Hyvönen, 2017; Tallerås, 2018; Ullah, Khusro, Ullah, & Naeem, 2018). The existence of so many and different library linked datasets further raises the concern if the published datasets can be linked to one another, and if they can ultimately support the *explore* user task (H. Park & Kipp, 2019; Patrício et al., 2020; Suominen & Hyvönen, 2017; Tallerås, 2018).

1.3. Main research question – Aim of the thesis

To serve the current user needs, library data should become part of the linked data universe and interoperate with the data existing in it. The inspection of the aims, core constructs, and linking mechanisms of the models under study in the literature review affirms that there are great differences among the models despite their library domain orientation. Moreover, pilot projects implementing linked data in libraries have used different models, or in cases where the same model is implemented, totally different choices were made regarding metadata element sets and vocabularies (Cagnazzo, 2017; Duchateau, Lumineau, & Aalberg, 2018; Frosterus, Dadvar, Hansson, Lappalainen, & Zapounidou, 2020; Hallo, Luján-Mora, Mate, & Trujillo, 2016; H. Park & Kipp, 2019; Smith-Yoshimura, 2016, 2018; Suominen & Hyvönen, 2017; Tallerås, 2017). Thus, even though linked data technologies are used, and technical interoperability is ensured, the meaning of the published library linked data is represented with different models and vocabularies. This is a semantic interoperability issue. It is evident that the observed lack of consensus in modeling and library linked data implementations will impede semantic interoperability (Cole, Han, Weathers, & Joyner, 2013).

Related studies undertaken by scholars focus on core constructs only, meaning core entities and inherent relationships (Baker, Coyle, & Petiya, 2014; Taniguchi, 2012, 2017a, 2018). At the same time, the preservation of bibliographic relationships as linking mechanisms in the linked data environment has not been thoroughly studied yet. Moreover, mappings between bibliographic data models have not yet been developed. Within this context, the existing literature presents a gap in studying the semantic interoperability between bibliographic models. Toward the goal of semantic interoperability and mappings, there is a need to compare bibliographic models to discover similarities and divergences in terms of modeling, granularity, constructs, and linking mechanisms.

The central research question of the thesis is: "**Is semantic interoperability between conceptual bibliographic data models feasible?**" This question can be further analyzed into the following:

- Is there some common ground between the bibliographic models? What are the similarities that support semantic interoperability, and what are the differences impeding it? How do models represent common real-world bibliographic cases? What are the core entities/classes and inherent relationships in each model? What are the bibliographic relationships they acknowledge and how do they represent them? Do they support the representation of families with their constructs?
- Is it possible to reconcile the identified differences? Can there be mappings? Is information included in core constructs (e.g., core entities/classes, inherent relationships and derivative bibliographic relationships) preserved after the mappings? Are there any losses of information or semantics after mappings?
- Are there any prerequisites or good cataloging practices that enable mappings?
- Based on the thesis' findings, can there be suggestions to stakeholders regarding the semantic interoperability between the models? Can there be suggestions to enable semantic interoperability between the implementations of the models?

To answer the research questions and to contribute to the study of semantic interoperability between bibliographic models, the thesis poses four objectives: 1) to study and to compare bibliographic models identifying similarities and differences, 2) to develop mappings between the models, 3) to assess the mappings using a testbed, and 4) to identify any possible prerequisites or good cataloging practices for better mappings.

1.4. Synopsis of the research design and methods

The thesis research has been conducted in three stages: 1) study of the models, 2) mappings, and 3) assessment of the mappings.

1.4.1. Study of the models

The study of the models focuses on FRBR and its consolidation LRM, FRBRoo, RDA, BIBFRAME, and EDM. The models selected for this investigation are the ones presenting granularity and being developed by reputable organizations, namely FRBR and its consolidation LRM developed by IFLA, RDA developed by the RDA Joint Steering Committee, FRBRoo developed by CIDOC and endorsed by IFLA, BIBRAME developed by the Library of Congress, and EDM developed by Europeana. Five models belong to the library domain, while EDM is a model developed and used in the cultural heritage domain. The models under study are inspected regarding their core constructs (core entities/classes and their inherent relationships), as well as their linking mechanisms in terms of bibliographic relationships and representation of bibliographic families. The inspection is based on representing real examples in the terms of each model and aims to discover how the core bibliographic entities and their relationships are captured and expressed by each model. The selected examples correspond to particular conceptualizations (common bibliographic patterns), such as the case of a single-volume monograph, which is the most common bibliographic description case. The bibliographic relationships included in the study, namely, derivative bibliographic relationship, equivalence relationship, and aggregates, are all identified by scholars as common ones (Bennett, Lavoie, & O'Neill, 2003; Neill, Žumer, & Mixter, 2015; Petek, 2007; Smiraglia, 1992, 1999; Smiraglia & Leazer, 1999; Tillett, 1987; Vellucci, 1995). This study concludes with similarities and differences between the models.

To organize the results about identified similarities and differences between the models, the Haslhofer and Klas categorization of metadata heterogeneities (Bernhard Haslhofer & Klas, 2010) is exploited. Haslhofer & Klas differentiate between semantic and structural heterogeneities either at the model level or at the instance level (Figure 1:1). Semantic heterogeneities involve the models' semantics (Bernhard Haslhofer & Klas, 2010). Due to the thesis' focus on the core entities and relationships of the models, instance-level heterogeneities are out of the scope of the current research and therefore they have not been taken under consideration (denoted with a red x mark in Figure 1:1). Domain conflicts involve differences between the domains on which each model focuses. There may be overlaps, inclusions, aggregation or even incompatibility between the models' domains. Terminological mismatches involve the use of different terms to express the same concept.

Structural heterogeneities appear due to model incompatibilities (Bernhard Haslhofer & Klas, 2010). The structural heterogeneities involving elements have not also been considered because the thesis inspects core modeling primitives (entities and classes) and not the exact elements used for their description (marked with a red x in Figure 1:1). The "domain representation conflicts" category has four types of conflicts (Bernhard Haslhofer & Klas, 2010):

i. Abstraction level incompatibilities. These may be observed when the "same real-world entities are arranged in different hierarchies" (Bernhard Haslhofer & Klas, 2010).
ii. Multilateral Correspondences. Due to abstraction level incompatibilities, one model's construct may correspond to multiple constructs of another model, and in reverse.
iii. Meta-level discrepancy. The same information is modeled with different constructs, e.g., in one model the information is represented as a discrete class, while in another model the same information is represented as an attribute. Four types of meta-level discrepancy are identified: content-value / attribute, entity / attribute, and content value / entity discrepancy.
iv. Domain coverage. These types of heterogeneities occur when real-world entities are represented in one model and not in another, despite of the models' focus on a common domain.

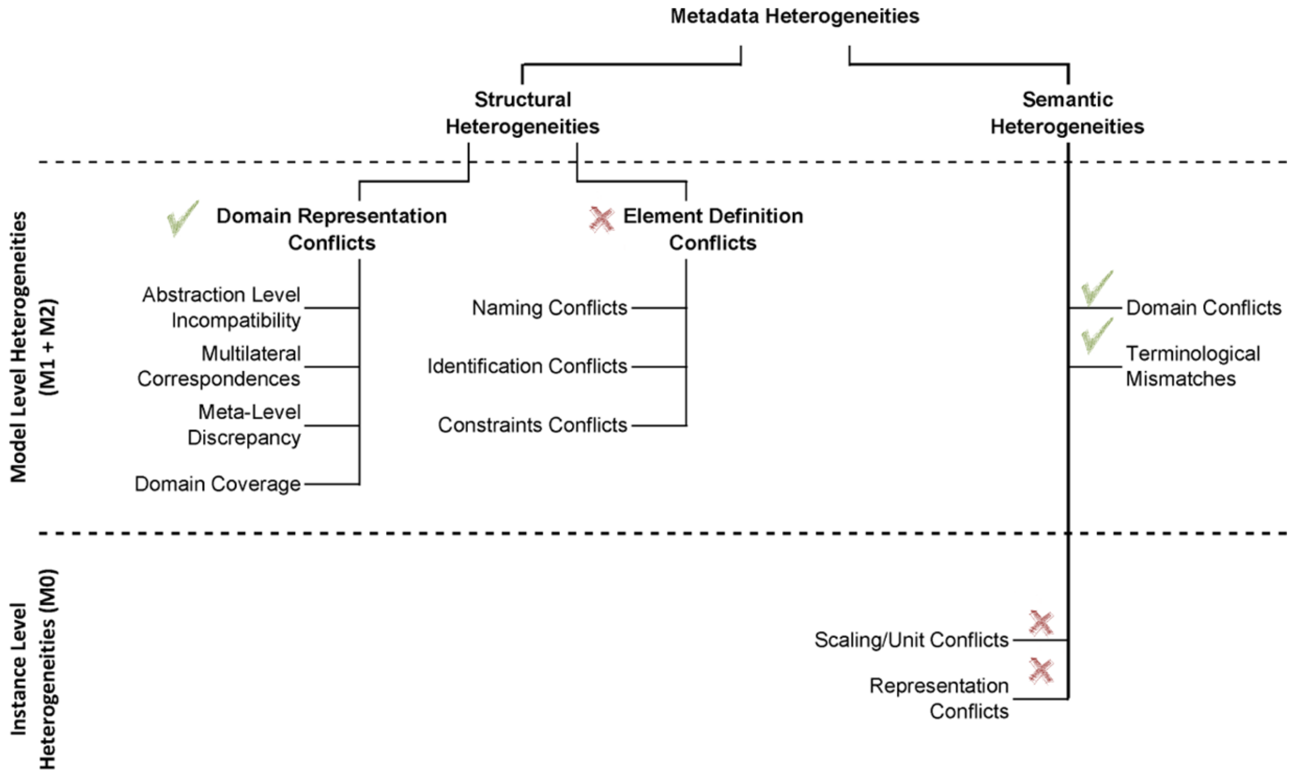


Figure 1:1. Structural and semantic heterogeneities according to the Haslhofer & Klas classification. Source of image: (Bernhard Haslhofer & Klas, 2010)*. On parts of the image, symbols have been added to declare which exact types of heterogeneities are taken under consideration by this thesis.*

Thus, the thesis studies semantic heterogeneities at the model level, and structural heterogeneities involving domain representation conflicts. Moreover, it extends this categorization by including the similarities between the models. Similarities are expected to enable mappings, while heterogeneities need to be tackled. Both identified similarities and heterogeneities are included in the "findings table" following the Haslhofer and Klas categorization for semantic (model-level only) and syntactic (domain representation conflicts only) heterogeneities.

1.4.2. Development of mappings

Haslhofer and Klas suggest that heterogeneities should be tackled to achieve interoperability and they identify three methods: (1) common use of a specific conceptual model; (2) development of a new meta-model with which all other models should comply or development of application profiles; and (3) mappings. The existence of many models in the same domain suggests that common adoption of one model or a meta-model is unlikely to happen. Thus, the thesis tests the application profile method by developing a BIBFRAME-EDM application profile and builds three mappings FRBR-BIBFRAME, RDA-BIBFRAME, and BIBFRAME-RDA. The mappings focus on core entities/classes, inherent relationships, and derivative bibliographic relationships. The derivative bibliographic relationship affects the explorability of data since bibliographic families most frequently start expanding through derivation (Smiraglia, 2005; Smiraglia & Leazer, 1999). The thesis approach in tackling each heterogeneity for the sake of mappings is added to the "findings table" (Figure 1:2) along with a "findings column" presenting the mappings' outcomes, as well as prerequisites and good cataloging practices for better mappings.

Category	Type	Heterogeneities	Thesis' 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

Figure 1:2. An excerpt from the "findings table" including the identified semantic heterogeneities and the thesis' approach toward reconciling them.

1.4.3. Assessment of mappings

The three mappings are assessed using Gold Datasets. These Gold Datasets have been created following the same set of principles and using each model's constructs to represent real-world bibliographic description cases. The cases involve eleven well-known literary works. For the development of the Gold Datasets, the publishing history of each work was thoroughly studied using literature and humanities-related resources. Thus, the Gold Datasets include eleven bibliographic families having members in various languages that also relate to other members of their family with derivative bibliographic relationships. For the assessment, a Gold Dataset is used as the source dataset being converted to an instance of the target model. Then, the generated dataset is compared to the corresponding Gold Dataset. As an example, in the RDA-BIBFRAME mapping, the Gold RDA dataset is converted to BIBFRAME, and the generated BIBFRAME dataset is later compared to the Gold BIBFRAME dataset. The assessment of mappings further enriches the "findings table" with more findings and prerequisites enabling mappings.

The Gold Datasets along with the datasets produced after the mappings are all uploaded on the following webpage http://libdata.tab.ionio.gr/models/si-mapping/si_project.html. This webpage presents the tools and data used, the mappings' results, and the published papers originating in the thesis' studies. It must be noted that both Gold Datasets and the datasets produced after the mappings have been uploaded on a Virtuoso RDF server and SPARQL queries can be submitted.

1.4.4. Presentation of findings

The discussion and thesis' conclusions are recorded in a new column of the "findings table". Thus, at the end of the thesis, the "findings table" presents the outcomes of the thesis' three stages and of the analysis that followed each one of them:

a) Similarities following the Haslhofer and Klas categorization.
b) Heterogeneities following the Haslhofer and Klas categorization.
c) The thesis' approach for reconciling heterogeneities.
d) Findings identified during the creation of the BIBFRAME-EDM application profile and the three mappings (FRBR-BIBFRAME, RDA-BIBFRAME, BIBFRAME-RDA), and the assessment of mappings. This column also includes prerequisites and good practices signified with the PGP acronym.
e) Conclusions, suggestions, and further work supported by the findings.

1.5. Contribution

The thesis contributes to the semantic interoperability between the models under study by providing answers to the posed research questions. The thesis has proved that there is some common ground (similarities) and important differences (heterogeneities). Identified heterogeneities have been reconciled using various approaches to build a BIBFRAME-EDM application profile and three mappings (FRBR-BIBFRAME, RDA-BIBFRAME, and BIBFRAME-RDA). The mappings and their following assessments determined several prerequisites and good practices that all enable semantic interoperability, proving that cataloging policy plays an important role in this context.

There are several important areas where this thesis makes an original contribution. To date, few studies have investigated aspects of the semantic interoperability issue in the bibliographic domain. The interoperability of well-known bibliographic models to non-bibliographic ones has been explored; OCLC has explored the compatibility of schema.org with FRBR and BIBFRAME (Godby, 2013; Godby & Vizine-Goetz, 2017; Godby, Wang, & Mixter, 2015), while the Europeana community has developed the EDM-FRBRoo application profile (Doerr et al., 2013). Most published studies tend to focus only on the interoperability of the models' core constructs (core entities/classes and inherent relationships) (Baker et al., 2014; Taniguchi, 2012, 2017a, 2018) ignoring bibliographic relationships. Some studies identify interoperability issues by observing differences regarding the models' implementations (Cagnazzo, 2017; Rasmussen Pennington & Cagnazzo, 2019; Tallerås, 2018), while others compare bibliographic models having MARC21 fields as a point of reference (H. Park & Kipp, 2019; Taniguchi, 2017a). The present research studies, for the first time, five bibliographic models (FRBR, LRM, RDA, FRBRoo, and BIBFRAME) and a cultural heritage one (EDM) taking under consideration core constructs (core entities/classes and inherent relationships), bibliographic relationships and families. The inclusion of bibliographic relationships and families in the thesis' research should make an important contribution to the support of the new *explore* user task in the library linked data environment. Using a path-oriented approach and exploring the representation of common bibliographic description cases, the thesis provides a comparative overview of the models and identifies similarities enabling interoperability, as well as heterogeneities impeding it.

The thesis investigates the reconciliation of the identified heterogeneities by developing the BIBFRAME-EDM application profile and three mappings (FRBR-BIBFRAME, RDA-BIBFRAME, and BIBFRAME-RDA). All built mappings and the application profile are the first ones to be developed in the library linked data domain. Moreover, they contribute to the identification of prerequisites for the preservation of semantics after conversion. The assessment stage focuses on the mappings between bibliographic models and uses three Gold Datasets, i.e., Gold FRBR, Gold RDA, and Gold BIBFRAME. To date, no other published Gold Datasets exist focusing on derivative relationships.

The thesis provides new insights into the semantic interoperability in the library linked data domain. It makes an original contribution to the understanding of how modeling constructs and modeling decisions may determine the semantic interoperability of the bibliographic models' instances. This understanding will help libraries in formulating cataloging policies that enable the preservation of semantics after conversions.

1.6. Thesis' structure

Chapter 2 presents the literature review of the thesis with references to cataloging history, standards and core concepts, bibliographic relationships and families. Semantic Web and Linked Data, the W3C Library Linked Data Incubator Group definitions and Library Linked Data initiatives are also considered. The literature review introduces current bibliographic models and differences between them in terms of core entities and bibliographic relationships. Semantic interoperability of bibliographic data in previous projects and studies are reviewed. The chapter concludes with the gaps identified in the literature survey. The four objectives of the thesis, that is the study of models, the development of mappings, the assessment of mappings, and the identification of prerequisites, are met in the next chapters.

Chapter 3 relates to the first objective of the thesis, the study of models. Chapter 3 investigates the representation of real-world bibliographic description cases in all six selected models, FRBR, LRM, FRBRoo, RDA, BIBFRAME, and EDM. The cases are single-volume monographs (related to core entities and inherent relationships), common bibliographic relationships (derivative, equivalence, and aggregates) and bibliographic families. By comparing the different representations, similarities and heterogeneities are identified and recorded in a “findings table” using the Haslhofer & Klas categorization.

Chapter 4 relates to the second objective of the thesis, the development of mappings. Chapter 4 presents the BIBFRAME-EDM application profile (testing the meta-model agreement method), and three mappings: FRBR-BIBFRAME, RDA-BIBFRAME, and BIBFRAME-RDA. The selected approach for reconciling each identified heterogeneity and the mappings’ findings are all included in the “findings table”. To serve the objective of identifying prerequisites for semantic interoperability, the “findings table” is enriched with prerequisites that proved during the development of mappings to enable semantic interoperability. Prerequisites and good practices are signified with the PGP acronym.

Chapter 5 relates to the third objective of the thesis, the assessment of mappings. Chapter 5 begins with the presentation of the Gold Datasets (Gold FRBR, Gold RDA, and Gold BIBFRAME), namely, the selection of bibliographic families included in the datasets, and the decisions taken for their development. Afterward, it presents the assessment of the three mappings; the occurrences of core entities and bibliographic relationships in the Gold Datasets are exhibited and they are later compared to the corresponding instances in the datasets produced after each mapping. During the creation of the Gold Datasets, and the assessment of mappings, more prerequisites and good practices enabling semantic interoperability are identified and, thus, included in the “findings table”.

Chapter 6 includes the discussion and conclusions. It begins with an overview of the thesis. Later, it discusses the thesis’ findings against the current literature. The importance of the findings is presented in relation to different stakeholders, i.e., scholars studying semantic interoperability issues, models’ editorial groups, libraries and cataloging agencies, and software developers. Along with the importance of findings, recommendations based on the thesis findings are proposed to the same stakeholders. The limitations of the thesis follow. The chapter then goes on with the future work that this hopefully new scholar wishes to conduct. In this final chapter all findings, approaches, conclusions, recommendations, and future work aspirations are displayed in a tabular format organized once again in the “findings table” that uses the Haslhofer & Klas categorization. The chapter and thesis conclude with the final thesis statement.

2. Literature review

The literature review starts with the user tasks and the role of technology in their evolution. Bibliographic relationships and bibliographic families are highlighted to describe their potential as linking mechanisms and enablers of the *explore* user task.

The literature review exhibits core Semantic Web and Linked Data concepts. After, the Library Linked Data domain is articulated using the definitions provided by the W3C Library Linked Data Incubator Group ("W3C Library Linked Data Incubator Group," 2012). The review proceeds with bibliographic conceptual models used in library linked data projects to present differences between them regarding core constructs (entities and inherent relationships) and bibliographic relationships. Later, some examples are given presenting how well-known bibliographic models have been used in library linked datasets. To date, there have been few investigations regarding semantic interoperability in the bibliographic domain. The review reports semantic interoperability projects and related studies by scholars.

The review concludes with the identification of gaps in the related literature.

2.1. Background

2.1.1. The user tasks and the evolution of technology

To service users' information needs is the mission of libraries. Scholars and librarians have analyzed the bibliographic universe and the users' searching behavior to determine the tasks the latter perform. These tasks defined the objectives of the Library Catalog as a tool and the evolution of library standards. In this context, technology has always had a significant influence. Library catalogs implement each era's current technological solutions and library standards to support users' needs, defined as user tasks. Interestingly, technological developments have fueled libraries with new user tasks and new technology standards that eventually impact the *Library catalog* as a tool.

Even though library catalogs exist from antiquity and the first printed Library Catalog for the University of Leiden Library was published in 1595, the publication of Panizzi's 91 rules for the British Museum Library Catalog in 1841 (Panizzi, 1841) is considered the starting point for library standards. Panizzi in his set of rules and in defending his work against critics (Panizzi, 1985) set the foundations of cataloging. Panizzi identified the role of catalogs and the user tasks they should serve. Moreover, he identified important pieces of information that need to be captured in a Library Catalog and provided guidelines for their systematic recording. Regarding technology, this is the analog era of book catalogs.

In the US, another acclaimed scholar, Charles Amni Cutter, elevated past ideas and good practices to the status of principles. In his 1876 "Rules for a Dictionary Catalog" book (Cutter, 1904), he provided a set of three catalog objectives, which he called "objects of the catalog" (Table 2-1). The first objective, known as the *finding* function, described core search tasks using a book's known attributes, i.e., search for a specific author, title, or subject. The second objective illustrated the *collocation* function, which was a new idea at the time. With the *collocation* function a library gathered all available information related to an author, a subject, or a specific literature genre. Thus, users could find through collocation unknown books to them that exhibited a known characteristic, e.g., a known author. For the third objective, the *selection* function, the Catalog provided all important information about the publication details of an edition or the literary/topical character of the book to help users decide what best fits their needs. Cutter's catalog objectives dominated the library theory and they were later updated and expanded according to the needs of each time. Most important, Cutter's objectives provided the framework for the development of cataloging standards. In the first decade of the 20th century card catalogs started replacing book catalogs. Cutter's rules and objectives were used to produce standardized cards and card catalogs. Card catalogs have been proven really successful, considering that the last card was published in 2015 (OCLC, 2015).

Table 2-1. The objects of the catalog, according to C.A.Cutter (Cutter, 1904).

1. To enable a person to find a book of which either: (A) the author is known (B) the title is known. (C) the subject is known 2. To show what the library has: (D) by a given author (E) on a given subject (F) in a given kind of literature 3. To assist in the choice of a book: (G) as to its edition (bibliographically) (H) as to its character (literary or topical)

During the 20th century, the library theory focused on standardization (Denton, 2007). Cutter's commonly accepted *Catalog Objectives* induced the identification of the inhabitants of the bibliographic universe. Scholars identified bibliographic entities that need to be captured to meet the catalog objectives. Pettee (Pettee, 1936), Lubetzky (Lubetzky, 1953, 1969, 1986) and Verona (Verona, 1959) differentiated between a *work* and its *manifestations* (known also as the 'content versus carrier' issue). Indian librarian S.R. Ranganathan differentiated between "expressed thought" and *manifestations* (Ranganathan, 1955). Moreover, these bibliographic entities needed to be properly described. American librarian Seymour Lubetzky revised Cutter's objectives (Lubetzky, 1953, 1969, 1986) and worked for standardization at both local and international level. His revised objectives (Table 2-2) simplified Cutter's objectives, clearly differentiated between two bibliographic entities, work and publication, and referred to *editions* in terms of new editions and translations. The second *collocation* objective suggested that all editions are arranged under author using an author-title heading.

Table 2-2. The objects of the catalog, according to S.Lubetzky (Lubetzky, 1969).

The objectives which the catalog is to serve are two: First, to facilitate the location of a particular publication, i.e., of a particular edition of a work, which is in the library. Second, to relate and display together the editions which a library has of a given work and the works which it has of a given author.

Lubetzky's revised objectives were adopted in the Paris principles in 1961 (*Statement of Principles adopted by The International Conference on Cataloguing Principles Paris, October 1961*, 1961) with slight but important modifications (Table 2-3). Paris Principles omitted the entity *work* from the first *finding* function seeming to not fully comprehend Lubetzky's differentiation (Yee, 1994). Nevertheless, Lubetzky's work further impacted the Anglo-American Cataloging Rules in 1967 (Spalding et al., 1967) and the International Standard Bibliographic Description in 1974 (International Federation of Library Associations, 1974). These developments were all related to descriptive cataloging and could be considered as pertaining to the *finding* objective only.

The second objective, the *collocation* one, was not fully adopted by the AACR rules; the rules regarding uniform titles, the collocation mechanism used in the then catalogs for *works* published under different titles, were made optional.

Table 2-3. The functions of the catalog in the Paris Principles (*Statement of Principles adopted by The International Conference on Cataloguing Principles Paris, October 1961*, 1961)*.*

2. Functions of the Catalogue The catalogue should be an efficient instrument for ascertaining 2.1. whether the library contains a particular book specified by (a) its author and title, or (b) if the author is not named in the book, its title alone, or (c) if author and title are inappropriate or insufficient for identification, a suitable substitute for the title; and 2.2. (a) which works by a particular author and (b) which editions of a particular work are in the library.

Even though Lubetzky himself referred to the collocation of a *work's* different editions and Wilson in 1968 (Wilson, 1968) described a set of different texts realizing the same work as the members of a family, the exact nature of bibliographic relationships was studied much later by Tillett (Tillett, 1987) and Smiraglia (Smiraglia, 1992, 2005). During the 1960s and 1970s, library cataloging practices and standardization remained constrained by the technology at hand. Bibliographic information was recorded on cards in card catalogs. These cards had specific dimensions (Figure 2:1) and described *manifestations* using the structure determined by the ISBD and the AACR rules. The production of cards became automated using the MAchine-Readable Cataloging format (known as MARC) (Avram, 1975) in the late 1960s.

Figure 2:1. A typical catalog card describing Kazantzakis' Askitiki. The data on the card is human-understandable. Photo taken from the Aristotle University of Thessaloniki old card catalog.

MARC was a record structure for the storage of bibliographic information on magnetic tapes. In the 1960s and the 1970s, magnetic tape storage was "the state of the art in data processing" (Hopkinson, 1984). Storage of and access to data was made sequentially (Figure 2:2). Thus, the MARC structure used fields, subfields, and indicators to separate different "sub-elements of data" (Hopkinson, 1984), and make data machine-readable.

In the 1970s, libraries started using MARC to transition bibliographic data to electronic records (Figure 2:3). Electronic catalogs replicated the card's record structure and they enabled keyword searching. Commercial vendors started developing Integrated Library Systems (ILSs) that used database technology, were more user-friendly (Figure 2:3) and enabled advanced search tasks. The MARC structure could be described as "a series of tagged literals or tagged text strings" (J.-R. Park, Andrew Brenza, & Richards, 2020). Due to this characteristic structure, MARC data was stored in a record-based format limiting the processability of bibliographic data in relational database settings. Thus, relational database potential for the querying and manipulation of entities and their relationships was not fully exploited, despite the fact that ILSs used database technology.


```
00837nam a22002051a
45000010007000000030007000070050017000140080041000310350014000720400015000860900025001O
1100006300126245006900189250001600258260005100274300002300325942001200348999001100360095
2026000371132474GrThAP20190520163505.0980519s1964    gr         000 1 gre d  ahzn133718
aAUHMcAUHM  aPA5610.K39bA94 19641 aΚαζαντζάκης, Νίκος,d1883-1957937025410aΑσκητική
salvatores dei /cN. Καζαντζάκη.  a3η εκδ.  aΑθήνα :bΚαζαντζάκη,c1964.  a95 σ. ;c21 εκ.  2lcc012
d132474  00102lcc406PA5610 K39 A94  01964708main9430986ab222bb222d2000-06-04g0.00kB3B -
Π.Κτήριο Φιλοσοφικής (204)l6m3oPA5610.K39A94
1964p2220011849q2022-01-31r2017-07-25s2017-07-25w2015-09-09yStandardz c. 1xAE.KB=180021
```


Figure 2:2.The raw MARC record describing Kazantzakis' Askitiki. The raw MARC record was created using the MARCedit tool (Reese, 2013). *The data in the MARC record is machine-readable and remains human-understandable.*

Ασκητική salvatores dei /

Main Author:	Καζαντζάκης, Νίκος, 1883-1957
Format:	Book
Language:	Greek
Published:	Αθήνα : Καζαντζάκη, 1964.
Edition:	3η εκδ.
Collection :	Main

NO IMAGE AVAILABLE

Holdings Description Similar Items Staff View

LEADER	00821nam a22002051a 4500	
001	132474	
003	GrThAP	
005	20190520163505.0	
008	980519s1964 gr 000 1 gre d	
035		**|a** hzn133718
040		**|a** AUHM **|c** AUHM
090		**|a** PA5610.K39 **|b** A94 1964
100	1	**|a** Καζαντζάκης, Νίκος, **|d** 1883-1957 **|9** 370254
245	1 0	**|a** Ασκητική salvatores dei / **|c** Ν. Καζαντζάκη.
250		**|a** 3η εκδ.
260		**|a** Αθήνα : **|b** Καζαντζάκη, **|c** 1964.
300		**|a** 95 σ. ; **|c** 21 εκ.
942		**|2** lcc **|0** 12
999		**|d** 132474
952		**|0** 0 **|1** 0 **|2** lcc **|4** 0 **|6** PA5610 K39 A94 01964 **|7** 0 **|8** main **|9** 430986 **|a** b222 **|b** b222 **|d** 2000-06-04 **|g** 0.00 **|k** B3B - Π.Κτήριο Φιλοσοφικής (204) **|l** 6 **|m** 3 **|o** PA5610.K39A94 1964 **|p** 2220011849 **|q** 2022-01-31 **|r** 2017-07-25 **|s** 2017-07-25 **|w** 2015-09-09 **|y** Standard **|z** c. 1 **|x** AE.KB=180021

Figure 2:3. The MARC21 record describing Kazantzakis' Askitiki in catalog. The data in this record is easier to understand by humans in contrast to the raw MARC record. Even though data in this MARC record is machine-readable, it remains human-understandable only. The screenshot is taken from the Aristotle University of Thessaloniki Library Catalog.

During the 1980s, libraries transitioned to online catalogs and proceeded with AACR updates to better serve online searches. By the end of the decade, it was really evident that the legacy AACR rules should be replaced. Thanks to the use of relational databases, a significant study was initiated, and libraries experienced two major changes. Barbara Tillett contributed to the Lubetzky's overlooked *collocation* objective by studying bibliographic relationships as a collocation mechanism. In her thesis' introduction, she described bibliographic relationships as an element of a bibliographic entity-relationship model (Tillett, 1987). Tillett's thesis formed the basis for Smiraglia's studies regarding derivative bibliographic relationships and bibliographic families (Smiraglia, 1992, 2005). Both Tillett's and Smiraglia's studies are further presented in 2.1.2 Bibliographic relationships and families. The changes that libraries witnessed involved users and materials. Users started implementing new information seeking practices in the new online catalog environments (Coyle, 2017; Sridhar, 2004). Libraries acquired new types of materials (Borgman, 1997), such as bibliographic databases and electronic journals that could not be described with the legacy AACR rules (the then current version was AACR2R) despite the updates during the 1980s. Thus, it was evident in the late 1980s that the then current standards and tools could not satisfy new user information seeking practices and new descriptive needs.

In August 1990, national libraries representatives met in Stockholm and agreed on the development of a new framework for understanding bibliographic record purposes and achieving a consensus among national libraries for bibliographic records exchange. In the resolutions of the 1990 Stockholm Seminar on Bibliographic Records, there was one asking for a clear delineation of "the functions performed by the bibliographic record with respect to various media, various applications, and various user needs" (IFLA Study Group on the Functional Requirements for Bibliographic Records, 2009). This resolution provoked the reexamination of the then cataloging rules by the IFLA Study Group on Functional Requirements for Bibliographic Records (Tillett, 1994). The Study Group recognized the "theoretical and practical implications of bibliographic records" (Tillett, 1994), considered the sound cataloging theory of the 19th and 20th centuries regarding the inhabitants of the bibliographic universe and exploited relational database modeling concepts. In 1997, at the Toronto "International Conference on the Principles & Future Development of the AACR", the procedures for the new model seemed mature (Biswas & Rath, 2014) and the revision of the AACR rules to follow up the new developments was decided. The new model "Functional Requirements for Bibliographic Records – FRBR" was published in 1998 (IFLA Study Group on the Functional Requirements for Bibliographic Records, 1998). After the 1997 Toronto conference, the AACR rules evolved into the Resource Description and Access (RDA) standard. The history and the evolution of the RDA standard is aligned with the FRBR model.

FRBR is considered a milestone in the history of cataloging theory (Denton, 2007) and has influenced other conceptual models too. FRBR used entity-relationship modeling primitives. Its publication signaled the future of library catalogs would be entity-based, and not records-based. Among the entities identified by FRBR, there are four core entities for the representation of "products of intellectual or artistic endeavor" (IFLA Study Group on the Functional Requirements for Bibliographic Records, 1998), namely, *Work, Expression, Manifestation,* and *Item*. The entity *Work* contains the ideas that form a distinct intellectual or artistic creation, obviously inspired by the views of Pettee (Pettee, 1936), Lubetzky (Lubetzky, 1953, 1969, 1986) and Verona (Verona, 1959). The *Expression* entity contains the set of signs (e.g., alphanumeric notation, music notation, image, etc.) to realize a *Work,* while the *Manifestation* entity refers to physical embodiment of an *Expression* (e.g., a publication). The distinction between "expressed thought" and its *manifestations* was first made by Indian librarian Ranganathan (Ranganathan, 1955). Exemplars of *Manifestations* are represented with *Item* entity. The four entities consist the Group 1 of FRBR entities. Group 2 entities, namely *Person* and *Corporate Body,* are the ones that relate to the Group 1 entities, e.g., a *Person* authored a *Work.* Group 3 entities serve as subjects for the *Work* entity. Group 3 entities are *Concept, Object, Event,* Place, and any of the Groups 1 and 2 entities. Group 2 and Group 3 entities were refined in subsequent reports (IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR), 2009, 2013; IFLA Working Group on the Functional Requirements for Subject Authority Records, 2010). FRBR also identified all the entities' attributes, and the relationships between them to enable the support of four newly defined user tasks, namely, *find, identify, select, obtain* (Table 2-4). The *find* user task implicated that users' search criteria are met. These criteria involved either an entity's attribute or a relationship. Thus, with the *find* user task the *collocation* objective, stated in earlier sets of Library Catalog Objectives (Cutter, 1904; Lubetzky, 1953, 1969, 1986; *Statement of Principles adopted by The International Conference on Cataloguing Principles Paris, October 1961*, 1961), was also met (Tillett, 2004). The *identify* user task enabled users to check within similar entities the one(s) they were originally searching for. The *select* user task involved the provision of more information to help users select the entity that best fits their requirements. Examples include content type, physical format, etc. All three user tasks (find, identify, select) could be applied to any of the Group 1 entities. The fourth user task, to *obtain*¸ could be applied only to the *Manifestation* and *Item* entities. It involved the physical or electronic access to the sought and selected entity.

Table 2-4. User tasks as defined in the FRBR (IFLA Study Group on the Functional Requirements for Bibliographic Records, 1998, 2009).

The tasks are defined in relation to the elementary uses that are made of the data by the user:

- to find entities that correspond to the user's stated search criteria (i.e., to locate either a single entity or a set of entities in a file or database as the result of a search using an attribute or relationship of the entity);
- to identify an entity (i.e., to confirm that the entity described corresponds to the entity sought, or to distinguish between two or more entities with similar characteristics);
- to select an entity that is appropriate to the user's needs (i.e., to choose an entity that meets the user's requirements with respect to content, physical format, etc., or to reject an entity as being inappropriate to the user's needs);
- to acquire or obtain access to the entity described (i.e., to acquire an entity through purchase, loan, etc., or to access an entity electronically through an online connection to a remote computer).

Almost in parallel to the FRBR development, during the 1990s, libraries started participating in digitization projects using new digital library systems and standards for resource description and discovery (Zeng & Qin, 2016). New metadata standards were created to describe the digitized objects, retain information about their contexts of creation and collection, and increase their findability. One of the first metadata standards was the Dublin Core Metadata Element Set ("DCMI: Dublin Core™ Element Set, v 1.0: Reference Description," 1998), a set of 15 basic elements for describing online resources. Bibliographic description evolved from records to resources, which could be anything, e.g., a book, a chapter, an artifact, a webpage, an online document, a whole collection. Metadata expanded serving varying goals, i.e., administrative, descriptive, preservation, technical, and use metadata (Gilliland, 2016). Metadata is not created at once and not by the same person. Instead, "metadata continues to accrue during the life of an information object or system" (Gilliland, 2016). In the 1990s, Web markup languages, such as HTML (Berners-Lee, 1991) and XML (Bray & Sperberg-McQueen, 1996), were developed. HTML was used for the creation of webpages. The XML language enabled the creation of metadata vocabularies and the use of these vocabularies for the description of data. Library digitization projects used the Dublin Core and developed other XML metadata structures to serve their needs, such as the MARCXML schema (Library of Congress' Network Development and MARC Standards Office, 2020), the Metadata Object Description Schema (MODS) (Library of Congress Standards, 2020), and the Metadata Encoding and Transmission Standard (METS) (Library of Congress Network Development and MARC Standards Office, 2019).

Due to the publication of the FRBR model and the technological developments, the Paris Principles were revised in 2009 formulating the *Statement of International Cataloguing Principles – ICP* (International Federation of Library Associations and Institutions, 2009). Interestingly, the 2009 ICPs (Table 2-5) used both FRBR and Web terminology, and considered Svenonius' critique on FRBR user tasks (Svenonius, 2009). The user tasks all referred to *resources;* further, in the *obtain* user task the term *data* is met for the first time. The user tasks described in the 2009 ICP were the same to the FRBR ones with an addition of a new fifth one, to *navigate.* To *navigate*, was a new user task proposed by Elaine Svenonius that could be met by exploiting the bibliographic relationships between *Works* to discover new *Works* related to the one initially sought (Svenonius, 2009). The 2009 ICPs used Svenonius' *navigate* task and rephrased it to explicitly include other FRBR entities too.

Table 2-5. Objectives and functions of the catalog in the 2009 Statement of International Cataloguing Principles (International Federation of Library Associations and Institutions, 2009).

4. Objectives and Functions of the Catalogue The catalogue should be an effective and efficient instrument that enables a user: 4.1. to **find** bibliographic resources in a collection as the result of a search using attributes or relationships of the resources: 4.1.1. to find a single resource 4.1.2. to find sets of resources representing all resources belonging to the same work all resources embodying the same expression all resources exemplifying the same manifestation all resources associated with a given person, family, or corporate body all resources on a given subject all resources defined by other criteria (language, place of publication, publication date, content type, carrier type, etc.), usually as a secondary limiting of a search result; 4.2. to **identify** a bibliographic resource or agent (that is, to confirm that the described entity corresponds to the entity sought or to distinguish between two or more entities with similar characteristics); 4.3. to **select** a bibliographic resource that is appropriate to the user's needs (that is, to choose a resource that meets the user's requirements with respect to medium, content, carrier, etc., or to reject a resource as being inappropriate to the user's needs); 4.4. to **acquire** or **obtain** access to an item described (that is, to provide information that will enable the user to acquire an item through purchase, loan, etc., or to access an item electronically through an online connection to a remote source); or to access, acquire, or obtain authority data or bibliographic data; 4.5. to **navigate** within a catalogue and beyond (that is, through the logical arrangement of bibliographic and authority data and presentation of clear ways to move about, including presentation of relationships among works, expressions, manifestations, items, persons, families, corporate bodies, concepts, objects, events, and places).

In the 2000s, there were many attempts to create FRBR catalogs. Even though the library community struggled to comprehend the notion of the FRBR entities, FRBR-ization projects were undertaken to resolve a major issue: the conversion of legacy library data from the MARC record structure format to FRBR. The extraction of entities and of descriptive information pertaining to each entity was not an easy task. The extraction was made using MARC fields and extended comparison of strings librarians had typed as fields' values (Aalberg, Haugen, & Husby, 2006; Decourselle, Duchateau, & Lumineau, 2015; Freire, Borbinha, & Calado, 2007; Hickey & O'Neill, 2009; Peponakis, Sfakakis, & Kapidakis, 2011). Despite the great number of FRBRization projects, FRBR catalogs were not achieved and the FRBR model continued to evolve with the addition of more tasks and entities needed for authority files and subjects (IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR), 2009, 2013; IFLA Working Group on the Functional Requirements for Subject Authority Records, 2010). The only true implementation of the FRBR model is the RDA. RDA rules are used to describe library materials in current ILSs still using the MARC record structure. As the library community struggled with the FRBR, the entity-relationship model (E-R model), on which FRBR is based, became obsolete and new technologies became available. These new technologies use conceptual models, enable descriptions in a machine-understandable way, and provide structure, meaning and trust to the World Wide Web. The sum of these technologies is called the Semantic Web.

Semantic Web was first envisioned by Tim Berners-Lee (Berners-Lee, 1998), the creator of the World Wide Web. The vision involves the representation of information in ways that both humans and software agents may understand and use. Linked data is a step further enabling the linking of these machine-understandable representations (Semantic Web and Linked Data are analytically presented in paragraph 2.2). In this context, the entities of the bibliographic universe, identified by scholars in the 20th century, may be described in a machine-understandable format (Figure 2:4) out of the library context and linked to other entities that may be of the bibliographic domain or not. Bibliographic relationships and families, collocation mechanisms developed by the library community, may serve the linking of bibliographic entities within and beyond the Library Catalog. Thus, library data may be linked to other data to serve new user tasks out of the library environment and in a wide variety of contexts.

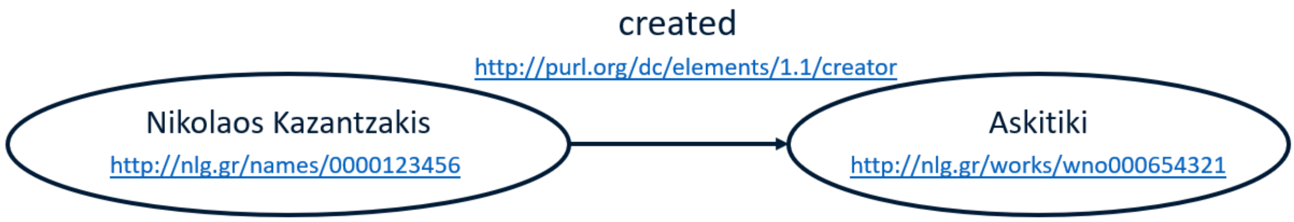


Figure 2:4. The representation of the assertion "Nikolaos Kazantzakis is the creator of Askitiki" in RDF, a core Semantic Web technology. This is a part of the data presented in the previous images of cards and MARC records. Yet, here the data is not in machine-readable format only; it is in a machine-understandable format, meaning that the data may be understood by both humans and software agents.

One new user task fueled by the World Wide Web and the Linked Data potential is the *explore* user task. The *explore* user task has been added in the current 2016 *International Cataloging Principles – ICP* (Galeffi et al., 2017) to describe a new function for the Library Catalog, that of enabling exploration beyond the catalog in non-library contexts (see point 6.5. in Table 2-6). The 2016 *ICPs* consider FRBR and its subsequent reports including all entities defined in them, e.g., the FRSAD *Thema* entity was added in the *find* user task (Table 2-6). The influence of current technological developments is explicitly stated by IFLA clarifying that the 2016 ICPs take "into consideration new categories of users, the open access environment, the interoperability and the accessibility of data, features of discovery tools and the significant change of user behaviour in general" ("IFLA -- Statement of International Cataloguing Principles (ICP) 2016," 2019).

Table 2-6. Objectives and Functions of the Catalog in the current International Cataloging Principles (Galeffi et al., 2017).

6. Objectives and Functions of the Catalogue

The catalogue should be an effective and efficient instrument that enables a user:

- 6.1. to **find** bibliographic resources in a collection as the result of a search using attributes or relationships of the entities:
  to find a single resource or sets of resources representing:
  - all resources realizing the same work
  - all resources embodying the same expression
  - all resources exemplifying the same manifestation
  - all resources associated with a given person, family, or corporate body
  - all resources on a given thema
  - all resources defined by other criteria (language, place of publication, publication date, content form, media type, carrier type, etc.), usually as a secondary limiting of a search result;
- 6.2. to **identify** a bibliographic resource or agent (that is, to confirm that the described entity corresponds to the entity sought or to distinguish between two or more entities with similar characteristics);
- 6.3. to **select** a bibliographic resource that is appropriate to the user's needs (that is, to choose a resource that meets the user's requirements with respect to medium, content, carrier, etc., or to reject a resource as being inappropriate to the user's needs);
- 6.4. to **acquire** or **obtain** access to an item described (that is, to provide information that will enable the user to acquire an item through purchase, loan, etc., or to access an item electronically through an online connection to a remote source); or to access, acquire, or obtain authority data or bibliographic data;
- 6.5. to **navigate** and **explore**
  - within a catalogue, through the logical arrangement of bibliographic and authority data and the clear presentation of relationships among entities
  - beyond the catalogue, to other catalogues and in non-library contexts.

It is clear that the Library Catalog as a tool evolves based on users' needs and seeking behavior. The evolution of the catalog's objectives, or user tasks, provoked the identification of bibliographic entities (Table 2-7), their attributes and their relationships. Thanks to the entity-relationship modeling, these conceptualizations were first expressed by the FRBR model, and now need to be represented with new technologies and models to facilitate the integration of library data into the Semantic Web. The explicit representation of bibliographic relationships and families using Semantic Web technologies will enable linking and will promote the development of a Library Linked Data Cloud that may interoperate with other data.

Table 2-7. User tasks: their definitions and the entities acknowledged in each one of them.

Definition in…	**Objectives / User tasks**	**Entities acknowledged**
Cutter, 1876	Finding Collocating Selecting	Book Edition Author Subject
Lubetzky, 1960	Find Collocate	Work Edition Publication Author
Paris Principles, 1961	Find Collocate	Work Edition Book Personal Author Corporate Body
FRBR, 1998, 2008	Find Identify Select Acquire/Obtain	**Group 1** (Work, Expression, Manifestation, Item) **Group 2 (**Person, Corporate Body) **Group 3** (Concept, Object, Event, Place)
ICP, 2009	Find Identify Select Acquire/Obtain Navigate	**Group 1** (Work, Expression, Manifestation, Item) **Group 2** (Person, Family, Corporate Body) **Group 3** (Concept, Object, Event, Place)
ICP, 2016	Find Identify Select Acquire/Obtain Navigate and Explore	**Group 1** (Work, Expression, Manifestation, Item) **Group 2** (Person, Family, Corporate Body) **Group 3** (Thema, Nomen)

2.1.2. Bibliographic relationships and families

Following the evolution of the objectives of catalogs and user tasks, it becomes evident that bibliographic relationships and families were perceived by scholars and the library community as linking and collocation mechanisms. Within the framework of the Semantic Web and to the support of the *explore* user task, it becomes evident that bibliographic relationships and families may facilitate the linking of bibliographic entities within and beyond the library catalog in totally new contexts. In this paragraph, the evolution of the concepts of bibliographic relationships and families is presented.

2.1.2.1. Bibliographic relationships

Cutter's second objective (Cutter, 1904), as rephrased by Lubetzky (Lubetzky, 1969) and later in the Paris Principles (*Statement of Principles adopted by The International Conference on Cataloguing Principles Paris, October 1961*, 1961), formulated the perception that bibliographic relationships may successfully serve the collocation of related bibliographic entities. Despite this common perception, bibliographic relationships were not studied until the last two decades of the 20th century.

Explicit definitions of bibliographic relationships were first provided in UNIMARC (IFLA Working Group on Content Designators, 1977, 1980), a MARC format developed by IFLA. Three types of relationships were identified: i) vertical relating the whole to its parts and the parts to its whole, ii) horizontal relationships between versions in different languages, formats, media, etc., and iii) chronological relationships between predecessors and successors, e.g., in serials. Goossens and Mazur-Rzesos (Goossens & Mazur-Rzesos, 1982) studied only the vertical/whole-part relationship and their findings could not be used for the study of bibliographic relationships in general. It is the works of Tillett and Smiraglia that have significantly contributed to the theory of information organization. Both empirically investigated the nature and extent of bibliographic relationships. Tillett performed a two-part investigation for her thesis (Tillett, 1987). First, she examined cataloging rules to reveal how the relationships were indicated in bibliographic records and to identify the typology of bibliographic relationships. Her study included all the rules ever published starting from Panizzi's 91 rules in 1841 till the second edition of the "Anglo-American Cataloguing Rules" in 1978. The main finding of her thesis has been the creation of a taxonomy of bibliographic relationships (Tillett, 1987). Tillett in her taxonomy often uses the *bibliographic item* term to describe the participants in a relationship. This term is not used in the list below, but it has been replaced by the FRBR/RDA entities, which may participate in each one of these relationships (IFLA Study Group on the Functional Requirements for Bibliographic Records, 2009; IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR), 2013; Picco & Ortiz Repiso, 2012). Tillett identified the following seven types of bibliographic relationships (Tillett, 1987), which are also depicted in Figure 2:5 with different colors to denote if they refer to content or structure:

1. Equivalence relationships. These may exist between either *Manifestations* or *Items*; examples include copies, reprints, microforms and reproductions.
2. Derivative relationships. These are called horizontal relationships in UNIMARC (IFLA Working Group on Content Designators, 1977, 1980), including either derivations of the same *Work* such as revisions and literal translations, or creation of new *Works* based on an earlier one. Examples of new derivative *Works* are adaptations, dramatizations and novelizations, free translations, imitations, and parodies. It must be noted that the treatment of literal translations as derivations of the same *Work* is a common approach in libraries, known as "realization approach". Other contexts such as the publishing industry perceive translations as new *Works* related to the original one. This approach is known as "derivation approach".
3. Descriptive relationships. These may exist between a description *Work* and its object of description, a *Work* or another bibliographic entity (*Expression, Manifestation, Item*). Examples include reviews, commentaries, annotated editions, etc.
4. Whole-part relationships. These are called vertical in UNIMARC (IFLA Working Group on Content Designators, 1977, 1980) or hierarchical by Goossens and Mazur-Rzesos (Goossens & Mazur-Rzesos, 1982) and they include a whole bibliographic entity (*Work, Expression, Manifestation, Item*) and its component parts.
5. Accompanying relationships. These relationships may exist between a bibliographic entity (*Work, Expression, Manifestation, Item*) and the bibliographic entity it accompanies (*Work, Expression, Manifestation, Item*). This relationship covers both cases where one entity is supplemented by the other (e.g., a teacher's guide to a textbook) and the case where the one bibliographic entity complements the other (e.g., a CD accompanying a textbook).
6. Sequential relationships. These relationships are called chronological relationships in UNIMARC. They may exist either between *Works* or *Expressions.* Typical examples are the sequels of a monograph or serials that have changed titles.
7. Shared characteristic relationships. These relationships may exist between two bibliographic entities (*Work, Expression, Manifestation, Item*) that share a common characteristic, e.g., common publisher, common place of publication, same language, etc.

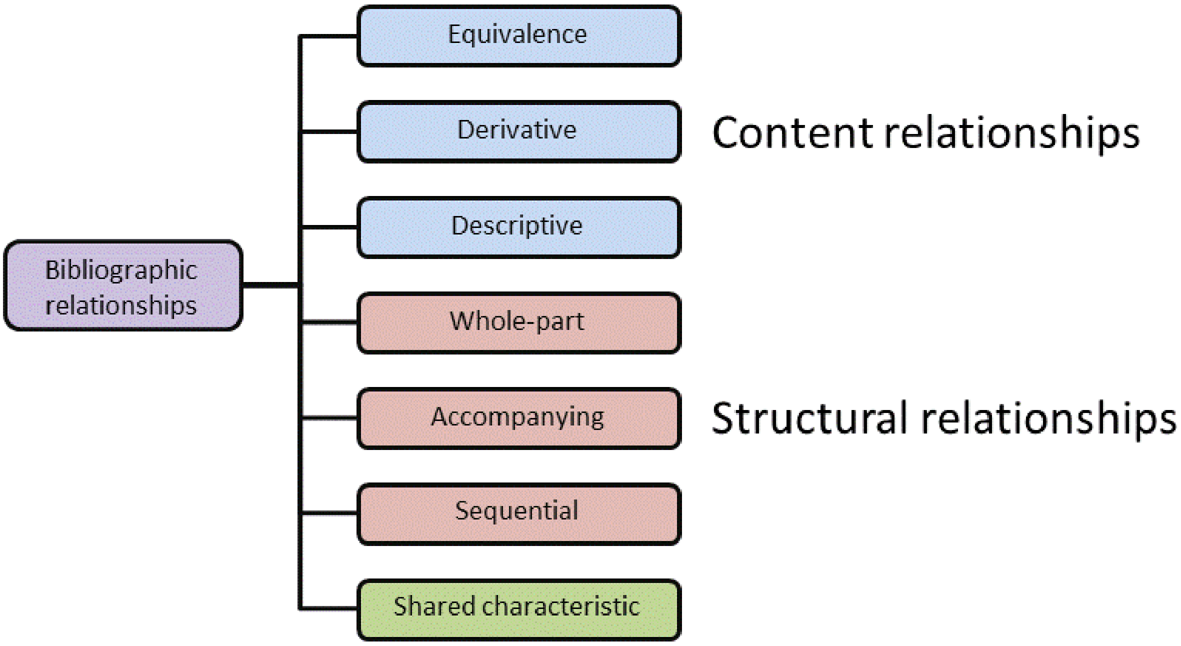


Figure 2:5. Tillett's taxonomy of bibliographic relationships. Equivalence, Derivative, and Descriptive relationships refer to the content of the related WEMI instances, whereas Whole-part, Accompanying, and Sequential refer to the structure of the related WEMI instances.

The second part of Tillett's thesis (Tillett, 1987) involved her empirical examination of the Library of Congress Catalog in terms of the occurrences of bibliographic relationships. She discovered that almost 75% of the total bibliographic records in the Library of Congress Catalog contains information regarding at least one bibliographic relationship type. She also discovered that some relationships are more frequent than others, e.g., whole-part relationships and derivative relationships. Tillett's study has been the first to analytically examine the types of bibliographic relationships (Noruzi, 2012).

The derivative relationship, as Tillett observed in her thesis, includes many types of derivations. Smiraglia focused on the derivative relationships and classified them into seven types (Smiraglia, 1992), applied between *Expressions* (Smiraglia, 2007a), either of the same *Work* or of different *Works* depending on the extent of change on the progenitor's ideational content, as follows:

1. Simultaneous derivations. They include (nearly) simultaneous editions.
2. Successive derivations. They include revisions, new editions.
3. Translations.
4. Amplifications. They include illustrated texts, musical settings, and criticisms, concordances and commentaries.
5. Extractions. They include abridgements, condensations and excerpts.
6. Adaptations. They include simplifications, screenplays, librettos, arrangements of musical works, and other modifications
7. Performances. They include sound or audiovisual recordings.

As Smiraglia's research progressed, new types of derivative bibliographic relationships were discovered. Smiraglia and Leazer in (Smiraglia & Leazer, 1999) add the "predecessor" and the "accompanying" derivative relationships. The *predecessor* relationship involves progenitor works having their own instantiation networks and being derived at the same time from earlier predecessor works. The *accompanying* relationship involves a derivation that physically accompanies the original work. With this relationship Smiraglia and Leazer describe the material that expands the theme and concepts of a work by accompanying it. They think that *accompanying relationship* resembles the amplification derivative relationship and should be included in

Smiraglia's classification. Material accompanying a work for marketing purposes without amplifying the accompanied work's theme or concepts is not considered for the *accompanying relationship*. Two more types of derivative relationships have been added due to Velucci's thesis on musical works, *notational transcription* and *musical representation* (Vellucci, 1995). In 2007, Smiraglia added one more relationship in his classification, *persistent works* (Smiraglia, 2007b). This relationship occurs in best-selling works that persist over time appearing in new editions.

All these types of derivative bibliographic relationships fall into two categories of derivation according to Smiraglia: derivation and mutation (Smiraglia, 2005). The former involves new instantiations of the original work that do not present any changes in the ideational or semantic content. Typical examples are simultaneous and successive editions. Mutation presents alteration of the progenitor's semantic or ideational content or both. Common types of mutation are translations and adaptations. Hence, Smiraglia's classification of derivation relationship types has evolved as follows (Smiraglia, 2009):

1. Derivations
    a. Simultaneous editions
    b. Successive editions
    c. Predecessors
    d. Amplifications
    e. Extractions
    f. Accompanying materials
    g. Musical presentations
    h. Notational transcription
    i. Persistent works
2. Mutations
    a. Translations
    b. Adaptations
    c. Performances

Smiraglia refined the derivative relationship in Tillett's taxonomy. Further refinements by other scholars may be also made. According to Leazer in (G.-H. Leazer, 1993), there can be more refinements regarding derivations from performances, and equivalents. The equivalence relationship may be refined to include reformatting (e.g., microformatting and photoduplication) and republication (G.-H. Leazer, 1993).

2.1.2.2. Bibliographic families

To *navigate* and to *explore* (Galeffi et al., 2017) have been identified as key user tasks that bibliographic data and catalogs need to support. Bibliographic relationships may serve as a linking mechanism between bibliographic resources in a linked data environment. An extra linking and collocation mechanism could be provided with the representation of bibliographic families. In legacy bibliographic data and catalogs, the collocation of members of the same bibliographic family was somewhat achieved with the use of uniform titles. Uniform titles, despite the inconsistencies in their use, have served in more than identifying the progenitor work. Uniform titles with specific additions in their structure, such as, language, part, etc., could also serve the identification of members of the progenitor's family related to the progenitor (Delsey, Dullabahn, & Heaney, 1999; Weihs & Howarth, 2008) with "at least four types of bibliographic relationship: equivalence, derivative, whole-part, and sequential relationships" (Fattahi, 1997). Thus, collocation of bibliographic families' members depends on the alphabetic display of the uniform titles' structured literals created by catalogers. Despite the human factor related to the creation of these structured literals, Smiraglia and Leazer in (G. H. Leazer & Smiraglia, 1999) have provided proof that these linear descriptions of relationships result in insufficient control of bibliographic families' "robust and complex structures".

Central to the notion of a bibliographic family is the concept of *Work* and its role as a collocation mechanism. The definition of *Work* has been a matter of discourse in the 20th century. As Yee observed in (Yee, 1995), cataloging theorists have used diverse criteria to define it, such as "creativity and/or single personal authorship, content, text or symbol strings, medium, identity and representation, and interchangeability, as well as the concept of work as product". Lubetzky's research has been crucial on the matter for two reasons. First, he is probably among the first theorists that insisted on differentiating between the intellectual *Work* and the material object used to convey it (Lubetzky, 1969). Secondly, based on this differentiation, Lubetzky rephrased Cutter's second objective of the library catalog to facilitate the collocation of different editions (*Manifestations* in FRBR terms) of a *Work*. This rephrased second objective has been included in the Paris "International Cataloguing Principles" (*Statement of Principles adopted by The International Conference on Cataloguing Principles Paris, October 1961*, 1961). Wilson in (Wilson, 1968) went one step further by differentiating between *Works* and "texts" (*Expressions* in FRBR terms). Moreover, he recognized that a *Work* may start a family, "the composing of one or more texts that are the ancestors of later members of the family" (Wilson, 1968). The importance of "texts" in bibliographic description was recognized by other scholars too. Taniguchi proposed a focus on *Expression*-level cataloging that among others could serve as the basis for indicating bibliographic relationships (Taniguchi, 2002).

Smiraglia used the Wilson *bibliographic family* metaphor in his research. He characterized "this network of related works … a bibliographic family" considering that the accumulated *Works* "deliberately share ideational and semantic content" from an original *Work* which he called as the *progenitor* of the family (Smiraglia, 1992). All other works/members in the family somehow derive from the *progenitor* work and may be related to it through different types of relationships. Smiraglia found out that a bibliographic family usually starts with a derivation relationship, successive edition or translation in particular (Smiraglia, 2005; Smiraglia & Leazer, 1999). Through empirical studies, he concluded that *bibliographic families* range in complexity from a family with only one work, the *progenitor* one, to a family with many derivative works, where some of them may have their own derivations (Smiraglia & Leazer, 1999). Moreover, the size of a *bibliographic family* often depends on the age of the *progenitor* and its popularity within a specific culture, also called *canonicity* (Smiraglia & Leazer, 1999). Smiraglia finally redefined the term *bibliographic family* and replaced it with the *instantiation network* to include all realizations of a work in time that come into being either as events, e.g., performances, book readings, or as embodiments in physical form, e.g., a book (Smiraglia, 2005).

Broader to the concept of the *bibliographic family* is the *superwork* concept defined by Svenonius in (Svenonius, 2009) or by Ed O'Neill according to Yee in (Yee, 1993). Svenonius used "Domanovszky's criterion of descent from a common origin" but without Domanovszky's prerequisite that the same identity (in terms of the same author-title) must be preserved in all related *Works*. Hence, she defined that a "*superwork* may contain any number of works as subsets, the members of which, while not sharing essentially the same information content, are nevertheless similar by virtue of emanating from the same ur-work" (Svenonius, 2009). She notices that members of a *superwork* are not necessarily members of the same work. Both Smiraglia and Svenonius recognize that *bibliographic families* (or *instantiation networks*) or *superworks* may serve not only as a collocation mechanism, better than main entry and uniform title linear approaches (Carlyle, 1996), but also as navigating mechanisms to the bibliographic universe (Smiraglia & Leazer, 1999; Svenonius, 2009).

Bibliographic relationships and families are poorly represented in legacy catalogs (Fattahi, 1996; Frias & Rios Hilario, 2002; G. H. Leazer & Smiraglia, 1999; Mercun, Zumer, & Aalberg, 2017; Salaba, Merčun, & Aalberg, 2018). Smiraglia and Leazer in (G. H. Leazer & Smiraglia, 1999) have provided proof that the legacy catalog design may not serve the description and control of bibliographic families that are often complex ones and evolve in time. After the addition of the *explore* user task in the latest "International Cataloguing Principles" (Galeffi et al., 2017), one can claim that the representation of bibliographic relationships and families is a prerequisite. Legacy cataloging-generated implicit representations of families using uniform-titles have been proved as inadequate (Carlyle, Ranger, & Summerlin, 2008; G. H. Leazer & Smiraglia, 1999). Their explicit representation will ensure that bibliographic data may transcend the limits of flat bibliographic records and closed library catalogs, and that they may interact with third-party data openly available through the Web (Coyle, 2010a; Picco & Ortiz Repiso, 2012). Thus, the vision of a library linked data universe may become real, where researchers and web users will be able to easily navigate through bibliographic data using families and relationships as linking mechanisms.

2.2. Semantic Web, Linked, Data, and Library Linked Data

2.2.1. Semantic Web and Linked Data

Tim Berners-Lee is the creator of the World Wide Web. His vision was not a Web of hypertext and linked webpages, but a structured Web providing services based on meaning, which could be understood by both people and software agents (Berners-Lee, 1998). Two standards enabled his vision to be implemented, the XML standard provided structure to the then World Wide Web (Bray & Sperberg-McQueen, 1996) and the Resource Description Framework (RDF), announced in 1999 (Lassila & Swick, 1999), made possible machine-understandable assertions over the Web. Tim Berners-Lee shared his vision for the new Web, which he called "Semantic Web", two years later (Berners-Lee, Hendler, & Lassila, 2001). The Semantic Web is an extension of the current World Wide Web with the difference that it lies on the understanding of meaning by both humans and machines, enabling the collaboration between them (Berners-Lee et al., 2001).

XML and RDF are considered as cornerstone technologies for the Semantic Web architecture (Figure 2:6). XML is a mark-up language that enables the creation of semi-structured documents. The meaning to these structured documents is defined by RDF. In detail, RDF facilitates a) the identification of any piece of information as a resource using either an "International Resource Identifier - IRI" or a literal value, and b) the description of identified resources using statements in a "subject-predicate-object" form (Berners-Lee, 1998; Schreiber, Raimond, Manola, Miller, & McBride, 2014). Using this RDF triple structure (Figure 2:7), anything can be expressed in a machine-understandable format. Once concepts, ideas, people, places, monuments, can be identified with a URI, anything can be stated about them using RDF. Linked data technologies and principles (Berners-Lee, 2009) further enable the linking of these assertions into a unified universe where one can navigate and explore using the linked triples.

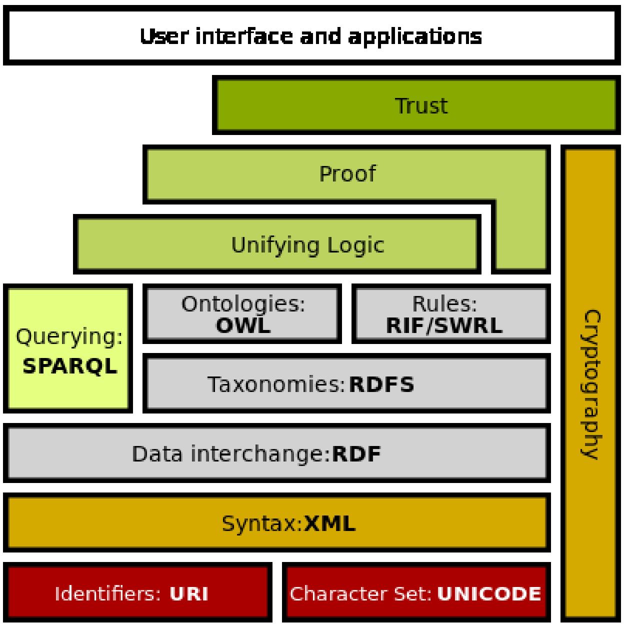


Figure 2:6. Semantic Web technologies stack. Architecture of the Semantic Web.

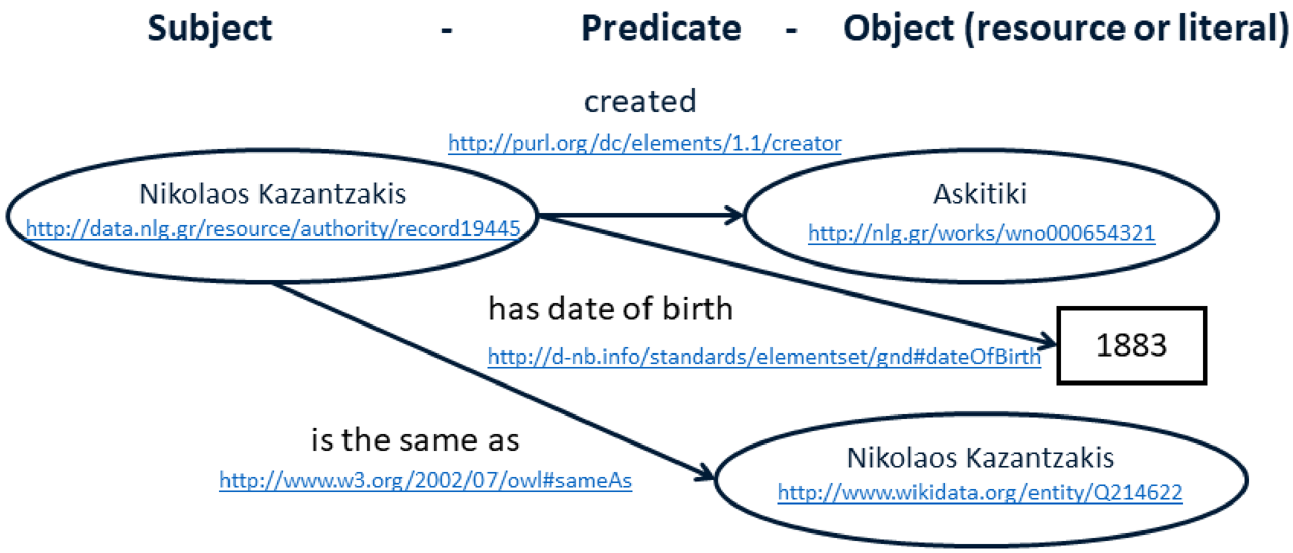


Figure 2:7. RDF triple structure (subject-predicate-object) and an example with three RDF statements providing information about i) Kazantzakis being the creator of "Askitiki"(a resource is used as the object of the triple), ii) Kazantzakis' birth date (a literal is used as the object of the triple), and iii) another resource also describing Kazantzakis. The third RDF statement asserts that the National Library of Greece resource for Nikos Kazantzakis describes the same person with the Wikidata resource. With the third RDF statement further exploration of data regarding Nikos Kazantzakis is made possible.

Linked Data involves the linking of data already on the web facilitating their exploration (Berners-Lee, 2009). According to Berners-Lee, "Linked Data is the Semantic Web done right" (Berners-Lee, 2008) and it can be achieved by following four rules (Berners-Lee, 2009):

1) Use URIs as names for things.
2) Use HTTP URIs so that people can look up those names.
3) When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL).
4) Include links to other URIs, so that they can discover more things.

These rules enable the unique identification of things and concepts regardless they are part of the digital world, e.g., a person named "Kazantzakis, Nikos" identified with the following National Library of Greece (NLG) URI http://data.nlg.gr/resource/authority/record19445 (Figure 2:7). These URIs can be processed by machines, but with the second rule humans may look up these names. As an example, the human-readable version of the "Kazantzakis, Nikos" resource is the webpage with the following HTTP URI https://data.nlg.gr/page/authority/record19445. More information regarding the "Kazantzakis, Nikos" resource can be provided with RDF statements. In Figure 2:7, the statement that "Nikos Kazantzakis was born in 1883" is expressed as an RDF statement. Nikos Kazantzakis is identified with the NLG URI, while the property describing "date of birth" is identified with a URI from the National Library of Germany. So far, the data is machine-understandable. What makes them *linked*, is the fourth rule. In the NLG dataset the resource with the URI http://data.nlg.gr/resource/authority/record19445 is defined as being the same (owl:sameAs) with another resource having the following URI https://www.wikidata.org/wiki/Q214622. When someone (either a human or a machine) visits this resource's webpage may find out more information about Nikos Kazantzakis in Wikidata. Moreover, the Wikidata webpage includes other RDF statements that may enable further exploration.

In 2010, Berners-Lee (Berners-Lee, 2009) develops a star-rating system toward linked open data (Figure 2:8) to encourage data owners to publish, distribute and reuse their data. The first star is gained once any type of data is made openly available over the Web. The next two stars are gained when data is published in structured format, proprietary or non – proprietary. The fourth star is gained when "things" in the published datasets are identified with URIs. At this point, RDF is used for describing data. The fifth star is gained when published datasets are linked to other data providing context. It must be noted that for linked open data, open licenses are a prerequisite.

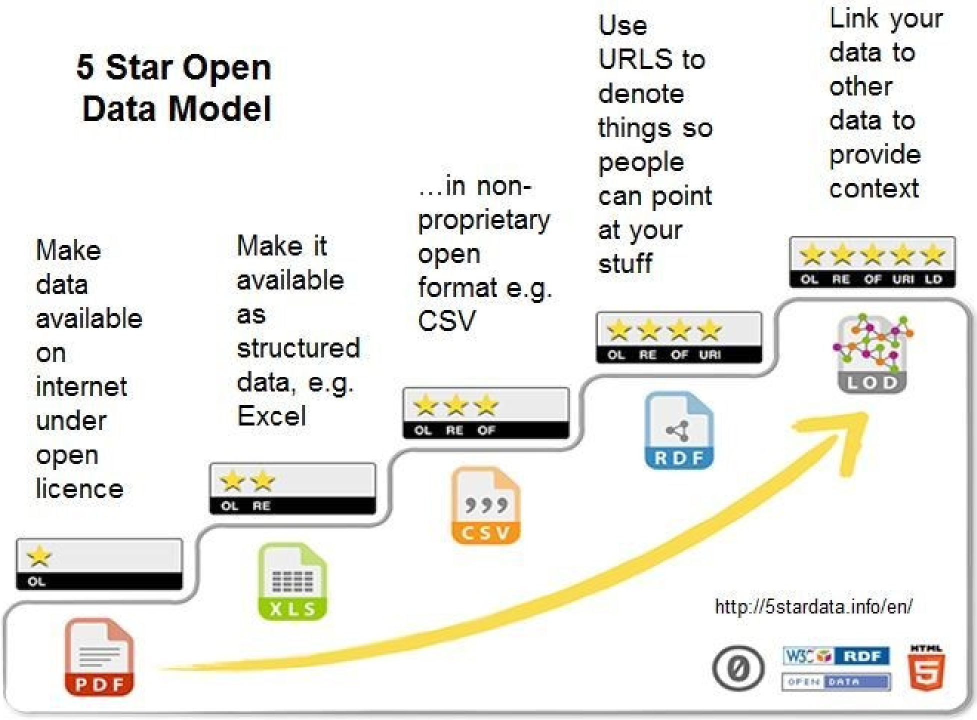


Figure 2:8. 5-star open data model. Linked Data licensed with an Open license are defined as the goal for creating the "network effect" from which both linked data publishers and consumers benefit.

2.2.2. Library Linked Data initiatives

The objective of integrating library data (often considered as of high-quality) into the web needs the development of several tools, from the most technical to the most conceptual ones. Libraries have always been places of innovation, and nowadays there are many libraries experimenting with linked open data technologies to further find out how the transition from closed systems with legacy data to a linked open data environment may happen. The most important initiative identifying the library linked data area of research and implementation has been the formation of the W3C Library Linked Data Incubator Group ("W3C Library Linked Data Incubator Group," 2012) and the publication of its reports (Baker et al., 2011; Isaac, Waites, Young, & Zeng, 2011; Suero, 2011) in 2011. The final W3C LLD Incubator Group report defined three types of library data, i.e., element sets, value vocabularies, and datasets. Linked data was defined as "data published in accordance with principles designed to facilitate linkages among datasets, element sets, and value vocabularies" (Baker et al., 2011).

Figure 2:9 presents a timeline with the most important initiatives in the web and libraries domains. Above the timeline, the linked data initiatives undertaken in the web domain section are depicted. These initiatives have been utilized in the library domain. The library domain initiatives are depicted below the timeline using different symbols, one per "type of library data", as defined in the W3C LLD Group report (Baker et al., 2011), i.e., models and metadata element sets, controlled vocabularies, and datasets. They are presented more analytically in the following paragraphs.

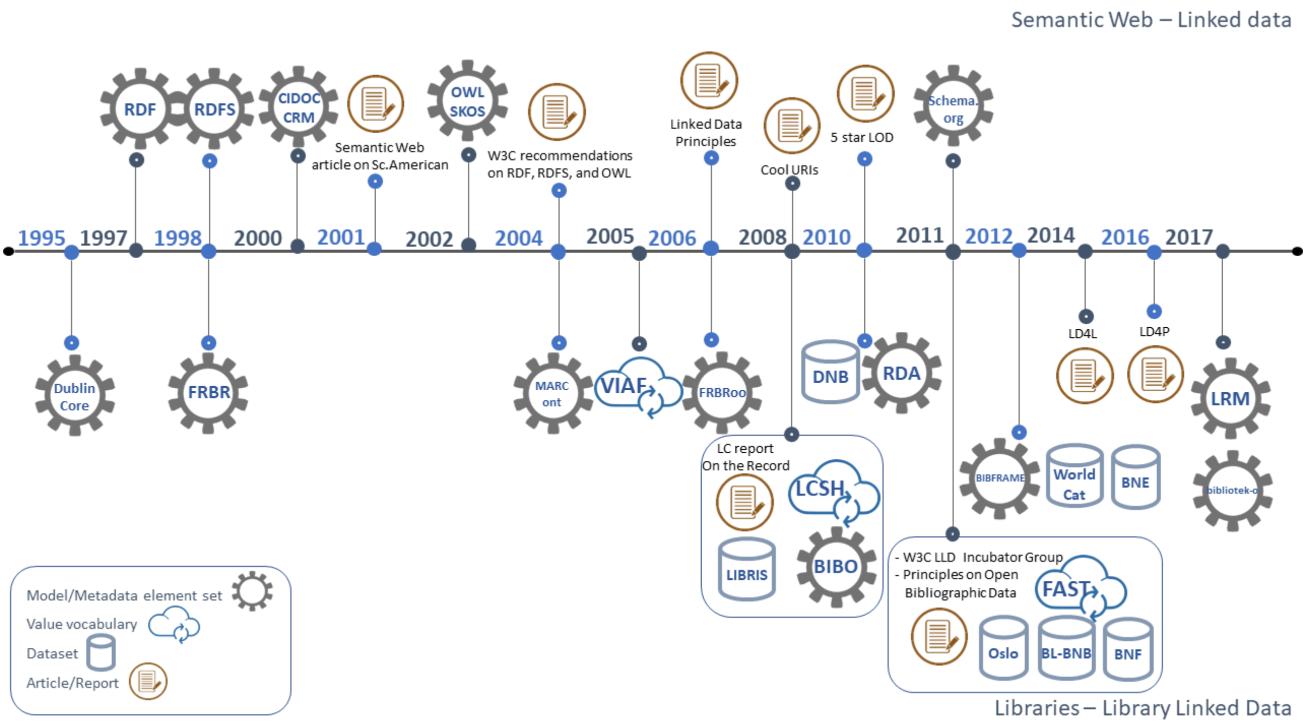


Figure 2:9. Timeline of Linked Data and Library Linked Data Initiatives. Inspired and using data from (Godby et al., 2015; Suominen & Hyvönen, 2017).

2.2.3. W3C Library Linked Data Incubator Group

The formation of the W3C Library Linked Data Incubator Group ("W3C Library Linked Data Incubator Group," 2012) and the publication of its three reports (one main and two complementing it) are considered as an influential initiative advocating the use of linked data technologies within the library community. This group was formed with the aim to contribute to the interoperability of legacy library data on the Web. Most importantly, its main report provided definitions to avoid disambiguation, described the aspirations of using linked data technologies in the library domain in terms of benefits for users, libraries, staff, developers, and vendors. The current situation was described, and the rights issues were lightly touched upon. The report concluded with different recommendations for interested parties, i.e., library leaders, standard bodies, developers, and staff. The key recommendations encouraged all these four parties to become early implementers and advocates for library linked data by 1) identifying candidate datasets "for early exposure", 2) participating in semantic web standardization efforts, developing models that are linked data compliant, and "disseminating best-practice design patterns" for Library Linked Data, 3) adhering to linked data principles in library systems and using URIs, and 4) preserving elements sets and value vocabularies (Baker et al., 2011). The use cases that were discussed within the Group Meetings were published in a separate additional report (Suero, 2011). Another complement report focusing on datasets, value vocabularies and metadata elements sets presented the then current efforts from the library community (Isaac et al., 2011).

The W3C Library Linked Data Incubator Group reports are relevant not just to libraries, but other cultural heritage and memory institutions too. The W3C LLD Incubator Group used the term "library" to encompass all these types of institutions, seen in this report as collections' curators, regardless of what these collections include. Patron and user data were out of scope in the report and the term "Library Data" referred to all types of data related to resources. Three types of "Library Data" were identified: element sets, value vocabularies, and datasets. Element sets are used to describe entities of interest, value vocabularies provide values for particular elements, and datasets are collections of descriptions regarding interesting entities of the bibliographic universe such as authors, books, subjects (Isaac et al., 2011). When any of these three types of data is expressed with linked data technologies, then it is considered “Library Linked Data”.

Besides the definitions and recommendations, the final report acknowledges that linked data may ensure the technical interoperability of published library data. Moreover, it raises some semantic interoperability concerns regarding all three types of data (element sets, value vocabularies, and datasets). The report highlights the need for alignments between value vocabularies, and between element sets. The report also raises the issue for persistent URIs and for authoritative element sets, as well as for better linking tools that do not depend on string matching only.

Following the LLD report’s types of data, the next paragraphs present some of the current library linked data efforts organized in three categories: value vocabularies, models, and datasets. Value vocabularies and name authority files have been among the first library linked data initiatives. Further, they serve as an infrastructure providing values for the description of the entities that each conceptual model identifies. The W3C Library Linked Data Incubator Group reports does not clearly differentiate between conceptual models and metadata element sets. The former establish the “entities of interest” and the relationships between them (Baca, 2016), while the latter provide the elements to describe the “entities of interest”. Yet, nearly all models are accompanied by their own vocabulary that may include elements developed solely for the model’s purposes or may combine elements from different element sets. Thus, the thesis focuses on the bibliographic conceptual models and proceeds with the presentation of well-known models. The models are presented with regard to the entities they acknowledge and to the linking mechanisms they provide in terms of bibliographic relationships (see paragraph 2.3 Bibliographic Conceptual Models). After, library datasets using models and value vocabularies follow to present the decisions taken so far by the implementing libraries (see paragraph 2.2.5 Library linked datasets).

2.2.4. Value vocabularies

Controlled/Value vocabularies define resources, e.g., instances of languages, countries, contributors, that may be later used as element values (Baker et al., 2011). The expression of legacy controlled vocabularies in RDF has been one of the first linked-data related initiatives in the library domain. The Library of Congress has started since 2009 the publication of “LC owned or maintained authorities and vocabularies” as RDF, permitting to be interlinked in a linked data environment (K. Ford, 2010). In 2011, after almost two years, the LC Linked data service (http://id.loc.gov/) had published six controlled vocabularies, the Library of Congress Subject Headings (LCSH), the Thesaurus of Geographic Materials, the MARC Code List for Relators and other three smaller preservation vocabularies (K. Ford, 2010). Nowadays, LC publishes nearly 100 controlled vocabularies through its id.loc.gov linked data portal.

Another important stakeholder has been the OCLC publishing the Virtual International Authority File (VIAF) among other vocabularies. VIAF is a collaborative project with OCLC aggregating name authority data from national libraries and converting them to linked data (OCLC, 2018). OCLC has also published the Faceted Application of Subject Terminology (FAST) and the Dewey Decimal Classification as linked data.

Other controlled vocabularies have been published within the framework of larger library linked data projects. Examples include the RAMEAU subject headings by the National Library of France (data.bnf.fr), LIBRIS vocabularies by the National Library of Sweden, name authority data (ULAN and TGN) by the Getty Institute (vocab.getty.edu), and the GND integrated authority file by the National Library of Germany (d-nb.info/datasets/authorities).

Even though some of the presented value vocabularies have been created in the context of a larger linked data project, value vocabularies may be developed in an autonomous way and may be used and re-used in a variety of linked data projects.

2.3. Bibliographic Conceptual Models

Bibliographic models provide abstract representations of bibliographic data by defining entities, their attributes, and their relationships between the defined entities. This paragraph presents bibliographic conceptual models in terms of aims, core entities, and the relationships between them.

2.3.1. FR family of models

The term “FR family of models” refers to three reports: the original Functional Requirements for Bibliographic Records - FRBR report presenting the general model and Group 1 entities, the Functional Requirements for Authority Data (FRAD) report (IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR), 2009, 2013) for the analysis of the Group 2 entities, and the Functional Requirements for Subject Authority Data (FRSAD) report (IFLA Working Group on the Functional Requirements for Subject Authority Records, 2010) for the analysis of the Group 3 entities. The FRBR model has been adopted by the Resource Description and Access content standard (see paragraph 2.3.3). The consolidated version of the FRBR model, known as Library Reference Model (LRM) was endorsed by the IFLA in August 2017 (see paragraph 2.3.4).

Gordon Dunsire expressed the FRBR family of models in RDF and made these RDF vocabularies available through the Open Metadata Registry (http://metadataregistry.org/).

2.3.1.1. FRBR model - Functional Requirements for Bibliographic Records and Group 1 entities

The Functional Requirements for Bibliographic Records - FRBR model has been published in 1998 (IFLA Study Group on the Functional Requirements for Bibliographic Records, 1998) and has been consolidated twice, resulting in the FRBR family of models in 2008 – 2010 (IFLA Study Group on the Functional Requirements for Bibliographic Records, 2009; IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR), 2009; IFLA Working Group on the Functional Requirements for Subject Authority Records, 2010) and the IFLA-Library Reference Model (IFLA LRM) in 2017 (Riva, Bœuf, & Žumer, 2017a).

FRBR is an entity-relationship model defining the bibliographic entities of the bibliographic universe, their attributes and the relationships between them. Three groups of key bibliographic entities have been identified. Group 1 includes the entities referring to intellectual or artistic products. Group 2 entities refer to agents that handle Group 1 entities in the framework of creation, publication, or custodial processes. Group 3 entities are used as topical terms of the Group 1 Work entity (see Figure 2:10).

The Group 1 entities are *Work, Expression, Manifestation, Item* (commonly referred to as WEMI) (IFLA Study Group on the Functional Requirements for Bibliographic Records, 2009). In their definition the differentiation between content and carrier was formally declared. The *Work* entity is the most abstract one that -along with the *Expression* entity- reflects the content. *Manifestation* and *Item* entities are more concrete accommodating the physical form of the content. In detail, the *Work* entity refers to the intellectual or artistic creation; it is the most abstract entity that may represent the set of ideas in the creator’s mind. The *Expression* entity refers to the realization of the *Work* entity. This realization is expressed as a set of signs representing the ideas, e.g., a text containing specific words in a specific sequence. The *Manifestation* entity is the one that embodies one or more *Expressions* of the same or of more *Works.* The *Manifestation* is the cataloging unit described in library catalog records. Exemplars of the *Manifestation* entity are described with the *Item* entity.

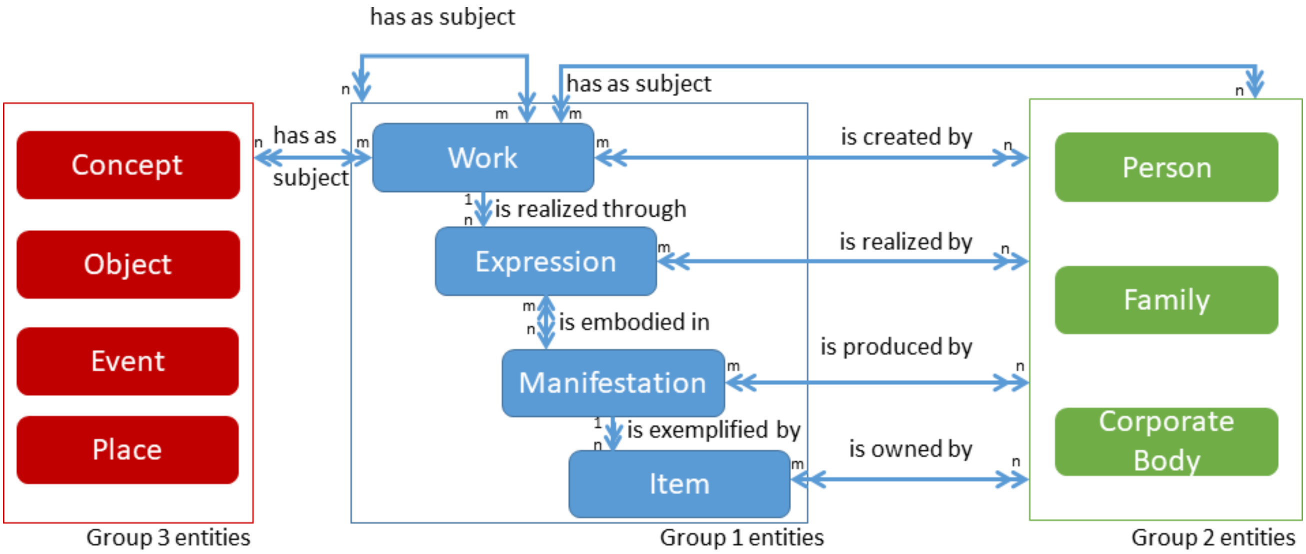


Figure 2:10. The FRBR model: Group 1,2,3 entities and inherent relationships.

2.3.1.2. FRAD and FRSAD reports – Group 2 and Group 3 entities

Group 2 and Group 3 entities were further analyzed in subsequent reports: the Functional Requirements for Authority Data (FRAD) report (IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR), 2009, 2013) for the analysis of the Group 2 entities, and the Functional Requirements for Subject Authority Data (FRSAD) report (IFLA Working Group on the Functional Requirements for Subject Authority Records, 2010) for the analysis of the Group 3 entities. Group 2 entities are *Person*, *Corporate Body*, and *Family*. Group 3 entities are *Concept, Object, Event, and Place.* Appellations for all Group 1, 2, and 3 entities could be represented with the *Name* entity. FRSAD added two more entities, *Thema* as a "super-entity" for all Group 1,2,3 entities, and *Nomen* to provide all different appellations of a *Thema* (see Figure 2:11). Each one of the entities of Group 1, 2, 3 is defined by a set of attributes. The number of attributes differs. As an example, the *Work* entity has 7 attributes (plus 5 for *Musical Works* and for *Cartographic Works*), while *Expression* has 12 attributes (plus 13 special attributes for serials, musical notations, images and objects), and Group 2 Person entity has 14 attributes.

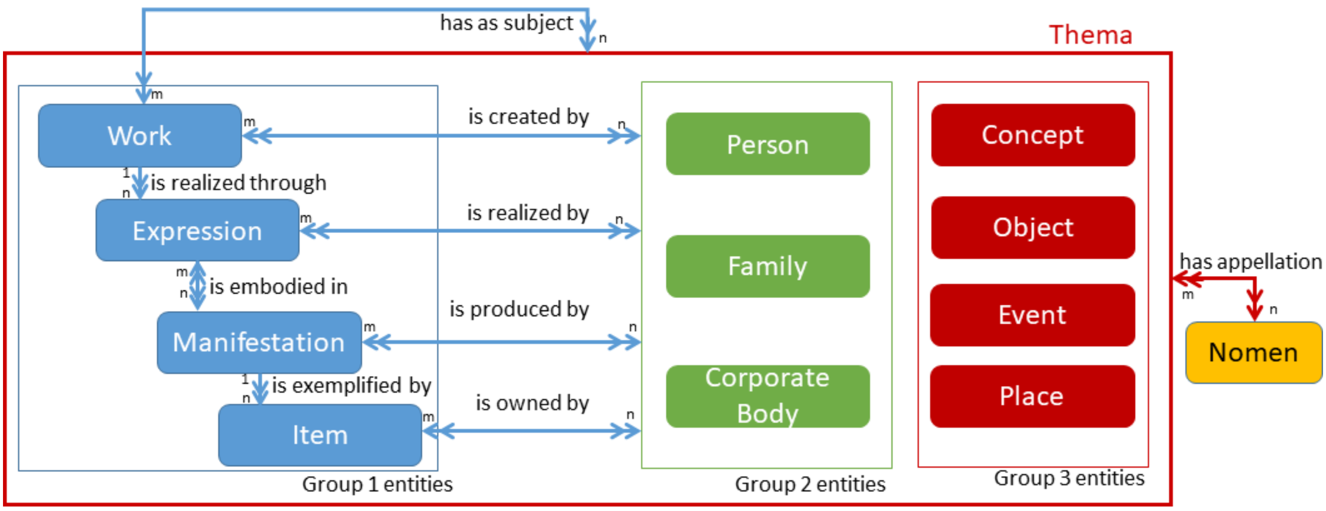


Figure 2:11. The FRBR model after the publication of FRSAD in 2010.

2.3.1.3. Relationships in the FRBR

The entities of the model are inter-linked by a set of relationships. There are two types of relationships: the inherent relationships and the bibliographic relationships (Tillett, 2004).

Inherent relationships are the ones provided purposely by the model. They exist between the Group 1 entities, and between Group 1 and Group 2 and Group 3 entities. A *Work* may *be realized through* one or more *Expression* instances*,* an *Expression* may *be embodied in* one or more *Manifestations* and one *Manifestation* may *embody* one or more *Expressions.* A *Manifestation* may *be exemplified by* one or more *Items*. These are the inherent relationships in Group1. Moreover, Group 1 entities are correlated with Group 2 entities through different relationships. The *Work* entity may be *created* by the Group 2 entities, while the *Expression* entity may be *realized* by them. The *Manifestation* entity may be *produced* by Group 2 entities and the *Item* entity may be *owned* by them. The *Work* entity may *have as a subject* a *Thema* whose *type* may be any of the Group 1, 2, or 3 entities. All entities and the inherent relationships among them are depicted in Figure 2:11.

Bibliographic relationships may exist between Group 1 entities within the scope of providing users with more information regarding the described entity, as well as with links to related *Works, Expressions, Manifestations, and Items.* In (IFLA Study Group on the Functional Requirements for Bibliographic Records, 2009), the relationships are organized according to the entities on which they may be applied: *Work-to-Work, Expression-to-Expression, Expression-to-Work, Manifestation-to-Manifestation, Manifestation-to-Item, Item-to-Item* (Figure 2:12). The FRBR report's goal was not to provide "higher level groupings for relationships, but … to show how the relationships operate in the context of the four primary entities in the model" (IFLA Study Group on the Functional Requirements for Bibliographic Records, 2009).

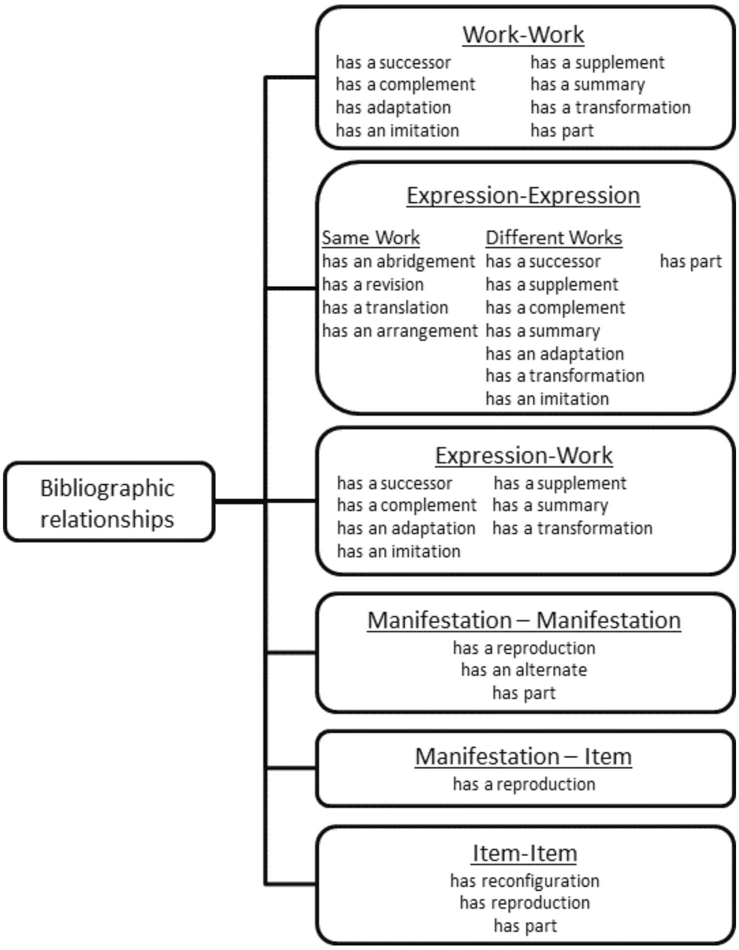


Figure 2:12. Hierarchy of relationships between FRBR Group 1 entities (IFLA Study Group on the Functional Requirements for Bibliographic Records, 2009).

The categorization of relationships changed in the subsequent Functional Requirements for Authority Data (FRAD) report (IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR), 2013). The FRAD report adopted Tillett's taxonomy of seven relationships (Tillett, 1987, 1991), and attempted for the first time to define on which exact WEMI entity each bibliographic relationship may be applied. The exact relationships were not provided in FRAD; examples for each type of relationship were given instead.

To compare the models in this paragraph with regard to the representation of bibliographic relationships, the relationships defined by the FRBR are once again presented in the following Figure 2:13 using the Tillett taxonomy of bibliographic relationships.

Tillett's taxonomy	FRBR entity where applied	Relationships
Equivalence	Manifestation Item	has a reproduction, has an alternate has reconfiguration, has reproduction
Derivative	Work Expression	W: has a summary, has adaptation, has a transformation, has an imitation E same Work: has an abridgement, has a revision, has a translation, has an arrangement E different Works: has a summary, has an adaptation, has a transformation, has an imitation
Descriptive	Work	has as subject
Whole-part	Work, Expression, Manifestation, Item	has part
Accompanying	Work, Expression	has a supplement, has a complement
Sequential	Work, Expression	has a successor

Figure 2:13. FRBR bibliographic relationships organized according to the Tillett taxonomy of bibliographic relationships.

2.3.1.4. Refinements of the FRBR and consolidation

The models of the FR family were developed within a decade by different working groups. There were different teams working on Group 2 and Group 3 entities, namely the IFLA Working on Functional Requirements and Numbering of Authority Records FRANAR, and the IFLA Working Group on the Functional Requirements for Subject Authority Records. Other working groups worked on specialized issues, i.e., the clarification of the *Expression* entity definition, and the modeling of aggregates within the FRBR framework. The first issue was studied by the Working Group on the Expression Entity. The group redefined the *Expression* entity (IFLA Working Group on the Expression Entity, 2007); the proposed changes were endorsed and incorporated in the second edition of the FRBR (IFLA Study Group on the Functional Requirements for Bibliographic Records, 2009). The second issue involved the handling of aggregates. The FRBR model treated aggregates as 'integral units' having two or more components (IFLA Study Group on the Functional Requirements for Bibliographic Records, 2009). It also defined that the components could operate similarly to FRBR WEMI entities at the integral unit level (IFLA Study Group on the Functional Requirements for Bibliographic Records, 2009). The FRBR model allows recursive relationships making the representation of aggregates complex (Hickey & O'Neill, 2009) and sometimes misleading (Peponakis, 2012). The FRBR Working Group on Aggregates considered the FRBR model approach, and after a long period of deliberations concluded differentiating between aggregates and components and their treatment (O'Neill et al., 2011). Aggregates consist of nonintegral parts which may stand on their own (Neill et al., 2015). In contrast, components are identifiable parts of a whole and can be represented with the whole/part relationship. According to the (O'Neill et al., 2011), aggregates are *Manifestations* embodying more than one *Expressions.* Three types of aggregates are defined: collections, augmentations, and parallels (Neill et al., 2015; O'Neill et al., 2011). In case, the aggregator's effort regarding the aggregation is significant and needs to be described, such information is given through the *Aggregating*

Work and its *Aggregating Expression* which is also embodied with the other aggregated *Expressions* in the *Aggregate Manifestation* (see *Figure 2:14*).

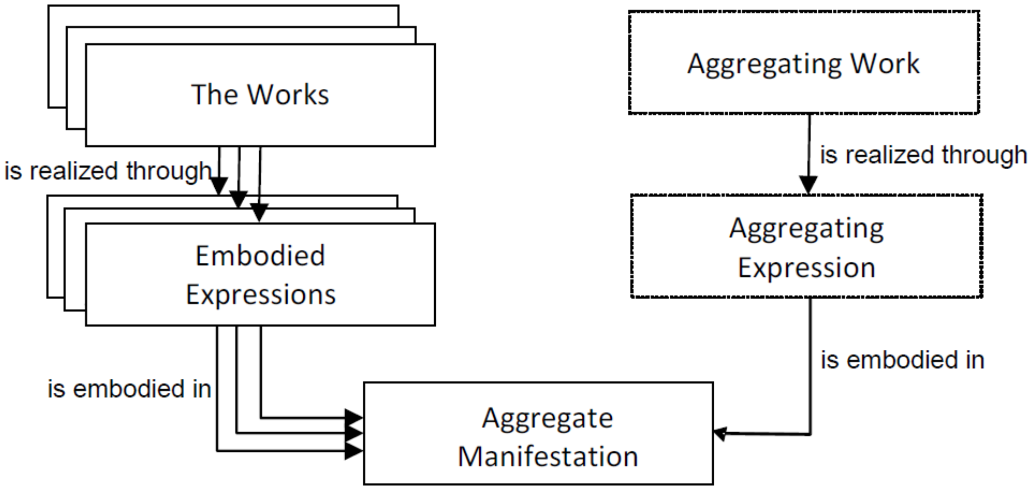


Figure 2:14. The general model of aggregates. Source: (O'Neill et al., 2011).

Despite the fact that the FRBR, FRAD and FRSAD reports conformed to the entity-relationship formalism, there were "different points of view and differing solutions for common issues" (Riva, Bœuf, et al., 2017a). The consolidation of the models and of the two amendments (clarification of the *Expression* entity and Aggregates report) in one coherent model resulted in the IFLA Library Reference Model (LRM) which was endorsed by IFLA in 2017 (Riva, Bœuf, et al., 2017a). The LRM model is presented in paragraph 2.3.4 Library Reference Model - LRM.

2.3.2. Functional Requirements for Bibliographic Records object-oriented version - FRBRoo

In the 1990s the library and museum communities developed almost simultaneously conceptual models for their data, namely the Functional Requirements for Bibliographic Records (FRBR) model and the CIDOC Conceptual Reference Model (CRM). In 2003 the International Working Group on FRBR and CIDOC CRM Harmonisation was formed with representatives from IFLA and ICOM-CIDOC (International Council for Museums – International Committee on Documentation). This Working Group worked for the harmonization of the two models to enable the semantic interoperability between the two communities' data (ICOM-CIDOC, 2019). The new model, named FRBRoo, extended the CIDOC-CRM model with FRBR concepts. The first version was published in 2009 and was based on the FRBR model. After the publication of the FRAD and FRSAD reports, the second version of the FRBRoo was based on all three models of the FR family. The current version is 2.4, and it has been endorsed by IFLA in 2016 (Working Group on FRBR/CRM Dialogue, Bekiari, Doerr, Le Boeuf, & Riva, 2016). The next version of FRBRoo is currently under development and it will be based on the consolidated LRM model. Therefore, its name will change to LRMoo (Riva & Žumer, 2018).

Even though, there have been some FRBRoo implementations, this model's aim is not to serve as a bibliographic data format, but rather as a tool in reusing legacy data by means of extracting the meaning of them and reusing them with other types of data in library and non-library environments (Le Boeuf, 2013).

2.3.2.1. FRBRoo entities

FRBRoo refines Group 1 entities, introduces temporal entities, events, and time processes. With the introduction of temporal entities, FRBRoo enables the representation of the processes that create intellectual or artistic products. Therefore, in FRBRoo intellectual or artistic products may be either described as static objects following the FRBR model (static representation), or represented in dynamic representations that also include the processes “through which ... [they] come into being” (Working Group on FRBR/CRM Dialogue et al., 2016). As an example, the full history of a work may be represented using events and other temporal entities. If such information is known, all processes starting from the initial conception of ideas and the realizations of these ideas using signs (Expression Creation event), to the production of the book product may be represented in FRBRoo. Such a decision lies in the aims of the cataloging institution.

FRBR entities are expressed as classes denoted by the combination of a name and an identifier starting with the letter F, e.g., *F1 Work*. FRBR attributes and relationships are expressed as properties. The model uses the CIDOC CRM properties, declared by the combination of a name and an identifier starting with the letter P, e.g., *P102 has title*, and defines new ones, declared by the combination of a name and an identifier starting with the letter R, e.g., *R3 is realised in (realises)*. For each property the domain and range classes, as well as cardinality constraints are given. FRBRoo introduces 54 new classes and 74 new properties to CIDOC-CRM classes and properties respectively.

Regarding core entities, FRBRoo analyzes Group 1 entities incorporating more interpretations of the FRBR *Work – Expression – Manifestation – Item* (WEMI) entities (see Figure 2:15). The *F1 Work* class is specialized to 4 *Work* subclasses: *F14 Individual Work, F15 Complex Work, F16 Container Work,* and *F21 Recording Work.* The *F14 Individual Work* class corresponds to a *Work* associated with only one complete set of signs, the *F15 Complex Work* class is closer to the FRBR’s interpretation of the *Work* entity. The *F16 Container Work* class may represent aggregates, while its *F19 Publication Work* subclass includes author’s original *work* plus the publisher’s contributions added during the publication process. The *F21 Recording Work* class corresponds to a work that captures an event (Le Boeuf, 2013) by recording sounds and/or images. Some of the *F1 Work* class’ subclasses have their own subclasses representing even more specialized interpretations of the work concept.

The *F2 Expression* has two subclasses: *F22 Self-Contained Expression*, and *F23 Expression Fragment*. The *F22 Self-Contained Expression* class corresponds to a set of signs that carries all the ideas of work it realizes, while the *F23 Expression Fragment* carries a fragment of the complete set of signs. Similarly to the *F19 Publication Work* that includes publisher’s intellectual contribution, the *F24 Publication Expression* class, subclass of the *F22 Self-Contained Expression* class, conveys all the signs corresponding to a specific publication, i.e., author’s signs, publisher’s signs, book cover designer’s signs, and other contributors’ signs (Le Boeuf, 2013).

There are two classes for *Manifestations:* the *F3 Manifestation Product Type* class representing a publication product, and the *F4 Manifestation Singleton* class representing a manifestation produced as a unique object, e.g., the manuscript submitted by an author to a publisher. There is only one class for *Item,* the *F5 Item* class.

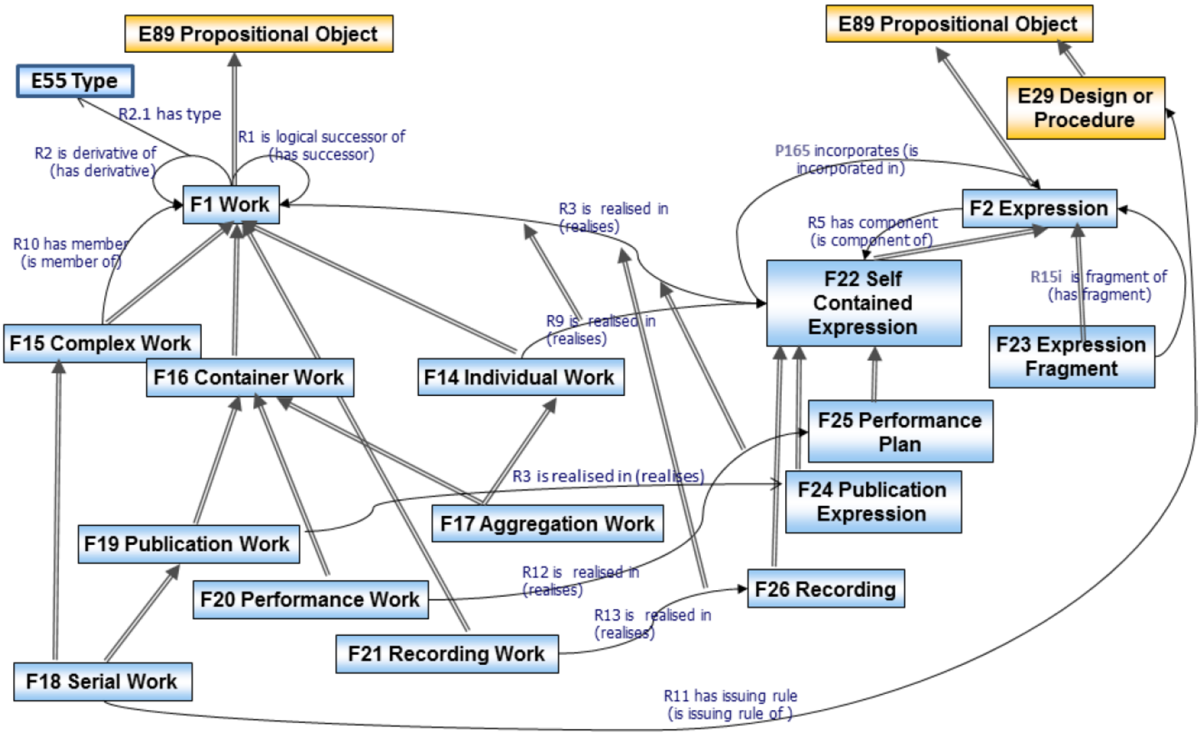


Figure 2:15. The static view of the Work and Expression classes in FRBRoo. The figure also presents the inherent relationships applied to the two classes. Source: (Working Group on FRBR/CRM Dialogue et al., 2016).

2.3.2.2. Bibliographic relationships

Regarding inherent and bibliographic relationships, FRBRoo represents them by using existing CIDOC-CRM properties or by extending these properties with subproperties to express relationships described in the FR family of models. It must be noted that some relationship categories are represented with one property; the exact type of relationship is represented as a *type* (Figure 2:16).

As an example, all derivative relationships are represented with the *R2 is derivative of / R2i has derivative* property. The exact type of derivation is represented by declaring that the *R2 is derivative of / R2i has derivative* property is of specific *type*, such as Abridgement, Adaptation, Arrangement, Imitation, Revision, Summary, Transformation, and Translation. This type is represented as a value of a vocabulary applied to the instances of the class *E55 Type*. Therefore, an *R2 is derivative of / R2i has derivative* property instance representing a revision can be represented by assigning the corresponding type with the following triple: *R2 is derivative of - has type - E55 Type = 'Revision'*. It must be noted that FRBRoo does not provide controlled vocabularies for the value of the *E55 Type* property permitting the implementing libraries to develop and use local vocabularies.

Tillett's taxonomy	FRBRoo class where applied	Properties
Equivalence	F3 Manifestation Product Type F4 Manifestation Singleton & F5 Item	P130 shows features of {P130.1 kind of similarity: E55 Type = "Reproduction"} P130 shows features of {P130.1 kind of similarity: E55 Type = "[type of alternate format]"}
Derivative	F1 Work, F2 Expression	R2 is derivative of {R2.1 has type E55 Type = "[type of derivation]"} [Examples of types of derivation]: Abridgement, Adaptation, Arrangement, Imitation, Revision, Summary, Transformation, Translation
Descriptive	F1 Work	P129 is about
Whole-part	F1 Work F2 Expression F3 Manifestation Product Type F4 Manifestation Singleton / F5 Item	R10 has member R15 has fragment & R5 has component CLP46 should be composed of P46 is composed of
Accompanying	F1 Work, F2 Expression	P16 used specific object {P16.1 mode of use E55 Type = " * "} * supplemented work, supplemented expression
Sequential	F1 Work	R1i has successor

Figure 2:16. FRBRoo properties for the representation of bibliographic relationships. Properties are organized according to the Tillett taxonomy of bibliographic relationships.

2.3.3. Resource Description & Access - RDA

The Resource Description & Access (RDA) is a cataloging standard developed between 2005-2009 by the Joint Steering Committee (JSC) in collaboration with groups from the library, the archival, and the publishing communities. First published in 2010 as a standard "designed for the digital environment" (Dunsire, 2007), the RDA is based on the FRBR models (FRBR, FRAD, FRSAD) within the scope of supporting "comprehensive coverage of all types of content and media, the flexibility and extensibility needed to accommodate newly emerging resource characteristics, and the adaptability needed for the data produced to function within a wide range of technological environments" (Joint Steering Committee for Development of RDA, 2009). The RDA takes also under consideration sound cataloging traditions and principles (Cutter, 1904; Lubetzky, 1969; Panizzi, 1841; *Statement of Principles adopted by The International Conference on Cataloguing Principles Paris, October 1961*, 1961), as well as the "Anglo-American Cataloguing Rules" (AACR) (Gorman & Oddy, 1998). RDA is also aligned with the RDA/ONIX Framework (Joint Steering Committee for Revision of AACR, 2006), the ISBDs (International Federation of Library Associations and Institutions. ISBD Review Group. & International Federation of Library Associations and Institutions. Cataloguing Section. Standing Committee., 2011; ISBD Review Group & Galeffi, 2015), the MARC 21 Formats (Authority and Bibliographic data) (Library of Congress, n.d.-b), and the Dublin Core (Dublin Core Metadata Initiative, 2012). It can be accessed online via the RDA Toolkit website (http://access.rdatoolkit.org/).

RDA may be considered as a model of its own (Taniguchi, 2013); it adheres to FRBR/FRAD models, but there are differences between the RDA and the FRBR/FRAD models (Peponakis, 2012; Riva & Oliver, 2012; Taniguchi, 2012), which are not referred as differences, nor are they explained in the RDA documentation. One core difference between the FRBR and the RDA is the addition of a new inherent relationship in the latter, namely the *manifestation of work* relationship relating a *Work* with a *Manifestation*.

2.3.3.1. RDA entities

The RDA defines the same entities, as the FRBR, namely, Group 1 entities (*Work, Expression, Manifestation*, and *Item*), Group 2 entities (*Person, Family, Corporate Body*), and the *Place* Group 3 entity (Figure 2:17). The other Group 3 entities (*Concept*, *Object*, and *Event*) are not yet developed in RDA. RDA definitions and guidelines are recorded in ten sections, four of which are dedicated to the recording of entities' attributes, and six sections are about recording relationships between RDA entities. The RDA entities, attributes and relationships define a vocabulary that has been implemented by Resource Description Framework (RDF). In detail, entities are defined as classes, attributes and relationships are represented in element sets as properties, and terms used with specific attributes are identified as property values in value vocabularies. Every element set, and value vocabulary has its own namespace and its recommended prefix ("RDA Registry | Data using the RDA vocabularies," 2017). There are 12 element sets; Classes, *Work* properties, *Expression* properties, *Manifestation* properties, *Item* properties, *Agent* properties, *Place* properties, Time-span properties, *Nomen* properties, RDA Entity properties, Unconstrained properties, and RDA/ONIX Framework elements. There are 41 value vocabularies, plus 14 RDA/ONIX Framework value vocabularies.

A *Work* may have one or more *Expressions.* An *Expression* may be manifested in one or more *Manifestations* and a *Manifestation* may manifest one or more *Expressions.* A *Manifestation* may have as exemplars one or more *Items.* A *Work* also may have one or more *Manifestations* and a *Manifestation* may embody one or more *Works.* All Group 1 entities are related to Group 2 entities with different relationships/properties. A *Work* may have as subject any other Group 1 entity (*Work, Expression, Manifestation, Item*) or Group 2 entities (*Person, Family, Corporate Body*). Group 3 entities have been defined as placeholders for future releases of the RDA [1].

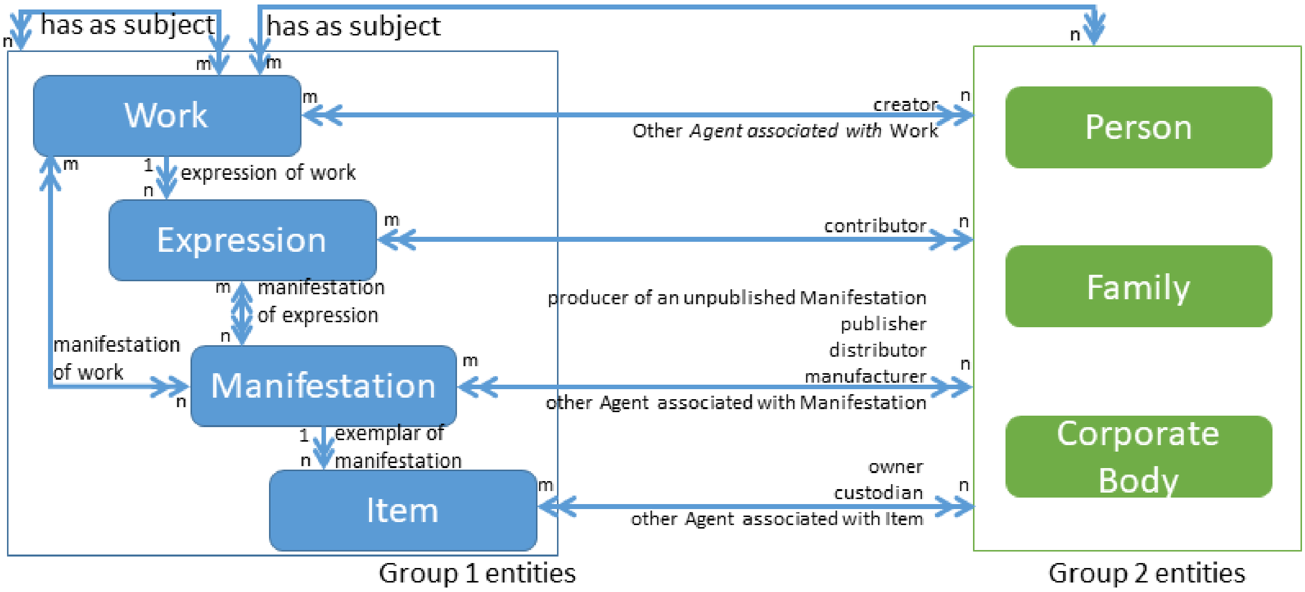


Figure 2:17. RDA entities and inherent relationships. The subject related entities are not fully developed in the RDA yet.

More than half of the RDA definitions and guidelines focus on the recording of relationships between entities/classes. Primary relationships between the Group 1 entities are described in RDA's Section 5 and a new one is added: the *rdaw:P10072 has manifestation of work* relationship between a *Work* and a *Manifestation* embodying it may be represented without representing the *Expression* that realizes the *Work* and actually being embodied in the *Manifestation.*

2.3.3.2. Bibliographic relationships

The recording of bibliographic relationships between Group 1 entities are described in Section 8 by RDA entity: Related Works, Related Expressions, Related Manifestations, and Related Items. Relationships between Group 1 entities and Agents are described in Section 6, and relationships between *Agents* are described in Section 9. Sections 7 and 10 are placeholder sections for subject relationships. Both of them will be developed in the future. The RDA manages to adopt both FRBR (IFLA Study Group on the Functional Requirements for Bibliographic Records, 2009) and FRAD (IFLA Working Group on Functional Requirements and Numbering of

[1] The last check on the RDA Toolkit (https://access.rdatoolkit.org/) was made on September 6, 2020. This check confirmed that the Guidelines on "Recording Attributes of Concepts, Objects, Events, and Places" have not been developed yet.

Authority Records (FRANAR), 2013) approaches, as well as Tillett's taxonomy of relationships. Guidelines are organized according to the entities involved (FRBR approach), while the exact relationships are presented in the appendix J. This "appendix is also first organized by entity, then by the broad FRAD categories which can apply to that entity, then by relationship types and sub-types" (Riva & Oliver, 2012). The hierarchy of relationships in RDA is depicted in Figure 2:18. It must be highlighted that RDA extends the FRAD-Tillett broad relationship categories providing many specialized relationships. In each one of these relationships-related sections, there is a high-level relationship designator for each type of relationship which is later narrowed with many specialized relationship designators represented as subproperties. As an example, Figure 2:19 presents the "Derivative work relationships". The derivative relationship between *Works* is represented with two high-level properties, the *rdaw:P10190 is based on work* and its inverse *rdaw:P10148 has derivative work* properties. The *rdaw:P10190 is based on work* is specialized into 17 relationships represented as subproperties, one of which is the *rdaw:P10142 is adaptation of work*. This subproperty is specialized even more with 14 subproperties.

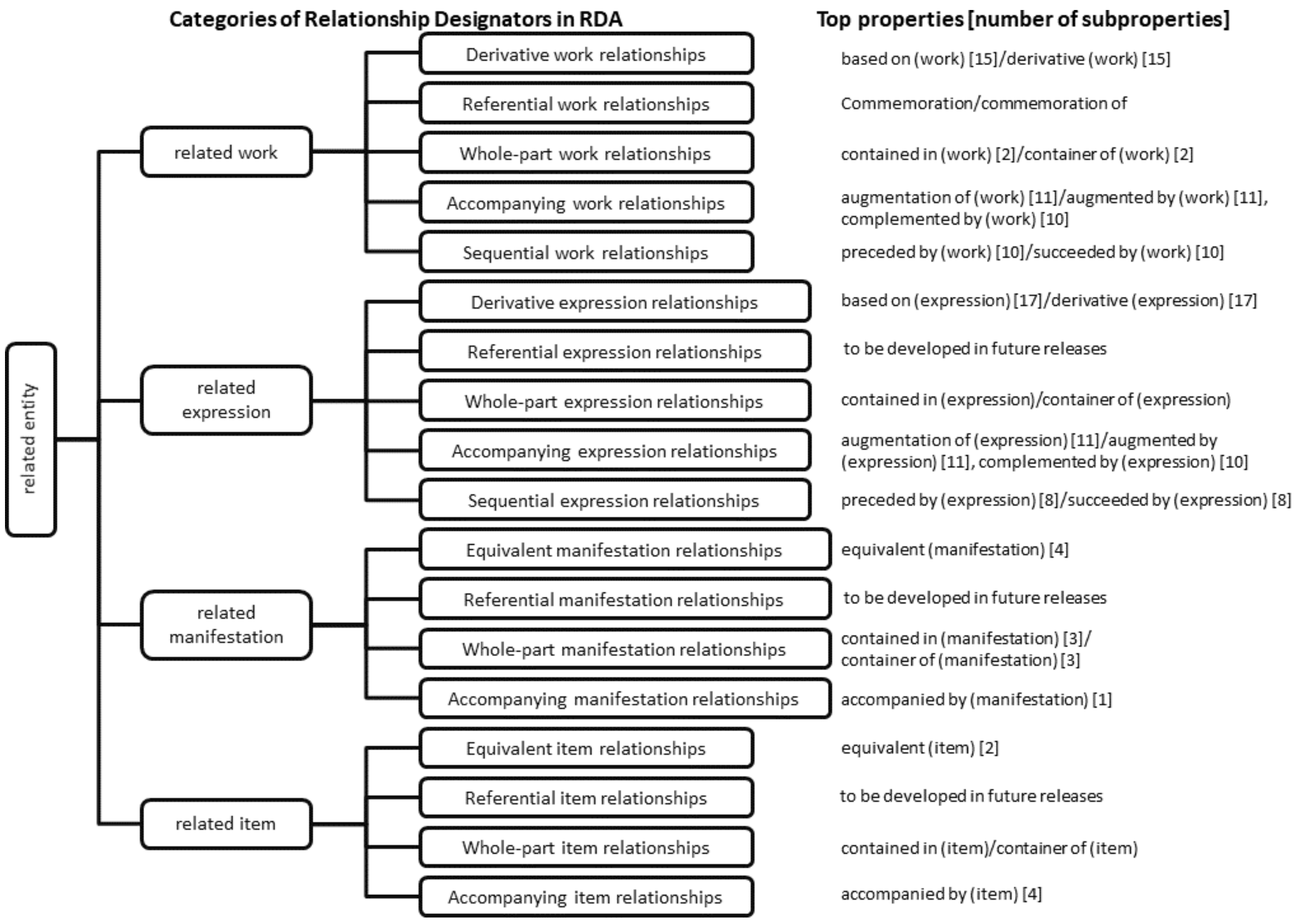


Figure 2:18. The hierarchy of bibliographic relationships in RDA. Top properties are listed along with the number of their subproperties.

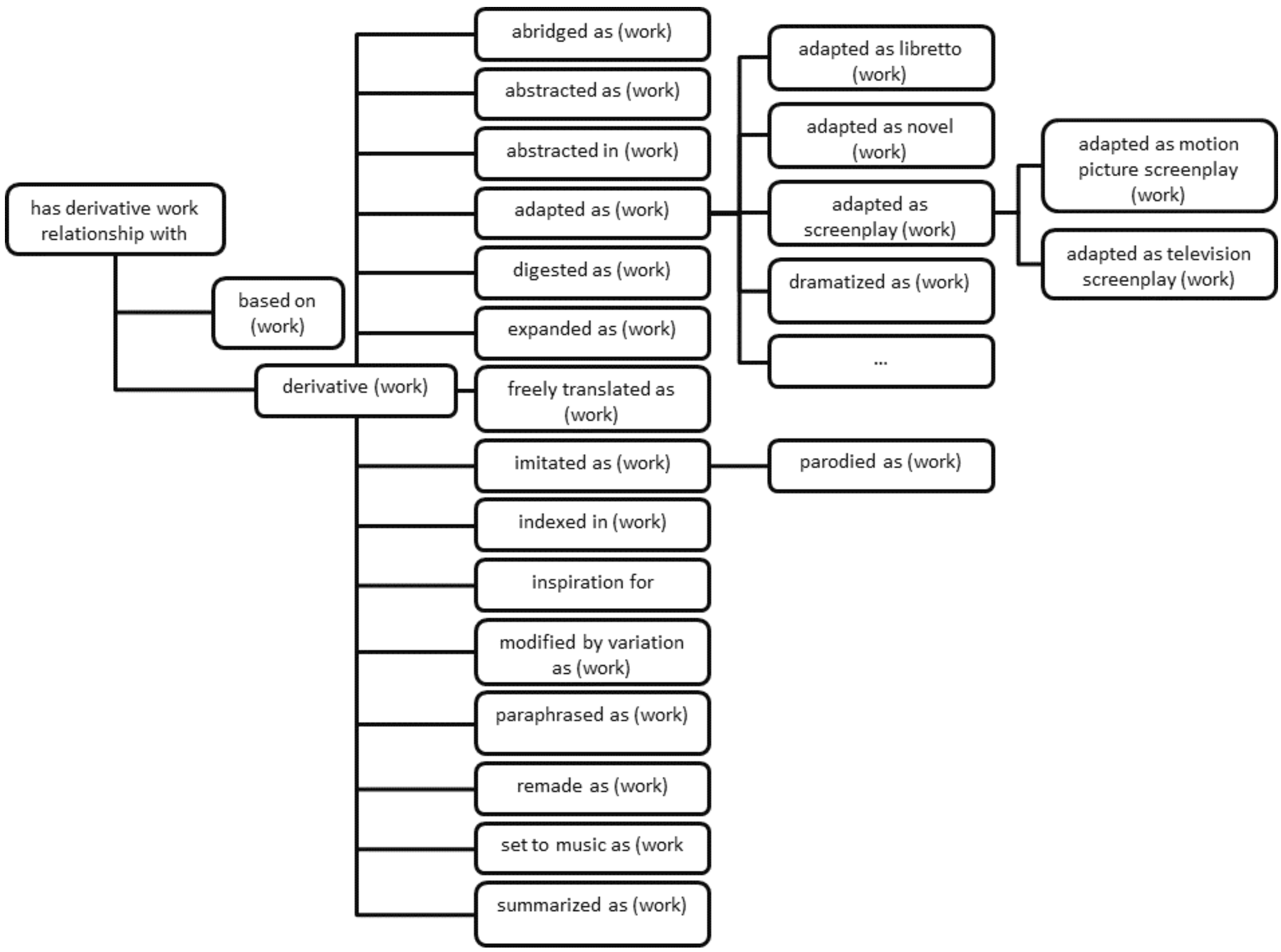


Figure 2:19. The RDA hierarchy of properties for the representation of derivations at the Work level.

2.3.4. Library Reference Model - LRM

The LRM model consolidates FRBR, FRAD, FRSAD and the two amendments (clarification of the *Expression* entity and Aggregates report) in one coherent model. The LRM has been designed to serve as a general high-level model that uses the entity-relationship formalism. Conceptualizations and techniques used in the FRBRoo (see paragraph 2.3.2) have been incorporated in LRM, such as the modeling of entities as classes, the representation of attributes and relationships as properties, and the extensibility of properties as a means of specialization. The model was endorsed by IFLA in 2017 (Riva, Bœuf, et al., 2017a).

2.3.4.1. LRM entities

Entities are organized in hierarchy with the *Res* entity as a superclass, and are identified with a unique ID having the following ID structure LRM-Ex, where x is a number, e.g., *LRM-E1 Res, LRM-E2 Work, etc*. Group 1 entities (*Work, Expression, Manifestation, Item*) remain with reworked definitions (Figure 2:20). There have been changes regarding Group 2 and Group 3 entities, correspondingly described in FRAD and in FRSAD. LRM introduces the *Agent* entity with two subclasses, *Person* and *Collective Agent.* The *Collective Agent* class may be used as the superclass for the deprecated *Corporate Body* and *Family* FRAD entities. It must be noted that the definitions of the entity *Person* differed between the FRBR and FRAD reports. LRM adopts the FRBR *Person* entity definition (Riva, Bœuf, & Žumer, 2017b). The *Concept, Object,* and *Event* Group 3 entities have been deprecated. New entities introduced are *Place, Time-Span.* The former FRSAD *Thema* entity has been generalized as *Res* (out of Resource) serving as the superclass of all entities in the LRM model. The former

FRAD *Name* and FRSAD *Nomen* entities have been merged into one, the *LRM-E9 Nomen* entity. In total, the 18 entities from the FRBR, FRAD, and FRSAD reports have been reduced to 11.

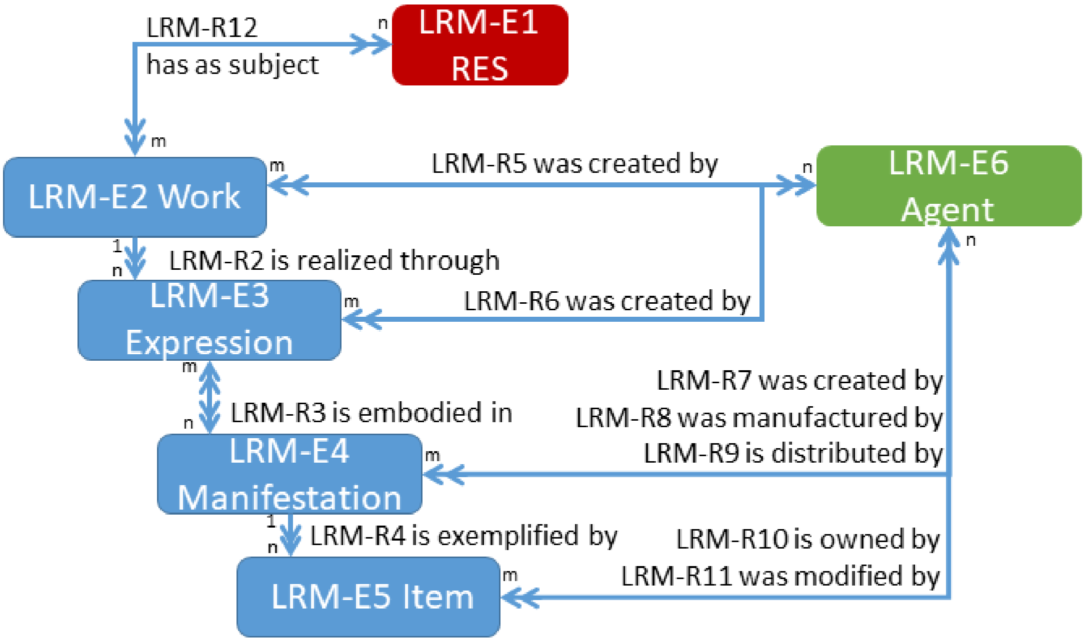


Figure 2:20. The Library Reference Model: core entities and inherent relationships.

Attributes and relationships are organized in hierarchy "following the entity hierarchy structure" (Riva, Bœuf, et al., 2017a). Due to the general character of the LRM model, the list of attributes and relationships is not exhaustive, and it is expected that cataloging rules will specify more specialized attributes and relationships, if needed. Yet, LRM has introduced some new attributes. Dispute has been raised regarding the *LRM-E2-A2 Representative expression* attribute (Glennan & James, 2018). This attribute may characterize an *LRM-E2 Work* by taking its value from attribute(s) that originate from a "representative or canonical" *LRM-E3 Expression* of the *LRM-E2 Work.* The *LRM-E2-A2 Representative expression* attribute may be multivalued taking its value from different *Expression* attributes depending on the *type* of *Work,* and on the selected cataloging rules. LRM does not define which attributes are to be used as "Representative expression" *Work* attributes. Another attribute added in the LRM is the *LRM-E4-A4 Manifestation statement* which may be used to transcribe from *Manifestation* exemplars information regarding the publication, such as place of publication, name of publisher, and date of publication. Attributes are "numbered sequentially within each entity" (Riva, Bœuf, et al., 2017a). As an example, an *LRM-E2 Work* attribute will have an ID with the following structure: *LRM-E2-Ax*, where x is a number. In total, 37 attributes are declared in LRM.

2.3.4.2. Bibliographic relationships

LRM identifies 36 relationships and their inverse ones. Similarly to the attributes modeling, relationships may be refined to represent more specialized ones. All relationships are numbered sequentially. Inherent relationships are numbered from LRM-R2 to LRM-R4. Bibliographic relationships may all be represented in LRM (Figure 2:21), but in many cases the names of relationships have changed, or the relationships have either merged with others or have been generalized (Riva, Bœuf, et al., 2017b). As an example, many *Expression*-to-*Expression* derivative relationships may be represented with the same *LRM-R24 is derivation of* relationship. Similarly to FRBRoo, the specific derivative relationship may be represented by assigning a *type* to the relationship, e.g., the *LRM-R24 is derivation of* property may be subtyped as abridgement, revision, translation, arrangement (Riva, Bœuf, et al., 2017b).

Tillett's taxonomy	LRM entity where applied	LRM properties for relationships
Equivalence	LRM-E4 Manifestation LRM-E5 Item	LRM-R27 has reproduction , LRM-R29 has alternate LRM-R28 has reproduction
Derivative	LRM-E2 Work LRM-E3 Expression	LRM-R21 is inspiration for , LRM-R22 is a transformation of LRM-R24 is derivation of (may be sub-typed using a vocabulary)
Descriptive	LRM-E2 Work	LRM-R12 has as subject
Whole-part	LRM-E2 Work LRM-E3 Expression LRM-E4 Manifestation	LRM-R18 has part LRM-R23 has part LRM-R26 has part
Accompanying	LRM-E2 Work	LRM-R20 accompanies / complements
Sequential	LRM-E2 Work	LRM-R19 precedes

Figure 2:21. LRM properties for the representation of bibliographic relationships. Properties are organized according to the Tillett taxonomy of bibliographic relationships.

It must be noted that LRM introduces a new relationship, the *LRM-R25 was aggregated by.* The addition of this relationship adheres to the FRBR amendment regarding aggregates (see 2.3.1.4). This relationship may be used to represent that an instance of the *LRM-E3 Expression* entity was aggregated by an *Aggregating Expression* that produced an *Aggregate Manifestation*.

2.3.4.3. Alignments following the LRM conceptualizations

Following the LRM conceptualizations, changes are expected in the design of other models, namely RDA (see 2.3.3 paragraph), and the FRBRoo which will be renamed to LRMoo.

With regard to RDA, the changes were considered in the RDA Toolkit Redesign and Restructure project. This project, known as the 3R Project, at first aimed to "add greater flexibility and utility to the Toolkit's display of instructions and RDA-related documents" (RDA Toolkit, 2016). After the publication of the LRM, the 3R Project included the alignment of the RDA with the LRM and the handling of issues, poorly developed in the RDA, such as aggregates and subjects. The 3R project has officially been completed in October 2019 (RDA Toolkit, 2019). The adoption of LRM as the conceptual model of the RDA introduces new entities and concepts in the latter, e.g., the *Collective Agent*, *Nomen*, and *Time-span* entities (see Figure 2:22), or the *Representative Expression* and the *Manifestation Statement* attributes. The changes are accessible in the beta RDA toolkit website (https://beta.rdatoolkit.org/), which is expected to switchover to official RDA status in December 2020 (RDA Toolkit, 2019).

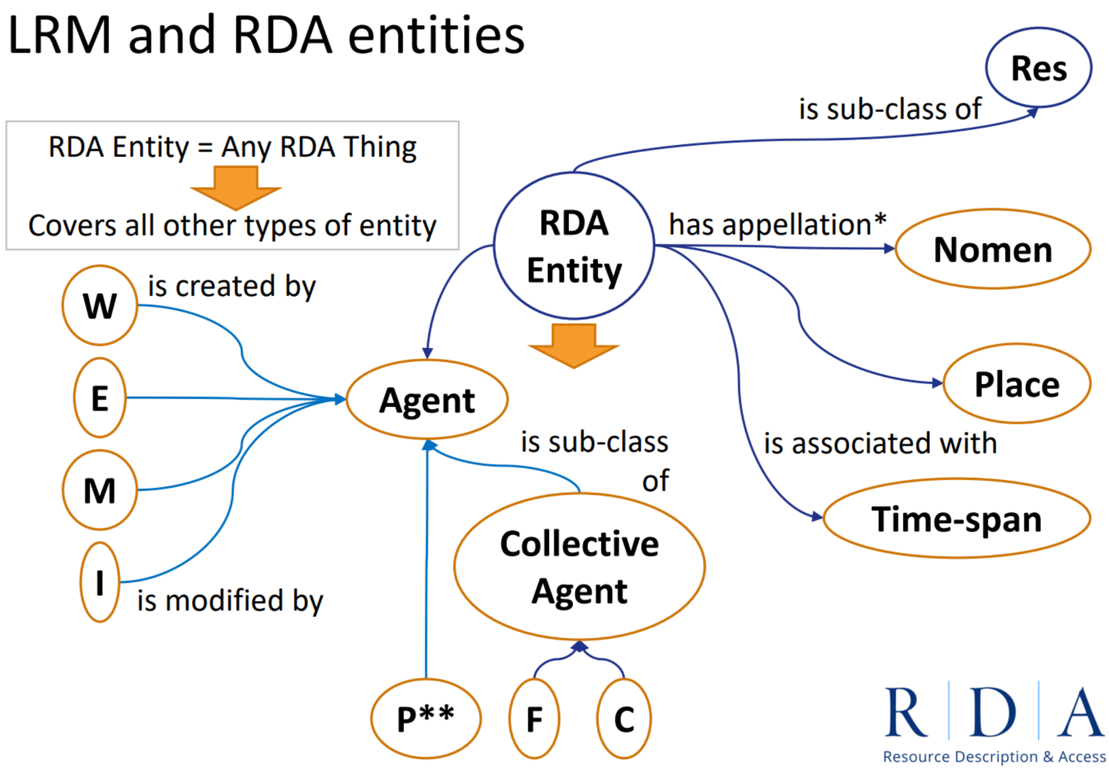


Figure 2:22. LRM and RDA entities. Source: (Dunsire, 2019).

2.3.5. Bibliographic Framework Initiative Data Model (BIBFRAME model)

Digital technologies evolved drastically in the late 20th century changing library collections and services. Therefore, in the 2000s there were talks, efforts within the library community regarding the future of bibliographic control. The Library of Congress participated in these discussions and in 2006 formed the "Working Group on the Future of Bibliographic Control". This Group submitted the "On the Record" report (Library of Congress Working Group on the Future of Bibliographic Control, 2008) where among other viewpoints it was recognized that the MARC21 has long-served as the primary bibliographic data carrier format between libraries. Among the recommendations were the replacement of MARC21 by a new "more flexible and extensible metadata carrier", and the integration of library standards and data into the web (Library of Congress Working Group on the Future of Bibliographic Control, 2008). Moreover, tests regarding the implementation of the Resource Description & Access standard in MARC21 have raised the MARC21's limitations as a an underlying data carrier (Library of Congress, 2011a). Therefore, in 2011 the Library of Congress announced the launch of an initiative for the analysis of the then present and future bibliographic environment (Library of Congress, 2011b). In the same year, Library of Congress published the "Bibliographic Framework Initiative General Plan" (Library of Congress, 2011a). With the aim of accommodating various communities' content rules and data models within the web environment, two prerequisites were stated regarding the technologies that shall be used: "Linked Data principles and mechanisms, and the Resource Description Framework (RDF) as [the] basic data model" (Library of Congress, 2011a).

The Bibliographic Framework model, known as BIBFRAME, was created in collaboration with Eric Miller's Zepheira company, and officially announced in 2012 (Miller, Ogbuji, Mueller, & MacDougall, 2012). The current version is the second one, published in 2016 (Library of Congress, 2016b). BIBFRAME 2.0 is formally expressed using the Web Ontology language and several RDF conventions have been followed (Library of Congress, 2017). Aiming for simplicity, the BIBFRAME model avoids "proliferation of properties by defining a single general property", generally avoids specifying explicit domain and range for properties (Library of

Congress, 2017), and thus, even though usual domain and range are noted for each property, they are not to be perceived as formal constraints. The BIBFRAME model has been designed for simplicity and ease of use (Schreur, 2018). Its definitions remain purposely flexible enabling different BIBFRAME interpretations and implementations.

The BIBFRAME community is a vibrant one, supported by the Library of Congress and other funded projects, and it already demonstrates various early implementors, extensions for specific communities (e.g., art resources, music resources), the bibliotek-o derivation, and tools, such as the BIBFRAME Editor (editing tool), the MARC21-BIBFRAME comparison viewer, and the MARC21-BIBFRAME transformation software.

2.3.5.1. BIBFRAME classes

The development of the BIBFRAME model is not task-fueled as FRBR or other models inspired by FRBR. The starting point for its development has been the replacement of MARC21 and the integration of bibliographic data in the web. Key principles and library practices of the library community were taken into account. The definition of the *bf:Work* and the *bf:Instance* classes adheres to the distinction between content and carrier, differentiating between the abstract content and its physical embodiments. Information entities (authorities in legacy library systems) have been included in the BIBFRAME model as classes, e.g., *bf:Agent, bf:Place, bf:Event.* Bibliographic relationships have also been declared with a hierarchy of properties (Miller et al., 2012). BIBFRAME 2.0 (Figure 2:23) defines three core classes: *Work, Instance,* and *Item.* The *Work* (*bf:Work*) class reflects the content (both ideas and the signs used for their realization), the *Instance* (*bf:Instance*) class is the embodiment of the *Work.* The *Item* (*bf:Item*) class is the exemplar of the *Instance.* The *Agent* class refers to *Person*, *Family*, *Organization*, *Jurisdiction*, and *Meeting* agents that have a role in a resource. The *Event* class is used as the subject of the *bf:Work.* A *bf:Work* may have as a subject any other resource, such as *Work, Instance, Item, Agent, Event, Place,* topics, etc. (Figure 2:23).

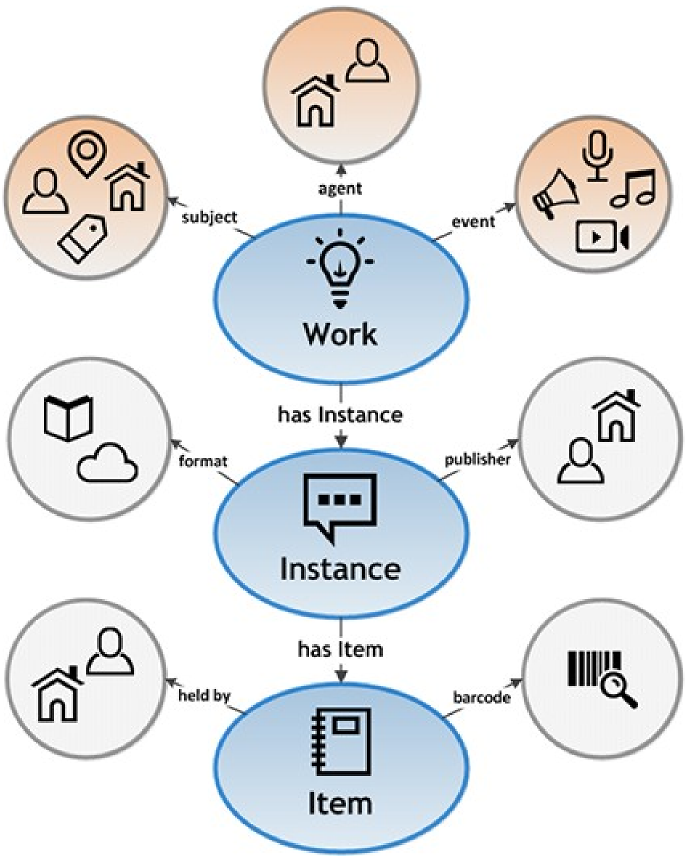


Figure 2:23. BIBFRAME 2.0 Model. Source: (Library of Congress, 2016b).

2.3.5.2. Bibliographic relationships

For the representation of relationships, BIBFRAME uses the *bf:relatedTo* property. All inherent and bibliographic relationships are represented with subproperties of the *bf:relatedTo* property. The subproperties for the representation of Inherent relationships are *bf:hasInstance* and *bf:hasItem*. BIBFRAME also declares the *bf:hasExpression* property either to relate two *bf:Works* where the one is an expression of the other, or to relate *bf:Works* under FRBR/RDA rules. The subproperties for the representation of other bibliographic relationships are: *bf:hasPart/ bf:part of* (representing the Whole-part relationships), *bf:accompaniedBy/bf:accompanies* (representing the Accompanying relationships), *bf:hasDerivative/bf:derivativeOf* (representing the Derivative relationships), *bf:hasEquivalent* (representing the Equivalence relationships), *bf:references/bf:referencedBy* (representing the Descriptive relationships), *bf:precededBy/bf:succeededBy* (representing the Sequential relationships). It must be noted that some of these subproperties have their own subproperties representing more specialized relationships. Most of the subproperties have an inverse one, while there is a small number of symmetric subproperties, e.g., *bf:issuedWith, bf:otherEdition.* The whole hierarchy of properties for the representation of bibliographic relationships in BIBFRAME is presented in Figure 2:24.

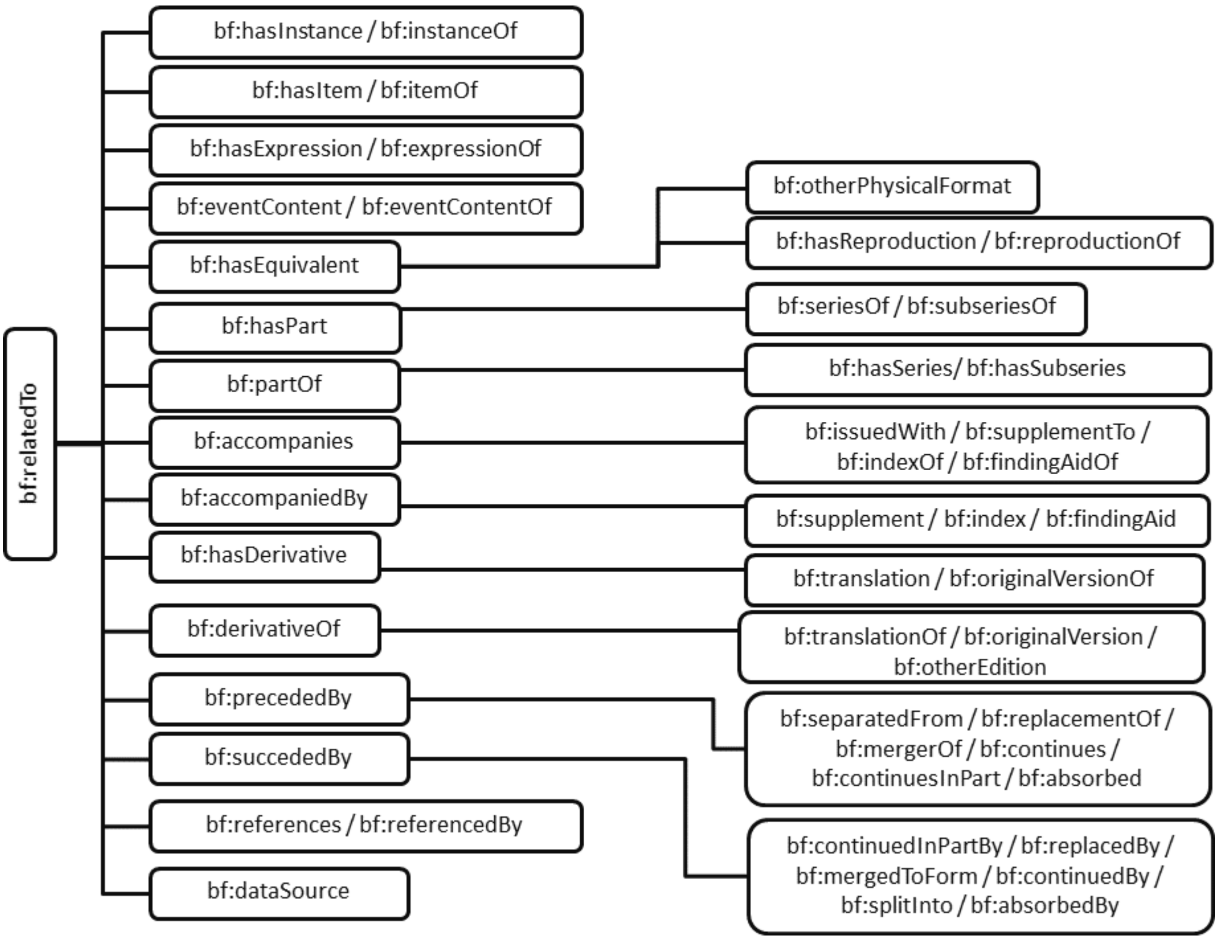


Figure 2:24. The hierarchy of properties for the representation of bibliographic relationships in BIBFRAME.

For comparison reasons, the BIBFRAME properties for the representation of relationships are once again presented in the following Figure 2:25 using the Tillett taxonomy of bibliographic relationships. The numbers in brackets present the number of subproperties.

Tillett's taxonomy	BF class where applied	BF properties for relationships
Equivalence	bf:Work, bf:Instance, bf:Item	bf:hasEquivalent [3]
Derivative	bf:Work, bf:Instance	bf:hasDerivative [2] / bf:derivativeOf [3]
Descriptive	bf:Work, bf:Instance, bf:Item	bf:subject, bf:references/bf:referencedBy
Whole-part	bf:Work, bf:Instance, bf:Item	bf:hasPart [2] / bf:part of [2]
Accompanying	bf:Work, bf:Instance, bf:Item	bf:accompanies [4] / bf:accompaniedBy [3]
Sequential	bf:Work, bf:Instance	bf:precededBy [6] / bf:succeededBy [6]

Figure 2:25. BIBFRAME properties for the representation of bibliographic relationships organized according to the Tillett taxonomy of bibliographic relationships. The numbers in brackets refer to the number of each property's subproperties.

2.3.5.3. *BIBFRAME's flexibility and BIBFRAME-inspired vocabularies*

BIBFRAME, within the scope of enabling flexibility in future implementations, purposely provides loose definitions and does not impose cardinalities. Yet, cardinalities and specific guidelines are needed in implementations. To support local cataloging practices, the use of BIBFRAME Profiles is proposed (BIBFRAME - Bibliographic Framework Initiative, 2014). A BIBFRAME Profile may be one document or a set of documents that provide guidance regarding the local cataloging policy constraining how a resource may be described with which properties and value vocabularies. Thus, a BIBFRAME Profile document may include domain/range constraints not specified in the official model to facilitate cataloging tools to implement the model according to a library's needs (BIBFRAME - Bibliographic Framework Initiative, 2014). Nevertheless, BIBFRAME profiles exist outside the BIBFRAME model and they serve local practices.

Varying interpretations of BIBFRAME's loose definitions lead to different representations of the same real-world bibliographic description cases. As an example, according to the definition of the *bf:hasExpression* property ("BIBFRAME ontology - hasExpression," 2016), the property can be used to relate two *bf:Works* that realize the same content in different sets of signs (Representation 1 in Figure 2:26) or in an FRBR-similar way (Representation 2 in Figure 2:26). With regard to the second representation, even though it is based on the definition of the *bf:hasExpression,* it is not clear if there is conflict against the *bf:Work* semantics. Due to BIBFRAME's lack of cardinalities, it is not known if a *bf:Work must* or *may* have one or more *bf:Instances.* Supposing that a *bf:Work may* have one or more *bf:Instances,* the second representation (Figure 2:26) can be further observed. Similarly to the FRBR, the left *bf:Work* in Figure 2:26 may represent only ideational content, while the second *bf:Work* (on the right) includes both the ideational content and the signs realizing it. This representation reminds the following FRBR statement: *Work - is realized through - Expression - is embodied in - Manifestation*. Yet, there is a difference; the left *bf:Work*, similarly to the FRBR *Work* entity, lacks information about signs and includes only ideas, contrary to the semantics of the *bf:Work* class. The thesis employs the term "*Expression*-agnostic *bf:Work"* for this specific use of the *bf:Work* class. Even though, the second representation is similar to the aforementioned FRBR statement and the left *Expression*-agnostic *bf:Work* may be considered similar to the FRBR *Work* entity, the right *bf:Work* cannot be considered equivalent to the FRBR *Expression* entity. The FRBR *Expression* entity includes only the signs used for the realization of ideas, whereas the *bf:Work* on the right represents both ideas and signs retaining its class' semantics. Both representation approaches presented in Figure 2:26 do not violate BIBFRAME semantics and are valid due to BIBFRAME's flexible nature. It is evident, though, that they may result in totally different implementations.

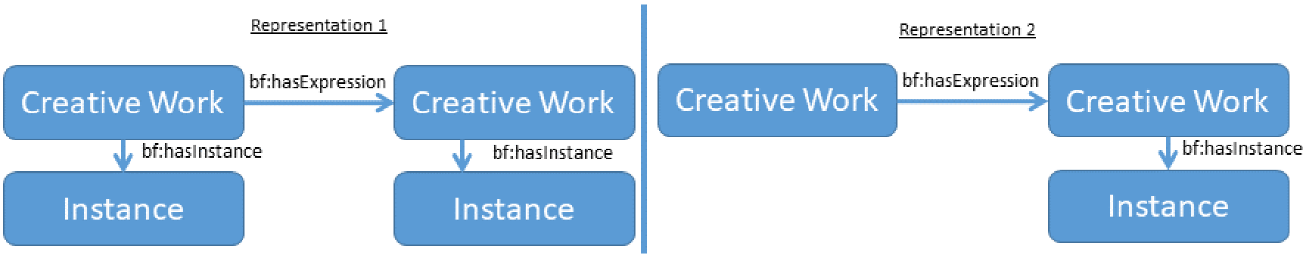


Figure 2:26. The loose definition of the bf:hasExpression property enables two different representations.

BIBFRAME has been designed to be simple. Yet, there are domains with special needs and have extended BIBFRAME to better support them. In 2015, the National Library of Medicine experimented with BIBFRAME by creating a core vocabulary (BIBFRAME Lite) (National Library of Medicine, 2015; Schreur, 2018). The Mellon-funded Linked Data for Production (LD4P) project has extended BIBFRAME 2.0 to support the cataloging of the following: art, rare books, performed music, moving image, and maps (Falcone, Greben, & Lorimer, n.d.; Schreur, 2018). Another LD4P BIBFRAME extension is the bibliotek-o ontology (check paragraph 2.3.7 for more information). BIBFRAME is still under development and updates are anticipated.

2.3.6. Europeana Data Model (EDM)

The Europeana Data Model (Europeana, 2017), has been developed for the cultural heritage domain in the framework of the Europeana aggregation portal (http://www.europeana.eu/). Europeana provides access to European cultural heritage resources. These resources, either born-digital or digitized, are provided by European memory institutions, i.e., libraries, museums, and archives. EDM is a data model developed according to Semantic Web principles to serve different communities (Isaac, 2013).

EDM re-uses elements from other namespaces (Europeana, 2017), such as Resource Description Framework and Resource Description Framework Schema (http://www.w3.org/2000/01/rdf-schema), OAI Object Reuse and Exchange (ORE) namespace (http://www.openarchives.org/ore/terms/), Simple Knowledge Organization System (SKOS) namespace (http://www.w3.org/2004/02/skos/core), etc.

The Europeana Data Model is constantly updated. The current version is 5.2.8. (Europeana, 2017).

2.3.6.1. EDM classes

Libraries and Museums provide to Europeana EDM descriptions of cultural heritage objects held in their collections, as well as links to the objects' digital surrogates. The objects are represented in EDM with the *edm:ProvidedCHO* class, and the digital surrogates are represented with the *edm:WebResource* class. The *ore:Aggregation* class is used to group the provided objects (*edm:ProvidedCHO*) with their digital representation(s), viz. one or more *edm:WebResource* instances (Figure 2:27). So, the EDM core classes are *edm:ProvidedCHO, edm:WebResource,* and *ore:Aggregation.*

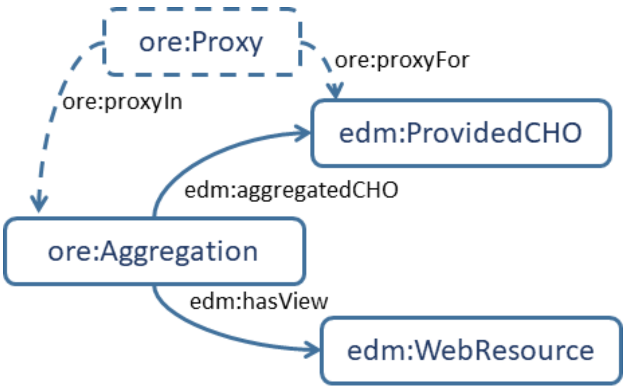


Figure 2:27. EDM core classes and inherent relationships. Representation using the ore:Proxy class is depicted with dashes.

EDM also facilitates an alternative representation that uses instances of the *ore:Proxy* class (Figure 2:27). The *ore:Proxy* class is used in Europeana as a placeholder for "cultural heritage objects within aggregations in order to make assertions about the corresponding cultural heritage objects while distinguishing the provenance of these assertions" (Europeana, 2017). This enhancement enables the existence of only one *edm:ProvidedCHO* instance for each European Heritage object, and of multiple instances of the *ore:Proxy* class including the descriptions provided by the different providers.

2.3.6.2. Bibliographic relationships

All bibliographic relationships as identified by Tillett can be represented in EDM, except for the accompanying relationship (Figure 2:28). The number of properties for each type of relationship is limited, probably due to the EDM's cultural heritage orientation and Europeana's role as aggregator of descriptions from a variety of providers. EDM either defines properties for the representation of bibliographic relationships, e.g., *edm:isDerivativeOf* (Derivative relationship) and *edm:isSuccessorOf* (Sequential relationship), or re-uses properties already identified in the third namespaces it employs, e.g., the Dublin Core *dcterms:hasFormat* (Equivalence relationship) and *dcterms:hasPart* (Whole-part relationship) properties.

Tillett's taxonomy	EDM class where applied	EDM properties for relationships
Equivalence	edm:WebResource	dcterms:hasFormat / dcterms:isFormatOf
Derivative	edm:ProvidedCHO	edm:isDerivativeOf, dcterms:hasVersion / dcterms:isVersionOf
Descriptive	edm:ProvidedCHO	dc:subject, dcterms:references / dcterms:isReferencedBy
Whole-part	edm:ProvidedCHO	dcterms:hasPart / dcterms:isPartOf, dc:source
Accompanying	edm:ProvidedCHO	-
Sequential	edm:ProvidedCHO	edm:isSuccessorOf, edm:isNextInSequence

Figure 2:28. Properties for the representation of bibliographic relationships in EDM. The properties are organized according to the Tillett taxonomy of bibliographic relationships.

2.3.6.3. EDM alignments and extensions

Moreover, EDM includes constructs to enable different modeling approaches, e.g., object-centric, event-centric, and representations that serve providers' specific interests, e.g., use of the *ore:Proxy* class. Providers are expected to use these constructs indirectly via more specialized constructs that conform to the general interoperability levels that EDM defines (Isaac, 2013). Different Europeana projects develop alignments and extensions to the EDM using the EDM constructs. The alignment of EDM with library metadata was studied within the Europeana Libraries project ("Europeana Libraries | Europeana Pro," n.d.).

Two reports were published in 2012 investigating how specific library material (monographs, multi-volume works and serials) could be described in EDM (Angjeli, Baumgartner, et al., 2012; Angjeli, Bayerische, et al., 2012). Both reports took under consideration the FRBR model. The alignment was not achieved, and the reports introduced the "edition" concept. According to the report, an *edm:ProvidedCHO* instance could be an "edition" incorporating information regarding three FRBR Group 1 entities, *Work, Expression, Manifestation.* The *edm:WebResource* class could be used as the web exemplar of the *edm:ProvidedCHO* class instance. The need for compliance with the FRBR was postponed and provoked the formation of a group working toward the integration of "FRBR entities in EDM using FRBRoo terms" (Angjeli, Bayerische, et al., 2012), and the development of an EDM-FRBRoo application profile (see paragraph 2.6.2).

The Europeana Data Model is being updated and there are ongoing projects regarding its extensions and its interoperability with other communities' models (Charles, Isaac, & Manguinhas, 2017).

2.3.7. Other conceptual models

This thesis focuses on the aforementioned models because they have been developed or endorsed by official institutions in the library and cultural heritage communities. Other models have been developed by research teams, or as part of greater projects involving the publication of linked data. Some of these models are presented below. They have been selected for a multitude reasons; some are among the first attempts for creating bibliographic conceptual models, others have been influenced by or extend the models presented in paragraphs 2.3.1 - 2.3.6, while others have been developed as part of greater linked data projects. The presentation starts from older models using "flat" modeling approaches (MarcOnt, BIBO, British Library Data Model), continues with models following conceptualizations expressed in well-known models (FaBiO, PRESSoo, bibliotek-o, DPLA-MAP), and concludes with the broader Schema.org that has been used by OCLC for publishing WorldCat data as linked data. It must be noted that the CIDOC-CRM model is not referred in this paragraph, because it is referenced in its bibliographic extension FRBRoo model (paragraph 2.3.2.). The presentation of each model shortly presents its core classes, its inherent relationships, as well as if bibliographic relationships are treated by its constructs.

2.3.7.1. MarcOnt ontology

One of the first bibliographic ontologies to investigate the conversion of legacy bibliographic data to RDF, was the MarcOnt ontology created in the early 2000s. This ontology was developed to integrate MARC21 data in digital libraries and semantic web environments (Kruk, Dabrowski, & Synak, 2009). The MarcOnt ontology was a social ontology conforming to MARC21, and reusing Dublin Core, and BibTEX metadata (Kruk, 2004). The main idea behind its development was that MarcOnt could serve as a transformation mechanism for converting bibliographic data from one model to another (Kruk & Zimmermann, 2005). The MarcOnt ontology adopted the MARC21 flat structure and it did not conform to FRBR conceptualizations. MarcOnt defined types of publications as core classes, e.g., *marcont:Book, marcont:Conference, marcont:Mastersthesis,* etc. A suite of technologies and tools (JeromeDL, Mediation Services, RDF Translator) were developed along with the ontology to support the conversion of MARC21 data in RDF. The project has ended and the site hosting it is accessible only through the Internet Archive's Way Back Machine.

2.3.7.2. Bibliographic ontology – BIBO

The Bibliographic Ontology was developed in 2009 by Bruce D'Arcus and Frédérick Giasson to enable the description of "citations and bibliographic references (i.e., quotes, books, articles, etc.) on the Semantic Web" (Bruce D'Arcus & Frédérick Giasson, 2016). BIBO's latest version is 1.3, published in OWL. BIBO's core class is the *bibo:Document*; examples of its subclasses are *bibo:Book, bibo:Manuscript, bibo:Thesis*, etc. The definition of the *bibo:Document* class being a "bounded physical representation of body of information" conforms to "flat" modeling approaches without differentiating between content and carrier. Due to BIBO's "flat" approach, no inherent relationships are recognized. The focus on the description of citations has induced the representation of citation-related properties, such as *bibo:cites*, *bibo:annotates, bibo:reviewOf.* The BIBO ontology has been reused in other frameworks, such the development of the British Library Data Model (see paragraph 2.3.7.3) and the publication of the British National Bibliography, the publication of bibliographic data as linked data by the Deutsche Nationalbibliothek (Biagetti, 2018; Deutche National Bibliothek, 2016), the publication of historic American newspapers as linked data (Library of Congress, n.d.-a), or the development of the Open Bibliography for Science, Technology, and Medicine (Jones et al., 2011).

2.3.7.3. British Library Data Model

The British Library has been one of the pioneers in developing a conceptual model adhering to Semantic Web principles. The British Library Data Model was published in 2011 within the framework of releasing the British National Bibliography (BNB) as linked data (Deliot, 2014). First, it was focused on monographs and did not adhere to FRBR conceptualizations because at the time of development (2010-2011) FRBR seemed "too complex" (Deliot, 2014) for implementation. After the diagrammatic model for Books, two more diagrams have been published, namely *Serial,* and *Forthcoming Book*. They are not different models, but different classes and properties are used for their description. This is expected, and it seems rational considering the different nature of serials and books. The BL Data Model reuses other element sets and vocabularies, such as the ISBD element set, the Bibliographic Ontology – BIBO (see paragraph 2.3.7.2), Dublin Core, Friend-Of-A-Friend vocabulary, Event Ontology, etc. Core classes described by the model are the *bibo:Book, bibo:MultiVolumeBook, bibo:Periodical,* and *bibo:Newspaper.* All classes' instances are described with object and data properties defined either in the British Library Terms RDF schema (http://www.bl.uk/schemas/bibliographic/blterms#), or in third-party vocabularies (ISBD, DC, BIBO, FOAF, etc.). Since the FRBR conceptualizations have not been adopted, the bibliographic resources described by the model may be characterized as "flat" and not granular. Inherent relationships are not defined and a few bibliographic relationships between bibliographic resources are used by the model. As an example, only one bibliographic relationship is used for *bibo:Book* and *bibo:MultiVolumeBook* instances, the whole/part relationship to the publishing series in which these instances have been published. Contrary, there are a lot of other relationships used to describe authorship, subjects and publication events. Publication has been modeled as an event to enable modeling of forthcoming publications (Deliot, 2014). The FRBRization of the BNB data will be revisited for future releases of the BNB (Deliot, 2014).

2.3.7.4. FRBR-aligned Bibliographic Ontology (FaBiO)

FaBiO is an ontology developed as a part of the "Semantic Publishing and Referencing Ontologies" (SPAR) (Peroni & Shotton, 2012) project. FaBiO classes are structured according to the FRBR; FRBR WEMI entities are defined as classes having their own specialized sub-classes. As an example, the *fabio:Work* class has 30 subclasses (e.g., *fabio:ArtisticWork, fabio:ReferenceWork, fabio:WorkCollection*), and the *fabio:Expression* class has 57 sub-classes (e.g., *fabio:Abstract, fabio:Excerpt, fabio:Oration*). FaBiO uses the FRBR terminology, but it seems not to clearly differentiate between the *fabio:Work* and *fabio:Expression* classes' meaning. According to (Biagetti, 2018), FaBiO developers have confused the FRBR *Work* and *Expression* bibliographic entities with the types of publications. FaBiO defines more inherent relationships than the FRBR, as expressed in the FRBR Core vocabulary (Davis, Newman, & D'Arcus, 2005). Besides the three inherent FRBR relationships (*frbr:realization, frbr:embodiment, frbr:exemplar*), FaBiO defines the *fabio:hasManifestation*, *fabio:hasRepresentation*, and the *fabio:hasProtrayal.* The first relationship relates a *fabio:Work* instance to a *fabio:Manifestation* instance, the second one relates a *fabio:Expression* instance with a *fabio:Item* instance, and the third relationship relates a *fabio:Work* instance with a *fabio:Item* instance (see Figure 2:29). Other bibliographic relationships are not defined in FaBiO.

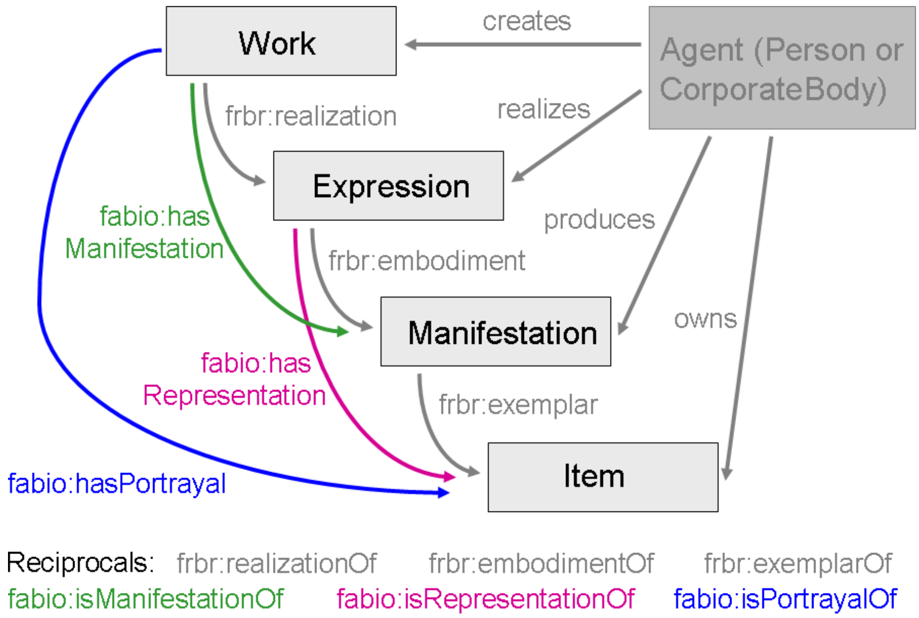


Figure 2:29. New Work-Expression-Manifestation-Item relationships in FaBiO. Source: (Shotton & Peroni, 2019).

2.3.7.5. PRESSoo

PRESSoo is another model inspired by the FR family of models. It is actually an extension to the FRBRoo model developed to serve the needs of describing serials and continuing resources using FRBR constructs (IFLA PRESSoo Review Group & Bœuf, 2017). The current 1.3 version was published in 2015 and its update is anticipated due to FRBRoo's imminent change into LRMoo (see paragraph 2.3.2).

2.3.7.6. Bibliotek-o ontology

bibliotek-o is a BIBFRAME extension jointly produced by the Linked Data for Libraries Labs (LD4L) and Linked Data for Production (LD4P) projects (Futornick & Younes, 2017). The LD4L project reviewed the first version of BIBFRAME considering it as inadequate for the representation of library material (Kovari, Folsom, & Younes, 2017). Therefore, the LD4L project supported the development of its own ontology. After the publication of BIBFRAME 2.0 in 2016, the LD4L ontology was abandoned and the development of the bibliotek-o ontology began. bibliotek-o embraces BIBFRAME at its core, but also utilizes "other external ontology fragments" (Kovari et al., 2017). Some of the differences between bibliotek-o and BIBFRAME are 1) the reuse of existing external value vocabularies, 2) the use of OWL axioms and RDF constructs, such as the domain and range constructs, and the preference to object properties and structured data over unstructured literals, and 3) the use of simpler representations over BIBFRAME multi-path representation patterns (Kovari et al., 2017). Moreover, editor tools and a "MARC21 to bibliotek-o convertor" are being developed to better support the use of bibliotek-o. The convergence of bibliotek-o and BIBFRAME is not anticipated. bibliotek-o is rather presumed as a framework for the study of alternative to BIBFRAME representations, and for the alignment with other communities' vocabularies and ontologies. Hopefully, the demonstrated results from the bibliotek-o experimentations will affect BIBFRAME's future updates and future modeling decisions (Futornick & Younes, 2017).

2.3.7.7. DPLA Metadata Application Profile – DPLA MAP

The Digital Public Library of America ("Digital Public Library of America," n.d.) is a project similar to Europeana. The description of the resources aggregated in the DPLA is made using the DPLA Metadata Application Profile. The DPLA-MAP is based on the EDM (presented in paragraph 2.3.6) with some differences. It defines four core classes and their inherent relationships. Other relationships between the core classes may exist. The ones depicted in Figure 2:30 are the recommended ones (Gueguen et al., 2017). Bibliographic relationships are not included in the model and the only property used to express a relationship between *dpla:SourceResource* instances is the generic *dc:relation* property.

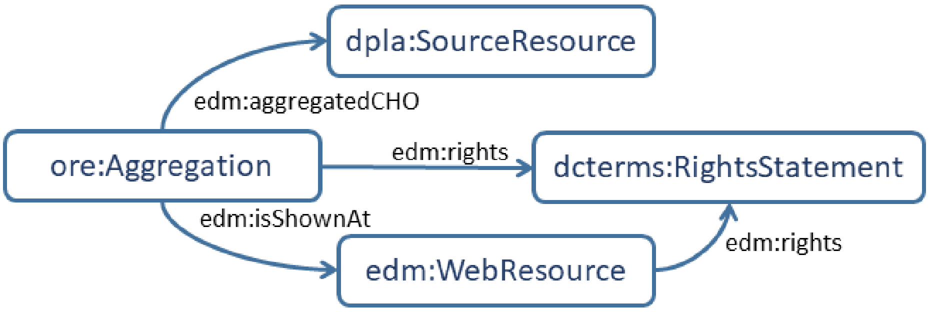


Figure 2:30. DPLA MAP core classes and their inherent relationships.

2.3.7.8. Schema.org, bib.schema.org and OCLC's model of Works

Schema.org has been developed by Google, Microsoft, Yahoo and Yandex to enable the mark-up of webpages with structured metadata ("Home - schema.org," 2019). Bib.schema.org is an extension of schema.org to represent bibliographic entities. OCLC has used schema.org and extended it with classes and properties from its own schemas available at the *bibliograph.net* site. OCLC Bibliograph.net extensions have been integrated into the bib.schema.org. OCLC uses the Schema.org and its extensions to publish WorldCat data as linked data since 2012 (Suominen & Hyvönen, 2017). It must be noted that the OCLC model descends from the British Library Data model (Godby et al., 2015) and its flat descriptive approach.

Schema.org's top class is *schema:Thing* and the *schema:CreativeWork* is one of its subclasses. The *schema:CreativeWork* has many subclasses, such as *schema:Book, schema:Article, schema:Game, schema:Movie*, etc. These subclasses have been defined according to content types (e.g., *schema:Map, schema:MediaObject*) or according to genre (e.g., *schema:Review, schema:ComicStory*), structure (e.g., *schema:Chapter*), etc. OCLC uses the *schema:CreativeWork* to model both content and publication products (Figure 2:31). The classes are not disjoint and one class' instance may be an instance of another class at the same time. A typical example is the representation of publication products, where the instances of the class *schema:CreativeWork* class may also be instances of the *schema:Product* or the *schema:IndividualProduct* classes. When a *schema:CreativeWork* is typed as a *schema:Product*, then it represents the publication. When a *schema:CreativeWork* is typed as a *schema:IndividualProduct*, then it represents one exemplar of the publication. The property relating a *schema:CreativeWork* class instance representing content with one or more *schema:CreativeWork* class instances representing publication products is the *schema:workExample* and its inverse *schema:exampleOfWork* properties. These properties may also be used to relate a *schema:CreativeWork* class with other *schema:CreativeWork* instances representing realizations and derivations of the former *schema:CreativeWork* instance. Therefore, the *schema:workExample/schema:exampleOfWork* properties may be either used for representing inherent relationships or derivative relationships.

Schema.org enables the representation of other bibliographic relationships too, i.e., derivative-translation (*schema:translationOfWork/schema:workTranslation*), derivative-adaptation (*schema:isBasedOn*), whole-part (*schema:isPartOf/schema:hasPart*), and descriptive (*schema:subjectOf/schema:about*). The top-level view of the OCLC's model is depicted in (Godby et al., 2015).

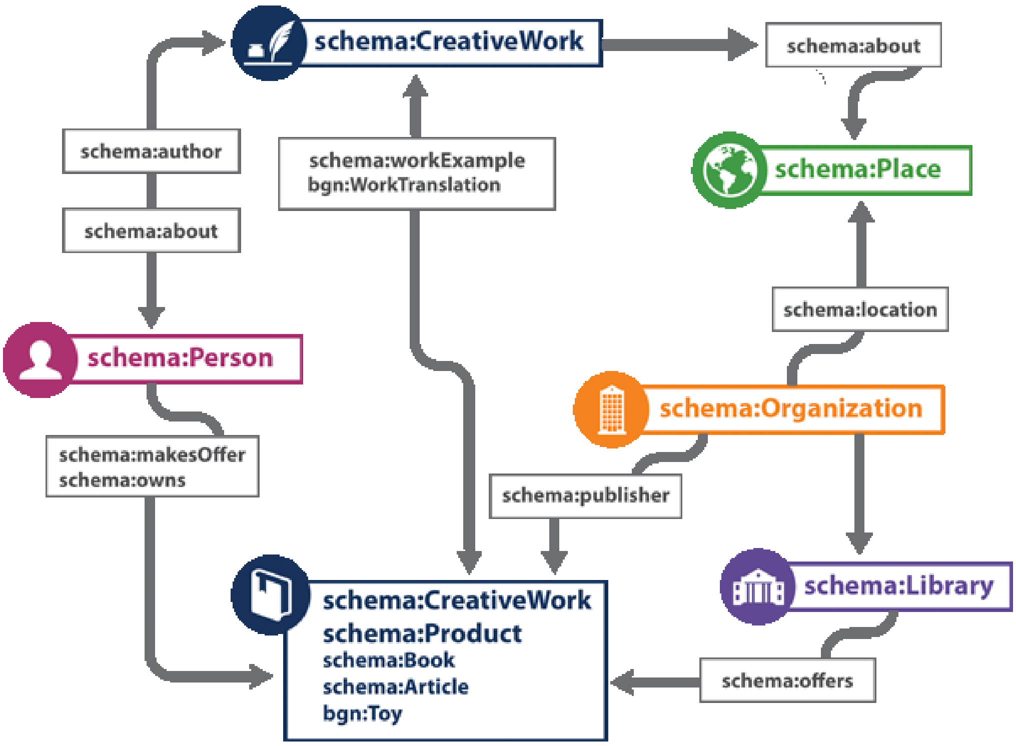


Figure 2:31. The top-level view of the OCLC model of Works. Source: (Godby et al., 2015).

2.4. Differences between the models regarding core entities and bibliographic relationships

The previous paragraph has attempted to provide an overview of the bibliographic conceptual models in terms of core entities/classes and inherent/bibliographic relationships. Besides the obvious abundance of models in the library domain, important differences have been identified between the models. These differences are displayed in the following two tables, Table 2-8 and Table 2-9.

Table 2-8 compares the number of core entities/classes between the models. The names of these core entities are given along with the number of subclasses, if such exist. The table clearly illustrates that there are differences between the models regarding the entities they acknowledge and describe. These differences consist different conceptualizations about the entities of the bibliographic universe and their attributes.

Table 2-8. Core entities in the studied bibliographic models: the number of entities/classes and their names. The numbers in brackets present the number of subclasses if they exist.

Model	Number of core entities/classes	Names of core entities/classes [number of subclasses]
FRBR	4	Work, Expression, Manifestation, Item
FRBRoo	5	F1 Work [4], F2 Expression [2], F3 Manifestation, F4 Manifestation Singleton, F5 Item
RDA	4	Work, Expression, Manifestation, Item
LRM	4	LRM-E2 Work, LRM-E3 Expression, LRM-E4 Manifestation, LRM-E5 Item
BIBFRAME	3	Work [11], Instance [5], Item
EDM	2	edm:ProvidedCHO, edm:WebResource

Table 2-9 uses Tillett's taxonomy of bibliographic relationships (Tillett, 1987) to present at which core entity each type of relationship may be applied and with how many relationships/properties. In case these relationship/properties are further specialized, then the number of subproperties is given in brackets. It must be noted that in BIBFRAME, contrary to other models, the same property is used for the representation of a relationship at different core classes. This the reason that the number of properties is given only once in each cell and not per core class as happens in the rest of the models.

First, it is apparent that the models do not agree at which level each type of relationship may be applied. Interestingly, there is no consensus, even between FRBR and other models inspired by it. As an example, the accompanying relationship can be applied to *Work* and *Expression* entities in FRBR, to all *WEMI* entities in RDA, and only to the *Work* entity in LRM. Another obvious point is the dissimilarity in the number of relationships/properties that each model defines for each type of relationship. The dissimilarity is noticeable especially in the case of the derivative relationship. FRBR defines 4 relationships for the representation of the derivative relationship between *Works,* while RDA defines one top-property with 15 subproperties, some of which have their own subproperties. FRBRoo and LRM enable the specialization of the derivative relationships they define by using values from controlled vocabularies. Some values are given by the models. Yet, they both clarify that local implementations can use local vocabularies and specialize even further the derivative relationships in question.

Table 2-9. Bibliographic relationships in the studied models. The entities on which these relationships are applied is presented, along with the number of properties and the number of subproperties if such exist. The number of subproperties is given in brackets.

Tillett's taxonomy	**FRBR**	**FRBRoo**	**RDA**	**LRM**	**BIBFRAME**	**EDM**
Equivalence	Manifestation-2 Item-2	F3 -1 F4-1 F5 -1	Manifestation-1[4] Item-1[2]	E4-2 E5-1	Work, Instance, Item-1[3]	WebResource-1
Derivative	Work-4 Expression-8	F1-1* F2-1*	Work-1[15] Expression-1[15]	E2-2 E3-1*	Work, Instance-1[3]	ProvidedCHO-2
Descriptive	Work-1	F1-1	Work-1	E2-1	Work, Instance, Item-2	ProvidedCHO-2
Whole-Part	Work-1 Expression-1 Manifestation-1 Item-1	F1-1 F2-2 F3-1 F4-1 F5-1	Work-1[2] Expression-1 Manifestation-1[3] Item-1	E2-1 E3-1 E4-1	Work, Instance, Item-1[2]	ProvidedCHO-2
Accompanying	Work-2 Expression-2	F1 -1* F2 -1*	Work-2[21] Expression-2[21] Manifestation-1[1] Item-1[4]	E2-1	Work, Instance, Item-1[4]	-
Sequential	Work-1 Expression-1	F1-1	Work-1[10] Expression-1[8]	E2-1	Work, Instance-[6]	ProvidedCHO-2
In case the code of entities/classes is used, please refer to the previous Table 2-8 for the entities' full names. * The model enables the specialization of the relationship/property using values from a controlled vocabulary						

Overall, the results in this paragraph indicate that there are important differences between the models with regard to core constructs, namely core entities and bibliographic relationships. These different conceptualizations are due to models' differing views of the bibliographic universe. Inevitably, different conceptualizations result in differing implementations that may not be easy to interoperate with other ones.

2.5. Library linked datasets

Datasets include full collections of bibliographic and/or authority data. They combine the selection of models and/or metadata element sets, of value vocabularies, and of third-party datasets for outward linking (Coyle, 2010b). The datasets published in the library domain so far vary in scale from small projects to great national and international efforts. Small projects usually focus on the descriptive needs of special collections, while national library projects are wider in scope offering metadata infrastructure and practices that may facilitate other library linked data projects. Thus, this paragraph focuses on presenting national-scale linked data projects.

Some well-known examples of national library projects are those undertaken by the National Library of Sweden, the British Library, the National Library of Spain (Biblioteca Nacional de España - BNE), the National Library of France, and the National Library of Germany. For each project, two pieces of information are recorded: the model it implements, and the metadata element sets it uses. All projects have developed custom element sets to serve special needs. In the following paragraphs only the third-party metadata element sets are referred.

The National Library of Sweden started experimenting with linked data since 2007 (Malmsten, 2008). Its union catalog, LIBRIS, was using linked data technologies and the FRBR model. The metadata element sets used were Dublin Core, FOAF, and SKOS for value vocabularies (Malmsten, 2008). In 2018, the new LIBRIS version adopted the BIBFRAME 2.0 model (presented in 2.3.5) with some local extensions (National Library of Sweden, 2019). The new LIBRIS uses the BIBFRAME vocabulary, as well as elements from the schema.org, Dublin Core, PROV models for the preservation of provenance information, SKOS for value vocabularies, and MADS for authority data (National Library of Sweden, 2019). Due to the wide use of the FOAF and BIBO element sets, there is an effort to coordinate the new LIBRIS with these elements sets too.

The British Library published in 2011 one of the first bibliographic datasets, the British National Bibliography including descriptions of “books (including monographs published over time), serial publications and new and forthcoming books” (The British Library, 2019). The British Library does not use the FRBR model (Deliot, 2014), but has developed three data models for the description of published books, serials, and forthcoming books. The models use elements from different element sets, some of which are Dublin Core, BIBO, ISBD elements, FOAF, RDA elements, and SKOS.

The National Library of Spain (Biblioteca Nacional de España - BNE) published its first set of library linked data in December 2011 including both bibliographic and authority data. The BNE data are modeled using the FRBR model. The metadata element sets used are FRBRer elements, ISBD elements, RDA elements, Dublin Core and BIBO. The current version of the dataset includes “practically all the library's materials, including ancient and modern books, manuscripts, musical scores and recordings, video recordings, photographs, drawings and maps” (“datos.bne.es 2.0,” 2019).

The National Library of France (Bibliothèque nationale de France - BnF) also published its first linked dataset in 2011 using the FRBR model. The BnF linked data aggregate data from the main library catalog, the Gallica digital library, and the archives and manuscripts catalogs (Simon, Di Mascio, Michel, & Peyrard, 2014). The metadata element sets used are: Dublin Core, RDA elements, FOAF, SKOS (Lapôtre, 2017). Due to the inclusion of archival material, the Encoded Archival Description standard (Library of Congress, 2020) is used.

The National Library of Germany (Deutsche National Bibliothek - DNB) first published its authority data as linked data in 2010. Its bibliographic data were published afterwards. It must be noted that data are modeled according to an internal model developed by the DNB (Deutche National Bibliothek, 2016). Some of the element sets used are Dublin Core, BIBO, RDA elements, ISBD elements, FOAF, schema.org.

Europeana is a European digital library project for aggregating descriptions of digital or digitized resources from national libraries, museums, archives and other cultural heritage institutions. Europeana published its first dataset in 2012. For the publication of this dataset, existing data in Europeana described with a Dublin Core Metadata Element Set extension, known as Europeana Semantic Elements, were converted to Europeana's Data Model (EDM is described in paragraph 2.3.6) ("Europeana Pro: Linked Open Data," 2019; B. Haslhofer & Isaac, 2011). Among the external element sets used in Europeana are Dublin Core and SKOS.

The Library of Congress in USA has been among the early implementers of linked data technologies. Initially, the LC focused on converting legacy controlled vocabularies to linked data (McCallum, 2017), then on the development of a new conceptual model called BIBFRAME (it is presented in 2.3.5). In 2019, the LC announced the publication of its catalog data using the BIBFRAME 2.0 model. BIBFRAME *Works* and *Instances* can be searched using the id.loc.gov search mechanisms.

OCLC WorldCat data have been published as dataset in 2012 using a set of schema.org terms (schema.org is presented in 2.3.7.8). The publication of this dataset has been the latest of a series of linked data projects. The other linked data projects involved the publication of name and subject authority data, i.e., VIAF and FAST . The WorldCat dataset is based on the schema.org and it also uses elements from the Dublin Core and the MADS element sets (Godby et al., 2015).

In summary, this paragraph shows that different models have been used in well-known library linked data projects. Even in cases where the same model was used, namely the FRBR model was used by both National Library of Spain and National Library of France, totally different vocabularies, meaning element sets and value vocabularies, were used.

2.6. Semantic interoperability of library data

Semantic interoperability has been highlighted in the Library Linked Data Incubator Group report (Baker et al., 2011) acknowledging that lack of alignments between vocabularies and between metadata element sets may hinder reuse of data. After nearly 8 years from the LLD Incubator Group reports, there have been some linked data implementations in the library domain. Researchers studying them have revealed many different approaches in terms of modeling, of element sets and value vocabularies, and of technologies (Koster, 2012; H. Park & Kipp, 2019; Smith-Yoshimura, 2016; Suominen & Hyvönen, 2017); this is a fact that may hinder semantic interoperability and contribute to the development of linked data silos.

Two semantic interoperability initiatives regarding library conceptual models are presented: OCLC's exploration for compatibility with FRBR and BIBFRAME, and the development of the EDM-FRBRoo application profile. The presentation of the unfinished LODLAM Patterns project and of related studies by scholars comparing conceptual models follows.

2.6.1. OCLC's compatibility with FRBR and BIBFRAME

OCLC implements the schema.org model for the WorldCat catalog. Thus, only one class, the *schema:CreativeWork,* is exploited for representing intellectual content and products embodying it. OCLC may trigger the identification of FRBR WEMI entities by using "instances of *schema:CreativeWork* with different properties" (Godby et al., 2015). As an example, an instance of the class *schema:CreativeWork* having a property with language information or an instance of a subclass of the *schema:CreativeWork* class revealing content type (e.g., *Schema:Book, schema:Movie*) represents an FRBR *Expression.* FRBR *Manifestations* are represented as instances of two classes *schema:CreativeWork* and *schema:ProductModel* and they are having properties implying that the description is about a product (Coyle et al., 2017). FRBR *Items* are also represented as instances of two classes *schema:CreativeWork* and *schema:IndividualProduct.* They have properties implying that the description is about an object, e.g., an *Item* in a library collection (Coyle et al., 2017). The accommodation of FRBR semantics in the OCLC model is depicted in Figure 2:32.

By contrast to the flat representation of schema.org in Figure 2:31, the representation in Figure 2:32 adheres to bibliographic domain's conceptualizations and descriptions. It must be noted, though, that Schema.org cannot be used for full bibliographic descriptions due to the use of few generic properties for the representation of attributes and relationships that are relevant in the bibliographic domain. A typical example is the lack of specialized properties for the representation of bibliographic relationships. Despite all that, OCLC has adopted Schema.org for the representation of its WorldCat's data. This decision may possibly hinder future exploration of WorldCat data using bibliographic relationships as linking mechanisms.

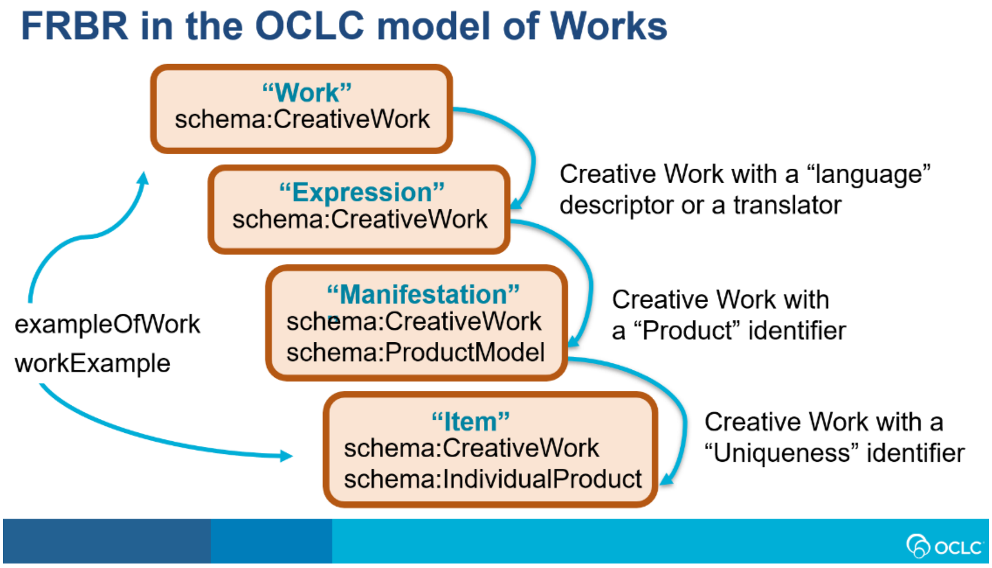


Figure 2:32. FRBR in the OCLC model of Works. Source: (Godby & Vizine-Goetz, 2017).

Soon after the publication of BIBFRAME 1.0 in 2012, two OCLC reports investigating the compatibility between BIBFRAME and OCLC's Model of Works were published (Godby, 2013; Godby & Denenberg, 2015). Both reports concluded that the two models are compatible. The BIBFRAME *bf:Work* is semantically close to the *schema:CreativeWork* class. Both classes refer to content, while content's realization(s) may be represented with use of specific properties. The *bf:Instance* class refers to material embodiments which are usually publication products. Publication products are represented in OCLC's Model of Works as instances of two classes: *schema:CreativeWork* and *schema:Product.* Material embodiments' exemplars are modeled using the *Item* class in BIBFRAME and an instance of two Schema.org classes, *schema:CreativeWork* and *schema:IndividualProduct.*

Due to the different scope of each model, BIBFRAME is to be used within libraries and other memory institutions, while Schema.org and OCLC's Model of Works are to be used for promoting the discovery of information resources by "general-purpose search engines" (Godby, 2013). This high-level alignment is depicted in Figure 2:33. In December 2018, the launch of a Bibframe2Schema.org Community Group was announced ("Bibframe2Schema.org Community Group," 2018). The objective of this group is to facilitate the mapping between the two models. In April 2020, a BIBFRAME to schema.org beta comparison viewer tool (https://bibframe2schema.org/compare) was announced, but no final reports have been published so far[2].

[2] The last checks on the Bibframe2Schema.org Community Group webpage (https://bibframe2schema.org/) and on the W3C bibframe2schema community webpage (https://www.w3.org/community/bibframe2schema/) were made on September 6, 2020. Both checks confirmed that no new report has been published.

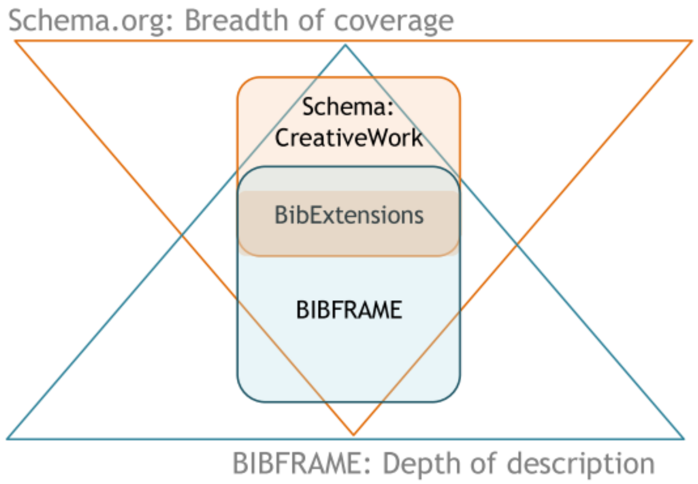


Figure 2:33. High-level alignment of SchemaBibEx and BIBFRAME. Source: (Godby, 2013).

2.6.2. Europeana Data Model alignments with library metadata and the EDM-FRBRoo application profile

Europeana experts investigated the alignment of the Europeana Data Model (EDM) with library metadata in 2011-2013. To align national library metadata with the EDM, three types of library materials were selected (monographs, multi-volume works, and serials) and a sample of metadata was studied by a Europeana's group of experts. Two reports were delivered, the "D5.1 Report on the alignment of library metadata with the Europeana Data Model (EDM) Version 2.0" (Angjeli, Bayerische, et al., 2012) that presented a modeling approach for describing the three selected materials and the "D5.2 Library domain metadata aligned with the Europeana Data Model Version 1.0" (Angjeli, Baumgartner, et al., 2012) that presented the validation of the model defined in D5.1 using the sample metadata.

Although the FRBR and the RDA were taken under consideration, the D5.1. alignment report adopted a records-based approach where different pieces of information about different entities are gathered in one description. It was recognized that the *edm:ProvidedCHO* class "implies nothing about the nature of the resource" (Angjeli, Bayerische, et al., 2012), and, hence, may represent any type of resource including all FRBR Group I entities (*Work, Expression, Manifestation, Item*). Moreover, the "edition" concept was introduced not as a new construct, e.g., a new EDM class, but as a new view of what the *edm:ProvidedCHO* class may represent. Thus, in this context, "edition" implies that an instance of the class *edm:ProvidedCHO* incorporates information regarding all three FRBR Group 1 entities, *Work, Expression,* and *Manifestation*, while the instances of the class *edm:WebResource* is to be used for the web exemplars (close to the FRBR *Item* entity) of the corresponding instances of the class *edm:ProvidedCHO*. The distinction between the WEMI entities "will lie only in the metadata used and in the relationships expressed" (Angjeli, Bayerische, et al., 2012). In the same report it was suggested that a further investigation was needed for the sake of granularity and the representation of the FRBR WEMI entities.

This provision induced the initiation of a new working group in July 2012. The scope of this Task Force group was to create an application profile that "allow a better representation of the FRBR group 1 entities" (Doerr et al., 2013). The group completed its tasks in April 2013 publishing the "EDM-FRBRoo Application Profile". The group used FRBRoo constructs and not FRBR ones. There were two reasons for this decision. First, FRBR as an entity-relationship model accommodates perceptions related to bibliographic records, whereas FRBRoo's object orientation is consistent with linked data design principles (Doerr et al., 2013). Secondly, FRBRoo further analyzes the WEMI entities and identifies more types of intellectual contribution as entities of their own special kind (Doerr et al., 2013). The Task Force group used three datasets, one per material type: monographs (*Don Quixote* case), plays (*Hamlet* case) and musical works (Brahms and Stravinsky cases). Each dataset was represented in FRBRoo and then the used FRBRoo classes and properties were mapped to EDM. This mapping developed the EDM-FRBRoo application profile consisting of two sets: 1) EDM classes and properties and 2) FRBRoo concepts represented as specializations of EDM superconcepts. As depicted in Figure 2:34, the EDM-FRBRoo Application Profile utilizes the EDM's extension mechanism, where the instances of the class *edm:InformationResource* are further defined using *edm:hasType* property instances and *skos:Concept* instances. The controlled vocabulary for the *skos:Concept* instances includes the concepts "FRBRWork", "FRBRExpression", and "FRBRPublicationExpression". The "FRBRPublicationExpression" concept originates from the FRBRoo model and was selected over FRBR *Manifestation* entity for the following reason. EDM selects digitized representations of real-objects and not the real-objects themselves. *Manifestations* represent real-world objects, whereas a "Publication Expression" represents the content, namely both creator's and publisher's intellectual efforts, embodied in the *Manifestation*.

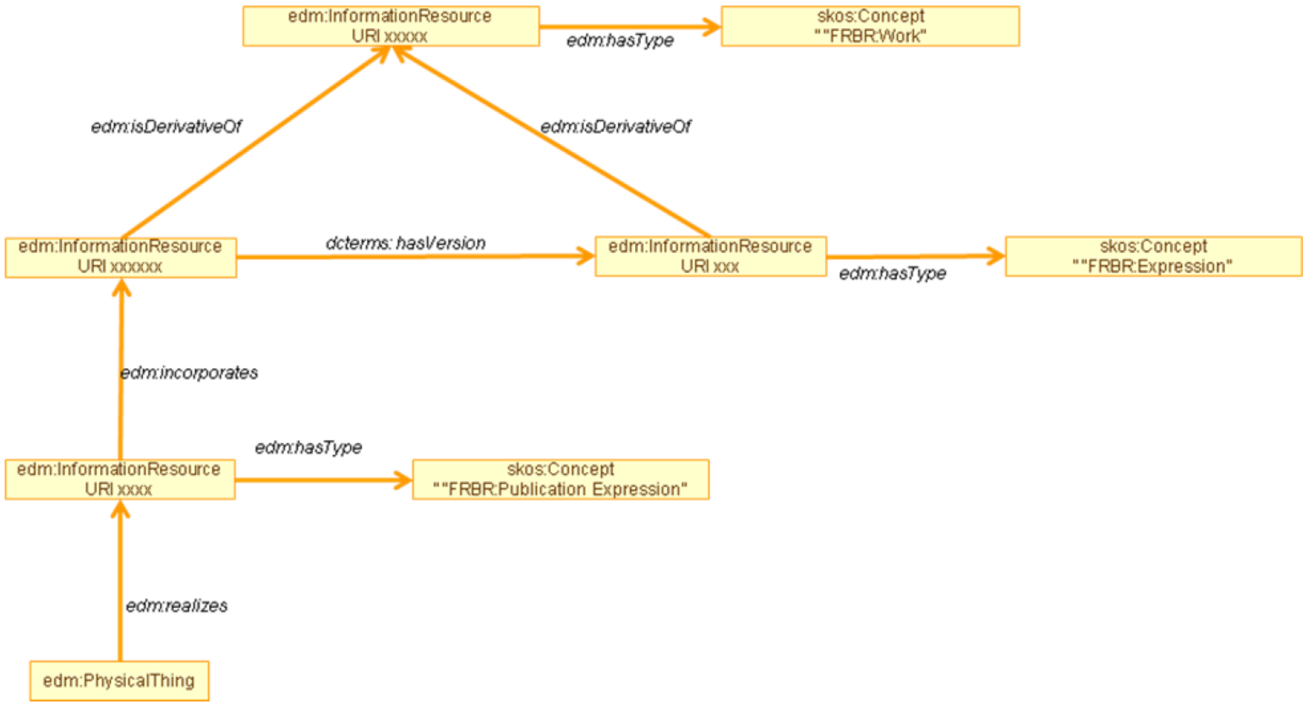


Figure 2:34. WEMI translated to EDM. Source: (Doerr et al., 2013).

2.6.3. LODLAM patterns - Linked Open Data Patterns for the Libraries, Archives, and Museums domain

One of the recommendations of W3C LLD Incubator Group Report for the standard bodies and organizations has been to "develop and disseminate best-practice design patterns tailored to library Linked Data" (Baker et al., 2011). The group acknowledged that expertise in library linked data should be shared to enable future library linked data projects. This recommendation was inspired by initiatives in the Linked Data domain, such as the Linked Data Patterns catalogue first published in 2010 (Dodds & Davis, 2012). This book shared patterns for the representation of common cases in linked data representations, such as the labelling of resources, or the use of ordered lists. The overview of the models in paragraph 2.3 has demonstrated that some models facilitate alternative representations without violating the models' semantics. Characteristics examples are the static and dynamic views in FRBRoo, the BIBFRAME alternative representations using the *bf:hasExpression* property in Figure 2:26, and the use of the *ore:Proxy* class in EDM (Figure 2:27).

Both, the recommendation by W3C LLD Incubator Group and the observation made in paragraph 2.3 regarding alternative representations, advocate for the discovery of the best or common representation patterns in the library linked data domain. This exactly had been the goal of the LODLAM project.

In 2013, Professor Richard Urban announced the “LODLAM Patterns” project (Urban, 2014). Inspired by the design patterns methodology implemented in software engineering and design communication, Urban launched the “LODLAM Patterns” project to identify patterns that could serve as “optimized solutions to common problems” (Urban, 2014). The first phase of the project would involve the identification of patterns and their storage in a representation pattern library that could serve as a tool for developers in the libraries-archives-museums domain. The second phase would involve the development of an ontology with concepts regarding the identified patterns.

The project aspired to discover the canon existing in representing LODLAM data. Among the aspirations of the project has been that the “LODLAM Patterns Library” would serve as a crosswalking tool to identify how a real-world description case or a common modeling issue would be represented by different metadata element sets or conceptual models. Initially, several element sets and models were included in the study, e.g., Dublin Core, MODS, EDM, BIBFRAME, etc., for the identification of patterns.

The sole publication about the project included one prototype pattern (the author used the term “proto-pattern”) as evidence of the method (Urban, 2013). The “proto-pattern” dealt with the *Surrogate identity* problem, that is the problem of differentiating between the metadata about an original resource and the metadata about a surrogate resource. The description of the proto-pattern used a specific structure, i.e., Problem, Context, Solution, Related Pattern, and Examples (Urban, 2013). The project received a little feedback and insufficient participation. The creation and the maintenance of patterns could not be guaranteed, thus, enforcing Professor Urban to pause the project.

The urge for a patterns catalog has been also highlighted in (Aalberg, Vennesland, & Farrokhnia, 2015). The authors focus on the implementation of the CIDOC-CRM model. Yet, they acknowledge a common trend among conceptual models’ documentation documents: they present a “sequential documentation of distinct classes and properties”. Even though, such sequential documentation is needed, implementers are mostly interested in “model fragments” to represent real-world cases. The authors suggest the creation of a pattern catalog and develop a prototype pattern catalog as a semantic wiki (named as “Ontology Pattern Semantic Wiki”). The “Ontology Pattern Semantic Wiki” is not available online and no newer reports have been published about it.

The number or the nature of the patterns included in either of the two projects are not known. Yet, the idea and the methodology of the two projects can be considered as promising for identifying the canon or the different patterns for representing common bibliographic description cases, such as a literal translation, a reproduction, a free translation, a dramatization of a novel, etc.

2.6.4. Studies by scholars

Semantic interoperability is not a straightforward problem; it relates to the common understanding of meaning which may involve varying issues, such as, modeling, standards, schemas, value vocabularies, mappings, multilingualism, tools, etc. (Zeng, 2019). There have been a number of studies regarding semantic interoperability between the bibliographic conceptual models. This paragraph excludes studies involving the conversion of legacy MARC data in FRBR, BIBFRAME or other models.

Reports and studies regarding existing library linked data projects have raised concerns regarding the proliferation of bibliographic models and vocabularies (Hillmann, Coyle, Phipps, & Dunsire, 2010; Jett, Cole, Page, & Downie, 2016; Lovins & Hillmann, 2017; Patrício et al., 2020), the interoperability of bibliographic models with Linked Data principles (Dunsire, 2012; H. Park & Kipp, 2019; Peponakis, 2016; Willer & Dunsire, 2013), and the interoperability between the bibliographic models themselves (Cagnazzo, 2017; Hallo et al., 2016; Nillson, 2010; H. Park & Kipp, 2019, 2015; Rasmussen Pennington & Cagnazzo, 2019; Smith-Yoshimura, 2016, 2018; Suominen & Hyvönen, 2017; Svensson, 2013; Tallerås, 2017, 2018). These studies revealed (a) the discordance of entity-relationship modeling adopted by the FRBR model to semantic web principles, (b) the lack of consensus among libraries regarding the linked datasets' underlying conceptual models, as well as (c) the abundance of element and value vocabularies for the same metadata. Such practices will likely impact semantic interoperability, which is already affected by 1) the plethora of conversion tools from MARC21 to FRBR, BIBFRAME or to other custom-made models generating bibliographic datasets with great differences between them (H. Park & Kipp, 2015), 2) the coincidental unavailability of tools for publishing linked data (Frosterus et al., 2020; Smith-Yoshimura, 2016, 2018; Taniguchi, 2017a; Ullah et al., 2018; Wahid, Warraich, & Tahira, 2018) causing the development of custom solutions and "locally developed routines" (Frosterus et al., 2020), and 3) the mixture of metadata elements (many of them are deprecated ones) used in library linked data projects to describe bibliographic resources (H. Park & Kipp, 2019). Toward the end of liberating library data from legacy silos, new models are developed and the danger of creating new linked data silos has already been expressed (H. Park & Kipp, 2019; Patrício et al., 2020; Suominen & Hyvönen, 2017; Tallerås, 2018).

Taniguchi focuses on the interoperability of well-known models, namely, FRBR, LRM, RDA, and BIBFRAME. He has compared RDA to the FRBR/FRAD models (Taniguchi, 2012) in terms of core entities and inherent relationships, as well as entities' attributes. Based on his findings, he proposed that, even though RDA adheres to the FRBR principles, it is a model of its own (Taniguchi, 2012). He also studied if BIBFRAME can be used to share bibliographic metadata created by various communities that use different conceptualizations, diverse vocabularies and constraints (Taniguchi, 2017b). He criticized the BIBFRAME's modeling decision of not formally specifying domain and range constraints proving that this policy results in the reuse of the same property with different classes. As an example, the same properties may be used with both *bf:Work* and *bf:Instance* classes; this fact results in the accommodation of most *bf:Work* properties by the *bf:Instance* class, and clearly "implies that Work and Instance classes are not exclusive in BIBFRAME" (Taniguchi, 2017a). Taniguchi used loose definitions about domain and range constraints and experimented with four definition methods. He proved that the number of properties used to describe an instance of the *bf:Work* or the *bf:Instance* class changes depending on the selected method. Lastly, Taniguchi questioned the applicability of BIBFRAME's main objective (Taniguchi, 2017b), that is "to accommodate any number of content models and specific implementations, but still enable data exchange between libraries" (Miller et al., 2012). The BIBFRAME policy regarding domain/range constraints is also referred in (Baker et al., 2014; J.-R. Park et al., 2020).

Taniguchi reexamined BIBFRAME from the RDA viewpoint trying to discover if BIBFRAME can be used for RDA metadata (Taniguchi, 2017a). He made important observations regarding classes, properties and mappings. Regarding classes, he observed that both models use more than one classes for the representation of bibliographic resources, even though the number and the semantics of classes are different. There is no corresponding RDA *Expression* class in BIBFRAME, which may challenge future mappings. Regarding properties, both those representing attributes of *bf:Work* and *bf:Instance* instances, and those representing bibliographic relationships, Taniguchi noticed that RDA is more granular than BIBFRAME and a possible mapping from BIBFRAME to RDA would be difficult due to many-to-one and many-to-many mappings. Moreover, he observed that BIBFRAME uses classes and literal values for denoting the role of an *bf:Agent,* whereas RDA uses properties which are considered as a more stable construct. Lastly, he criticized the plethora of MARC21 to BIBFRAME conversion tools resulting in different BIBFRAME representations that are likely difficult to merge in future processes (Taniguchi, 2017a). Contrary to BIBFRAME's flexible nature as an enabler for different representations, Taniguchi proposes that MARC21 data shall be converted to RDA using the RDA registry vocabularies, and, afterwards if needed, they may be converted to BIBFRAME. His most recent study explored the merging and mapping of LRM and BIBFRAME (Taniguchi, 2018). He identified several options for merging and mapping, all depending on different interpretations of the flexible BIBFRAME definitions.

Other surveys presenting or comparing bibliographic data models highlight the authors' views regarding semantic interoperability enablers. Baker, Coyle, and Petiya claim that interoperability may be achieved between different models' instances, as long as the models share vocabularies and some common views in terms of constraints (Baker et al., 2014), e.g., that core classes are disjoint. In this context, there are researchers suggesting a common agreement on what legacy data are needed for the future and an enrichment for this type of data to enable future conversions and interoperability (Alemu, Stevens, Ross, & Chandler, 2012; Bowen, 2010; Seeman & Goddard, 2014). In relation to enrichment of legacy data, scholars acknowledge the importance of controlled vocabularies and the use of URIs in legacy bibliographic records as precise and consistent tools for the semantic web (Hogan et al., 2012; Edward T. O'Neill & Žumer, 2014; Wallis, 2018; Zeng, 2019).

2.7. Gaps identified in the literature review

To service users' information needs is the mission of libraries. Nowadays, users are accustomed to online searching and to online materials; they consume online services daily. Semantic Web technologies and Linked Data offer the potential of advanced services based on meaning and linking of data. In this context, IFLA updated the user tasks that the Library Catalog must support (Galeffi et al., 2017). These are *find, identify, select, acquire* and *obtain,* and *navigate* and *explore.* To the support of these user tasks in the context of the Semantic Web, new models are needed. These new models may exploit the conceptualizations of the past, namely the entities of the bibliographic universe, as imagined by acclaimed scholars and explicitly described in the FRBR, and their relationships. The representation of bibliographic data according to these models' semantics will enable the integration of library data into the Semantic Web and will also provide relationships as linking mechanisms that facilitate the *navigate* and *explore* task.

The literature review has presented the abundance of bibliographic conceptual models (paragraph 2.3). It has also exhibited important differences between them regarding the numbers of bibliographic entities and bibliographic relationships (paragraph 2.4). The overview of well-known library linked datasets exhibited even more differentiations (paragraph 2.5). The aims of the presented library linked data projects differed and affected the decisions regarding the modeling of data. Some have used existing data models, such as the FRBR, while others have developed new ones. Even in cases where the same models were used, there were differences in terms of used metadata element sets and value vocabularies (Tallerås, 2017). In most of the cases including the conversion of MARC data, in-house conversion programs were mostly used generating really different datasets. Researchers have already highlighted the great diversity between existing library linked datasets in terms of data modeling, use of vocabularies, and tools (Duchateau et al., 2018; Frosterus et al., 2020; Hallo et al., 2016; H. Park & Kipp, 2019; Smith-Yoshimura, 2016, 2018; Suominen & Hyvönen, 2017; Tallerås, 2017). It is common sense that "at a minimum the lack of consensus can complicate interoperability" (Cole et al., 2013). Thus, even though linked data technologies are used, the meaning of the data in the datasets is not ensured. This is a semantic interoperability issue that needs to be resolved to avoid the development of library linked datasets that end up isolated and unused (Suominen & Hyvönen, 2017). The danger of linked data silos (Suominen & Hyvönen, 2017) is evident impacting the usage potential of the published datasets and most important hindering the vision of creating a unified bibliographic universe where users may seamlessly navigate and explore.

To enable a unified bibliographic universe, library data need to be integrated in it regardless of the conceptual model they implement. Yet, existing models present differences regarding bibliographic entities, granularity, or constructs they use to describe bibliographic information. Despite the differences, instances of the models need to provide links in the linked data environment. These links may be either links to international authority files, such as the VIAF, or bibliographic relationships between instances of entities. Relevant scholarly literature (paragraph 2.6.4) examines mostly core constructs, while the preservation of bibliographic relationships as linking mechanisms in the linked data environment has not been thoroughly studied yet. Early semantic interoperability projects focused on the compatibility of non-library models, namely schema.org and EDM, to library ones (paragraphs 2.6.1 & 2.6.2). Moreover, mappings between bibliographic data models have not been developed yet. Within this context, the existing literature presents a gap in studying the semantic interoperability between the bibliographic models used in publishing library linked data. The W3C LLD Incubator Group recommendation for "semantic alignment" between metadata element sets and for "more explicit conceptual connections" (Baker et al., 2011) has not been fulfilled as yet. Toward the goal of semantic interoperability and mappings, there is a need to compare the bibliographic conceptual models to discover similarities and divergences in terms of modeling, granularity, constructs, and linking mechanisms.

3. Study of representation of real-world bibliographic description cases [3]

To examine if semantic interoperability between the bibliographic conceptual models is achievable, the thesis proceeds with the investigation of how common bibliographic description cases are represented. The models selected for this investigation are the ones presenting granularity and being developed by reputable organizations, namely FRBR and its consolidation LRM developed by IFLA, RDA developed by the RDA Joint Steering Committee, FRBRoo developed by CIDOC and endorsed by IFLA, BIBRAME developed by the Library of Congress, and EDM developed by Europeana.

All these models offer constructs for the representation of real-world bibliographic description cases. The investigation starts from the simplest case, that of a single-volume monograph, and progresses to other common, but more demanding cases that involve bibliographic relationships, such as derivation and aggregates. The selection of these cases has been made based on studies made by researchers regarding the number of instances of each bibliographic relationship in legacy library catalogs.

The most common bibliographic description case is a single-volume monograph published once. Bennett, Lavoie, and O'Neill in (Bennett et al., 2003) defined this case as "elemental works" and found that more than three fourths (78%) of the WorldCat are "elemental works", followed by the 16% of another simple bibliographic description case, that of "simple works" with multiple publications of the same content.

Tillett in her study found out that books mostly exhibit whole-part, derivative, and equivalence relationships (Tillett, 1987). Regarding whole-part relationships, the aggregates case has been selected. They were wrongfully considered as a type of whole-part relationships and there was a long-time dispute regarding their nature (IFLA FRBR Working group on Aggregates, 2009; O'Neill et al., 2011). Differing views about their nature are still recorded in the literature (Coyle, 2016; Coyle et al., 2017; Fritz, 2016a, 2016b; Neill et al., 2015; Taniguchi, 2013; Wiesenmüller, 2012). Other reasons defending their inclusion are the following: First, aggregates are common; according to Benett, et. all (Bennett et al., 2003), 12% of WorldCat records describe aggregate manifestations, while a later study by Žumer and O'Neill revealed that one fifth (21,2%) of a sample with 1000 WorldCat records are aggregates (Neill et al., 2015). Moreover, aggregates are often difficult to identify in flat MARC records and in legacy library catalogs (Aalberg & Žumer, 2013; Hickey & O'Neill, 2009; Neill et al., 2015; Žumer & O'Neill, 2012). Their representation using the constructs of granular models may be proved challenging.

Studies have provided evidence affirming that there are significant proportions of derivative works in catalogs. Tillett examined notes in the Library of Congress catalog to find that the 16,4% of records of her sample contained derivative works (Tillett, 1987). This percentage is slightly smaller, 14,3%, when certain MARC fields are taken under consideration. Smiraglia discovered that nearly half (49,9%) of the works in his Georgetown University sample were derivative works (Smiraglia, 1992). Other studies performed by Smiraglia in other catalogs identified different percentages: 30,2% in (Smiraglia & Leazer, 1999), and 50-66% of theological works in (Smiraglia, 1999). Velucci in her study regarding musical works identified the largest proportion compared to other scholars (Vellucci, 1995). She found that the 85,4% of her sample, musical scores from the Sibley Music Library, exhibited the derivation relationship. A more recent study by Petek using a sample from the COBIB Slovenian Catalog found that a quarter (25,75%) of the works in COBIB are derivative ones (Petek, 2007).

[3] This chapter revisits and expands the study published in this paper: Zapounidou, S., Sfakakis, M., & Papatheodorou, C. (2017). Representing and integrating bibliographic information into the semantic web: A comparison of four conceptual models. *Journal of Information Science, 43*(4), 525-553. doi:10.1177/0165551516650410. The paper's study was performed in 2015-2016 and since then all models have been updated with significant changes. In this chapter the IFLA LRM and RDA models are also studied; these two models were not included in the published article.

Equivalence relationship has been the third one mostly exhibited in books, according to Tillett's study (Tillett, 1987). Tillett's thesis was published in 1987 and since then libraries have undertaken a wide variety of digitization projects (Coyle, 2006; Hahn, 2006; Lopatin, 2006; Powell, 2008; Spellman & Holley, 2011) and have made efforts to integrate digitized material into OPACs (Calhoun, 2006; Library of Congress Network Development and MARC Standards Office, 2002; Reed-Scott, 1999).

To *navigate* and to *explore* (Galeffi et al., 2017) have been identified as key user tasks that bibliographic data and catalogs need to support. Bibliographic relationships may serve as a linking mechanism between bibliographic resources in a linked data environment. An extra linking and collocation mechanism could be provided with the representation of bibliographic families.

The study of the representation of all these cases, namely, single-volume monographs, bibliographic relationships (derivative, equivalence, and aggregates), and bibliographic families, has identified some important findings regarding similarities and divergences between the models.

3.1. Representation of real bibliographic description cases

The first case is a single-volume monograph and then the study proceeds to common real bibliographic cases, such as translation (derivative relationship), adaptation (derivative relationship), reproduction (equivalence relationship), and aggregates.

At this point the concept of path is defined, which is frequently used in this thesis: A path is defined by a sequence of the triples "domain class – property – range class" (Kondylakis, Doerr, & Plexousakis, 2006). The name of the entity/class is depicted with a colored rectangle and below it in a white-filled rectangle the corresponding instance of the entity/class is written. Relationships are depicted as arrows with labels. The direction of the arrows is from the domain class/entity to the range class/entity.

3.1.1. Single-volume monographs, elemental and simple works

To study the representation of "elemental and simple works" in each model, a Miguel de Cervantes *Work* has been selected. The *Work* "El ingenioso hidalgo don Quixote de La Mancha" by Miguel de Cervantes is the first part of the *Work* known nowadays as "Don Quixote". Don Quixote is one of the most popular novels ("The Library 100," 2019) and, even though its first part is not an "elemental work", it could be used for studying the representation of simple bibliographic description cases, i.e., elemental and simple works, as well as for more complicated bibliographic description cases, e.g., translation, adaptation, aggregates, etc. Elemental works may be successfully represented in all models studied in this thesis. For readability reasons, the example of "El ingenioso hidalgo don Quixote de La Mancha" is referred as "First part" in the figures of this paragraph.

In **FRBR**, Cervantes' set of ideas specifying "El ingenioso hidalgo don Quixote de La Mancha" as a distinct intellectual creation is represented by a *Work* entity instance (Figure 3:1). The entirety of the *Work* entity ideas *is realized through* an *Expression* entity instance representing the text in Spanish. The publication of El ingenioso hidalgo don Quixote de la Mancha in 1605 designates a material embodiment (*Manifestation* entity) of the signs used in the Spanish language *Expression*. The *Manifestation* entity instance represents all the physical objects that bear the same content and publication's physical characteristics. An exemplar of this publication, a specific copy held in a library is represented with an *Item* entity instance. The **LRM** representation is the same to the FRBR one using LRM classes for FRBR entities and LRM properties for FRBR relationships (Figure 3:2).

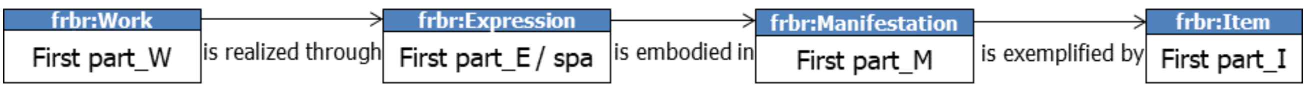


Figure 3:1. FRBR representation of a single-volume monograph, the Don Quixote's First Part entitled "El ingenioso hidalgo don Quixote de La Mancha".

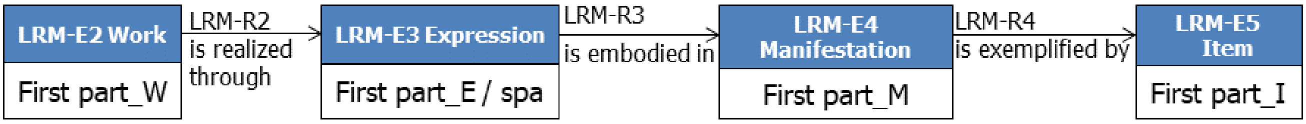


Figure 3:2. LRM representation of a single-volume monograph, the Don Quixote's First Part entitled "El ingenioso hidalgo don Quixote de La Mancha".

The **RDA** representation of the First Part is similar to the FRBR representation (*Figure 3:3*). FRBR entities are represented as classes, and inherent relationships are represented as properties. RDA enables the representation of an additional inherent relationship, the *rdaw:P10072 has manifestation of work,* that may relate an *rdac:C1001 Work* instance with its physical embodiment(s) represented as *rdac:1007 Manifestation* class instances.

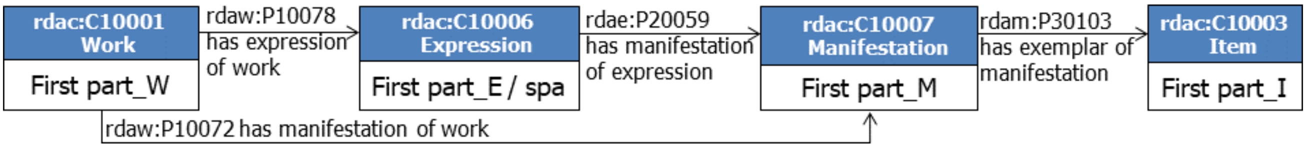


Figure 3:3. RDA representation of a single-volume monograph, the Don Quixote's First Part entitled "El ingenioso hidalgo don Quixote de La Mancha".

FRBR, its consolidation IFLA-LRM, and RDA all identify and represent the same distinct moments in the creation timeline of the *Work* example, namely, the set of Cervantes' ideas that form the *Work* with the title "El ingenioso hidalgo don Quixote de la Mancha" (*Work* entity); the original Spanish text (*Expression* entity); the publication (*Manifestation* entity); and an exemplification of that publication held at a library (*Item* entity).

The **object-oriented FRBR** (FRBRoo) enables the representation of more distinct moments in the creation timeline. Moreover, it provides more than one representation approaches, namely, a static view which is similar to FRBR and a dynamic view that also represents 'temporal entities', e.g., events, activities, to associate the entities represented in the static view to time-spans, locations and agents. Even though the FRBRoo static view is close to the FRBR, FRBRoo uses specialized classes that extend the FRBR entities' semantics (*Figure 3:4*). The First Part is one of Cervantes' distinct intellectual creations without having other works as parts. It is represented as an instance of the *F14 Individual Work*, which is a subclass of the class *F1 Work*. This *F14 Individual Work* instance *R9 is realised in* one *F22 Self-Contained Expression* which is incorporated (*P165 incorporates*) in an *F24 Publication Expression* instance. The *F24 Publication Expression* class represents the publisher's contribution incorporating both Cervantes' signs (represented in the *F22 Self-Contained Expression* instance) and the ones (textual or visual) used by the publisher for the production of the publication. The publication as a product is represented by an *F3 Manifestation Product Type* class instance that should carry the *F24 Publication Expression* with all the *F22 Self-Contained Expressions* that it may incorporate. An exemplar of this *F3 Manifestation Product Type* is represented with an *F5 Item* instance. As stated in the example (*Figure 3:4*), FRBRoo differentiates between the intellectual creation and the publication process; *F22 Self-Contained Expressions* are part of the intellectual creation and during the production process may be incorporated with other *F22 Self-Contained Expressions* into an *F24 Publication Expression.* A second differentiation between FRBR and FRBRoo is that the *F5 Item* class is explicitly stated as the actual carrier of the *F24 Publication Expression*.

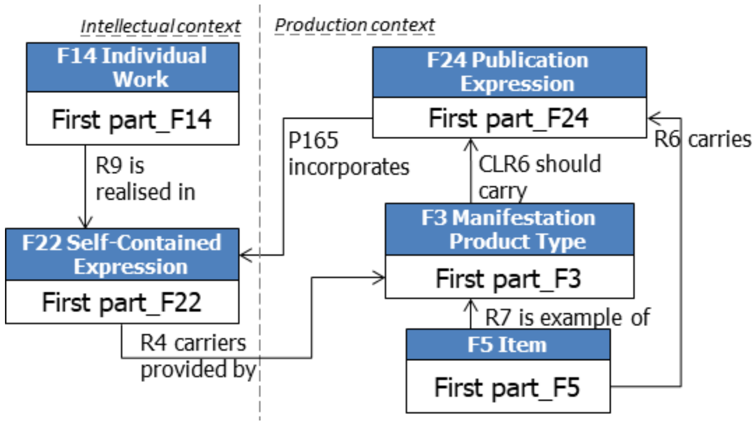


Figure 3:4. Static FRBRoo representation of a single-volume monograph, the Don Quixote's First Part entitled "El ingenioso hidalgo don Quixote de La Mancha".

In the FRBRoo dynamic view, entities from the static FRBRoo representation (*Figure 3:4*) may be represented along with associated 'temporal entities'. As presented in *Figure 3:5*, the dynamic view adds four events (depicted with different shapes) related to the conception of the ideas of the *Work* (*F27 Work Conception*), the creation of the signs (*F28 Expression Creation*), the creation of the publication signs (*F24 Publication Expression*), and the production of the publication products (*F32 Carrier Production Event*). It must be noted that the shapes used for the events consist of the upper part representing the class and the lower part representing not the instance of the event class, but the *E39 Agent* who carried out the event (*P14 carried out by/performed*). For readability reasons, the complete paths regarding agents are not depicted. An example regarding the agent participating in the *F27 Work Conception* event is presented in *Figure 3:6*. The exact role of the *F10 Person* instance in the *F27 Work Conception* event may be represented by using a literal value in the following path (*P14.1 in the role of - E55 Type="Creator"*). FRBRoo does not provide a full list of values for roles, nor does it refer to any related controlled value vocabulary.

Focusing on the First Part example, *Figure 3:5* represents that the *First Part_F14* instance was conceived in an *F27 Work Conception* event carried out by a *F10 Person* with the name Cervantes. The *First Part_F22* instance was created in an *F28 Expression Creation* event, also carried out by Cervantes at a particular location and time. Another *F10 Person* named Francisco de Robles has carried out an *F30 Publication Event* that created an instance of the First Part_F24 instance, which incorporates the *First Part_F22.* The production of the publication product and all of its exemplars occurred in an *F32 Carrier Production Event* instance carried out by an *F10 Person* named Juan de la Cuesta. Thus, dynamic FRBRoo representations may result to rich descriptions of 1) creation and production processes, 2) products of these processes (abstract or material), and 3) agents involved in these processes.

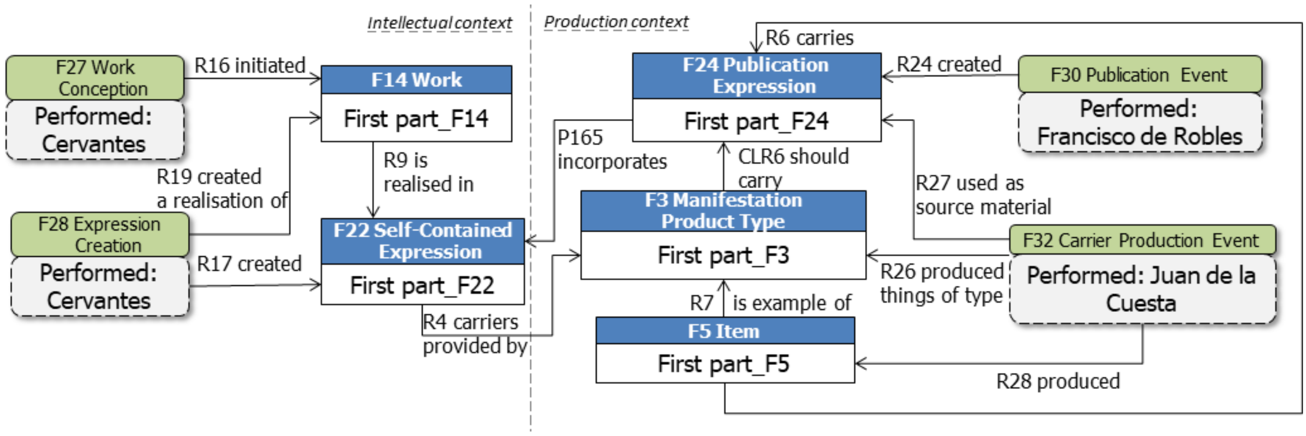


Figure 3:5. Dynamic FRBRoo representation of a single-volume monograph, the Don Quixote's First Part entitled "El ingenioso hidalgo don Quixote de La Mancha".

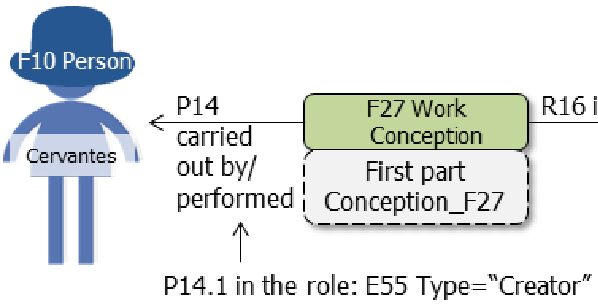


Figure 3:6. Dynamic FRBRoo representation of Cervantes' conception of the Don Quixote's First Part entitled "El ingenioso hidalgo don Quixote de La Mancha". The exact role of Cervantes is represented by typing the P14 carried out by / performed property.

In the **BIBFRAME 2.0** representation depicted in *Figure 3:7*, both the ideas of the "El ingenioso hidalgo don Quixote de La Mancha" and the signs realizing the ideas are represented by the class *bf:Text*, which is a subclass of the class *bf:Work.* With the *bf:Text* class the type of signs used for the realization of the ideas is also explicitly stated. The publication product is represented with a *bf:Instance* subclass, *bf:Print.* Using the *bf:Print* class the exact carrier type used for the publication product is explicitly stated. The *bf:Text* class instance has as its instance (*bf:hasInstance*) the *bf:Print* class instance. An exemplification of the *bf:Print* instance is represented with the *bf:Item* instance.

Figure 3:7. BIBFRAME representation of a single-volume monograph, the Don Quixote's First Part entitled "El ingenioso hidalgo don Quixote de La Mancha".

BIBFRAME provides flexible definitions that may cause ambiguity. As an example, BIBFRAME provides the *bf:hasExpression* property to either relate different *bf:Work* containing different expressions of the same ideas, or to enable an FRBR-like representation. In the latter case, depicted in *Figure 3:8*, the first *bf:Work* instance includes -similarly to the FRBR *Work*- only the ideas and ignores the signs realizing it. We define this case an *Expression*-agnostic *bf:Work*.

BIBFRAME does not clarify if the *bf:Work* including the signs (the *bf:Text* instance in *Figure 3:8*), includes just the signs -similarly to the FRBR *Expression* entity- or keeps both ideas and signs according to the *bf:Work* definition. Most likely, the *bf:Text* instance includes both ideas and signs. Thus, even though the representation in *Figure 3:8* is similar to the FRBR one in *Figure 3:1*, there is a significant difference between information described by the FRBR *Work* and *Expression* entities and the two *bf:Work* instances (*bf:Work* and *bf:Text*).

Figure 3:8. Alternative BIBFRAME representation of a single-volume monograph, the Don Quixote's First Part entitled "El ingenioso hidalgo don Quixote de La Mancha". The first bf:Work instance is an Expression-agnostic one.

EDM enables the representation of born-digital or digitized European Cultural Heritage Objects available online. Therefore, in the following EDM representation (*Figure 3:9*), a digitization of the First Part example is described. The First Part is represented as an *edm:ProvidedCHO* class instance incorporating in a rather flat approach all information regarding ideas, signs, and publication. A digital exemplar of the First Part_CHO is represented with an *edm:WebResource* class instance. An *ore:Aggregation* class instance is used to group the description of the cultural heritage object (First part_CHO) and its digital exemplar (First part_URL). In the *Figure 3:9*, the *ore:Aggregation* instance is identified by using a combination of the institution providing the *edm:ProvidedCHO* instance (XPROV) and the unique identifier of the *edm:ProvidedCHO* instance in the provider's system (7791).

Figure 3:9. EDM representation of a single-volume monograph, the Don Quixote's First Part entitled "El ingenioso hidalgo don Quixote de La Mancha".

An alternative EDM representation could use the *ore:Proxy* class to preserve provenance and contextual information regarding the First part_CHO. There can exist many *ore:Proxy* instances, each one of which may preserve its provider's metadata about the same *edm:ProvidedCHO* (the First part_CHO in the example). Use of the *ore:Proxy* class will enable the use of one *edm:ProvidedCHO* class instance for each European Cultural Heritage Object and the contextualization of multiple descriptions made by different providers. In Figure 3:10, there are two *edm:WebResource* instances provided for the same First part_CHO by two different providers. Each provider's metadata is preserved in an *ore:Proxy* class instance.

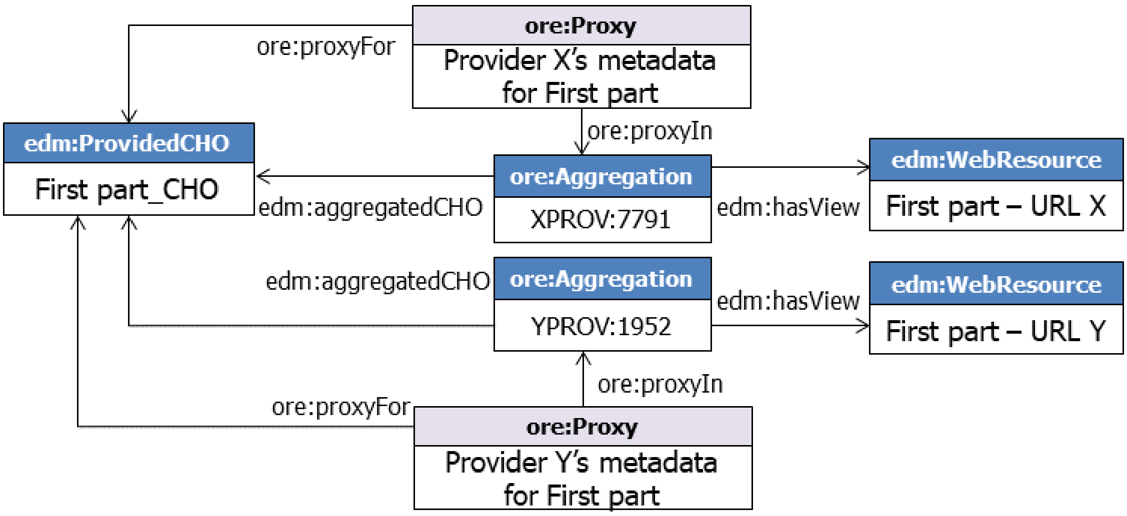


Figure 3:10. EDM representation using ore:Proxy class instances to accommodate providers' metadata regarding the same edm:ProvidedCHO instance, a single-volume monograph, the Don Quixote's First Part entitled "El ingenioso hidalgo don Quixote de La Mancha".

A second alternative representation in EDM may use the EDM-FRBRoo application profile. The EDM-FRBRoo application profile may provide additional information on the occasion of an *edm:ProvidedCHO* (*Figure 3:11*). In this representation, the *edm:ProvidedCHO* instance may describe either the publication information (similar to the FRBR *Manifestation* entity), or as in typical EDM representations (see *Figure 3:9*) all information regarding ideas, signs, and publication. Either way, ideas and signs may be explicitly represented using typed *edm:InformationResource* instances. The *edm:InformationResource* instance representing the ideas in the El ingenioso hidalgo don Quixote de La Mancha single-volume monograph is typed (*edm:hasType*) as a *skos:Concept* instance with the literal value “FRBRWork”. The *edm:InformationResource* instance representing the signs is typed (*edm:hasType*) as a *skos:Concept* instance with the literal value “FRBRExpression”. The inherent relationship between the ideas and the signs realizing them is represented with an instance of the *edm:isDerivativeOf* property relating the two typed *edm:InformationResource* instances.

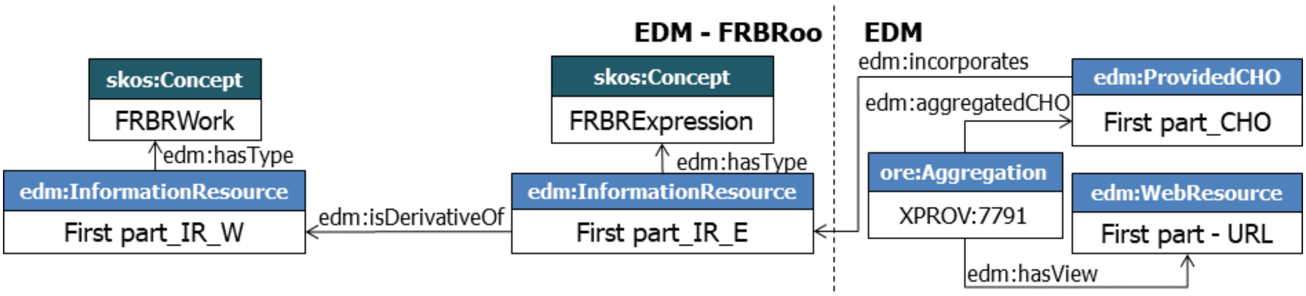


Figure 3:11. EDM representation using the EDM-FRBRoo application profile to represent abstract FRBR entities on the occasion of the edm:ProvidedCHO instance describing a single-volume monograph, the Don Quixote’s First Part entitled “El ingenioso hidalgo don Quixote de La Mancha”.

3.1.2. Bibliographic relationships

In the linked data cloud, relationships are a key linking mechanism. In the bibliographic universe, bibliographic relationships provide exploration opportunities by representing how one *Work* has influenced another or how it has evolved over time through its editions, translations, transformations, etc. Such *instantiation networks* are really common (Smiraglia, 2005), as presented in the literature review (paragraph 2.1.2.1) and in the beginning of this chapter. The representation of three bibliographic relationships is studied: derivative, equivalence, and aggregates. It must be noted that many models enable alternative representations using the models’ constructs. Some of these alternative representations are also included in this section’s investigation.

3.1.2.1. Derivative relationships

The Don Quixote example is used once more. Even though Cervantes wrote two parts, they were later published together forming an aggregate. Nowadays, Don Quixote is considered as one *Work,* and this is how it is represented in the paragraph’s figures. Don Quixote is among the most popular novels (“The Library 100,” 2019) having a really great bibliographic family or instantiation network with many translations, adaptations and transformations. Don Quixote has been translated as a whole, and its parts have also their own derivations published separately. For the study of derivative relationships two translations have been selected. The first is a free translation/adaptation, where the French translator has changed the ending to later publish his own sequel. The French translation by Filleau de Saint-Martin was really popular in its time having its own derivations (“Cervantes project: Cervantes collection,” n.d.). John Phillips translated the Filleau de Saint-Martin’s French text in English offering to the Don Quixote’s publishing history the worst English translation ever (Stavans, 2008). The Filleau de Saint-Martin translation is denoted as “French translation” in the figures, while John Phillips translation is denoted as “English translation”.

In **FRBR**, the Filleau de Saint-Martin free French translation entitled “Histoire de l'admirable don Quichotte de la Manche” constitutes a new *Work* adapting the original “Don Quixote”. The John Phillips English translation is a new realization of the adaptation *Work* and not of the original one. Therefore, in *Figure 3:12* the original “Don Quixote” *Work* is related with the *has adaptation* relationship to the French translation *Work.* The French translation *Work* is realized into two *Expressions*, the French one and the English one, which are related with a *has a translation* relationship.

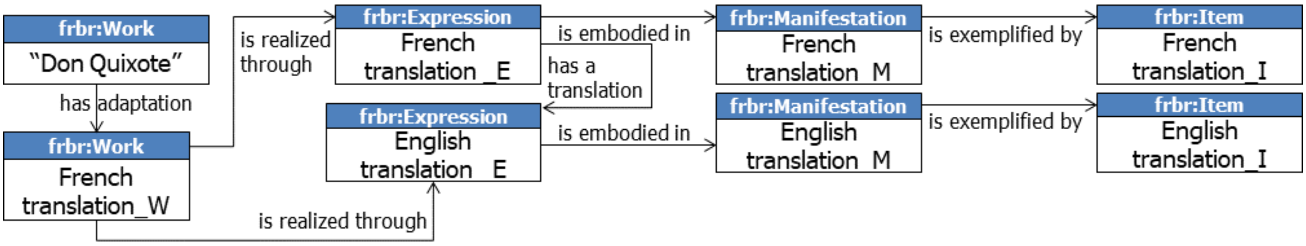


Figure 3:12. FRBR representation of derivative (translation-adaptation) relationships of Don Quixote.

At this point, it must be noted that the FRBR represents literal translations as *Expressions* of the same *Work.* In this realization-based approach, a translation relationship may be represented only when both original and derivative signs are known. In case the original signs used for the translation are not known, then the relationship is not represented, but the derivative *Expression* containing the translation remains as a realization of its *Work.* As an example using the *Figure 3:12*, if it was not known that the Philips *English translation_E* used the *French translation_E*, then the Philips *English translation_E* would remain as realization of the *French translation_W* (using the *is realized through* relationship), but the translation relationship (*has a translation*) would not be represented. This realization-based approach for the representation of literal translations is common among librarians and is also enabled by the LRM and RDA models that follow.

The **LRM** representation (Figure 3:13) is similar to the FRBR one. There is a difference regarding relationships. LRM identifies less relationships than FRBR. FRBR relationships have been either merged or generalized. As an example, four *Work* to *Work* derivative relationships have been merged into the *LRM-R22 is a transformation of* property; namely, *is derivation of, is adaptation of, is transformation of,* and *is an imitation of.* Thus, for the case of adaptation, the *LRM-R22 is a transformation of* property is used to relate the two *LRM-E2 Works,* where the one (French translation_W) is the free translation of the other (Don Quixote). The translation relationship between the *LRM-E3 Expressions* is represented with an instance of the *LRM-R24 is derivation of* property typed as “translation”. The exact nature of the derivative relationship is represented by “typing” the *LRM-R24 is derivation of* property. Yet, a controlled vocabulary with terms that can be used for typing the *LRM-R24 is derivation of* property has not been created in the context of LRM, nor the LRM proposes the use of a related third-party vocabulary. Use of free text for typing the *LRM-R24* property will probably cause inconsistencies in LRM instances. Even though, the *LRM-R24* property can be typed to represent more specialized derivative relationships, the *LRM-R22 is a transformation of* cannot be sub-typed. There is no reference neither in the official LRM document (Riva, Bœuf, et al., 2017a), nor in the mappings document (Riva, Bœuf, et al., 2017b) regarding the use of sub-types for the *LRM-R22* property.

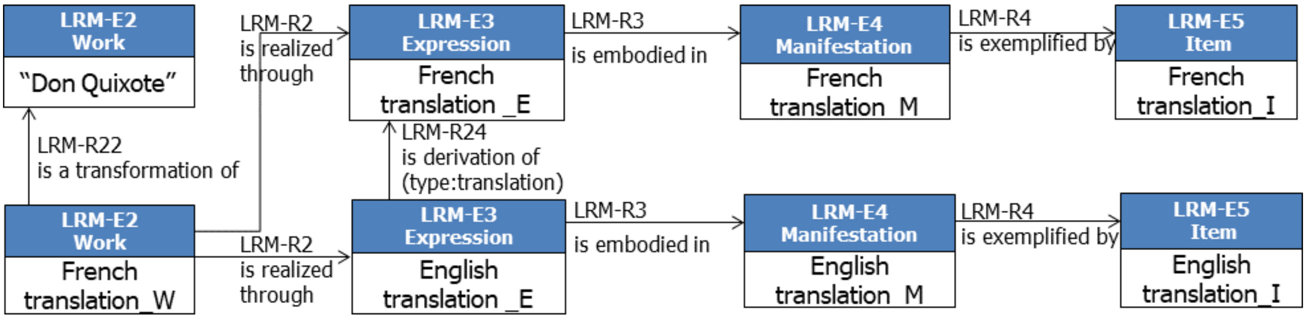


Figure 3:13. LRM representation of derivative (translation-adaptation) relationships of Don Quixote using the realization approach.

Even though all examples in the official LRM document (Riva, Bœuf, et al., 2017a) follow the realization approach, it is clearly stated that the LRM model may accommodate the derivation approach too. According to the derivation approach, which is common in the publishing industry, each translation is not a new realization (*Expression*), but a new *Work.* This alternative representation is depicted in *Figure 3:14*. Each translation is considered as new *LRM-E2 Work.* The translation relationship is represented with the *LRM-R24 is derivation of* property instance typed as translation. The *LRM-R24 is derivation of* property relates the *LRM-E3 Expression* instances of the two translation *LRM-E2 Works*, namely the French translation and the English translation *Expressions.*

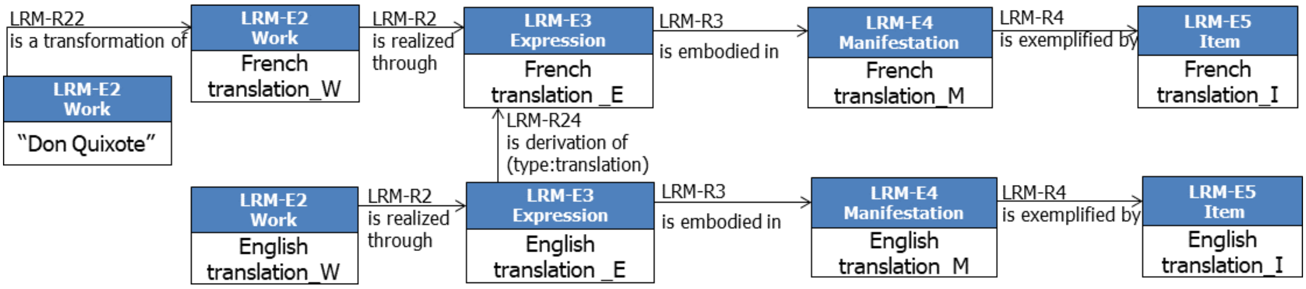


Figure 3:14. LRM representation of derivative (translation-adaptation) relationships of Don Quixote using the derivation approach.

The **RDA** representation (*Figure 3:15*) adheres to the realization approach and is similar to the FRBR representation in *Figure 3:12*. RDA has refined the relationships identified in FRBR and identifies in an analytic way the exact nature of the relationship. Free translation is not represented as a generic adaptation, but the exact nature of adaptation is represented using the *rdaw:P10099 is freely translated as work* property.

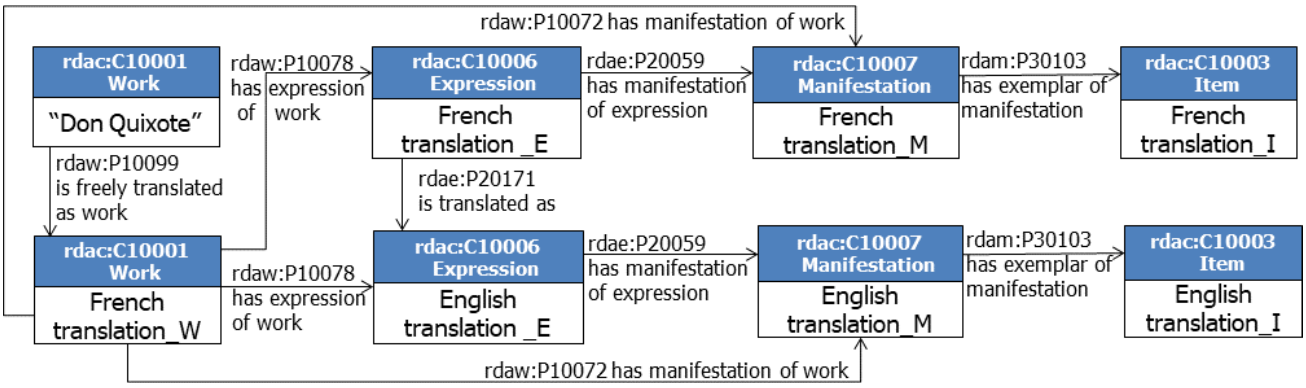


Figure 3:15. RDA representation of derivative (translation-adaptation) relationships of Don Quixote.

FRBRoo refines FRBR entities, and, thus, different work-related classes are used in the representation (*Figure 3:16*). Don Quixote is represented with a general *F15 Complex Work* instance class instance representing the dominance of Cervantes' concepts about the adventures of Don Quixote. An *F15 Complex Work* instance may have other *F15 Complex Works* or *F14 Individual Works* as members provided that they are dominated also by the same initial concepts (Bekiari, Doerr, Le Bœuf, & Riva, 2015). The “French translation” is also represented as an *F15 Complex Work* instance related to the *F15* Don Quixote instance with an instance of the *R2 is derivative of* property typed as adaptation. The “French translation” *F15 Complex Work* instance has two members, an *F14 Individual Work* instance representing the complete set of Filleau de Saint-Martin’s ideas for the “Histoire de l'admirable don Quichotte de la Manche”, and another *F14 Individual Work* instance representing the complete set of John Philip’s ideas. The *F14 Individual Work* instance representing the English translation is related to the French translation *F14 Individual Work* with the *R2 is derivative of* property typed as translation. There are two important issues to be noted here. First, all derivative relationships are represented between *F1 Works* with the *R2 is derivative of* property. The exact nature of the derivative relationship is represented by “typing” the *R2 is derivative of* property. The *R2 is derivative of* property may be “typed” with the *R2.1 has type* property. For consistency, the *R2.1 has type* property may have as values only the following ones: “Abridgement”, “Adaptation”, “Arrangement”, “Imitation”, “Revision”, “Summary”, “Transformation”, “Translation”. Secondly, the translation relationship between the two *F22 Self-Contained Expressions* (*Figure 3:16*) is implied through the representation of the relationship between the *F14 Individual Work* instances that the two *F22 Self-Contained Expression* instances realize.

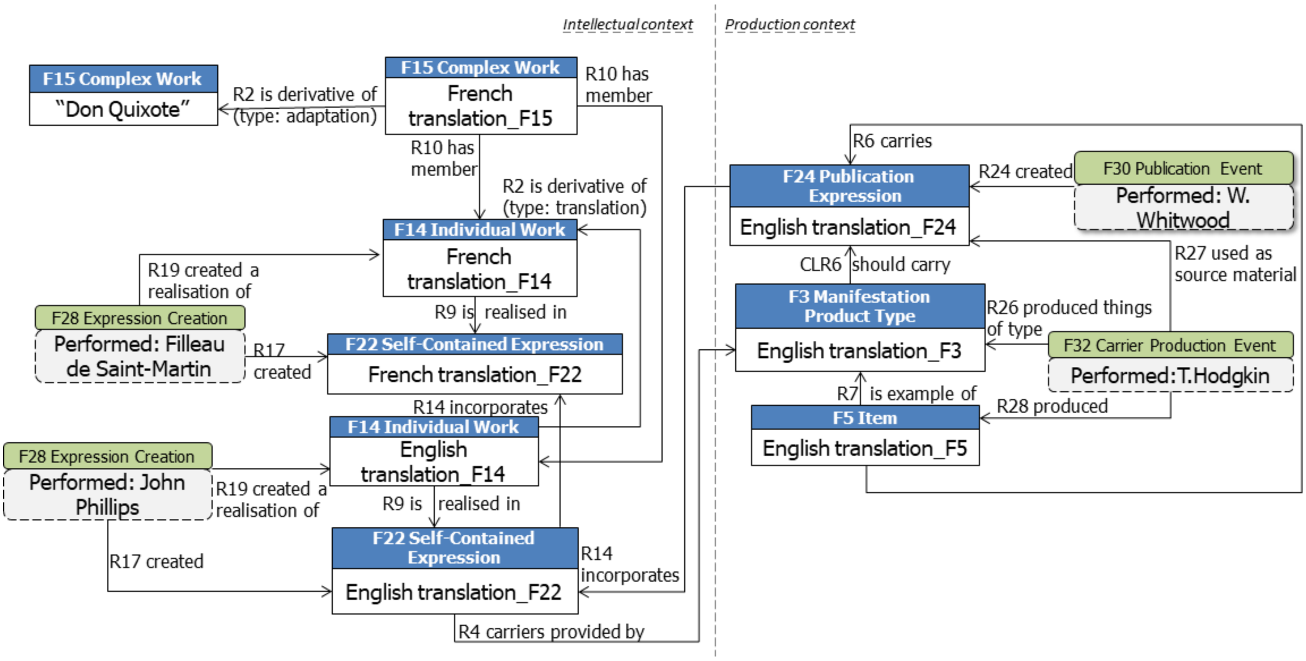


Figure 3:16. FRBRoo representation of derivative (translation, adaptation) relationships by specializing the R2 is derivative of property with the respective property type. The derivation approach is used.

Figure 3:16 presents a derivation approach in representing translations as new *Works*. FRBRoo may also enable the representation of the realization approach using its constructs. The alternative realization approach treats literal translations as new realizations of the same *Work* (*Figure 3:17*). The free “French translation” is not treated by the realization approach, since free translations are actually adaptations that change both content and signs resulting in the creation of new *Works.* Both “Don Quixote” and “French translation” are represented as *F1 Work* instances related to one another with an *R2 is derivative of (type:adaptation)* property instance. The “French translation” *F1 Work is realized in* two *F22 Self-Contained Expression* instances representing the French and English texts. These two *F22 Self-Contained Expression* instances are related with the *R14 incorporates* property. Therefore, even though the translation relationship is represented between two realizations of the same *Work,* the specificity of the relationship is lost and a more generic property (*R14 incorporates*) is used instead.

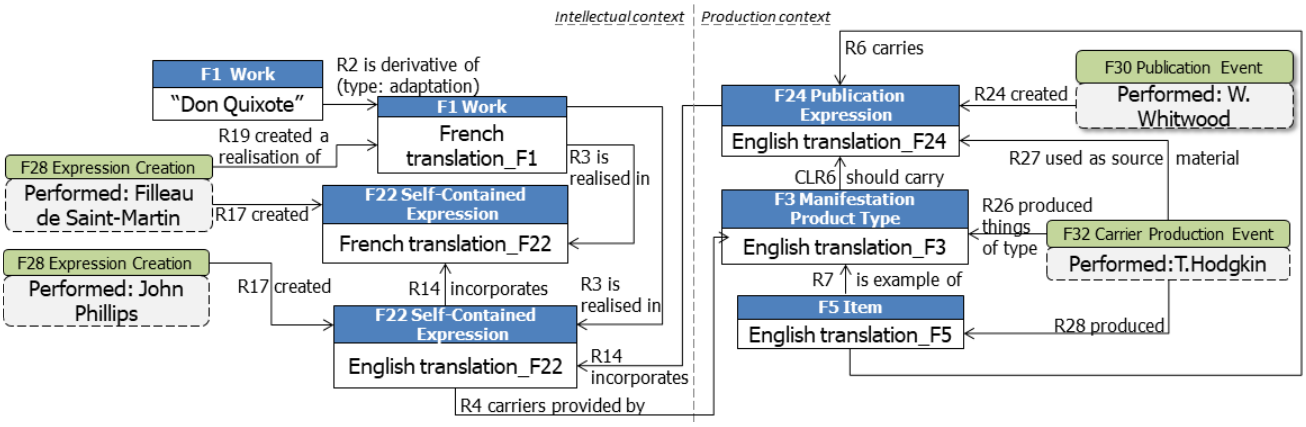


Figure 3:17. FRBRoo alternative representation of derivative (translation) relationships using the realization approach.

In **BIBFRAME**, the *bf:Work* class is used to represent both author's ideas and the signs realizing them. Hence, any change on the intellectual content or on the signs triggers the creation of a new *bf:Work.* The original *bf:Work* is related to its derivative *bf:Work* with a property depending on the type of derivation. The "Don Quixote" example is once again represented using the BIBFRAME constructs (*Figure 3:18*). Since all three *bf:Works* are textual ones, the *bf:Text* class, subclass of *bf:Work,* has been used for their representation. BIBFRAME does not provide many properties for the representation of derivative relationships. Adaptations, free translations, transformations may be represented using the *bf:hasDerivative* property. There is a specialized property for the case of literal translation, namely, the *bf:translation* property. Hence, the Don Quixote *bf:Text* instance is related to its free translation into French (French translation *bf:Text* instance) with a *bf:hasDerivative* property. The French translation *bf:Text* instance is related to the *bf:Text instance* representing its literal translation into English with the *bf:translation* property.

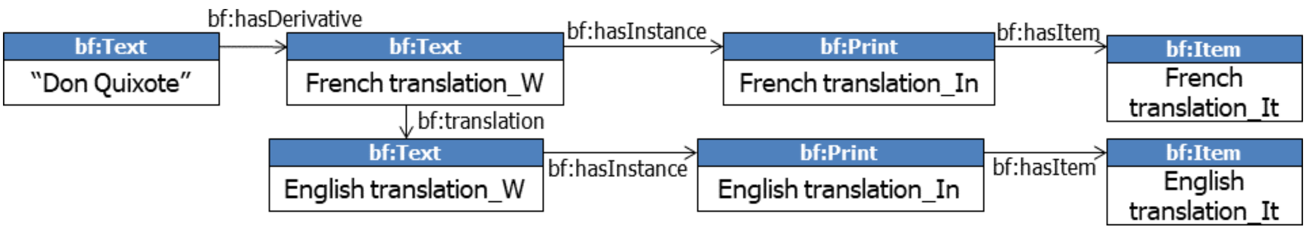


Figure 3:18. BIBFRAME representation of derivative (translation, adaptation) relationships.

BIBFRAME seems to adhere to the derivative approach enabling the representation of the derivative relationship (translation or adaptation in *Figure 3:18*) only when both signs are known. In case they are not known, it is very likely that some *bf:Works* are rendered "orphan". This issue is further analyzed in section 3.2 that examines the representation of bibliographic families and progenitors.

The **Europeana Data Model** enables the descriptions of material objects and their digital representations. The Don Quixote example has been represented in all previously mentioned models in an abstract manner, in terms of ideas and signs, *Work* and *Expression* entities in FRBR or *bf:Work* in BIBFRAME. Such representation is not possible in EDM. Thus, in the following *Figure 3:19* only the French and English translations are represented with two *edm:ProvidedCHO* class instances. Don Quixote is not represented, nor the adaptation relationship with the French translation *edm:ProvidedCHO* instance. Derivations may be represented in EDM using a generic property, the *edm:isDerivativeOf* one*.* The French translation *edm:ProvidedCHO* instance is related to the English translation *edm:ProvidedCHO* instance with an inverse of the *edm:isDerivativeOf* property instance. Even though the translation relationship is represented, the specificity of the relationship is lost.

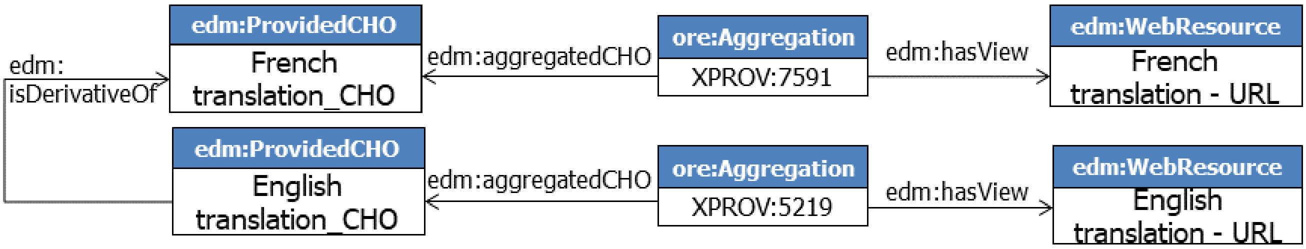


Figure 3:19. EDM representation of derivative (translation) relationships.

Using the **EDM-FRBRoo application profile**, an alternative representation may be created for the same example (*Figure 3:20*). The EDM-FRBRoo application profile enables the representation of more bibliographic resources being somehow related to *edm:ProvidedCHO* instances. Hence, the relationship of the "French translation" with the "Don Quixote" may be represented, despite the absence of a "Don Quixote" *edm:ProvidedCHO* instance. The adaptation relationship between Don Quixote and its French free translation is represented between the two *edm:InformationResource* instances that have been typed as "FRBRWork". The translation relationship between the French translation and its literal English translation may be also represented in EDM-FRBRoo either between the corresponding *edm:InformationResource* instances (typed as "FRBRWork"), or between the *edm:InformationResource* instances (typed as "FRBRExpression") using the *edm:isDerivativeOf* property. These two *edm:isDerivativeOf* property instances are depicted in *Figure 3:20* with dashed arrows for readability reasons. It must be noted that the *edm:isDerivativeOf* property is used in EDM-FRBRoo for representing two different types of relationships. When relating a typed "FRBRWork" *edm:InformationResource* instance with another typed "FRBRExpression" *edm:InformationResource* instance, then the inherent relationship of realization is represented. When relating *edm:InformationResource* instances of the same type (either "FRBRWork", or "FRBRExpression"), then a derivative relationship (translation and adaptation) is represented. This adds some ambiguity and may need more clarifying rules in future mappings. Another note at this point is that currently EDM (Europeana, 2017), defines that the *edm:isDerivativeOf* property is to be used for the representation of translation, summarization, and abstraction, while the *dcterms:isVersionOf* property is to be used for other editions and adaptations. Such a separation is not included in the EDM-FRBRoo application profile documents and test cases.

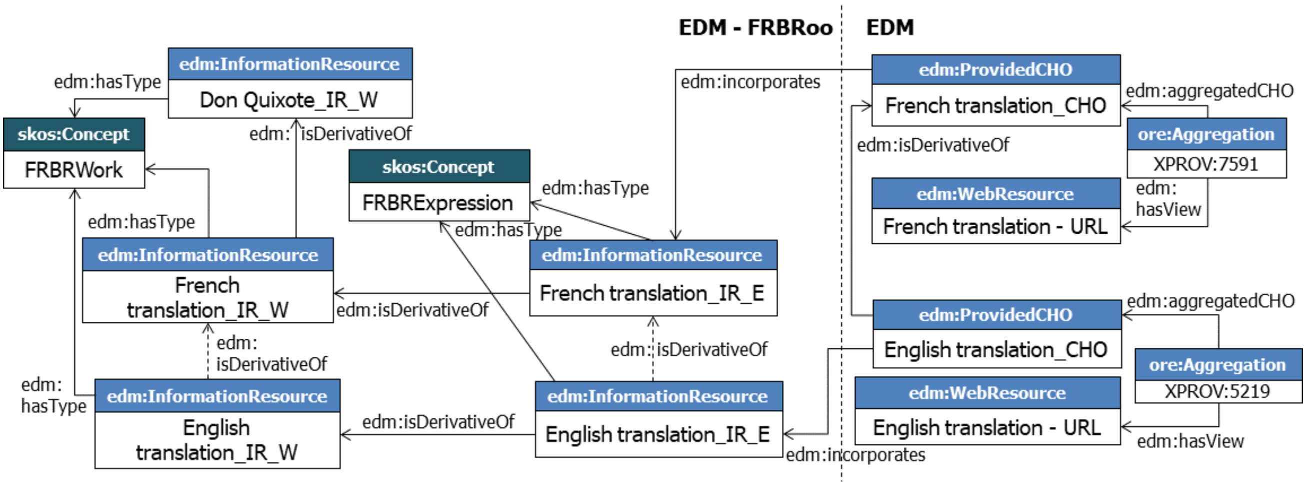


Figure 3:20. EDM alternative representation of derivative (translation, adaptation) relationships using specialized EDM classes.

3.1.2.2. Equivalence relationship

Reproduction enables the creation of exact copies of a certain material embodiment. These copies may be facsimiles, reprints, photocopies, microforms, digitizations, etc. To study the equivalence relationship, the "Don Quixote" example is used. The English translation by John Phillips, already presented in the examples of the previous paragraph, has been digitized by the National Library of Spain. In the following paragraphs the representation of this digitization case is presented using each model's constructs.

In **FRBR**, the equivalence relationship is represented between *Manifestation* instances (*Figure 3:21*). Even though, reproductions are produced by a specific *Item*, this *Item* is considered as a representative exemplar of the *Manifestation*. The result of a reproduction is a new *Manifestation*, regardless the physical characteristics of the original *Manifestation* are preserved. The intellectual content and the authorship status must be preserved though (Tillett, 1991). In case the reproduction is about a characteristic *Item,* e.g., a book with handwritten annotations by its owner, then the relationship is represented between the specific *Item* instance and the produced *Manifestation.* This alternative representation is depicted in *Figure 3:21* with the long dash dotted arrow. Another representation of the equivalence relationship in FRBR involves the reproduction of *Item* that preserves intellectual content, authorship, and physical characteristics. For this type of reproduction another *Item* is represented (English translation_RepI in *Figure 3:21*) and is related to the original *Item* with the inverse of the *has a reproduction* relationship. The relationship is depicted with the dashed arrow in *Figure 3:21*.

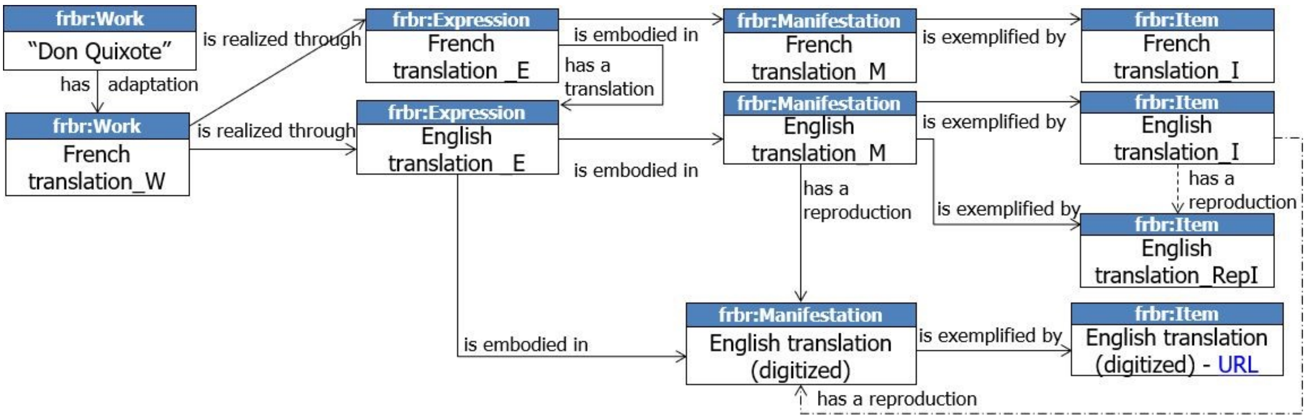


Figure 3:21. FRBR representation of equivalence relationship expressed between Manifestation instances. Two alternative representations are depicted. The dashed arrow depicts the reproduction of an Item resulting in another Item that preserves original Item's physical characteristics too. The long dash dotted arrow depicts the reproduction of a characteristic Item resulting in a Manifestation instance that does not preserve the original Item's physical characteristics.

The **LRM** representation is similar to the FRBR. The only difference is that LRM does not enable the representation of reproduction between two *LRM-E5 Item* instances. In LRM, a reproduction process always ends up with a new *LRM-E4 Manifestation* instance (*Figure 3:22*). The equivalence relationship is represented either between two *LRM-E4 Manifestation* instances (the original one and the reproduced one), or between the characteristic *LRM-E5 Item* instance and the reproduced *LRM-E4 Manifestation* instance.

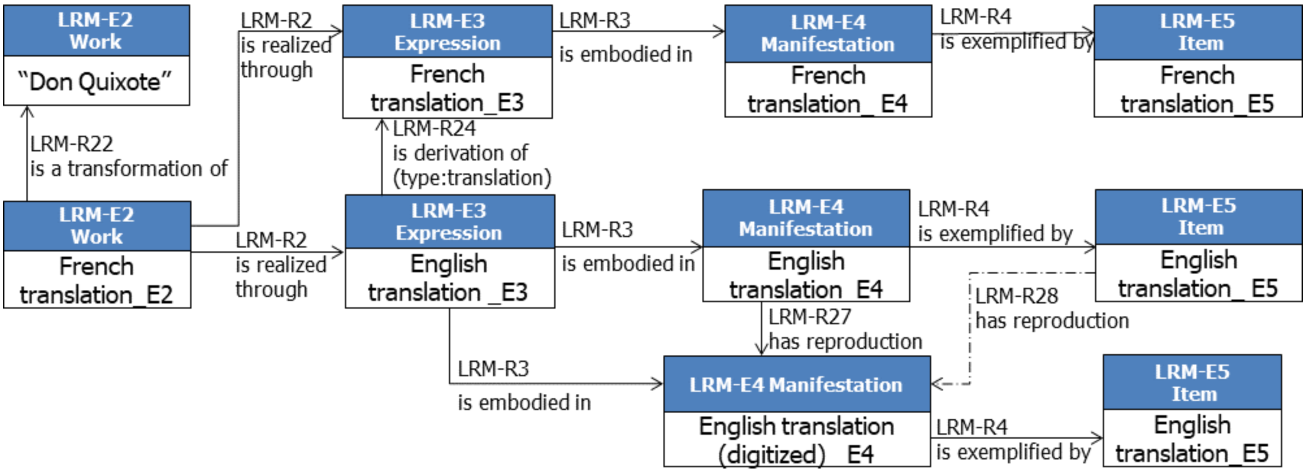


Figure 3:22. LRM representation of equivalence relationship expressed between LRM-E4 Manifestation instances. The alternative representation of reproducing a characteristic LRM-E5 Item is depicted with the long dash dotted arrow.

In **RDA**, the equivalence relationship may be represented between *rdac:C1007 Manifestation* instances (*Figure 3:23*). In case the reproduction happens using a characteristic *rdac:C1003 Item,* then the equivalence relationship is represented between this *rdac:C1003 Item* and the reproduced *rdac:C1007 Manifestation* instance.

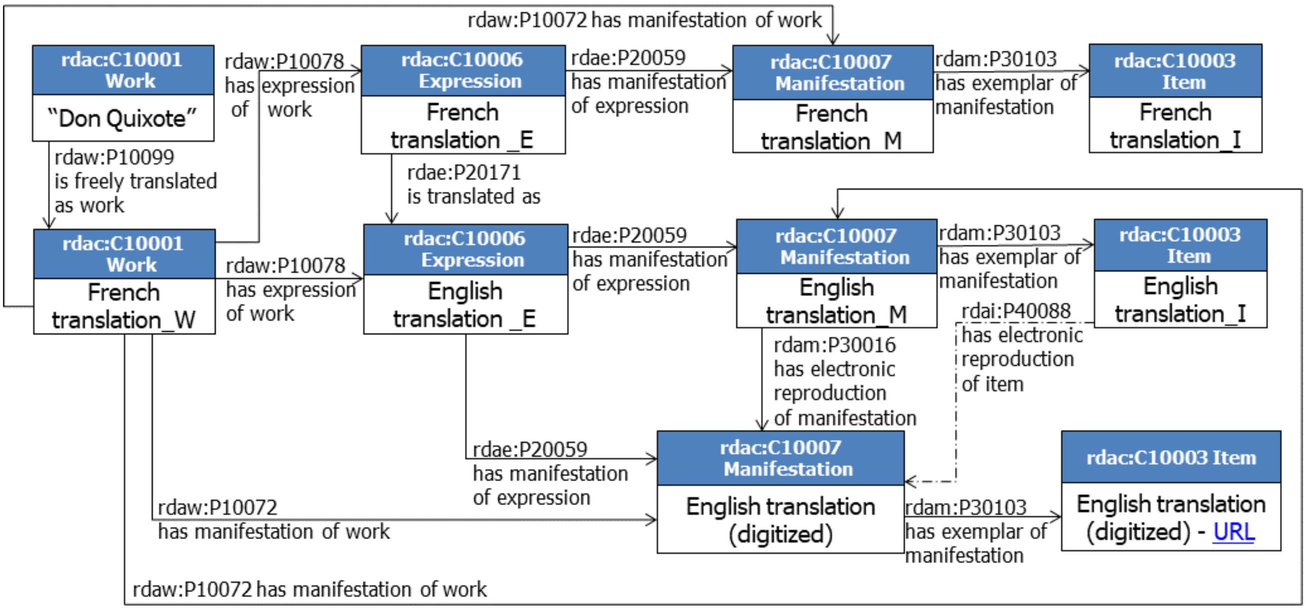


Figure 3:23. RDA representation of equivalence relationship expressed between rdac:C10007 Manifestation instances. The alternative representation of reproducing a characteristic rdac:C10003 Item is depicted with the long dash dotted arrow.

FRBRoo provides multiple options for the representation of the equivalence relationship, i.e., reproduction in the case of the English translation example. These options involve representation of the relationship at the *Manifestation* or *Item* level in an either static or dynamic way. Not all options may be used for the representation of the selected example. Yet, they are all presented to justify the selection of the proper representation of the example using the FRBRoo constructs.

In the **static representation** approach, the equivalence relationship is represented by typing the *P130 shows features of /P130i features are also found on {P130.1 kind of similarity: E55 Type = "Reproduction"}* property. This relationships can be used to relate either two *F3 Manifestation Product Type* instances (case 1 in *Figure*

3:24), or between an *F5 Item* instance and an *F3 Manifestation Product Type* instance (case 2 in *Figure 3:24*), or between two *F5 Item* instances (*Figure 3:25*). The representation of reproduction between two *F3 Manifestation Product Type* instances actually means that, even though a specific example of the original *F3 Manifestation Product Type* instance has been used, this specific example (an *F5 Item* instance) is thought to be a representative exemplar of the *F3 Manifestation Product Type* instance. The representation of reproduction between the *F5 Item* instance and the *F3 Manifestation Product Type* instance reveals that the *F5 Item* instance is not thought as a representative exemplar of the original *F3 Manifestation Product Type* instance, but it has some characteristic, e.g., annotations or a dedication, that differentiates it from other *F5 Item* instances-examples of the original *F3 Manifestation Product Type* instance. The representation of reproduction between *F5 Item* instances (*Figure 3:25*) describes a reproduction of a specific *F5 Item* instance that produces a reproduction (English translation_RepF5 in *Figure 3:25*) carrying the same physical characteristics with the original *F5 Item* instance.

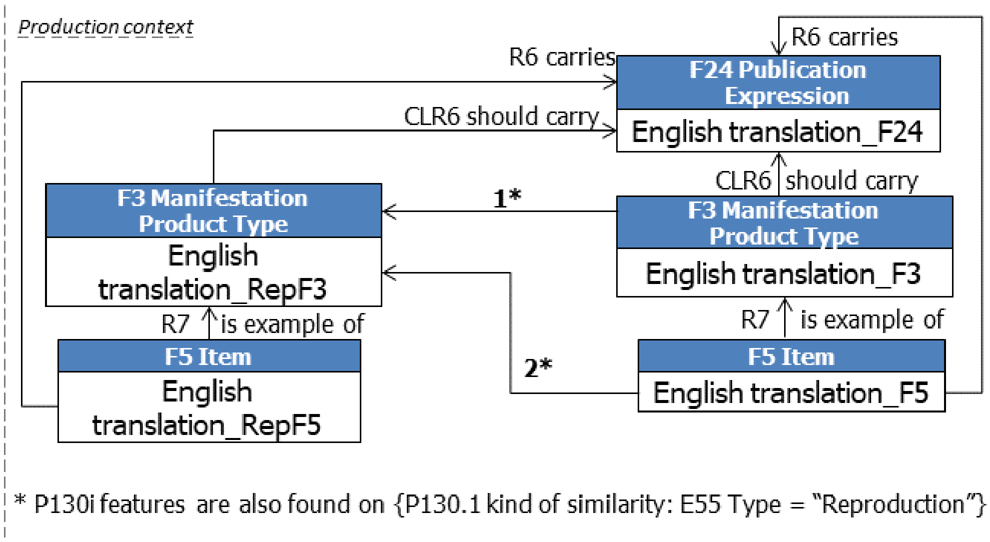


Figure 3:24. FRBRoo static representation of equivalence relationship expressed between F3 Manifestation Product Type instances (relationship depicted with number 1). The alternative representation between an F5 Item instance and an F3 Manifestation Product Type instance (relationship depicted with number 2) is also presented.

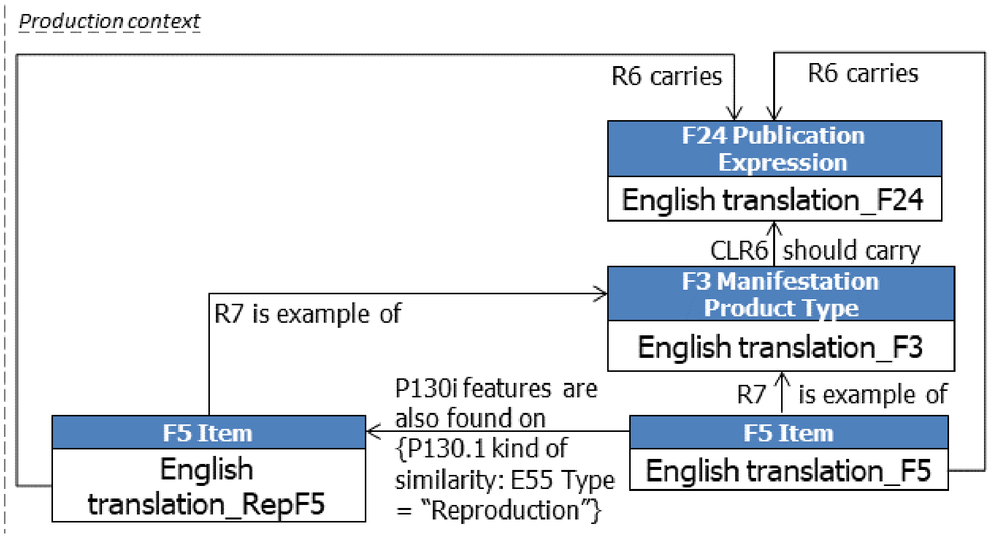


Figure 3:25. FRBRoo static representation of equivalence relationship expressed between F5 Item instances.

FRBRoo enables the **dynamic representation** of reproduction using instances of the *F33 Reproduction Event* class. It must be noted that this class is used to represent reproductions taking place for 'fair use'. Reproductions in multiple copies for commercial use may be represented using the *F32 Carrier Production Event* class. The *F33 Reproduction Event* and the *F32 Carrier Production Event* classes are not disjoint. Therefore, there might be cases where a production plan has been implemented for 'fair use' purposes. Such cases may be represented using both classes' instances (Working Group on FRBR/CRM Dialogue et al., 2016). *Figure 3:26* presents the dynamic representation of reproducing an *F3 Manifestation Product Type* instance to produce another *F3 Manifestation Product Type* instance. *Figure 3:27* presents the reproduction represented between *F5 Item* instances using the FRBRoo dynamic approach. The *F33 Reproduction Event* class produces only *E84 Information Carrier* instances (*F5 Item* is a subclass of the *E84 Information Carrier* class). The *E84 Information Carrier* (or the *English translation_RepF5 F5 Item* instance in *Figure 3:27*) instance does not have to carry the same physical characteristics of the original *F5 Item* used in the *F33 Reproduction Event*. Therefore, the *R31 is a reproduction of* property has a semantic difference to the *P130 shows features of /P130i features are also found on {P130.1 kind of similarity: E55 Type = "Reproduction"}* property as used in *Figure 3:25*. The *R31 is a reproduction of* is the shortcut of the "fully developed path from *E84 Information Carrier* through *R30 produced (was produced by), F33 Reproduction Event R29 reproduced (was reproduced by)* to *E84 Information Carrier"* (Working Group on FRBR/CRM Dialogue et al., 2016). Therefore, the produced *F5 Item* instance (English translation_RepF5) may carry different characteristics to the original *F5 Item.* The use of the *P130 shows features of /P130i features are also found on {P130.1 kind of similarity: E55 Type = "Reproduction"}* property in *Figure 3:25* conveys that the English translation_RepF5 has the same physical characteristics to the original English translation_F5.

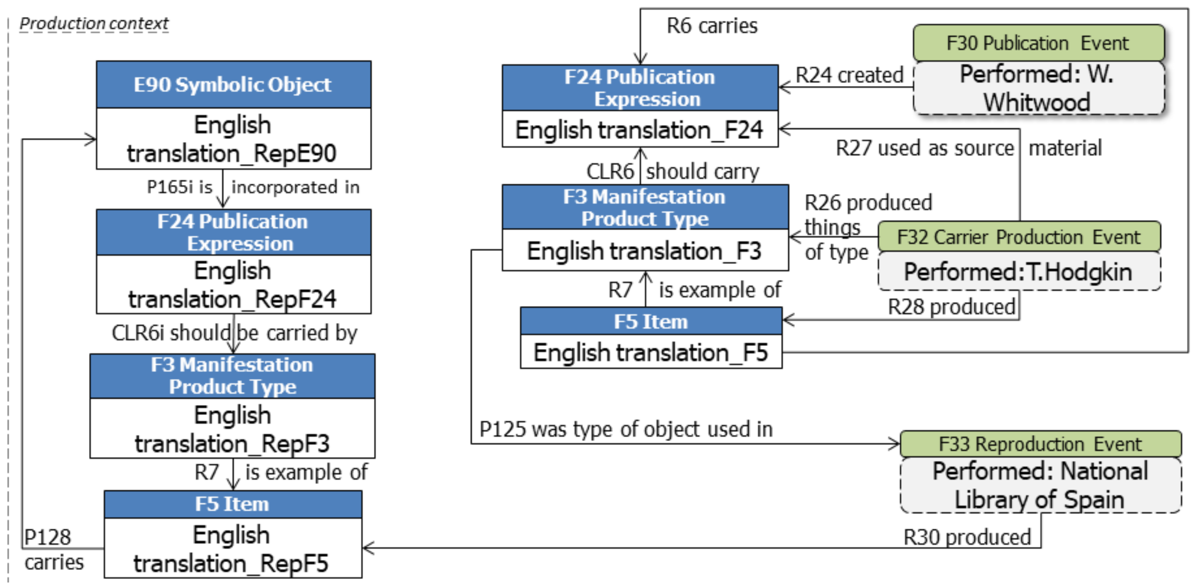


Figure 3:26. FRBRoo dynamic representation of equivalence relationship expressed between F3 Manifestation Product Type instances.

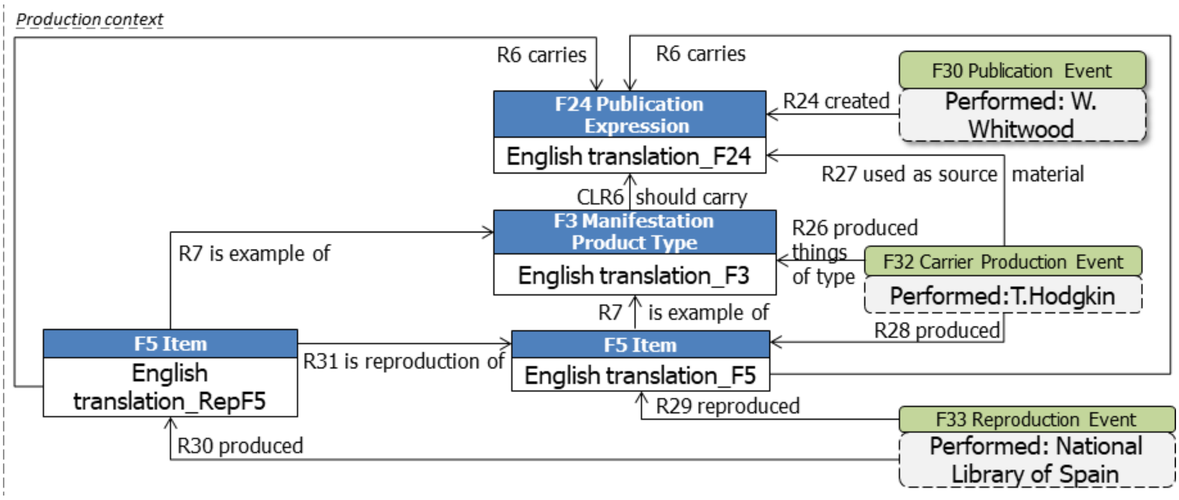


Figure 3:27. FRBRoo dynamic representation of equivalence relationship expressed between F5 Item instances.

To conclude the FRBRoo representation of equivalence relationship, the example of the digitized English translation may be represented using either the static approach as presented in *Figure 3:24* (case 1), or the dynamic approach as presented in *Figure 3:26*. The reproduced English translation_RepF5 is a digitization of a print publication which is considered as a representative of the English translation_F3. Therefore, the reproduction is represented at the *Manifestation* level, as presented in *Figure 3:24*'s the first case, and in *Figure 3:26*. The representations at the *Item* level as presented in *Figure 3:25*, and in *Figure 3:27* cannot describe the example successfully. The representation, as depicted in *Figure 3:25*, has been discarded for the representation of the selected case (reproduction of the John Philips English translation) due to an additional reason. The English translation_RepF5 (*Figure 3:25*) does not present the same physical characteristics as the original English translation_F5. Therefore, the *P130 shows features of /P130i features are also found on {P130.1 kind of similarity: E55 Type = "Reproduction"}* cannot be used to represent the equivalence/reproduction relationship.

In **BIBFRAME**, reproduction is represented between *bf:Instance* class instances with the *bf:hasReproduction* property (Figure 3:28). Due to domain and range restrictions, the reproduction using a unique *bf:Item* instance cannot be represented contrary to all the previously presented models (FRBR, LRM, RDA, and FRBRoo).

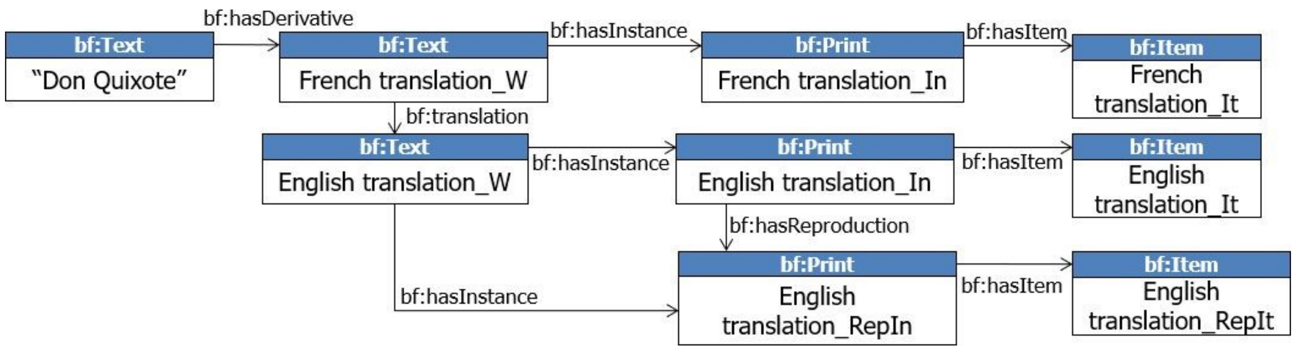


Figure 3:28. BIBFRAME representation of equivalence relationship.

In **EDM**, each digital reproduction is represented with a new *edm:ProvidedCHO* class instance (Figure 3:29). Equivalence is represented in terms of different formats between the related *edm:ProvidedCHO* instances, e.g., a digitized publication in pdf format and daisy format. The property used to represent the relationship is the *dcterms:hasFormat* and its inverse *dcterms:isFormatOf* property (Europeana, 2017).

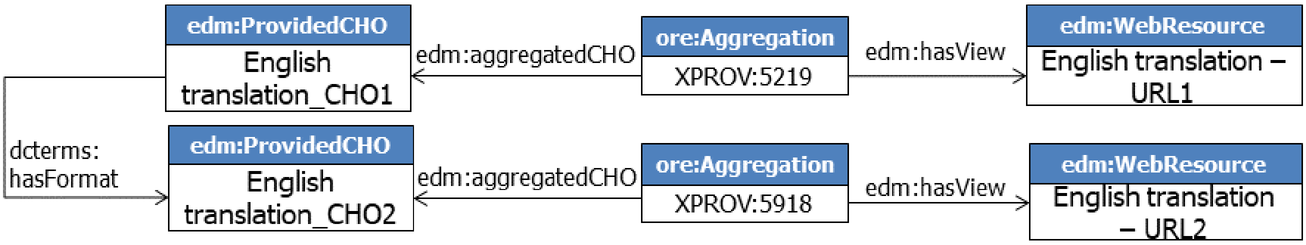


Figure 3:29. EDM representation of equivalence relationship.

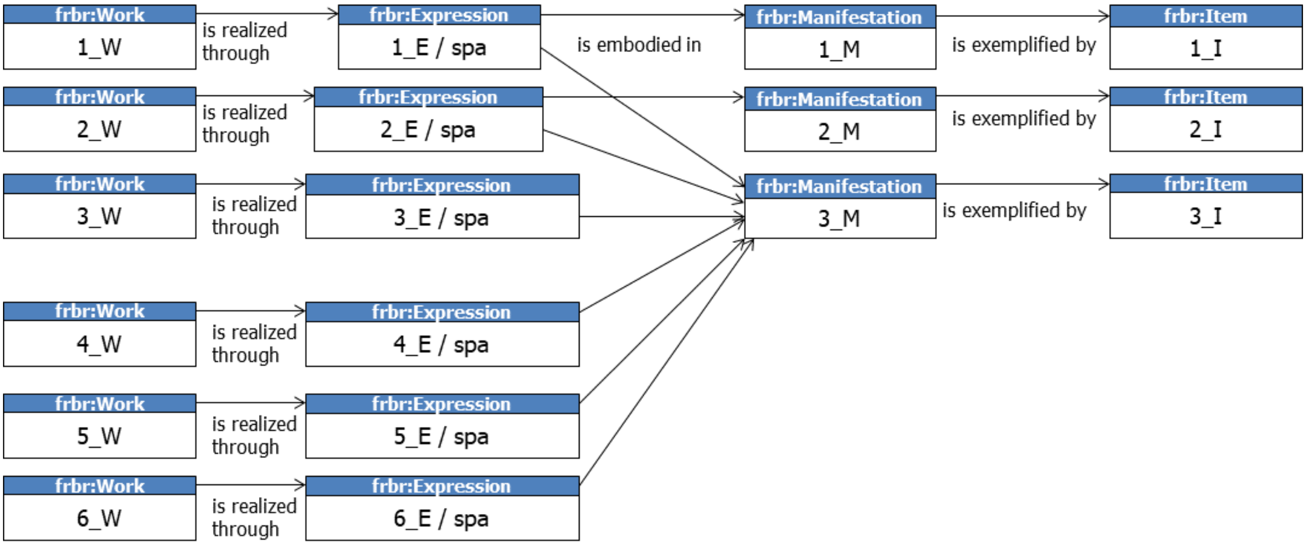
3.1.2.3. *Aggregates*

The example used for the representation of aggregates is an annotated edition of Cervantes' Don Quixote. The publication title is "Don Quijote de la Mancha" and it was published in Madrid, 2004 by Alfaguara for the Real Academia Espanñola and the Asociación de Academias de la Lengua Espanñola. This annotated edition (denoted as *Work 3* in the figures of this paragraph) includes the following nonintegral parts in this sequence:

- Introduction by editor Fransisco Rico
- a preface entitled "Un Novela Para El Siglo XXI" by the famous writer Mario Vargas Llosa (denoted as *Work 4* in the following figures),
- a study on the Don Quixote character, entitled "La Invención Del "Quijote"" by Francisco Ayala (denoted as *Work 5*),
- a second study on the Don Quixote character, entitled "Cervantes Y El "Quijote"" by Martin de Riquer,
- a general note on the text by the editor Francisco Rico,
- Don Quixote first part (El Ingenioso Hidalgo Don Quijote) with annotations by the editor Fransico Rico (denoted as *Work 1*),
- Don Quixote second part (Segunda parte del ingenioso caballero Don Quijote de la Mancha) with annotations by the editor Fransico Rico (denoted as *Work 2*),
- a study regarding the language used by Cervantes entitled "La Lengua De Cervantes Y El "Quijote"" by Jose Manuel Blecua, Guillermo Rojo, Jose Antonio Pascual, Margit Frank, and Claudio Guillen (Work 6), and a
- Glossary

It must be noted that for clarity reasons this paragraph's figures i) depict some and not all aggregated nonintegral parts, and ii) use abbreviated names for the aggregated nonintegral parts. At the bottom of each figure all abbreviations are analytically presented.

In **FRBR**, after the publication of the IFLA FRBR Working Group on Aggregates (O'Neill et al., 2011), aggregates are represented as *Manifestations* embodying multiple *Expressions.* The effort of Francisco Rico is worth mentioning. Therefore, an aggregating *Work* (3_W) and an aggregating *Expression* (3_E) are also represented. It must be noted that not all nonintegral parts included in the aggregate *Manifestation* (3_M) are represented in the following *Figure 3:30* for the sake of providing straightforward figures.



1_W: Work instance for the 1st part of Don Quijote by Cervantes
2_W: Work instance for the 2nd part of Don Quijote by Cervantes
3_W: Work instance for the aggregating *Work* edited by F.Rico
4_W: Work instance for the preface by M.V.Llosa
5_W: Work instance for the study of Don Quixote character by F. Ayala
6_W: Work instance for the Cervantes' language by J.M.Blecua, et.al.

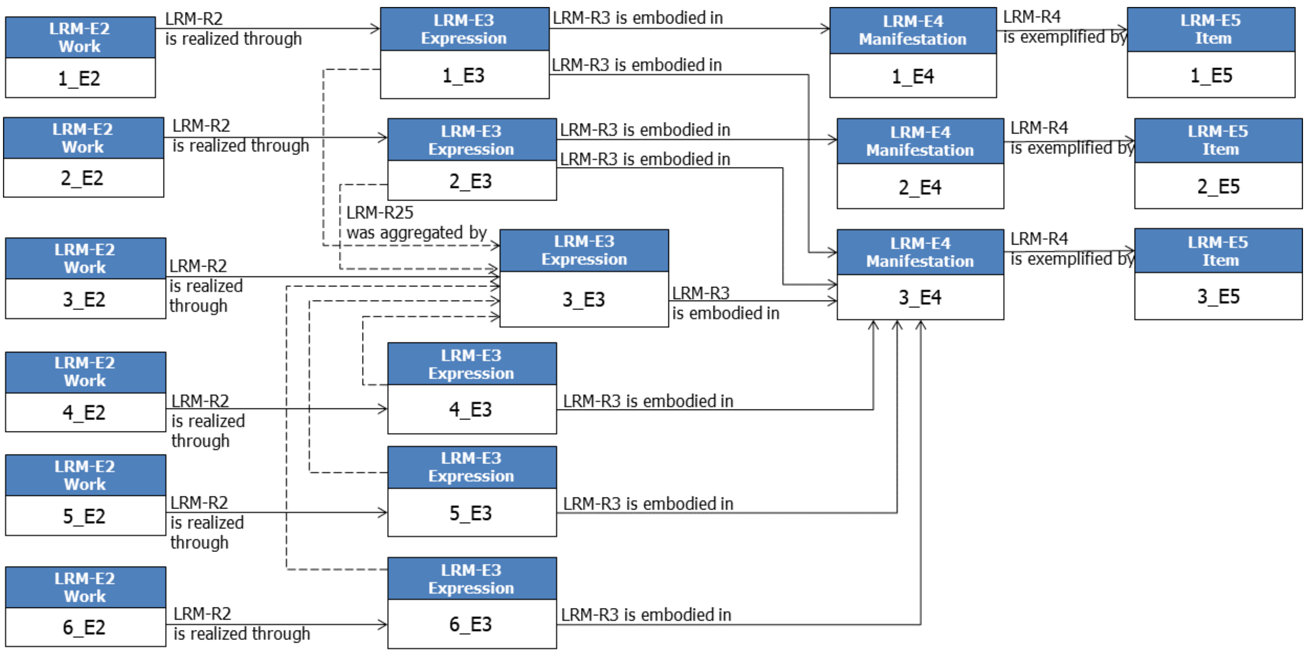
1_E/spa: An *Expression* of the *1_W Work* in Spanish
2_E/spa: An *Expression* of the *2_W Work* in Spanish
3_E/spa: The aggregating *Expression* of the *3_W Work* in Spanish
4_E/spa: An *Expression* of the *4_W Work* in Spanish
5_E/spa: An *Expression* of the *5_W Work* in Spanish
6_E/spa: An *Expression* of the 6_*W Work* in Spanish

1_M: A *Manifestation* of the *1_E* Expression
2_M: A *Manifestation* of the *2_E* Expression
3_M: The aggregate *Manifestation* of the *3_E* Expression

1_I: An exemplar of the *1_M Manifestation*
2_I: An exemplar of the *2_M Manifestation*
3_I: An exemplar of the *3_M Manifestation*

Figure 3:30. FRBR representation of aggregates. For readability purposes the label of the is embodied in relationship is written only once.

The **LRM** representation is slightly different to the FRBR one. The aggregating *LRM-E2 Work* (3_E2) and the aggregating *LRM-E3 Expression* (3_E3) are represented, and the aggregated *Expressions* are embodied in the same aggregate *Manifestation* (3_E4). The process of aggregating *LRM E3 Expressions* demands intellectual effort (represented with the 3_E2 *Work* instance) and takes place at the signs level. Therefore, it is represented by relating the aggregated *Expressions* with the aggregating one (3_E3) with instances of the *LRM-R25 was aggregated by* property. For readability purposes, the *LRM-R25 was aggregated by* property is depicted with long dashed arrows and its label is only written once (*Figure 3:31*).



1_E2: E2 Work instance for the 1st part of Don Quijote by Cervantes
2_E2: E2 Work instance for the 2nd part of Don Quijote by Cervantes
3_E2: E2 Work instance for the aggregating Work edited by F.Rico
4_E2: E2 Work instance for the preface by M.V.Llosa
5_E2: E2 Work instance for the study of Don Quixote character by F. Ayala
6_E2: E2 Work instance for the study on Cervantes' use of language by J.M.Blecua, et.al.

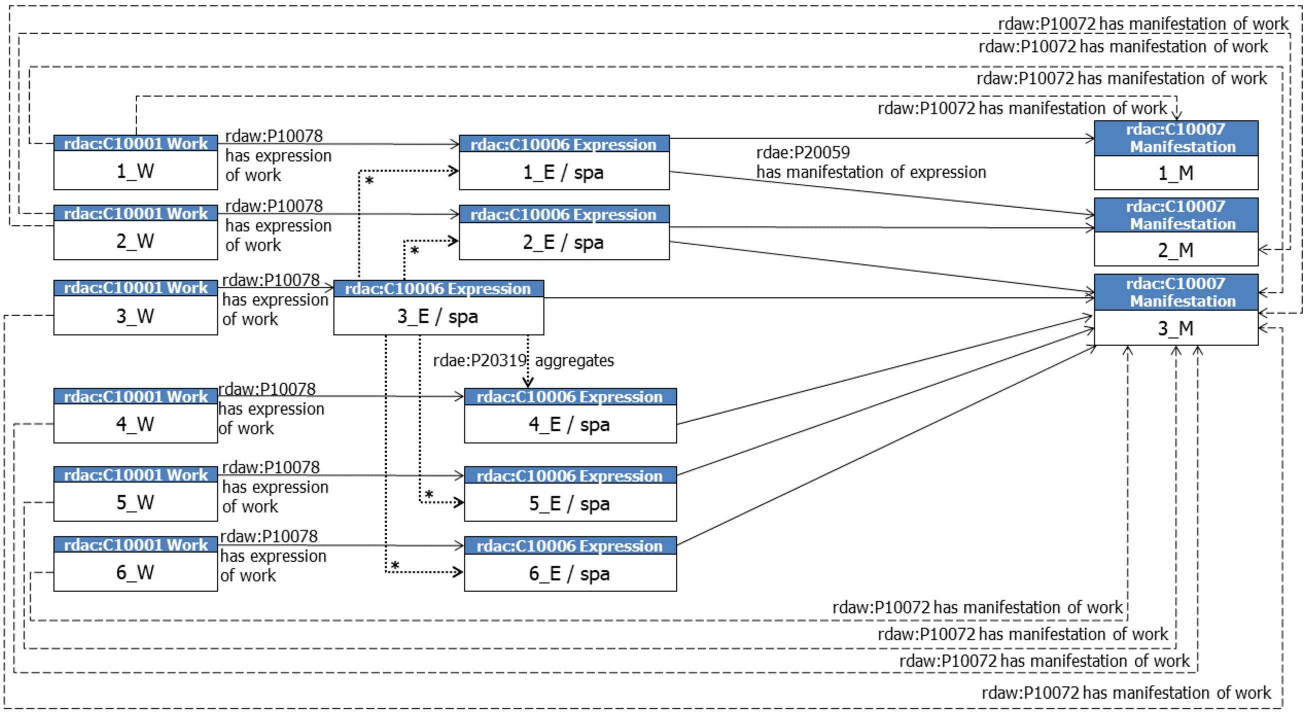
1_E3: An E3 Expression of the 1_E2 Work
2_E3: An E3 Expression of the 2_E2 Work
3_E3: The aggregating *Expression* of the *3_E2 Work*
4_E3: An E3 *Expression* of the *4_E2 Work*
5_E3: An E3 *Expression* of the *5_E2 Work*
6_E3: An E3 *Expression* of the *6_E2 Work*

1_E4: An *E4 Manifestation* of the *1_E3* Expression
2_E4: An *E4 Manifestation* of the *2_E3* Expression
3_E4: The aggregate *E4 Manifestation* of the *3_E3* Expression

1_E5: An exemplar of the 1_E4 *Manifestation*
2_E5: An exemplar of the 2_E4 *Manifestation*
3_E5: An exemplar of the 3_E4 *Manifestation*

Figure 3:31. LRM representation of aggregates. For readability purposes, the LRM-R25 was aggregated by relationship is depicted with long dashed arrows and its label is only written once.

RDA, due to the 3R Project (see the 2.3.4.3 paragraph) and the RDA conforming to the IFLA LRM (RDA Toolkit, 2019), represents aggregates according to the LRM. This means that aggregates are represented as *Expressions* embodied in a *Manifestation,* and in case an Aggregating Expression (the 3_E in *Figure 3:32*) is represented, the *rdae:P20319 aggregates* property is used to relate it to the *Expressions* it aggregates (namely, 1_E, 2_E, 4_E, 5_E, and 6_E *Expressions* in *Figure 3:32*). The only difference between the LRM representation in (*Figure 3:31*) and the RDA one (*Figure 3:32*) is the use of instances of the *rdaw:P10072 has manifestation of work* property relating the *rdac:C1001 Work* and *rdac:C10007 Manifestation* instances. This property represents an inherent relationship in the context of RDA.



* rdae:P20319 aggregates

1_W: Work instance for the 1st part of Don Quijote by Cervantes
2_W: Work instance for the 2nd part of Don Quijote by Cervantes
3_W: Work instance for the aggregating *Work* edited by F.Rico
4_W: Work instance for the preface by M.V.Llosa
5_W: Work instance for the study of Don Quixote character by F. Ayala
6_W: Work instance for the Cervantes' language by J.M.Blecua, et.al.

1_E/spa: An *Expression* of the *1_W Work* in Spanish
2_E/spa: An *Expression* of the *2_W Work* in Spanish
3_E/spa: The aggregating *Expression* of the *3_W Work* in Spanish
4_E/spa: An *Expression* of the *4_W Work* in Spanish
5_E/spa: An *Expression* of the *5_W Work* in Spanish
6_E/spa: An *Expression* of the 6*_W Work* in Spanish

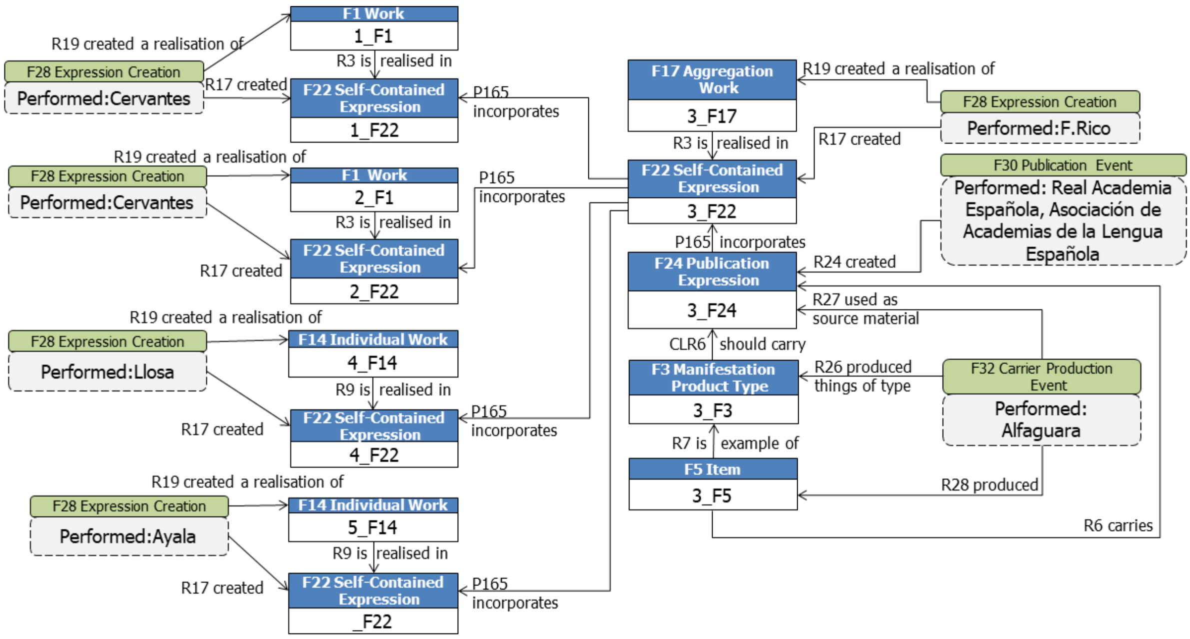
1_M: A *Manifestation* of the *1_E* Expression
2_M: A *Manifestation* of the *2_E* Expression
3_M: The aggregate *Manifestation* of the *3_E* Expression

1_I: An exemplar of the *1_M Manifestation*
2_I: An exemplar of the *2_M Manifestation*
3_I: An exemplar of the *3_M Manifestation*

Figure 3:32. RDA representation of aggregates. For readability purposes, the rdac:C10003 Item instances are excluded from the representation, the rdae:P20059 has manifestation of expression property is written only once, the rdae:P20319 aggregates instances are depicted with an asterisk, and the rdaw:P10072 has manifestation of work relationship is depicted with long dashed arrows.

FRBRoo represents aggregating works with a *F1 Work* subclass, i.e., the *F16 Container Work* class. This class may offer the framework for the representation of works that aggregate full/fragments of sets of signs originating from various works. Depending on the type of aggregation, different *F16 Container Work* subclasses may be used. The *F17 Aggregation Work* subclass is used to represent a *Work* that aggregates existing expressions of other works and adds value to them by selecting and arranging them. The *F19 Publication Work* subclass is used to represent works that establish all features of publication, such as layout, graphics, etc. The *F18 Serial Work* class, subclass of the *F19 Publication Work,* is used to represent works that establish all features of serials with their characteristic constraints regarding frequency, numbering, etc. The *F20 Performance Work* subclass is used to represent works that establish all features of a performance or a series of like performances.

The annotated edition may be represented as an *F17 Aggregation Work* instance (3_F17 in *Figure 3:33*), due to Fransisco Rico's effort to aggregate and arrange expressions from different works. The *F17 Aggregation Work* is a subclass of the *F14 Individual Work* class too, and as such, it can be only realised in a *F22 Self-Contained Expression* class instance which may incorporate other *F2 Expression* class instances. This *F22 Self-Contained Expression* (3_F22) includes not just the signs realizing the ideas behind F.Rico's aggregation, but all the aggregated expressions too, namely, 1_F22, 2_F22, 4_F22, and 5_F22. FRBRoo enables the representation of aggregates at the signs level with one *F22 Self-Contained Expression* instance (3_F22 in *Figure 3:33*) incorporating other *F2 Expression* instances. Even though the *F16 Container Work* and its subclasses (3_F17 instance in *Figure 3:33*) is similar semantically to the "Aggregating Work" of the FRBR WG on Aggregates report (O'Neill et al., 2011), there is an important difference. The *F22 Self-Contained Expression* that represents the signs used by the aggregator (Fransisco Rico in the example of *Figure 3:33*) cannot be considered as semantically equivalent to the FRBR "Aggregating Expression". The *F22 Self-Contained Expression* (3_F22) realizes the *F17 Aggregation Work* (3_F17) and at the same time incorporates all aggregated *F22 Self-Contained Expressions* (1_F22, 2_F22, 4_F22, and 5_F22).



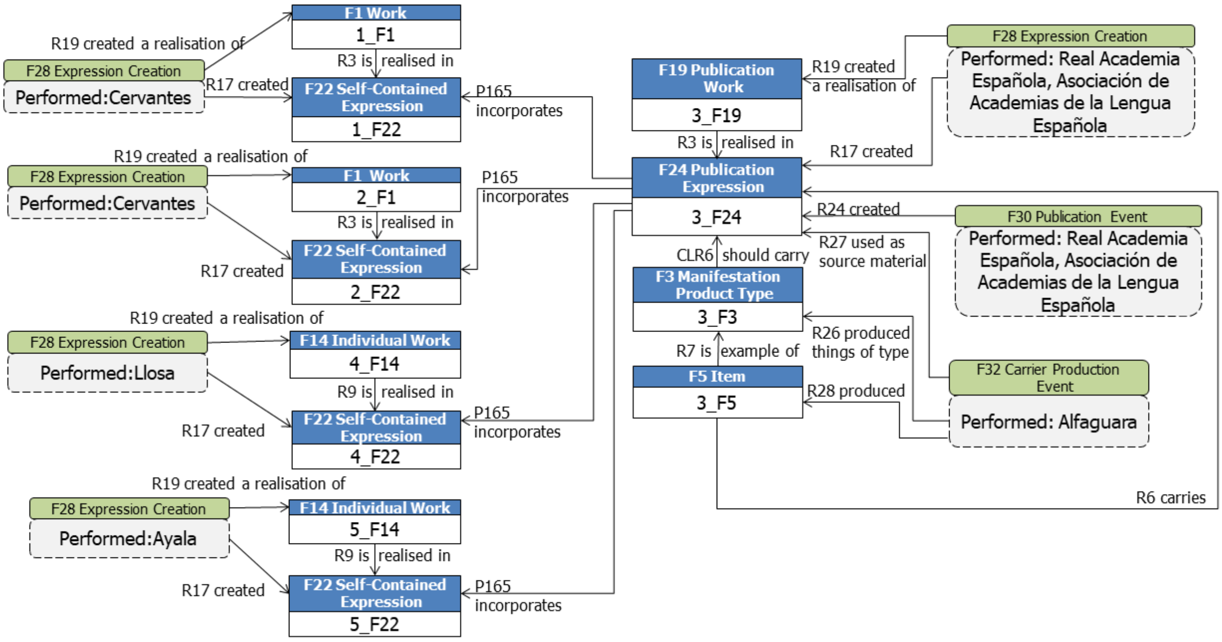
1_F1: Work instance for the 1st part of Don Quijote by Cervantes
2_F1: Work instance for the 2nd part of Don Quijote by Cervantes
4_F14: Work instance for the preface by M.V.Llosa
5_F14: Work instance for the study of Don Quixote character by F. Ayala

1_F22: An *Expression* of the *1_F1 Work*
2_F22: An *Expression* of the *2_F1 Work*
4_F22: An *Expression* of the *4_F14 Work*
5_F22: An *Expression* of the *5_F14 Work*

3_F17: *Aggregation Work* instance for the aggregating *Work* edited by F.Rico
3_F22: The Aggregating *Expression* of the *3_F17 Work* including the aggregated F22 Self-Contained Expressions
3_F24: The *Publication Expression* incorporating the *3_F22 Expression*
3_F3: The *Manifestation* instance carrying the *3_F24 Publication Expression*
3_F5: An exemplar of the *3_F3 Manifestation*.

Figure 3:33. FRBRoo representation of aggregates. For readability purposes only the F3 Manifestation Product Type instance representing the aggregate Manifestation is depicted (3_F3).

In case, the decision for the aggregation was taken by a publisher, which is rather usual in the publishing world, then the representation of the example would be different. An *F19 Publication Work* instance would be used to represent the "Aggregating Work", and an *F24 Publication Expression* instance would be used to represent the "Aggregating Expression". The *F24 Publication Expression* instance (Aggregating Expression) both realizes the *F19 Publication Work* and incorporates all aggregated *F22 Self-Contained Expressions* (1_F22, 2_F22, 4_F22, and 5_F22). This alternative representation is depicted in *Figure 3:34*.



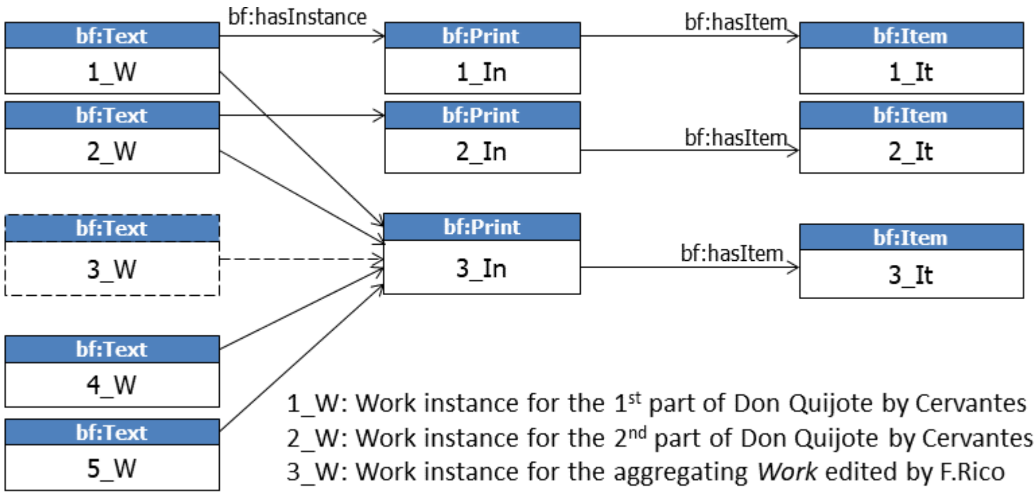
1_F1: Work instance for the 1st part of Don Quijote by Cervantes
2_F1: Work instance for the 2nd part of Don Quijote by Cervantes
4_F14: Work instance for the preface by M.V.Llosa
5_F14: Work instance for the study of Don Quixote character by F. Ayala

1_F22: An *Expression* of the *1_F1 Work*
2_F22: An *Expression* of the *2_F1 Work*
4_F22: An *Expression* of the *4_F14 Work*
5_F22: An *Expression* of the *5_F14 Work*

3_F19: *Publication Work* instance for the aggregating *Work* edited by F.Rico
3_F24: The Publication *Expression* of the *3_F19 Work* that aggregates 1,2,4, and 5 F22 Self-Contained Expressions, and the publisher's signs
3_F3: The *Manifestation* instance carrying the *3_F24 Publication Expression*
3_F5: An exemplar of the *3_F3 Manifestation*

Figure 3:34. Alternative FRBRoo representation of aggregates. The aggregation happened by the publisher and triggered the representation of the aggregating work as an F19 Publication Work instance realized by an F24 Publication Expression instance that incorporates all aggregated expressions (1_F22, 2_F22, 4_F22, and 5_F22).

The representation of aggregates has not been described in any official **BIBFRAME** document yet. In BIBFRAME 1.0 the sentence that "Each BIBFRAME Instance is an instance of one and only one BIBFRAME Work" (Miller et al., 2012) implied that there was a 1:1 relationship between *bf:Work* and *bf:Instance* classes. In BIBFRAME 2.0, this sentence was erased and no other reference regarding cardinality constraints between the two classes was made. In the BIBFRAME mailing list, a question triggered a conversation regarding this matter. In this online conversation, Ray Denenberg described that the model offers two ways of representing aggregates (Denenberg, 2017b). The first representation approach enables the representation of aggregates as embodiment of different *bf:Works* into the same *bf:Instance* instance (*Figure 3:35*). The second approach enables the representation of an aggregating *bf:Work* having as its parts other *bf:Works.* In this latter representation, there is an 1:1 relationship between the a *bf:Instance* and its *bf:Work* (*Figure 3:36*).



1_W: Work instance for the 1st part of Don Quijote by Cervantes
2_W: Work instance for the 2nd part of Don Quijote by Cervantes
3_W: Work instance for the aggregating *Work* edited by F.Rico
4_W: Work instance for the preface by M.V.Llosa
5_W: Work instance for the study of Don Quixote character by F. Ayala

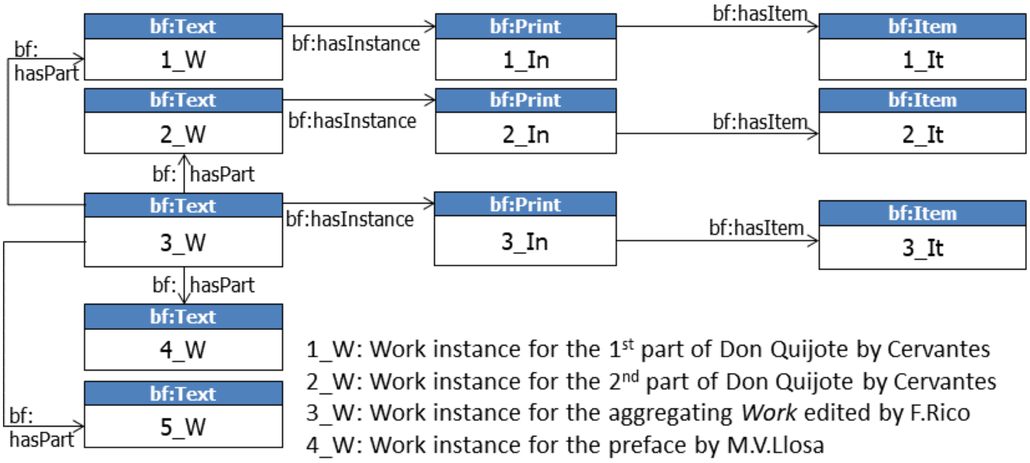
1_In: The *Instance* of the *1_W Work*
2_In: The *Instance* of the *2_W Work*
3_In: The aggregate *Instance* of the *3_W Work*

1_It: An *Item* exemplifying the 1_In Instance
2_It: An *Item* exemplifying the 2_In Instance
3_It: An *Item* exemplifying the 3_In Instance

Figure 3:35. BIBFRAME representation of aggregates at the bf:Instance level. The aggregating 3_W is depicted with a long-dashed rectangle. Its representation depends on the aggregator's effort. For readability reasons, the bf:hasInstance property is written only once.

The first approach (*Figure 3:35*) adheres to the FRBR WG on Aggregates report that treats aggregates as *Manifestations* (*Figure 3:35*). All aggregated *bf:Works* have as an instance the aggregate *bf:Instance* named 3_In. The aggregating *bf:Work* instance (3_W) may be either represented or not depending on the amount of effort expended by the aggregator. The availability of choice regarding the representation of the aggregating *bf:Work* is expressed by depicting the 3_W in *Figure 3:35* with a long-dashed rectangle. In case it is represented, then the 3_W aggregating *bf:Work* includes only the aggregator's idea for the aggregation and the signs realizing these ideas. For readability reasons, the *bf:hasInstance* property is written only once.

The second approach (*Figure 3:36*) is closer to the original FRBR report (IFLA Study Group on the Functional Requirements for Bibliographic Records, 2009) that treats aggregates as “integral units” with parts. The aggregating *bf:Work* (3_W) has other *bf:Work* instances as parts, namely, 1_W, 2_W, 4_W, and 5_W. It must be noted, similarly to the 3_F22 Expression in *Figure 3:33*, or to the 3_F24 Expression in *Figure 3:34*, the aggregating *bf:Work* (3_W) represents 1) aggregator's ideas and signs, and 2) all aggregated *bf:Works'* ideas and signs too. As a result, the aggregate *bf:Instance* (3_In) is a *bf:Instance* of only one *bf:Work,* the 3_W one. Yet, the use of the *bf:hasPart* property does not make clear if the 3_W is an aggregate or a *bf:Work* having more than one parts.



1_W: Work instance for the 1st part of Don Quijote by Cervantes
2_W: Work instance for the 2nd part of Don Quijote by Cervantes
3_W: Work instance for the aggregating *Work* edited by F.Rico
4_W: Work instance for the preface by M.V.Llosa
5_W: Work instance for the study of Don Quixote character by F. Ayala

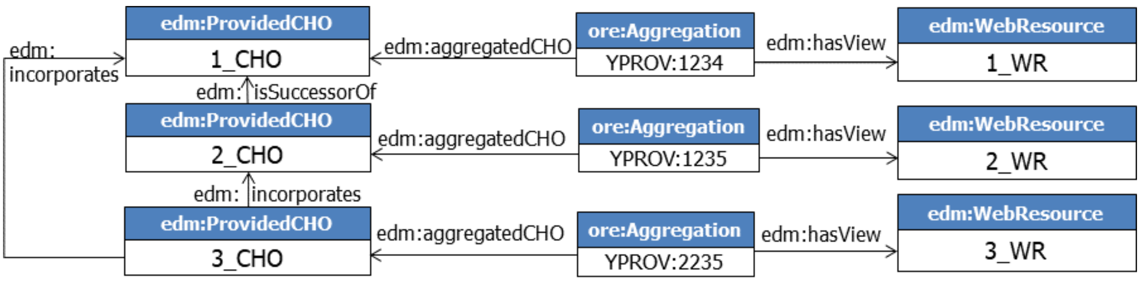
1_In: The *Instance* of the *1_W Work* 1_It: An *Item* exemplifying the 1_In Instance
2_In: The *Instance* of the *2_W Work* 2_It: An *Item* exemplifying the 2_In Instance
3_In: The *Instance* of the *3_W Work* 3_It: An *Item* exemplifying the 3_In Instance

Figure 3:36. BIBFRAME representation of aggregates at the bf:Work level. The aggregating 3_W has as parts all other aggregated bf:Works.

EDM explicitly differentiates between aggregates and whole/part relationships in the definition of the *edm:incorporates* property. The EDM definition refers that (Europeana, 2017):

> "… incorporated resources do in general not form proper parts. Incorporated resources are not part of the same resource, but are taken from other resources, and have an independent history. Therefore edm:incorporates is not a sub-property of dcterms:hasPart".

As a result, aggregates may be represented using *edm:incorporates* property instances to relate an aggregating *edm:ProvidedCHO* instance with the *edm:ProvidedCHO* instances it aggregates. Since, no web-accessible digital resource for the aggregated Novela (denoted with number 4 in the figures of this paragraph) and Invención (denoted with number 5 in the figures of this paragraph) resources has been found, they have been excluded from the representation in *Figure 3:37*.

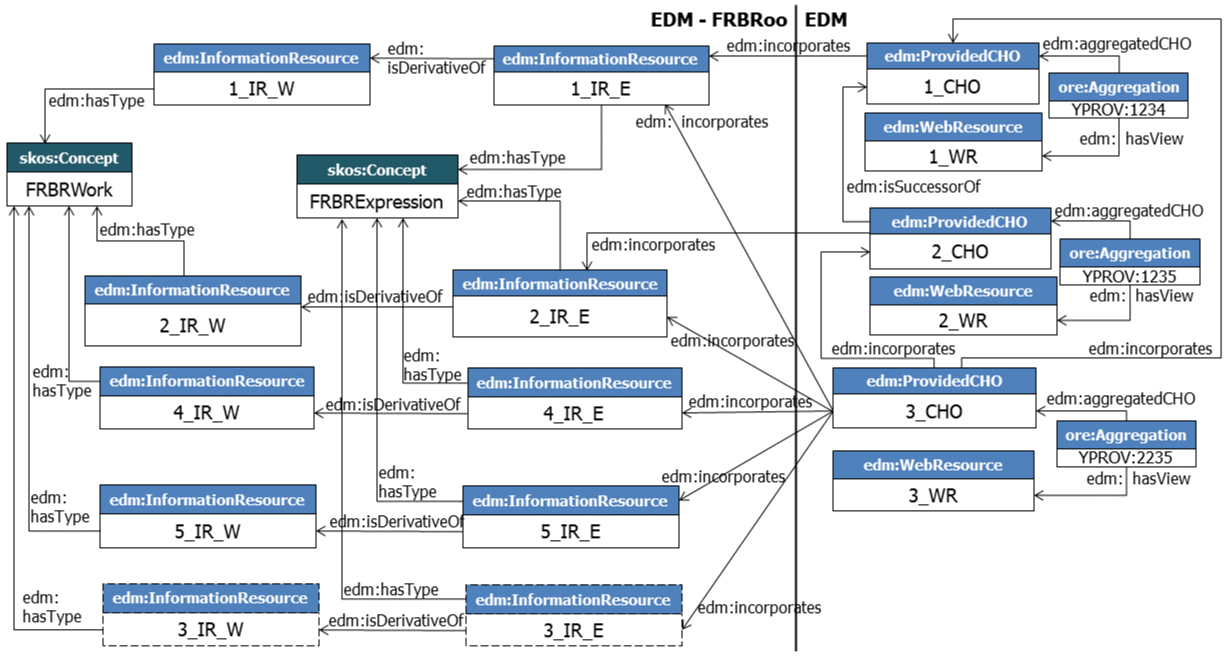


1_CHO: ProvidedCHO instance for the 1st part of Don Quijote by Cervantes
2_CHO: ProvidedCHO instance for the 2nd part of Don Quijote by Cervantes
3_CHO: ProvidedCHO instance for the aggregating ProvidedCHO, edited by F.Rico

1_WR: The *WebResource* instance of the *1_CHO*
2_WR: The *WebResource* instance of the *2_CHO*
3_WR: The *WebResource* instance of the *3_CHO*

Figure 3:37. EDM representation of aggregates.

The EDM-FRBRoo application profile is more flexible than EDM regarding the representation of bibliographic entities and of the relationships between them. All aggregated *Works* may be represented as *edm:InformationResource* class instances (also typed as "FRBRWork" using *skos:Concept* instances) regardless the existence of a web-accessible resource. Thus, Novela (denoted with number 4 in the figures of this paragraph) and Invención (denoted with number 5 in the figures of this paragraph) are also represented at the ideas and signs level using *edm:InformationResource* instances (4_IR_W, 5_IR_W, 4_IR_E, and 5_IR_E). All aggregated *Expressions* are incorporated (*edm:incorporates* instances) into the same aggregate *edm:ProvidedCHO* class instance (3_CHO). In the EDM_FRBRoo part of the *Figure 3:38*, the aggregating 3_IR_W is also represented with its own 3_IR_E realization. For readability purposes, they are both depicted with long-dashed rectangles.



<u>InformationResource instances typed with the *skos:Concept* value "FRBRWork"</u>
1_IR_W: 1st part of Don Quijote by Cervantes
2_IR_W: 2nd part of Don Quijote by Cervantes
3_IR_W: aggregating *InformationResource*, edited by F.Rico
4_IR_W: preface by M.V.Llosa
5_IR_W: study of Don Quixote character by F. Ayala

<u>InformationResource instances typed with the *skos:Concept* value "FRBRExpression"</u>
1_IR_E: 1st part of Don Quijote by Cervantes
1_IR_E: 2nd part of Don Quijote by Cervantes
1_IR_E: aggregating *InformationResource*, edited by F.Rico
1_IR_E: preface by M.V.Llosa
1_IR_E: study of Don Quixote character by F. Ayala

1_CHO: ProvidedCHO instance for the 1st part of Don Quijote by Cervantes
2_ CHO: ProvidedCHO instance for the 2nd part of Don Quijote by Cervantes
3_ CHO: ProvidedCHO instance for the aggregating ProvidedCHO, edited by F.Rico

1_WR: The *WebResource* instance of the *1_CHO*
2_WR: The *WebResource* instance of the *2_CHO*
3_WR: The *WebResource* instance of the *3_CHO*

Figure 3:38. EDM-FRBRoo representation of aggregates as Manifestations following the FRBR WG on Aggregates report. The Aggregating Work and its aggregating Expression are represented with long-dashed rectangles.

3.2. Representation of bibliographic families and the progenitor Work

The representation of bibliographic families is really important in the linked data universe. Their representation may serve as a collocation mechanism, and as a navigating mechanism to related bibliographic resources (Smiraglia & Leazer, 1999; Svenonius, 2009).

In **FRBR** and **LRM**, as well as in all other FRBR-inspired models (RDA and FRBRoo), the progenitor of a family is represented at the *Work* level. The progenitor (instance of the class *Work*) may have many derivations represented either as new *Work* instances or as new *Expressions.* For instance, in the FRBR, LRM, and RDA representations (depicted correspondingly in *Figure 3:39*, in *Figure 3:40*, and in *Figure 3:41*), Cervantes' progenitor *Work* entitled "Don Quixote" is represented as *Work* having two realizations (Spanish text with Mateo de Bastida as editor, and Lorenzo Franciosini's Italian translation), and two adaptations. The first one is a libretto adaptation by Henri Cain in French, and the second one is Filleau de Saint-Martin's free translation. Each one of these two adaptations has its own realization(s). Cervantes' Don Quixote is related with bibliographic relationships to 'close descendants', such as realizations and literal translations, and to more distant members of the bibliographic family it has inspired, such as adaptations. According to Smiraglia, a bibliographic family usually starts with a derivation (Smiraglia, 2005; Smiraglia & Leazer, 1999). Therefore, when a progenitor *Work* participates in a large number of bibliographic relationships, the extent of inspiration it has motivated could be considered large, and its bibliographic family is expected to be correspondingly big.

Both FRBR's consolidation, the **LRM** (*Figure 3:40*) and the **RDA** (*Figure 3:41*) provide similar constructs for the representation of bibliographic families, namely bibliographic relationships relating the progenitor *Work* to 'close descendants' or to 'distant relatives' that may have their own bibliographic families.

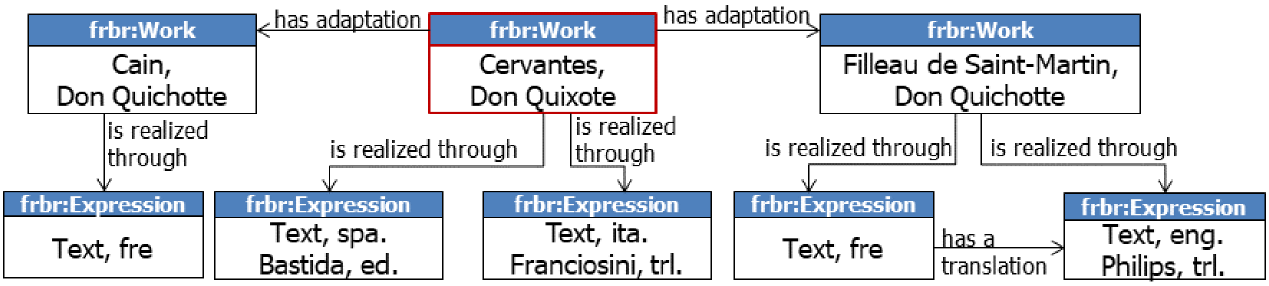


Figure 3:39. FRBR representation of the progenitor "Don Quixote" Work and some of the members of its bibliographic family. For readability reasons, the progenitor Work is marked with a red outline, and Manifestations and Items are excluded from the representation.

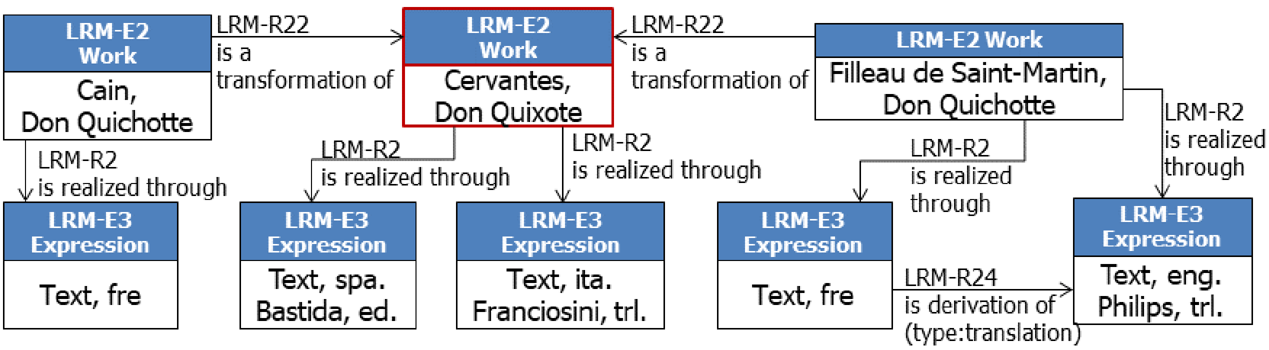


Figure 3:40. LRM representation of the progenitor "Don Quixote" Work and some of the members of its bibliographic family. For readability reasons, the progenitor LRM-E2 Work is marked with a red outline, and LRM-E4 Manifestations and LRM-E5 Items are excluded from the representation.

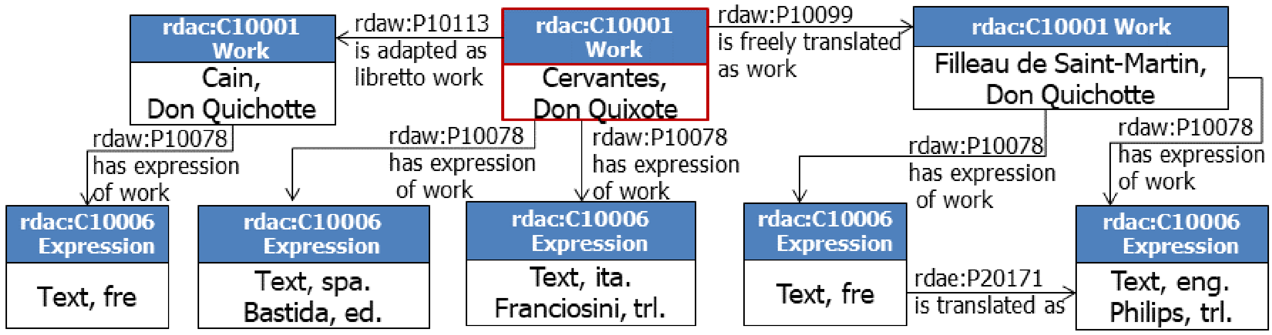


Figure 3:41. RDA representation of the progenitor "Don Quixote" Work and some of the members of its bibliographic family. For readability reasons, the progenitor rdac:C10001 Work is marked with a red outline, and rdac:C10007 Manifestations and rdac:C10003 Items are excluded from the representation.

FRBRoo presents a slight difference (*Figure 3:42*). The *F15 Complex Work* class may be used to represent the progenitor of a bibliographic family. An *F15 Complex Work* instance expresses the dominance of a set of ideas, as it was conceived by the creator(s). 'Close descendants' are considered as conceptual (not structural) members (Working Group on FRBR/CRM Dialogue et al., 2016) of the *F15 Complex Work*, and may be represented as *F14 Individual Work* class instances with their own *F22 Self-Contained Expressions.* In case there is a *Work* that uses the progenitor's initial concept, and significantly changes it, then a new *Work* is represented which may serve as a progenitor of its own bibliographic family with its own members (another *F15 Complex Work* instance).

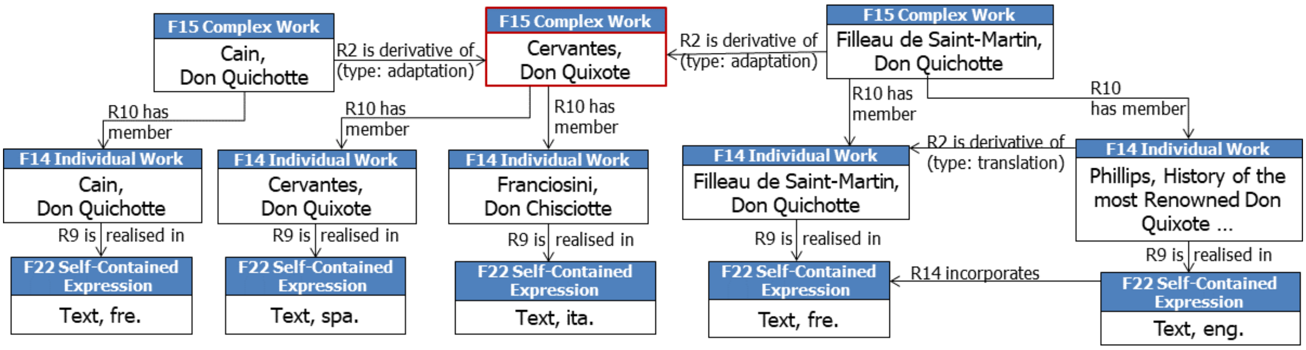


Figure 3:42. FRBRoo representation of the progenitor "Don Quixote" Work and some of the members of its bibliographic family. The derivation approach has been used for the representation of translated Works. For readability reasons, the progenitor F15 Complex Work is marked with a red outline, F3 Manifestation Product Type instances, F5 Item instances, and events used in the dynamic view of classes are excluded from the representation.

Contrary to the all the previous models, i.e., FRBR, LRM, RDA, and FRBRoo, **BIBFRAME** does not provide a construct for the clustering of members of the same bibliographic family. FRBR, LRM, RDA, and FRBRoo enable the clustering of 'close descendants' as realizations of the same *Work*. Moreover, they all enable the representation of a bibliographic relationship, e.g., adaptation, either at the ideas (*Work*), or the signs (*Expression*) level, depending on the information at hand regarding the original and the produced signs (*Expressions*) used in the bibliographic relationship. If both original and produced signs are known, then the relationship is represented at the *Expression* level. Otherwise, it is represented at the *Work* level.

BIBFRAME's *Work* class represents both ideas and signs. Therefore, the representation of bibliographic relationships between *bf:Work* instances demands the prior knowledge of which exactly realizations (signs) have been used in the situation that the bibliographic relationship describes. The representation of bibliographic relationships at the more abstract (ideas) level is not possible. In the example used in this section's figures, there was only one case, namely, the English translation of Fillaeu de Saint-Martin's adaptation by John Phillips, where it was known which exactly original signs have been used to produce the signs of the J.Phillips English translation. All other *bf:Works* representing other members of the "Don Quixote" family are not related to one another due to the absence of information regarding the origins of each *bf:Work's* signs. Thus, BIBFRAME's stricter approach regarding the representation of bibliographic relationships renders many *bf:Works* orphan (*Figure 3:43*).

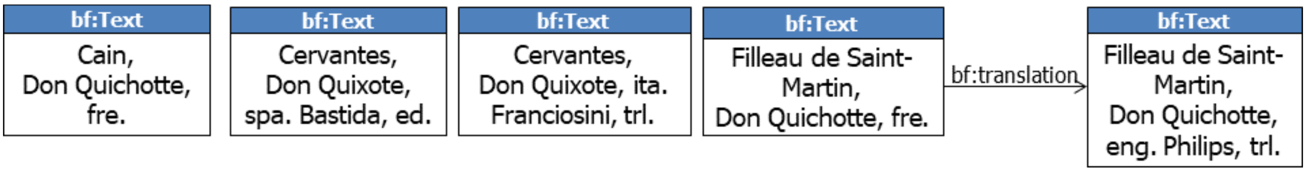


Figure 3:43. BIBFRAME representation of members of the "Don Quixote" bibliographic family. For readability reasons, bf:Instances and bf:Items are excluded from the representation.

It has already been mentioned that BIBFRAME enables varying representations due to its flexible (or sometimes ambiguous) definitions. Regarding the clustering of a bibliographic family's members there can be three alternative representations. Two of these representations use the *bf:hasExpression* property, and the third one uses the *bflc:Hub* class, a new class developed by the Library of Congress for BIBFRAME experimentations with their data.

The first representation (Figure 3:44) using the *bf:hasExpression* relates an *Expression*-agnostic *bf:Work* instance with other *bf:Work* instances that include its ideas and different sets of signs realizing them. This representation is somewhat close to the FRBR representation. One can observe that the *bf:Work* instances used as the object of the *bf:Work-bf:hasExpression-bf:Work* statement include both ideas and signs. Thus, even the representation shows some similarities with the FRBR representation, there is a semantic difference; the *bf:Work* instances used as the objects of the *bf:Work-bf:hasExpression-bf:Work* statements cannot be considered semantically equivalent to the FRBR *Expression* entity instances. In FRBR (*Figure 3:39*) the *Expression* instances carry only the signs used to realize the ideas of the progenitor *Work*. By contrast, in *Figure 3:44*, the *bf:Work* instances used as the objects of the *bf:Work-bf:hasExpression-bf:Work* statements carry both the ideas of their progenitor *bf:Work* (the *bf:Work* instances used as the subjects of the *bf:Work-bf:hasExpression-bf:Work* statements) and the signs used to realize these ideas.

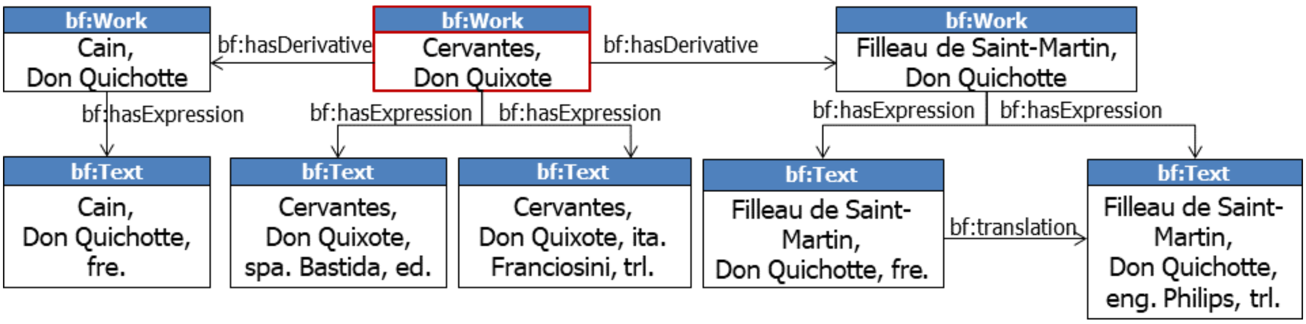


Figure 3:44. Alternative BIBFRAME representation of members of the "Don Quixote" bibliographic family. In this representation the bf:hasExpression property is used to relate Expression-agnostic bf:Works with other bf:Works containing their realizations. For readability reasons, the progenitor bf:Work is marked with a red outline, and bf:Instances and bf:Items are excluded from the representation.

The second representation (*Figure 3:45*) also uses the *bf:hasExpression* relating in pairs all *bf:Work* instances realizing the same ideas. All related *bf:Works* contain some pieces of identical information. These pieces of information refer to Cervantes as the creator of a set of ideas entitled "Don Quixote". In this representation, the progenitor is implied by relating all *bf:Works* that contain sets of signs realizing it. It can be safely assumed that in a large bibliographic family, the number of *bf:hasExpression* instances will be correspondingly large relating its members in pairs.

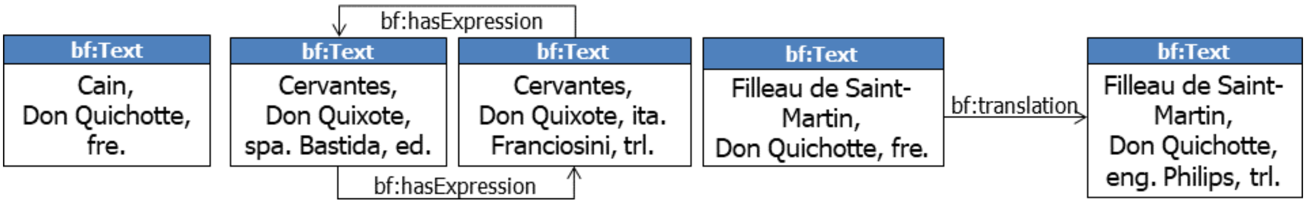


Figure 3:45. Alternative BIBFRAME representation of members of the "Don Quixote" bibliographic family. In this representation the bf:hasExpression property is used to relate in pairs all bf:Works containing realizations of the same progenitor. For readability reasons, bf:Instances and bf:Items are excluded from the representation.

The third representation uses the newly-developed *bflc:Hub* class that has not been incorporated in the official BIBFRAME yet remaining a local construct developed and used by the Library of Congress. Uses of the *bflc:Hub* class have been presented in the 3rd European BIBFRAME Workshop (K. Ford, 2019a, 2019b; McCallum, 2019). One of them has been the use of the class as semantically equivalent to the RDA *Work* class. This representation is presented in Figure 3:46. Despite the similarity to the representation in *Figure 3:44*, there are two semantic differences between this representation and the RDA one. First, even though the *bflc:Hub* class clusters some *bf:Work* instances, it may not be considered as semantically equivalent to the RDA *rdac:C10001 Work* class. As already stated in Kevin Ford's presentations (K. Ford, 2019a, 2019b), the *bflc:Hub* class has been designed as an aggregation mechanism to "collect like or related Things". In Figure 3:46, it is used to aggregate all *bf:Works* that realize the same set of ideas. Nevertheless, *bflc:Hub* may be used to aggregate other "like Things". In Figure 3:47, the *bflc:Hub* is used to relate all *bf:Works* containing realizations of Cervantes' Don Quixote in Italian. In this representation the *bflc:Hub* class is not semantically equivalent to the *rdac:C10001 Work* class either. Regarding the second difference, despite the resemblance of the BIBFRAME representation to WEMI representations in FRBR and RDA, the *bf:Work* instances are not equivalent to FRBR/RDA *Expressions.* They still carry both ideas and signs according to the semantics of the *bf:Work class.*

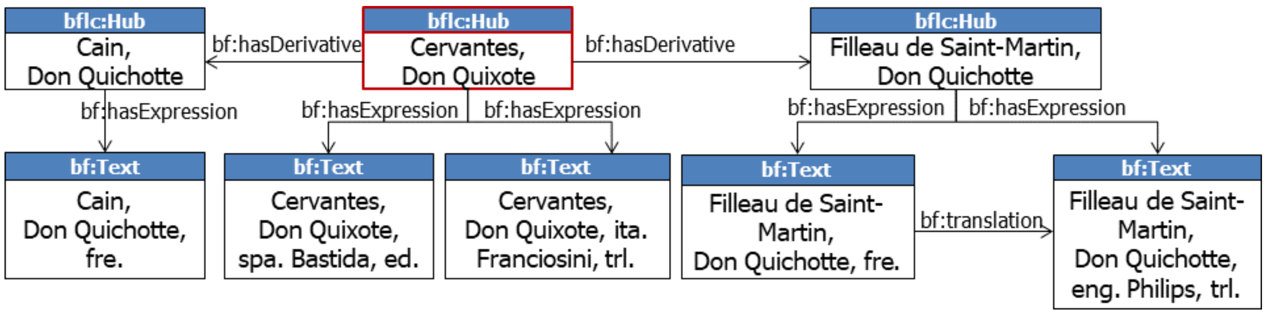


Figure 3:46. Alternative BIBFRAME representation of members of the "Don Quixote" bibliographic family. In this representation the bflc:Hub class is used to aggregate bf:Works containing realizations of the same progenitor. For readability reasons, the progenitor bflc:Hub is marked with a red outline, and bf:Instances and bf:Items are excluded from the representation.

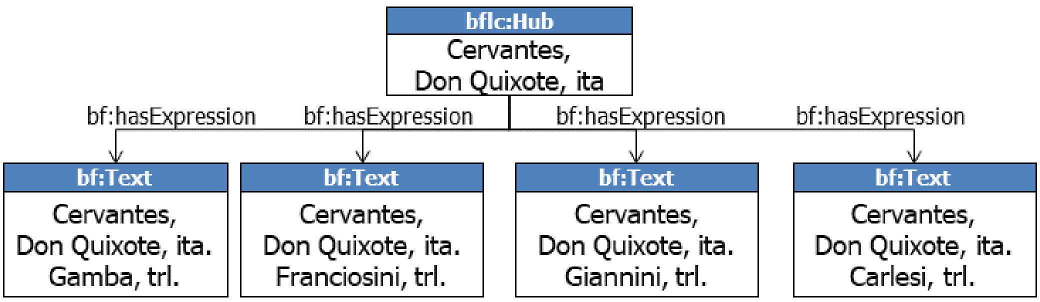


Figure 3:47. Use of the bflc:Hub class to aggregate all bf:Works realizing Don Quixote in Italian.

According to the **EDM** library alignment report (Angjeli, Bayerische, et al., 2012), the *edm:ProvidedCHO* class accommodates ‘editions’, thus, incorporating information regarding three FRBR Group 1 entities, i.e., *Work, Expression,* and *Manifestation.* As a result, two characteristic problems observed in legacy library catalogs are also present in EDM: i) data exhibiting a flat structure, and ii) difficulty in collocation of related resources. Due to the EDM’s domain, that of cultural heritage, and Europeana’s aggregating role, the clustering of related *edm:ProvidedCHO* instances occurs on a provider basis using the *ore:Aggregation* class (*Figure 3:48*).

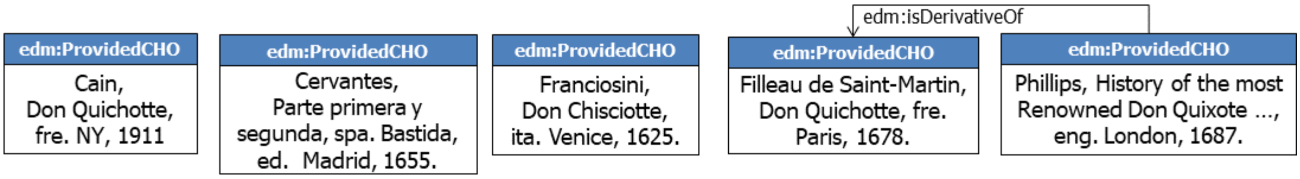


Figure 3:48. EDM representation of some of the members of the "Don Quixote" bibliographic family. For readability reasons, ore:Aggregation and edm:WebResource instances are excluded from the representation.

3.3. Similarities and divergences between the models

The study of the representation of core entities and inherent relationships, as well as of bibliographic relationships, and bibliographic families has revealed similarities and divergences between the models. In the following paragraphs these similarities/differences are presented and based on the Haslhofer & Klas (Bernhard Haslhofer & Klas, 2010) categorization of metadata interoperability techniques (Figure 1:1). All identified similarities and heterogeneities are summed up into one table at the end of the chapter.

3.3.1. Single-volume monographs, elemental and simple works

The investigation suggests that all models studied may describe the case of single-volume monographs. In this case important similarities and heterogeneities have been identified. The former are expected to facilitate future mappings and therefore the interoperability between the models, while the latter are expected to impede it.

An obvious semantic similarity is that all studied models may be characterized as belonging more or less to the bibliographic domain. Therefore, they all capture the same or similar pieces of information regarding bibliographic products. EDM is in the cultural heritage domain that is wider than the bibliographic one. Due to their common/similar domains, common terminology has been observed. As an example, the four FRBR WEMI disjoint entities are used in all FRBR-inspired models, i.e., the consolidated FRBR – IFLA LRM, RDA, and FRBRoo. Another example is the naming of exemplars, all studied models (except EDM) use the term “Item” for exemplars. Despite all that, terminological mismatches have been also observed. The most important ones are the use of Entity-Relationship (E-R) modeling terms contrary to RDF terms and the naming of the BIBFRAME *Work* class that is semantically different to the FRBR *Work* class. FRBR and LRM use ER modeling terms (e.g., entity, attribute, relationship), while FRBRoo and BIBFRAME use RDF terms (e.g., class, property). Interestingly, the RDA uses both; the RDA documentation uses the ER modeling terms, while the RDA registry uses the RDF ones.

A syntactic similarity is that all models use the RDF model and RDF schema. RDF representation constructs are used to represent the bibliographic entities that each model identifies; entities are represented as classes, while properties are used to represent either the entities’ attributes, or the relationships (inherent or bibliographic) between them. FRBR has been expressed in RDF by Gordon Dunsire (Dunsire, 2015), RDA, FRBRoo, EDM and BIBFRAME are all developed in RDF. The representation of the IFLA LRM in RDF is anticipated (Riva, Bœuf, et al., 2017a). Use of a common syntactic language is an important contributor in both technical and semantic interoperability; each model’s exact semantics are defined in a common language so that semantic similarities and differences may be identified explicitly. An important abstraction level incompatibility relevant to the representation constructs is that each model enables different representation approaches for the same real-world case, even the simplest one that is single-volume monographs.

A major abstraction level similarity is that all models differentiate between content and carrier (*Table 3-1*). FRBR recognizes four core entities, where *Work* and *Expression* are at the intellectual level representing ideas and signs respectively. FRBR *Manifestation* and *Item* entities are at the material embodiment level representing the publication product and its exemplars respectively. These FRBR conceptualizations have remained in the consolidated IFLA LRM. They have also been used in RDA and FRBRoo. Even though FRBR has been a milestone in the evolution of bibliographic description and has been used as a point of reference by many of the studied models, there are differences in terms of varying interpretations of real-world entities and granularity. As an example, FRBRoo further refines the four FRBR entities and has added specialized classes for representing more distinct moments in the creation timeline. It differentiates between the signs created by the author (*F22 Self-Contained Expression*) and the signs used in the publication incorporating contributions by the publisher (*F24 Publication Expression*). It drops the *Manifestation* entity and defines the *F3 Manifestation Product Type* and the *F4 Manifestation Singleton* classes.

The distinction between content and carrier is present in BIBFRAME and EDM models also, even though both models are less granular than FRBR and all FRBR-inspired models. BIBFRAME defines three classes, where the *bf:Work* class accommodates information regarding ideas and signs. Thus, the *bf:Work* may be considered as equivalent to the union of two disjoint FRBR entities, i.e., *Work* and *Expression*. The *bf:Instance* and *bf:Item* classes are at the material embodiment level furnishing information regarding the publication product and its exemplars respectively. EDM defines two classes, where the *edm:ProvidedCHO* class contains information about ideas, signs, and the publication; the *edm:WebResource* class is used to represent the digital exemplars of the publication. Hence, the only class in all studied models that expands to both intellectual and material levels is the *edm:ProvidedCHO* class. The EDM-FRBRoo profile tries to deal with EDM descriptions, that contain information regarding three disjoint FRBR entities (*Work, Expression*, and *Manifestation*). The profile extends the semantics of the *edm:InformationResource* class by assigning newly-defined types (represented as two *skos:Concept* class instances having the literal values "FRBRWork" and "FRBRExpression")(*Figure 3:11*). *Table 3-1* presents each model's core entities or classes along with the intellectual or material embodiment level they belong to.

Table 3-1. Core entities/classes clustered according to intellectual and material embodiment characteristics, expressing semantic and structural commonalities and heterogeneities.

Level \ **Models**		**FRBR**	**LRM**	**RDA**	**FRBRoo**	**BIBFRAME**	**EDM**
Intellectual	Concepts / Ideas	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.

Cases of direct and multilateral correspondences may be observed in *Table 3-1*. Some direct correspondences may be observed, e.g., FRBR *Manifestation* entity with the RDA *rdac:C1007 Manifestation* class, and the BIBFRAME *bf:Instance* class. Another example of direct correspondence is the representation of exemplars, namely the FRBR *Item* entity, the *LRM-E5 Item,* the RDA *rdac:C1003 Item* class, the FRBRoo *F5 Item* class, the BIBFRAME *bf:Item* class, and the EDM *edm:WebResource* class. Multilateral correspondences occur when multiple classes of one model may correspond to a class in another model, and vice versa (Bernhard Haslhofer & Klas, 2010). The *Work* entity is represented as one entity/class in FRBR, IFLA LRM, and RDA. It corresponds to FRBRoo *F1 Work* class and its subclasses. Similarly, the *Expression* entity -represented as one entity/class in FRBR, IFLA LRM, and RDA- corresponds to the FRBRoo *F2 Expression* class and its subclasses. The union of the FRBR, IFLA LRM, and RDA *Work* and *Expression* entities corresponds to the BIBFRAME *Work* class. Likewise, the union of the FRBRoo *F1 Work* and *F2 Expression* classes and their subclasses coincide the BIBFRAME *Work* class. The EDM *ProvidedCHO* class equates to three disjoint classes, i.e., to the FRBR/IFLALRM/RDA models' *Work*, *Expression,* and *Manifestation* entities, and in the case of FRBRoo to *F1 Work* and subclasses, *F2 Expression* and subclasses, and either *F3 Manifestation Product Type* or *F4 Manifestation Singleton* class.

A *meta-level discrepancy* has also been identified between FRBR and BIBFRAME. Content type is represented in FRBR as a value of an *Expression* attribute, namely the *form of expression* attribute, e.g., *form of expression:*

alpha-numeric notation. In BIBFRAME, the same information is represented with *bf:Work* subclasses, e.g., the *bf:Text* subclass for alpha-numeric notation as depicted in *Figure 3:7*.

3.3.2. Bibliographic relationships

An important similarity is that all models offer constructs for the representation of bibliographic relationships. The way each model represents a relationship is described using statements in Table 3-2. It must be noted that this table does not include all possible representations that each model may support for each studied bibliographic description case. The representation approaches that adhere to common perceptions in the library community are only included in it, in the hope that these representation approaches are likely to present similarities between the models rendering mappings easier. Studying the statements in *Table 3-2* reveals that there is an abstraction-level consensus between the models regarding the representation of the derivative/adaptation, the equivalence/reproduction, and the aggregates. The derivative relationship of adaptation is represented in all models at the most abstract entity/class level, namely i) the *Work* entity in FRBR, IFLA LRM, RDA, and FRBRoo, ii) the *Work* class in BIBFRAME, and iii) the *edm:ProvidedCHO* class in EDM. The equivalence/reproduction relationship is represented at the material embodiment level, i.e., *Manifestation* in FRBR, IFLA LRM, RDA, *F3 Manifestation Product Type* in FRBRoo, *bf:Instance* in BIBFRAME, and *edm:ProvidedCHO* class in EDM. Aggregates are represented as embodiments of different signs in the same publication adhering to the "aggregates-as-manifestations" approach proposed by the FRBR WG on Aggregates (O'Neill et al., 2011).

In contrast to the other relationships presented in Table 3-2, the derivation/translation relationship presents an *abstraction-level incompatibility.* In FRBR, IFLA LRM, and RDA translation is represented at the signs level, meaning between *Expressions*. This is the realization approach that is common in the libraries; translations are considered as new realizations of the same *Work.* On the other hand, FRBRoo implements the derivation approach where a translation is considered as a new *Work* (concepts level). Translation is represented as type of a derivative relationship (*R2i has derivative*) relating two *F14 Individual Work* instances. There are two things that must be noted. First, FRBRoo enables the representation of translation at the signs level (*Expressions*) according to the realization approach. Yet, the relationship is represented with a generic property, the *R14 incorporates* one (see *Figure 3:17*). Secondly, the IFLA LRM may support the representation of translation according to both derivation and realization approaches. Similarly to the FRBRoo, the IFLA LRM constructs enable the representation of translation between *Works*; yet, in the model specification, the examples follow the realization approach that is most common among libraries and librarians' perceptions. BIBFRAME represents the translation relationship between *bf:Work* instances that are at both concepts and signs level (see *Table 3-2*).

Another *abstraction-level incompatibility* can be observed in the FRBRoo representation of aggregates. All other models represent aggregates at the material embodiment level. The FRBRoo representation uses the signs level, where multiple *F2 Expressions* may be incorporated in either *F22 Self-Contained Expressions* or *F24 Publication Expressions.* In case the *F22 Self-Contained Expression* class is used (example in *Figure 3:33*), then this class instance represents the aggregating *Expression* that, contrary to the FRBR WG on Aggregates report's aggregating *Expression,* represents the signs created by the aggregator's effort and the signs from all aggregated *Expressions.* Closer to the "aggregates-as-manifestations" approach is the representation of aggregates using the *F24 Publication Expression* class (see *Figure 3:34*). This class represents the signs used in a publication including all incorporated *F2 Expressions* and the publisher's signs created and used for this specific *F24 Publication Expression* and the *F3 Manifestation Product Type* instance carrying it.

Even though *multilateral correspondences* are not observed in *Table 3-2,* there exist regarding the representation of derivative relationships in BIBFRAME and EDM. The BIBFRAME *bf:hasDerivative* property is a generic one that may be used for the representation of all derivative relationships, except for translation, e.g., revision, abridgement, adaptation, summarization, etc. As an example, the *bf:hasDerivative* property can be used for multiple derivation cases, such as adaptation, summarization, transformation, revision, etc. Therefore, it corresponds to many properties in RDA, such as *rdae:P20203 is derivative (expression), rdae:P20166 is abridged as (expression), rdae:P20153 is adapted as (expression),* and *rdae:P20211 is revised as*. Similarly in EDM, the *edm:isDerivativeOf* property may be used for the representation of translation, summarization, abstraction, while the *dcterms:isVersionOf* may be used for other editions and adaptations.

Meta-level discrepancies appear when there are structural representation mismatches between the models. Even though, the same semantics are captured, they are represented in a different way. In some of the models the relationships are represented with a specific property, e.g., *rdaw:P10155 is adapted as work* for the representation of adaptation in RDA, while in others the relationship is represented by typing a more generic property, e.g., the *R2i has derivative* property typed as “adaptation” or as “translation”. Such discrepancies can be observed in the IFLA LRM representation of translation, and the FRBRoo representations for the adaptation, translation, and reproduction cases. It must be noted that the FRBRoo, only in the case of fair-use reproduction, enables the representation of reproduction using just a property, the *R31 is reproductionOf* property relating *F5 Item* instances (see *Figure 3:27*).

Another notable example regarding meta-level discrepancies is the representation of bibliographic relationships in RDA in contrast to FRBRoo and LRM. RDA uses a great number of specific properties to represent each bibliographic relationship, e.g., *rdae:P20110 is adapted as libretto (expression)*. FRBRoo and LRM assign generic properties with different “type values”, to represent specific bibliographic relationships, e.g., *LRM-R24 is derivation of (type: abridgement).* It must be noted that in some properties there are predefined “type values”, while for others there is not. As an example, in LRM the *Work*-level *LRM-R22 is a transformation of* property may be refined but no “type values” are suggested. Contrary, the *Expression*-level *LRM-R24 is derivation of* property may be refined using the following four “type values”: abridgement, revision, translation, musical arrangement. In FRBRoo the *R2 is derivative of* property may be “typed” with the *R2.1 has type* property. For consistency, the *R2.1 has type* property may have as values only the following ones: “Abridgement”, “Adaptation”, “Arrangement”, “Imitation”, “Revision”, “Summary”, “Transformation”, “Translation”. The *P130 shows features of* property may be typed with the *P130.1 kind of similarity: E55 Type* path where the *E55 Type* class may have as values “Reproduction” or any other type of alternate format. Thus, even though the LRM and FRBRoo provide an extension mechanism to represent different derivation cases, there is not a consensus about which derivative relationships are going to be used as “type values”. The existence of a common controlled vocabulary regarding bibliographic relationships could be used for typing the corresponding properties in both models.

Table 3-2. 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.*

3.3.3. Bibliographic families and progenitor *Works*

Structural similarities/heterogeneities have been identified at the abstraction level. In FRBR and FRBR-inspired models, the progenitor of a bibliographic family is implicitly represented as a *Work* with many relationships to other *Works* and with *Expressions* clustered under it (e.g., *Figure 3:39*, *Figure 3:40*, *Figure 3:41*, *Figure 3:42*). Contrary to the other FRBR-inspired models, in FRBRoo the progenitor *Work* may be explicitly represented with the *F15 Complex Work* class (see the example in *Figure 3:42*). Instances of the *F15 Complex Work* class instances may have other *F1 Work* instances as members. And the author in these member-*Works* may even be different from the progenitor's one. In BIBFRAME, clustering of *bf:Works* is possible only when there is information about a bibliographic relationship referring to original and derivative signs. Otherwise, *bf:Works* become 'orphan'. Alternative representations enabled by the ambiguous BIBFRAME definitions have been presented. These representations -when seen as graphs- may misguide to believing that there are similarities between them and the corresponding FRBR WEMI graphs. Yet, the thesis has demonstrated that there are important semantic differences regarding the entities/classes involved in these graphs.

The thesis has some reservations regarding the newly-developed *bflc:Hub* class. Despite of the inexistence of a *bflc:Hub* official definition ("LC Linked Data Service: BIBFRAME Ontology LC Extension," 2017), it seems that this class will be used for more than one uses (K. Ford, 2019a, 2019b; McCallum, 2019) raising the possibility of more structural heterogeneities. Both multilateral correspondences and meta-level discrepancies are expected to be observed. A multilateral correspondence is the possible use of the *bflc:Hub* class to abstractly represent sets of ideas (similar to the FRBR *Work* entity semantics), but also to aggregate, e.g., the example of all English translations of Pippi Longstocking in (K. Ford, 2019a). Thus, in this case the *bflc:Hub* class is used to represent the semantics of two disjoint FRBR classes, i.e., a *Work* identified with the following author-title string "Lindgren, Astrid, 1907-2002. Pippi Långstrump" and a subset of this *Work's Expressions* having as shared characteristic the English language: "Lindgren, Astrid, 1907-2002. Pippi Långstrump. English". Staying on the example of "Lindgren, Astrid, 1907-2002. Pippi Långstrump. English", another structural heterogeneity may be identified, that of *meta-level discrepancy*. The information of language regarding textual realizations is represented at the *Expression* level in FRBR, LRM, RDA, and FRBRoo, at the *bf:Work* level in BIBFRAME, and at the *edm:ProvidedCHO* class in EDM. Many *Expressions* of the same *Work* may be realized in the same language by different contributors, e.g., different translators of the same *Work.* The shared characteristic of language is represented as a common value of an *Expression* attribute in FRBR, LRM, RDA, and FRBRoo, while in BIBFRAME is represented with an instance of the *bflc:Hub class.* This meta-level discrepancy proves that differing representation approaches for the same real-world information (here the language used in a realization) regulate cataloging policies impacting on the interoperability of data. Selecting a specific representation approach is a matter of a cataloging policy. Thus, catalogers should take under consideration common representation approaches to select the one that will be implemented by their library's cataloging policy. With regard to the representation of bibliographic families, the *bflc:Hub* seems to enable the clustering of *bf:Works* under various perceptions, e.g., all *bf:Works* realizing the same ideas in a specific language, or all *bf:Works* translated by the same person. Differing uses of the *bflc:Hub* class for the clustering of *bf:Works* will likely result in fragmentation of a bibliographic family's *bf:Works* in many *bflc:Hub* instances and may ultimately hamper the explorability and the interoperability of data. The interoperability of the *bflc:Hub* class remains an open issue to be further studied once the Library of Congress publishes the class' official definition.

Lastly, a *domain conflict* has been observed between the EDM, which is a cultural heritage model created to serve the Europeana aggregation service, and the other library conceptual models. In EDM, descriptions are organized per their provider using *ore:Aggregation* class and *ore:Proxy* instances. Such primitive does not exist in the other studied models. The further study of this heterogeneity was excluded from the mappings as out of scope.

3.4. Conclusions

This chapter has investigated the representation of common bibliographic description cases using the studied models' constructs. The cases studied are single-volume monographs, bibliographic relationships, and bibliographic families. Regarding the bibliographic relationships, the thesis focused on the derivative relationship, the equivalence relationship, and aggregates.

Each case has been successfully represented using each model's constructs. Nevertheless, each case may be represented with more than one way in each model. The thesis has used the term 'representation approach' for these different ways. Hence, the path-oriented approach was used to explicitly describe the representation(s) supported by each model's constructs. The selection of a specific representation approach is a matter of policy adopted by each library/cataloguing agency. Despite the variety of approaches in modeling the studied real-world bibliographic description cases using a model's constructs, there are some approaches that are met more often in the library environment. Such an example is the realization approach regarding the representation of translations as new realizations of the same *Work.* The thesis has identified these representation approaches presenting a common perception for each studied case (single-volume monographs, bibliographic relationships, and bibliographic families). Under the assumption that the common perceptions in the identified representation approaches shall enable the interoperability between the models, they have been selected for inclusion in two tables, namely Table 3-1 and Table 3-2. Table 3-1 presents each model's core entities or classes, while Table 3-2 presents the representation of relationships in each model. The comparison revealed similarities as well as important semantic and structural heterogeneities.

All semantic and structural heterogeneities identified during the study of how single-volume monographs, bibliographic relationships (derivative, equivalence, and aggregates) and bibliographic families are represented in each model, are presented in Table 3-3. The assumption that representation approaches expressing common perceptions regarding the bibliographic world will likely enable the interoperability between the studied models is going to be tested in the development of mappings. Even if the desired mappings developed successfully, new research questions would emerge with regard to representation approaches. What are the representation approaches that each model enables for the representation of other bibliographic description cases not studied by this thesis? Which ones of them disclose common perceptions among librarians? Can these common perceptions be implemented in cataloging policies so that the produced data becomes interoperable? Will future cataloging policies be oriented to representation approaches of common bibliographic description cases? Will future cataloging policies consider the RDF graph representation of the produced library data for the Semantic Web?

In the next chapter of the thesis, mappings are developed considering the representation approaches, and the semantic and structural heterogeneities identified in the present chapter. Each identified heterogeneity is a problem to be solved for the sake of interoperability. The thesis' approach in solving them along with the mappings created, are presented in chapter 4.

Table 3-3. Semantic and structural similarities/heterogeneities among the studied models, FRBR, LRM, RDA, FRBRoo, BIBFRAME, and EDM. The table is inspired by the Haslhofer and Klass' work on metadata interoperability (Bernhard Haslhofer & Klas, 2010).

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	
Structural	Abstraction level (in)compatibilities	<u>Representation constructs</u>	
		RDF as common syntactic language. Entities/classes with properties as attributes or as relationships. Relationships either inherent or bibliographic	Different representation approaches enabled by each model. There might be differences even between datasets using the same model.
		<u>Core entities/classes</u>	
		Content vs carrier	Different abstractions FRBR, RDA, LRM: four entities - WEMI BIBFRAME: three classes - WII FRBRoo: drops Manifestation/author vs publisher's signs (F24 Publication Expression) EDM: almost no granularity / EDM-FRBRoo: typed *edm:InformationResource* instances
		<u>Representation of relationships</u>	
		Adaptation – most abstract entity/class Equivalence/reproduction – embodiment level Aggregates-most models as signs embodied in manifestation	Translation: signs level (FRBR, LRM, RDA), concepts level (FRBRoo - derivation approach), and both concepts/signs (BIBFRAME). Aggregates - FRBRoo at signs level.
		<u>Bibliographic families</u>	
		FRBR, LRM, and RDA: clustering using progenitor *Work*	FRBRoo: *F15 Complex Work* with other *F14 Individual Works* as members. BIBFRAME: clustering possible only if there is known connection between original and derivative signs / alternative representations with *bf:hasExpression / bflc:Hub* EDM: provider-oriented clustering.
	Direct / Multilateral correspondences	<u>Classes</u>, e.g., FRBR *Manifestation / rdac:C10007 Manifestation / bf:Instance* <u>Inherent relationships</u>, e.g., FRBR *is embodied in/ rdae:P20059 / bf:hasInstance* <u>Bibliographic relationships</u>, e.g., FRBR *has a translation* /BIBFRAME *bf:translation*	<u>Classes</u>, e.g., *bf:Work* equals the FRBR *Work-is realized through-Expression* path *edm:ProvidedCHO* equals three disjoint FRBR entities (WEM) <u>Relationships</u>, e.g., *bf:hasDerivative* property (adaptation, summarization, transformation, etc.) edm:isDerivativeOf (translation, summarization, abstraction) dcterms:isVersionOf (versions, editions and adaptations)
	Meta-level matches / discrepancies	Information about the same real-world objects is captured/represented using same constructs, e.g., embodiment is captured using classes in both RDA (*rdac:C10007*) & BIBFRAME (*bf:Instance*).	BIBFRAME uses classes where other models use properties, e.g., content type represented with attributes/properties in FRBR/RDA, and with *bf:Work* subclasses in BIBFRAME. RDA uses specific properties, while FRBRoo & LRM generic properties that can be 'typed' with specific values.
	Domain coverage		EDM. The providers' descriptions are really important and represented with ore:Proxy & ore:Aggregation classes.

4. Mappings for achieving semantic interoperability between the models [4]

The heterogeneities observed in the previous chapter need to be overcome for the scope of interoperability. Haslhofer & Klas in (Bernhard Haslhofer & Klas, 2010) identify three methods to achieve interoperability between models: (1) common use of a specific conceptual model; (2) development of a new meta-model with which all other models need to comply, or development of application profiles; and (3) model reconciliation through crosswalks, mappings, or instance transformations that untangle identified structural and semantic heterogeneities. The two first methods are likely to fail; common use of a specific model has not been achieved in the library domain (Baca, 2016; Register et al., 2009) even in the previous years that the AACR2 rules and the MARC formats were the prevalent cataloguing tools used worldwide. Moreover, nowadays with the development of a plethora of digital projects, and of library linked data projects, there have been created so many models and application profiles that the danger of incompatible silos has already been stated (Suominen & Hyvönen, 2017). The second method of introducing one upper level, that of a meta-model, has not proven to be successful. The FRBR model, nearly after 20 years its publication, is considered a significant milestone in the evolution of cataloguing theory (Denton, 2007). It has surely influenced most models that were developed afterwards, and in some way, it can be considered as a bibliographic conceptual model universally accepted. Yet, there are differences between the models, as already presented in Chapter 3, and these differences exist even between FRBR-inspired models, such the RDA and FRBRoo. Another type of meta-model agreement suggested in (Bernhard Haslhofer & Klas, 2010), that of developing application-profiles, has been considered in the thesis as a plausible way to achieve semantic interoperability in a specific context. The example of the EDM-FRBRoo application profile (Doerr et al., 2013) is a proof that the EDM may accommodate FRBR semantics by using FRBRoo classes and properties, and by exploiting the EDM extension mechanism. The thesis tests the meta-model agreement method by developing a BIBFRAME-EDM application profile.

[4] This chapter revisits and expands the studies published in the following papers:

- Zapounidou S., Sfakakis M., & Papatheodorou C. (2014). Integrating library and cultural heritage data models: the BIBFRAME - EDM case. In: *Proceedings of the 18th Panhellenic Conference on Informatics. PCI 2014*. ACM. doi:10.1145/2645791.2645805.
- Zapounidou S., Sfakakis M., & Papatheodorou C. (2014). Library Data Integration: Towards BIBFRAME Mapping to EDM. In: Closs S., Studer R., Garoufallou E., Sicilia MA. (eds). *Metadata and Semantics Research. MTSR 2014. Communications in Computer and Information Science, vol 478*. Springer, Cham. doi:10.1007/978-3-319-13674-5_25.
- Zapounidou, S., Sfakakis, M., & Papatheodorou, C. (2017). Representing and integrating bibliographic information into the semantic web: A comparison of four conceptual models. Journal of Information Science, 43(4), 525-553. doi:10.1177/0165551516650410.
- Zapounidou S., Sfakakis M., & Papatheodorou C. (2017). Preserving Bibliographic Relationships in Mappings from FRBR to BIBFRAME 2.0. In: Kamps J., Tsakonas G., Manolopoulos Y., Iliadis L., Karydis I. (eds) Research and Advanced Technology for Digital Libraries. TPDL 2017. Lecture Notes in Computer Science, vol 10450. Springer, Cham. doi:10.1007/978-3-319-67008-9_2.
- Zapounidou S., Sfakakis M., & Papatheodorou C. (2019). Assessing the Preservation of Derivative Relationships in Mappings from FRBR to BIBFRAME. In: Garoufallou E., Sartori F., Siatri R., Zervas M. (eds) Metadata and Semantic Research. MTSR 2018. Communications in Computer and Information Science, vol 846. Springer, Cham. doi:10.1007/978-3-030-14401-2_22.
- Zapounidou S., Sfakakis M., & Papatheodorou C. (2019). Mapping Derivative Relationships from RDA to BIBFRAME 2. *Cataloging & Classification Quarterly 57* (5): 278-308. doi:10.1080/01639374.2019.1650152.
- Sfakakis M., Zapounidou S., & Papatheodorou C. (2020). Mapping derivative relationships from BIBFRAME 2.0 to RDA. *Cataloging & Classification Quarterly 58.* doi: 10.1080/01639374.2020.1821856.

Mappings are the third method suggested in (Bernhard Haslhofer & Klas, 2010), and they include four phases: (1) mapping discovery; (2) mapping representation; (3) mapping execution; and (4) mapping maintenance. The identification of heterogeneities in Chapter 3, may serve the first phase of mapping discovery. This chapter presents the mapping representation for three cases: 1) FRBR-BIBFRAME, 2) RDA-BIBFRAME, and 3) BIBFRAME-RDA.

It must be noted that the mappings in this chapter focus on core entities, inherent relationships, and derivative relationships. The mapping of each model's core representation mechanisms (core entities and inherent relationships) is studied. Derivative relationships are studied as well, due to their virtue of being one of the core representation mechanisms that enable the expansion of a bibliographic family and the exploration among the family members.

4.1. Meta-model agreement: investigation toward a BIBFRAME-EDM application profile

A BIBFRAME-EDM application profile can be implemented only for digital or digitized resources, since Europeana does not include descriptions about resources in legacy formats. Therefore, the BIBFRAME path "*bf:Work-bf:hasInstance-bf:Instance-bf:hasItem-bf:Item-bf:electronicLocator-URI*" is considered prerequisite for the BIBFRAME-EDM application profile. If this prerequisite path exists, then the mapping from BIBFRAME to EDM is feasible (Figure 4:1). It must be noted that a full BIBFRAME-EDM application profile was not developed; rather, the feasibility of creating one was investigated in terms of mapping BIBFRAME representations of core entities and inherent relationships, derivative relationship (adaptation, translation), equivalent relationship, and aggregates to EDM.

Figure 4:1. The BIBFRAME path that is a prerequisite condition for the BIBFRAME-EDM application profile.

The methodology selected for the creation of the BIBFRAME-EDM application profile takes under consideration the report for the alignment of library metadata with the EDM (Angjeli, Bayerische, et al., 2012), and the EDM-FRBRoo application profile (Doerr et al., 2013). The methodology involves the following steps:

1. Selection of library material types. The case of single-volume monographs was selected to investigate the applicability of the profile for the representation of BIBFRAME's core entities and inherent relationships in EDM.
2. Requirements. It is required that the profile will use the EDM definitions described in the alignment report (Angjeli, Bayerische, et al., 2012), and that the profile will be extensible to serve the needs for representing other library materials.
3. Selection of test case. The case of Don Quijote was selected using the representations presented in Chapter 3**.** In detail, the test case includes the French free translation of Don Quijote by Fillaeu de Saint-Martin and its translation in English by John Phillips. Both texts have been digitized by the National Library of Spain and are available online.
4. Representation of the test case in BIBFRAME, and in EDM.
5. Study all the possible EDM representation approaches (e.g., library alignment report, use of ore:Proxy class, etc.).
6. Mapping to EDM using a path-oriented approach.

The BIBFRAME representation of the test case (Figure 4:2) presents the French free translation as a *bf:Work* which is considered as an adaptation of the "Don Quixote" *bf:Work* instance and is related to it with a *bf:hasDerivative* property instance. The literal translation in English by John Phillips is represented as another *bf:Work* instance, and it is related to the "French translation_W" with a *bf:translation* property instance. The same test case using EDM constructs is depicted in Figure 4:3. It can be observed that there is no *edm:ProvidedCHO* class instance for Don Quixote due to the inexistence of a digital or digitized *bf:Item* instance. Two *edm:ProvidedCHO* class instances are represented, one for the French translation, and one for the English one. They are related with the *edm:isDerivativeOf* property instance. The prerequisite path exists in both French and English translation enabling the mapping of their representation to EDM. The *bf:Text* instance entitled "Don Quixote" does not present the prerequisite path and it is excluded from the mapping.

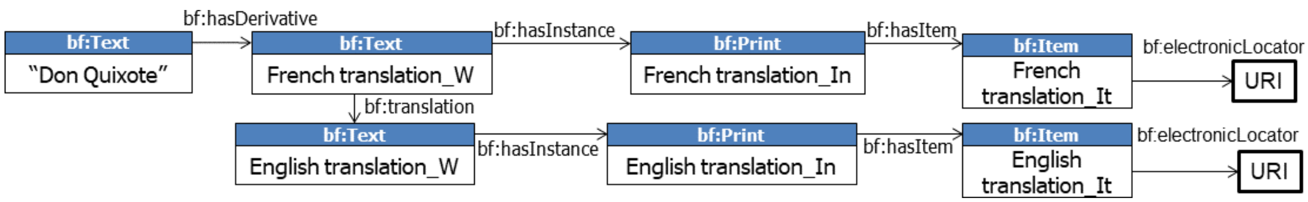


Figure 4:2. BIBFRAME representation of the selected test case.

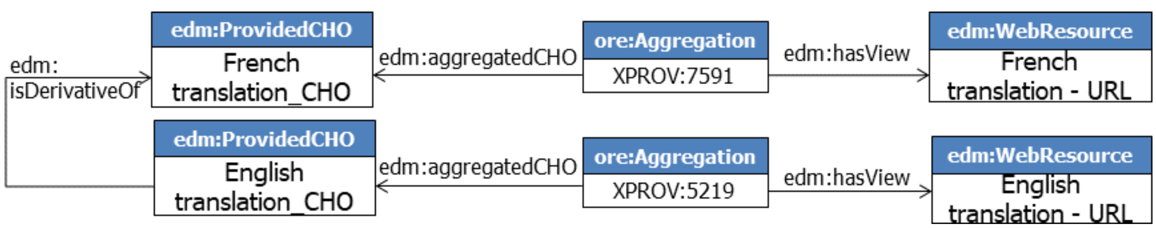


Figure 4:3. EDM representation of the test case.

4.1.1. Using the library metadata alignment report

According to the library metadata alignment report (Angjeli, Bayerische, et al., 2012), the *edm:ProvidedCHO* class is to incorporate information regarding three disjoint FRBR entities, namely *Work, Expression, and Manifestation.* Hence, the *bf:Work* class, which may be considered as equal to the union of the FRBR *Work* and *Expression* entities, along with the *bf:Instance* class, which is equivalent to the FRBR *Manifestation* entity, will be mapped to the *edm:ProvidedCHO* class. The *bf:Item-bf:electronicLocator-rdfs:literal(URI)* path will be mapped to an *edm:WebResource* class instance. This core mapping is visually displayed in Figure 4:4. The mapping of the BIBFRAME prerequisite path seems feasible prompting the investigation of all cases investigated in Chapter 3**.** The paths used for the BIBFRAME representation of adaptation, translation, reproduction, and aggregates were mapped to EDM ones. All these mappings are analytically presented in Table 4-1.

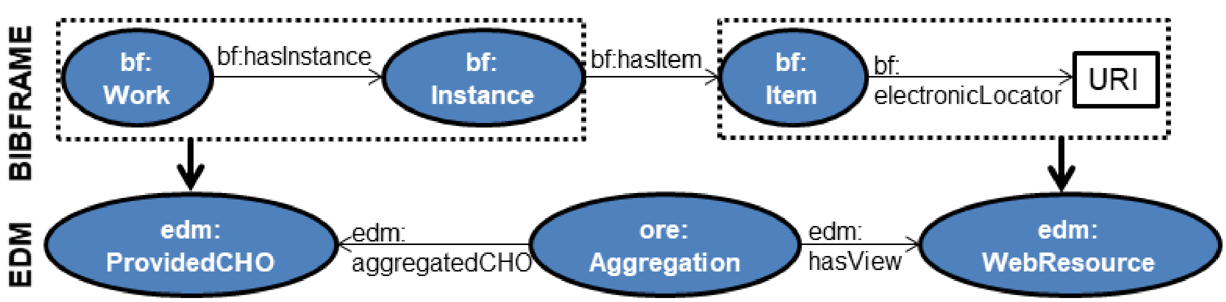


Figure 4:4. Mapping the BIBFRAME prerequisite path to EDM. This mapping adheres to the library metadata alignment report (Angjeli, Bayerische, et al., 2012).

Table 4-1 presents the mappings of four BIBFRAME paths describing five bibliographic description cases. First, the mapping of the BIBFRAME prerequisite path to EDM. This path is also valid for the representation of aggregates in BIBFRAME with more than one *bf:Work* instances having the same *bf:Instance*. Secondly, the mapping of two derivative relationships, the adaptation and the translation ones, and thirdly, the mapping of reproduction are attempted. Conforming to the library metadata alignment report (Angjeli, Bayerische, et al., 2012), all cases may be mapped to EDM. Yet, there are some important points to focus on.

With regard to aggregates, an EDM path (*edm:ProvidedCHO-inverse of edm:aggregatedCHO-ore:Aggregation-edm:hasView-edm:WebResource)* is to be created for each *bf:Work* instance aggregated to the same aggregate *bf:Instance.* Even though, each one of the produced EDM paths is correct, the information regarding aggregation is lost and may be implicitly represented due to the common URI value shared among the *edm:WebResource* instances of the produced EDM paths.

BIBFRAME represents all derivative-non translation relationships with one property *bf:hasDerivative* and its inverse *bf:derivativeOf*. Translation is represented with the *bf:translation/bf:translationOf* properties. Derivative relationships are all represented in EDM with *dcterms:hasVersion / dcterms:isVersionOf* properties. Translation is represented with the *edm:isDerivativeOf* property. Therefore, in case of translation, exact semantics is lost after mapping from BIBFRAME to EDM.

Since EDM includes descriptions only for cultural heritage objects available online, a condition for the mapping of the BIBFRAME reproduction representation to EDM is that both *bf:Instance* instances participating in the relationship are at the same time participating in a what we have called BIBFRAME prerequisite path, namely, *bf:Work-bf:hasInstance-bf:Instance-bf:hasItem-bf:Item-bf:electronicLocator-URI.*

Table 4-1. Mapping BIBFRAME representation paths to EDM ones.

BIBFRAME	**EDM**
bf:Work-bf:hasInstance-bf:Instance-bf:hasItem-bf:Item-bf:electronicLocator-URI	i. Map *bf:Work-bf:hasInstance-bf:Instance* to an *edm:ProvidedCHO* instance ii. Map *bf:hasItem-bf:Item-bf:electronicLocator-URI* to an *edm:WebResource* instance iii. Create the following path *edm:ProvidedCHO-inverse of edm:aggregatedCHO-ore:Aggregation-edm:hasView-edm:WebResource*
bf:Work-bf:translation-bf:Work	*edm:ProvidedCHO-inverse of edm:isDerivativeOf-edm:ProvidedCHO*
bf:Work-bf:hasDerivative-bf:Work	*edm:ProvidedCHO- dcterms:hasVersion -edm:ProvidedCHO*
bf:Instance-bf:reproduction-bf:Instance	If both instances of the class *bf:Instance* participate in a prerequisite path (defined above), then relate the two *edm:ProvidedCHO* instances with the *dcterms:hasFormat* property resulting in the following path *edm:ProvidedCHO-dcterms:hasFormat-edm:ProvidedCHO* else, ignore.

EDM enables alternative representations using the *ore:Proxy* class, or the *edm:InformationResource* class. Both representation approaches were included in this investigation along with one combining them both.

4.1.2. Use of the *ore:Proxy* class

As in the previous basic mapping, the BIBFRAME prerequisite path must be present to enable the mappings. The *ore:Proxy* class is to be used to contextualize the *edm:ProvidedCHO* descriptions that Europeana aggregates. Each provider's metadata can be accommodated in an *ore:Proxy* class instance. Thus, the mapping from BIBFRAME to EDM changes; the *bf:Work-bf:hasInstance-bf:Instance* path is now mapped to the *ore:Proxy* class instance and not to the *edm:ProvidedCHO* instance (Figure 4:5). The *edm:ProvidedCHO* instance, for which more than one description (*ore:Proxy* instances) may exist from different providers, is presupposed to exist in Europeana. Otherwise, it must be created automatically using the first provided *ore:Proxy* instance. In case multiple *ore:Proxy* instances for the same real-world object are provided, then matching and merging is needed to result in one *edm:ProvidedCHO* instance with multiple *ore:Proxy* instances. Though, the issue of how Europeana will recognize and merge identical objects' descriptions provided by different providers "remains open" (Isaac, 2013). Relationships between *bf:Works* will be mapped to properties relating *edm:ProvidedCHO* instances (according to Table 4-1).

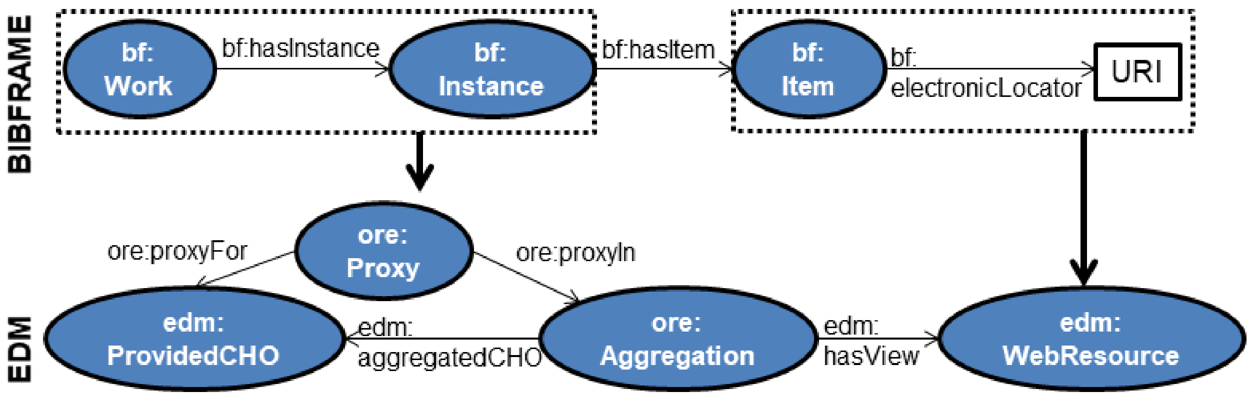


Figure 4:5. Mapping of the BIBFRAME prerequisite path to EDM using the ore:Proxy class.

4.1.3. Use of the *edm:InformationResource* class

Implementing the EDM-FRBRoo application profile approach, the *edm:InformationResource* class may be typed to represent *bf:Work* semantics (typed as *skos:Concept* instance with the literal value "BFWork"). In this representation approach (Figure 4:6), the *edm:ProvidedCHO* class may not represent "edition" as defined in the library metadata alignment report (Angjeli, Bayerische, et al., 2012), but the embodiment only. Thus, EDM may present more granularity and the BIBFRAME semantics may be better preserved after the mapping. Relationships between *bf:Work* instances will be mapped in this scenario to properties relating *edm:InformationResource* instances typed as "BFWork". In this representation, *bf:Work* instances that do not have a digital or digitized *bf:Instance* may also be represented at the *edm:InformationResource* level without being incorporated in some *edm:ProvidedCHO* instance.

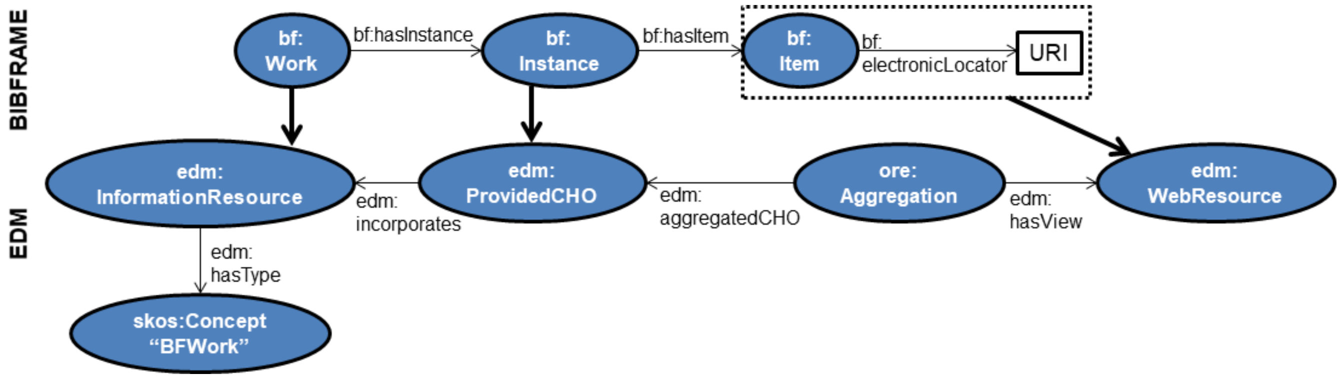


Figure 4:6. Mapping of the BIBFRAME prerequisite path to EDM using typed edm:InformationRecource *class instances.*

The representation of the BIBFRAME test case (Figure 4:2) using typed *edm:InformationResource* class instances is depicted in Figure 4:7. This representation is different to the representation depicted in Figure 4:3, in terms of preserving the "Don Quixote" *bf:Work* semantics, despite this *bf:Work* instance's absence from a prerequisite path.

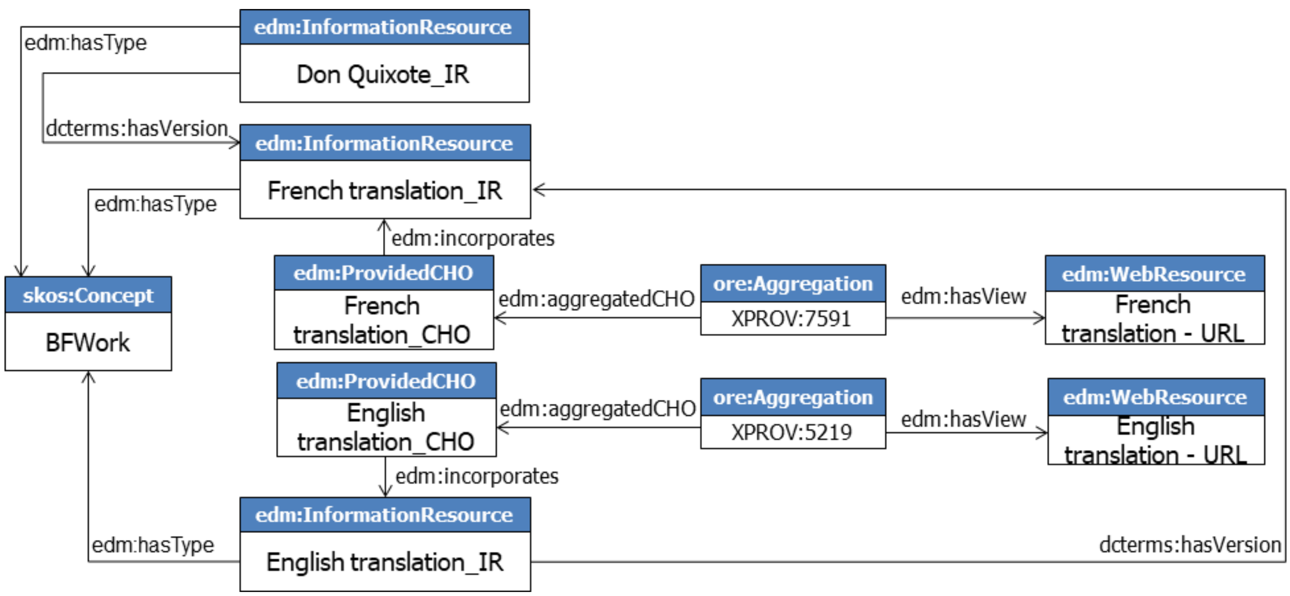


Figure 4:7. "Don Quixote" test case in EDM using typed edm:InformationRecource class instances.

4.1.4. Use of the *edm:InformationResource* and *ore:Proxy* classes

Regarding the representation of textual information resources in Europeana, a very interesting scenario would be the contextualization of descriptions and the preservation of BIBFRAME semantics at the same time. Thus, an EDM representation using both *edm:InformationResource* and *ore:Proxy* classes was investigated. The mapping of the BIBFRAME prerequisite path to EDM using the *edm:InformationResource* and *ore:Proxy* classes is presented in Figure 4:8. The *bf:Work* class is mapped to the *edm:InformationResource* class (typed as "BFWork"), the *bf:Instance* class is mapped to the *ore:Proxy* class, and the *bf:Item-bf:electornicLocator-rdfs:literal("URI")* path is mapped to the *edm:WebResource* class. Since *bf:Works* are mapped to *edm:InformationResource* class instances typed as "BFWork", relationships between *bf:Works* are mapped using the corresponding properties of Table 4-1 to relationships between the *edm:InformationResource* class instances.

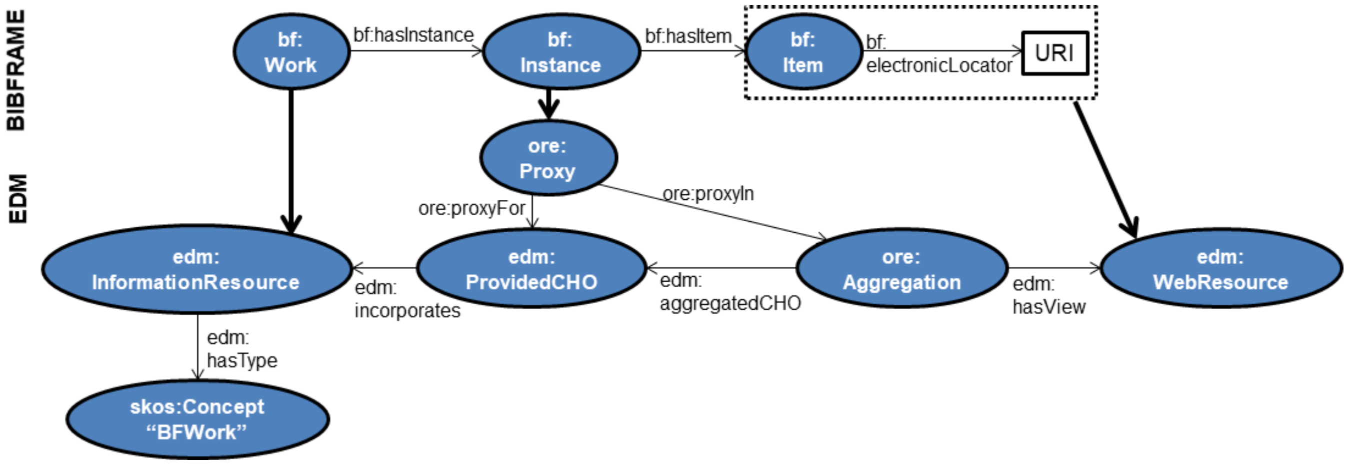


Figure 4:8. Mapping of the BIBFRAME prerequisite path to EDM using typed edm:InformationRecource and ore:Proxy instances.

The representation of the BIBFRAME test case (Figure 4:2) using typed *edm:InformationResource* and *ore:Proxy* instances is depicted in Figure 4:9. This representation is different to the representations depicted in Figure 4:3 and in Figure 4:7, in terms of preserving i) the "Don Quixote" *bf:Work* semantics, and ii) the provider's view regarding the two *edm:ProvidedCHO* instances (*French translation_CHO* and *English translation_CHO*).

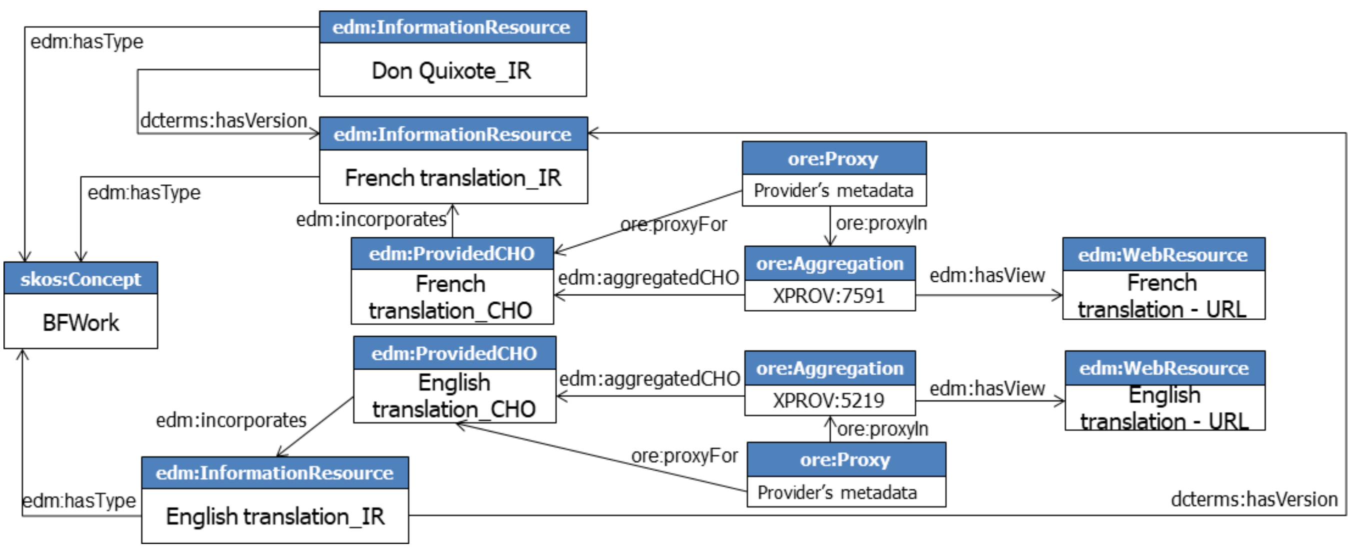


Figure 4:9. "Don Quixote" test case in EDM using typed edm:InformationRecource class instances, and ore:Proxy class instances.

4.1.5. Findings

The investigation toward the BIBFRAME-EDM application profile revealed that the BIBFRAME core entities and inherent relationships may be mapped to EDM. Due to the nature of EDM, a model for describing Europeana's aggregated digital resources, the following BIBFRAME path *"bf:Work-bf:hasInstance-bf:Instance-bf:hasItem-bf:Item-bf:electronicLocator-URI"* was identified as a prerequisite for the implementation of the application profile.

There are different representation approaches that can be implemented using the EDM constructs. The decision about which one is selected is a matter of policy. The implemented path-oriented approach enabled the mapping of the test case from BIBFRAME to the different EDM representations, i.e., library metadata alignment report, use of the *ore:Proxy* class, use of the *edm:InformationResource* class, and use of both *edm:InformationResource* and *ore:Proxy* classes. The last two representation approaches exploit the EDM's extension mechanism (assigning a new type to the *edm:InformationResource* class instances with *skos:Concept* instances) adding BIBFRAME semantics in EDM representations.

4.2. Model reconciliation: FRBR-BIBFRAME mapping

FRBR is considered as a milestone in the history of cataloguing (Denton, 2007) and the investigation regarding the model reconciliation method forwarded with the FRBR-BIBFRAME mapping. BIBFRAME is the model currently being developed and used by the Library of Congress to convert its MARC21 records to linked data. The investigation focuses on the mapping of core entities, inherent relationships, and derivative relationships.

Regarding the mapping of FRBR to BIBFRAME, the thesis uses the Library of Congress approach as explicitly stated in (BIBFRAME - Bibliographic Framework Initiative, 2014; McCallum, 2017, 2018). According to this approach, a BIBFRAME *Work* is considered equivalent to the union of the following two FRBR entities, *Work* and *Expression.* Other possible mappings enabled by loose BIBFRAME definitions and alternative interpretations of the classes, have not been generated. A typical example would be the mapping of FRBR representations to BIBFRAME representations using *Expression*-agnostic *bf:Work* instances (check Figure 3:44), or *bflc:Hub* instances (Figure 3:46).

4.2.1. Mapping of core entities and inherent relationships

FRBR uses four core entities for representing content and its embodiments, whereas BIBFRAME uses three. In FRBR, content ideas are represented with the *Work* entity, and signs used to realize these ideas are represented with the *Expression* entity. In BIBFRAME these two pieces of information are both represented with one class, the *bf:Work class.* Physical embodiment is represented in FRBR with the *Manifestation* entity; copies are represented with the *Item* entity. Similarly, BIBFRAME represents physical embodiments with the *bf:Instance* class, and exemplars with the *bf:Item* class. The mapping is presented in Table 4-2 and is depicted in Figure 4:10. There are two issues that need to be taken under consideration regarding the mapping of core entities and inherent relationships. The first issue involves the case where a FRBR *Work* has more than one *Expressions*, and, consequently, participates in more than one *Work-is realized through-Expression* paths*.* The second issue involves the mapping to specific *bf:Work* or *bf:Instance* subclasses; BIBFRAME, contrary to the FRBR, defines 11 *bf:Work* subclasses, and 5 *bf:Instance* subclasses.

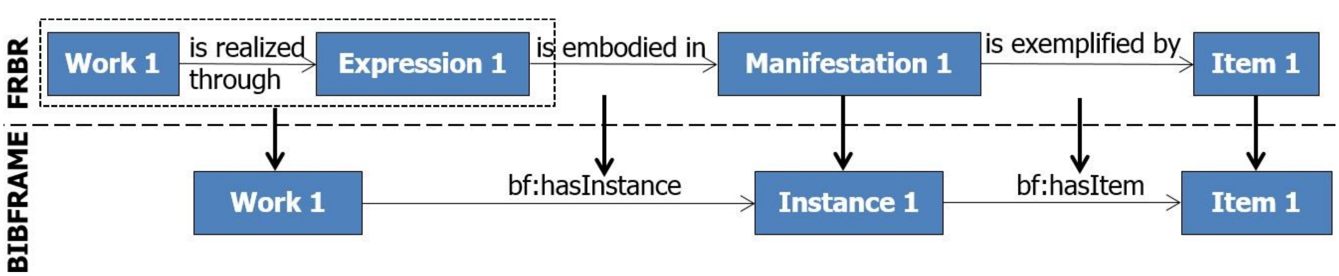


Figure 4:10. Mapping FRBR core entities and inherent relationships to BIBFRAME.

Table 4-2. Rules for mapping FRBR core entities and inherent relationships to BIBFRAME.

FRBR	BIBFRAME
Work-is realized through-Expression	The FRBR *Work* has more than one *Expressions* If yes Step 1. For each *Expression,* map the whole path to a *bf:Work* instance Step 2. Relate in pairs all the mapped *bf:Work* instances with the *bf:hasExpression* property If no Map the whole path to a *bf:Work* instance
Manifestation	*bf:Instance*
Item	*bf:Item*
is embodied in	*bf:hasInstance*
is exemplified by	*bf:hasItem*

Regarding the first issue, when an FRBR *Work* is realized through more than one *Expressions,* then it participates in multiple *Work-is realized through-Expression* paths. After the mapping though, the information that the mapped *bf:Works* contain realizations of the same ideas is lost. To avoid loss of common origin, the *bf:hasExpression* property is used (Table 1). With the use of the *bf:hasExpression* property, the sharing of common ideas is preserved; the information about the exact progenitor is lost, though. In the following Figure 4:11, the FRBR *Work is realized through* three different *Expressions.* Each FRBR *Work-is realized through-Expression* path is mapped to a new *bf:Work* instance. If the three new mapped *bf:Work* instances are not related to one another, the information that they realize the same set of ideas is lost. Using the *bf:hasExpression* property this piece of information is preserved.

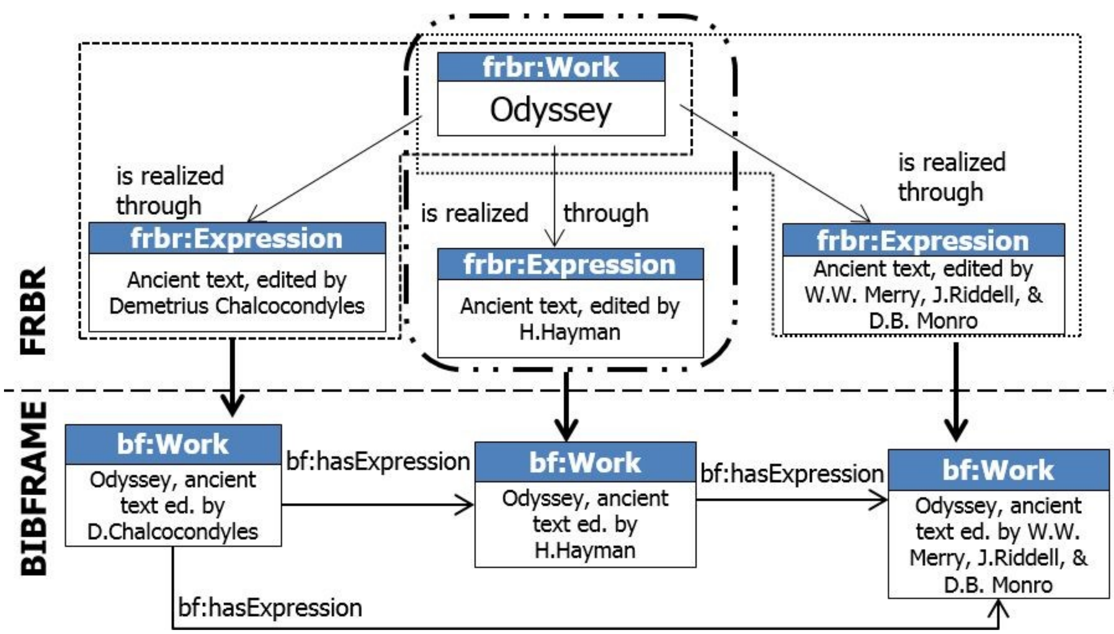


Figure 4:11. Mapping three realizations of the FRBR Work 'Odyssey' to BIBFRAME. Use of the bf:hasExpression property preserves the relationship of the three bf:Work instances being realizations of the same set of ideas.

The second issue regarding the mapping of the FRBR *Work-is realized through-Expression* to one of the *bf:Work* subclasses depends on the value of an *Expression* attribute, namely, the *form of expression* attribute. This attribute may take literal values or values from a controlled vocabulary to describe the type of signs used to realize the ideas of a *Work*. In this mapping the values from the *LC Content Types Scheme* ("LC Linked Data Service: Authorities and Vocabularies - Content Types Scheme," 2014) are used (Table 4-3). The values of this controlled vocabulary may trigger the mapping to all *bf:Work* subclasses, except the *bf:MixedMaterial* one. Moreover, certain values may even trigger the mapping to some *bf:Instance* subclasses (Table 4-3). Similarly to the *bf:Work* class, the specialization by attribute may be also exploited for the mapping of the FRBR *Manifestation* entity to the *bf:Instance* class and subclasses. In this case the *form of carrier* attribute, a *Manifestation* attribute, is used to trigger the mapping (Table 4-4). The use of the *Carriers Scheme* ("LC Linked Data Service: Authorities and Vocabularies - Carriers Scheme," 2014) values do not enable the specialization to all *bf:Instance* subclasses, only for *bf:Electronic*, and *bf:Print.* The use of the *bf:Tactile* subclass is triggered by certain values of the LC Content Types Scheme used in instances of the *form of expression* attribute (see Table 4-3). Hence, there are no values in the *Carriers Scheme* that may trigger the mapping of a *Manifestation* entity instance to either the *bf:Archival* or the *bf:Manuscript* subclasses.

Table 4-3. Different values for the form of expression attribute trigger the mapping of the FRBR 'Work – is realized through – Expression' triple to different bf:Work subclasses.

If (Expression - form of expression – (contentTypes:	**then map (Work-is realized through-Expression) to bf:Work subclass**	**and Manifestation to bf:Instance subclass**
sounds OR performed music OR spoken word))	bf:Audio	
cartographic dataset OR cartographic image OR cartographic moving image OR cartographic tactile image OR cartographic tactile three-dimensional form OR cartographic three-dimensional form))	bf:Cartography	
computer dataset))	bf:Dataset	bf:Electronic
text))	bf:Text	
tactile text))	bf:Text	bf:Tactile
notated music))	bf:NotatedMusic	
tactile notated music))	bf:NotatedMusic	bf:Tactile
notated movement))	bf:NotatedMovement	
tactile notated movement))	bf:NotatedMovement	bf:Tactile
still image))	bf:StillImage	
tactile image))	bf:StillImage	bf:Tactile
three-dimensional moving image OR two-dimensional moving image))	bf:MovingImage	
three-dimensional form))	bf:Object	
tactile three-dimensional form))	bf:Object	bf:Tactile
computer program))	bf:Multimedia	

Table 4-4. Different values for the form of carrier attribute trigger the mapping of the FRBR Manifestation entity to different bf:Instance subclasses.

If (Manifestation - form of carrier - (carriers:	**map *Manifestation* to *bf:Instance* subclass**
computer tape reel OR online resource OR computer disc))	bf:Electronic
volume))	bf:Print

4.2.2. Mapping of derivative relationships

FRBR defines more derivative relationships, for example translation, abridgement, imitation, adaptation, comparing to BIBFRAME that identifies only translation and derivation in general (see the *bf:hasDerivative/bf:derivativeOf* properties in the BIBFRAME hierarchy of properties in Figure 2:24). Derivative relationships are represented in FRBR either between *Works* or *Expressions.* In BIBFRAME, derivation is expressed between *bf:Works.* A considerable disparity between the FRBR and the BIBFRAME representations must be noted. In FRBR, derivative relationships between *Works* convey that the exact signs used for the derivation remain unknown to the cataloger. Contrary to this, the representation of derivative relationships between *bf:Works* indicates the exact opposite, that catalogers acknowledge the exact signs used in the derivation process. Similarly to the BIBFRAME semantics, in FRBR the knowledge of the exact sets of signs used to produce a derivation is indicated by representing the derivative relationship between *Expressions.* Consequently, only derivative relationships represented between FRBR *Expressions* may be mapped to BIBFRAME. Table 4-5 presents derivative relationships defined by the FRBR to be used between *Expression* instances that realize either the same *Work* (i.e., cases of translation, revision, and abridgement) or different *Works* (i.e., cases of adaptation, and dramatization).

Table 4-5. Derivative relationships between FRBR Expressions mapped to BIBFRAME.

FRBR	**BIBFRAME**
has a translation	bf:translation
has a revision	bf:hasDerivative
has an abridgement	
has an adaptation	
has a transformation	
has an imitation	

Figure 4:12 presents the mapping of two derivative relationships, a *has a translation* relationship between two *Expressions* of the same *Work,* and a *has an adaptation* relationship between two *Expressions* realizing two different *Works.* The FRBR *Work* named "Odyssey" is realized through two *Expressions,* one in ancient Greek edited by Jean de Sponde and one in English, namely the notorious Chapman English translation. Chapman did not know ancient Greek very well; he used a 1583 publication including the ancient Greek text edited by Jean de Sponde (Spondanus), and in parallel the Andreas Divus' Latin translation (Underwood, 1998). Thus, the translation relationship is represented between the Divus Latin translation and the Chapman English translation. The relationship is mapped to the *bf:translation* property. The Chapman English translation was used by Charles Lamb

to create an adaptation of 'Odyssey' for children (P. Ford, 2006). This relationship is represented with the *has an adaptation* relationship. This relationship is mapped to the *bf:hasDerivative* property.

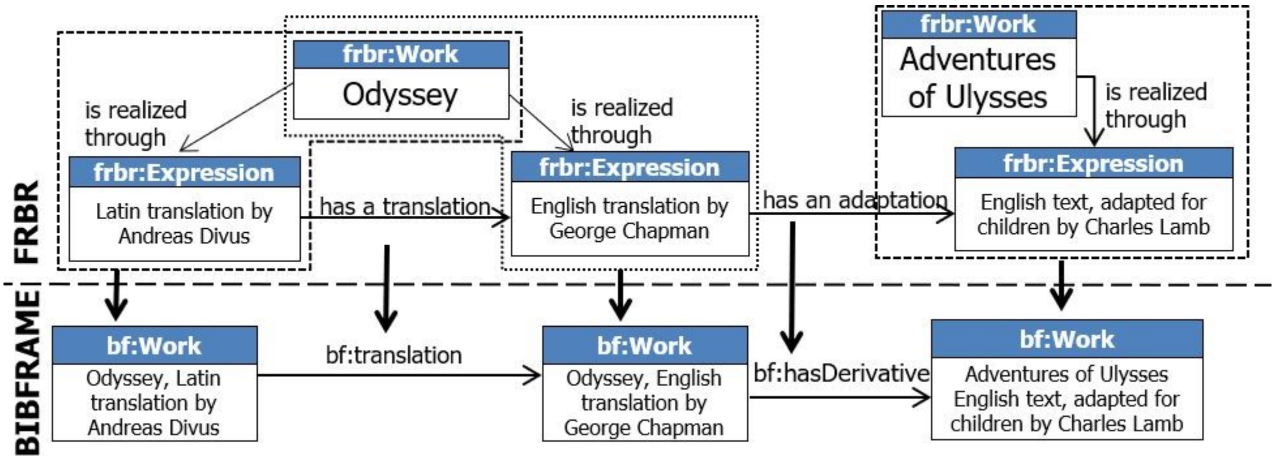


Figure 4:12. Mapping derivative relationships (translation, adaptation) from FRBR to BIBFRAME.

The mapping of derivative relationships at the *Work* level from FRBR to BIBFRAME concludes to the generation of extra and invalid derivative relationships in BIBFRAME. Due to the dual nature of the *bf:Work* class including both ideas and signs, the attributes of an FRBR *Work* and its relationships shall be mapped to all *bf:Work* instances that have been produced by the *Work-is realized through-Expression* paths in which the FRBR *Work* participates. As an example, Figure 4:13 is presented. This figure presents members of the Wuthering Heights bibliographic family. The FRBR Wuthering Heights *Work* created by Emily Brontë is realized through three *Expressions*, the original text in English, an Armenian translation, and a French translation. The Armenian translation has used the original text as its source, whereas there is no such piece of information regarding the source text used for the French translation. All three *Expressions* are mapped along the Wuthering Heights *Work* to three *bf:Work* instances. All three mapped three *bf:Work* instances carry the ideas, the attributes, and the relationships of the Wuthering Heights FRBR *Work.*

The Wuthering Heights *Work* has been adapted as musical play by (B.J.Taylor). It is not known which Wuthering Heights *Expression* has been used to create the adaptation; so, the *has an adaptation* relationship is represented at the ideas level, between the two FRBR *Works*. The B.J.Taylor adaptation *Work* is realized through two *Expressions*, one in English and another one in German. The two *Work-is realized through-Expression* paths in which the B.J.Taylor adaptation *Work* participates are mapped to two *bf:Work* instances. These two *bf:Works* bear the ideas, the attributes, and the relationships of the B.J.Taylor adaptation *Work.* The *has an adaptation* relationship is mapped to the *bf:hasDerivative* property (Table 4-5). Therefore, six instances of the *bf:hasDerivative* property will link the three *bf:Works* carrying the ideas of the Wuthering Heights FRBR *Work* to the two *bf:Works* carrying the ideas of the B.J.Taylor adaptation *Work* (Figure 4:13).

The representation of the relationship between FRBR *Works* conveys a derivative relationship at an abstract level. When mapping the relationship to BIBFRAME, the relationship is represented at a less abstract level, between *bf:Work* instances that carry both ideas and signs. As presented in Figure 4:13, the mapping of *Work-Work* relationships from FRBR to BIBFRAME results in the generation of many relationships (property instances) that in many cases are false. As an example, after the mapping the Armenian translation is presented to have two derivatives *bf:Works,* the B.J.Taylor Musical Play in English and the B.J.Taylor Musical Play in German. It may be concluded that trying to map this relationship to BIBFRAME will possibly result in extra and false relationships.

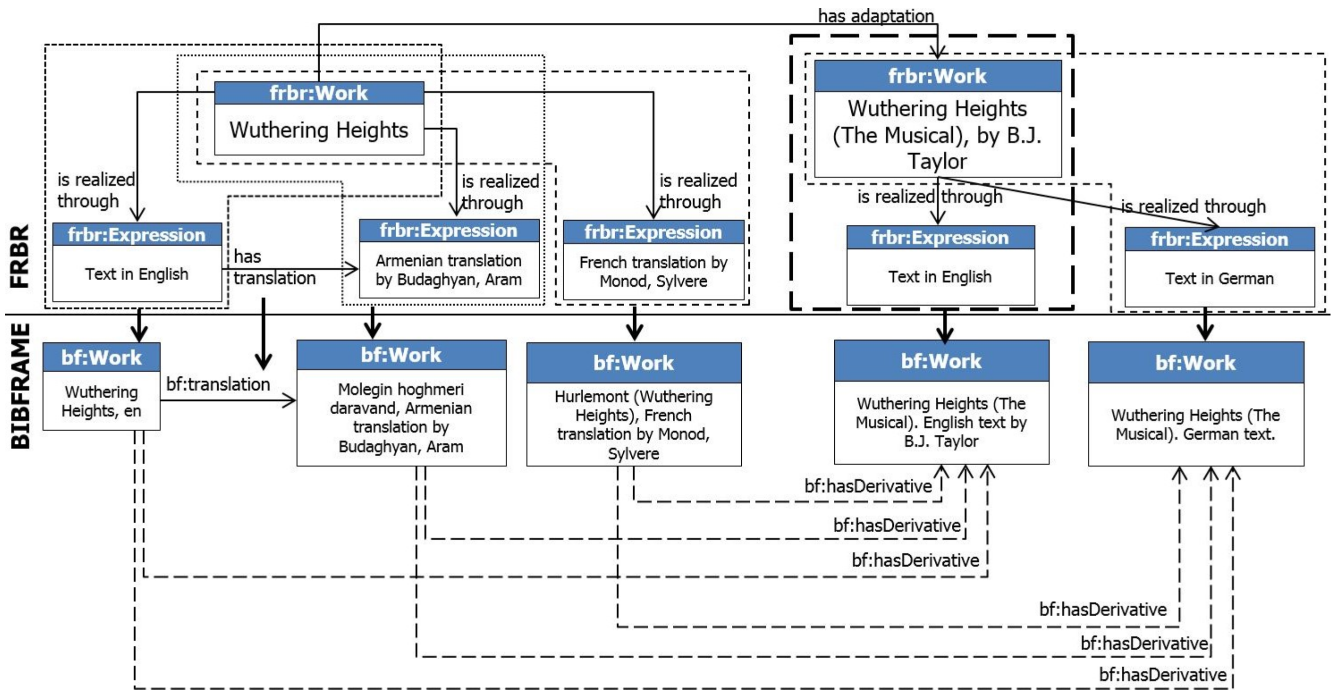


Figure 4:13. Mapping Work to Work derivative relationships from FRBR to BIBFRAME produces redundant and erroneous relationships.

4.2.3. Findings

FRBR and BIBFRAME are different in terms of granularity. FRBR uses four core entities for representing content and its embodiments, whereas BIBFRAME uses three. When an FRBR *Work* entity instance *is realized through* more than one *Expression* instances, the mapping uses *bf:hasExpression* property instances to relate the generated *bf:Works,* and to preserve the information that these *bf:Works* share the same ideational content. With the use of the *bf:hasExpression* property the existence of shared ideational content among the related *bf:Works* is preserved. Otherwise, these *bf:Works* would have remained unrelated to one another and the exploration of a bibliographic family's members would not be possible.

The mapping of the FRBR *Work* and *Expression* entities to the *bf:Work* class is feasible taking under consideration a specific *Expression* attribute, i.e., *form of expression*, to map to specialized *bf:Work* subclasses. The values used for this attribute were taken from the *LC Content Types Scheme* ("LC Linked Data Service: Authorities and Vocabularies - Content Types Scheme," 2014). Thus, a different vocabulary will trigger different mappings. Similarly, the mapping of the FRBR *Manifestation* entity to *bf:Instance* subclasses is triggered by the values of the *form of carrier* attribute taken from the *Carriers Scheme* ("LC Linked Data Service: Authorities and Vocabularies - Carriers Scheme," 2014).

Regarding the mapping of derivative relationships BIBFRAME provides less properties compared to FRBR. Thus, the mapping of all FRBR derivative relationships, except for *has a translation*, is made to the generic BIBFRAME *bf:hasDerivative* property. An important finding is that BIBFRAME enables the representation of bibliographic relationships at the signs level only. The mapping of derivative relationships at the *Work* level from FRBR to BIBFRAME concludes to the generation of extra and invalid derivative relationships in BIBFRAME.

4.3. Model reconciliation: RDA – BIBFRAME mapping

Even though FRBR is a conceptual model that has influenced cataloging and the conceptualizations defined in other bibliographic conceptual models, its semantics have been implemented mostly through the vocabulary of the RDA content standard. Therefore, the investigation about model reconciliation proceeded with the RDA to BIBFRAME mapping. As it has already been presented in Chapters 2 and 3, the RDA adheres to FRBR and extends it defining more types of relationships. Moreover, RDA is used by libraries worldwide to enable the identification of bibliographic entities in legacy MARC records and to prepare the ground for future conversions of legacy data to linked data. BIBFRAME is currently developed and used by the Library of Congress to convert legacy MARC21 to linked data. Considered that RDA is the de facto content standard in MARC21 records, and Library of Congress' influential role in the development of library standards, it is most likely that in the near future there will be the need for transforming RDA data to BIBFRAME, and vice-versa. Based on this assumption, the investigation progressed to the mapping of RDA core entities, inherent relationships, and derivative relationships to BIBFRAME.

4.3.1. Mapping of core entities and inherent relationships

An RDA *Work* is not semantically equivalent to a BIBFRAME *Work*. The former is used for the representation of the "ideational content", while the signs realizing this content are represented with the use of another class, namely, *Expression.* Contrary to this, the *bf:Work* class includes both ideas and signs. Implementing the Library of Congress rationale (BIBFRAME - Bibliographic Framework Initiative, 2014; McCallum, 2017, 2018), the RDA path *rdac:C10001 - rdaw:P10078 - rdac:C10006 (Work - has expression of work – Expression)* is mapped to the *bf:Work* class (Figure 4:14 and Table 4-6). A direct consequence of this mapping is that the mapped *bf:Work* class instance will convey both RDA *Work* and *Expression* instances' semantics, attributes, and relationships. The RDA *Manifestation* class is mapped to the BIBFRAME *Instance* one. The RDA *Item* class is mapped to the BIBFRAME *Item* class. The inherent relationships *has manifestation of expression* and *has exemplar of manifestation* are correspondingly mapped to the *bf:hasInstance* and the *bf:hasItem* properties. The mappings are displayed in Figure 4:14 and analytically presented in the first mapping rule (Table 4-6).

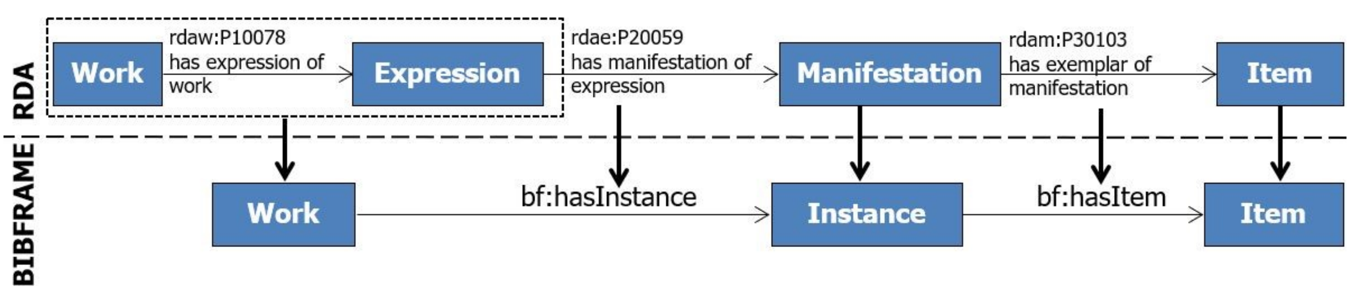


Figure 4:14. Mapping RDA core entities and inherent relationships to BIBFRAME.

Table 4-6. RDA to BIBFRAME mapping rule 1 – Mapping core entities and inherent relationships.

RDA	**BIBFRAME**
rdac:C10001 - rdaw:P10078 - rdac:C10006 *(Work - has expression of work – Expression)*	Step 1 Create a *bf:Work* as the union of the *Work* and *Expression.* Step 2 (Applied only when RDA *Work* has several *Expressions,* When more than one *bf:Work* instances created from step 1, link them with *bf:hasExpression* relationship (see details in Table 4-7).
rdac:C10007 (Manifestation)	bf:Instance
rdae:P20059 (has manifestation of expression)	bf:hasInstance
rdac:C10003 (Item)	bf:Item
rdam:P30103 (has exemplar of manifestation)	bf:hasItem

In case an RDA *Work* has more than one *Expressions,* the mapping rule is extended (Step 2 in Table 4-6). In RDA all *Expressions* of a *Work,* regardless if they are related to one another with some kind of bibliographic relationship, they remain members of the same *family* clustered under the same *Work* they all realize. After the mapping of all these *Expressions* to BIBFRAME, the result is different *bf:Work* instances that are not related to one another. This mapping resulting in 'orphan' *bf:Works* is presented as Pattern A in Figure 4:15. The extension of the first mapping rule involves exactly this: the prevention of 'orphan' *bf:Works* that remain separated from other *bf:Works* realizing the same ideational content. The second mapping rule (Table 4-7) uses the *bf:hasExpression* property to cluster all *bf:Works* that carry the same abstract notions. This property, according to the BIBFRAME definition, correlates two *bf:Works* where the latter is an expression of the former. The definition further clarifies that this property may be used "to relate Works under FRBR/RDA rules" ("BIBFRAME ontology - hasExpression," 2016). The clustering may be applied by coupling all mapped *bf:Work* instances. This pattern is depicted in Figure 4:15 as pattern B. It must be noted that the number of *bf:hasExpression* property instances depends on the number of *Expressions.* For a number of N *Expressions,* N*(N-1)/2 *bf:hasExpression* property instances will be used to relate in pairs the mapped *bf:Works*. As an example, mapping an RDA Work with 5 *Expressions* will generate five *bf:Work* instances that will be related in pairs with 10 *bf:hasExpression* property instances. For an RDA *Work* with 10 *Expressions,* 45 instances of the *bf:hasExpression* property are needed to couple the mapped *bf:Work* instances! A simpler representation can be applied when using the *bf:hasExpression* property as an OWL transitive one (Sean Bechhofer et al., 2004). With a transitive *bf:hasExpression* the *bf:Work* instances may be related successively, and, since *W_E1* instance has *W_E2* as its expression, and *W_E2* has *W_E3* as its expression, then it can be inferred that *W_E1* has *W_E3* as its expression too. This pattern is depicted in Figure 4:15 as pattern C. With this pattern much less *bf:hasExpression* property instances are needed. As an example, an RDA Work with 5 *Expressions* will be mapped to five *bf:Works* that will be related with only 4 *bf:hasExpression* property instances. For the RDA Work with 10 *Expressions,* 9 *bf:hasExpression* property instances will be eventually used to cluster the mapped *bf:Works.* The second mapping rule that extends the first one and uses the *bf:hasExpression* property, either as non-transitive, or as transitive, is presented in Table 4-7.

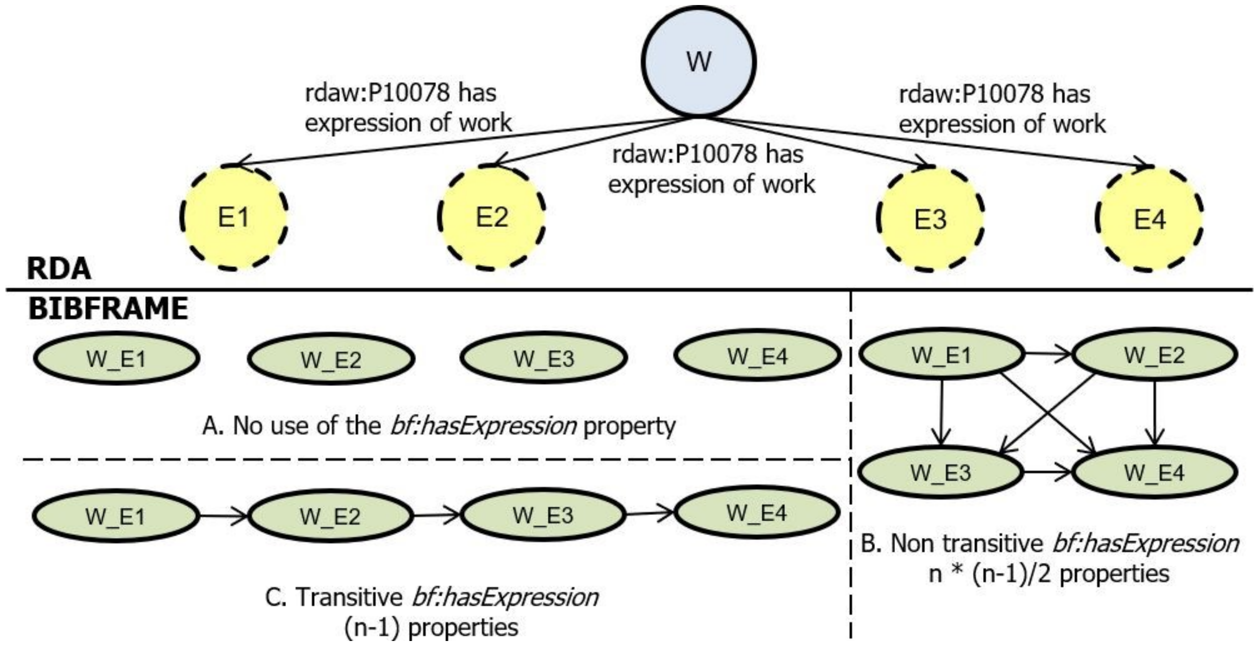


Figure 4:15. Mapping an RDA Work with four Expressions to BIBFRAME. Three patterns are presented: A. No use of the bf:hasExpression property, B. use of the bf:hasExpression property as non-transitive, and C. use of the bf:hasExpression property as transitive.

Table 4-7. RDA to BIBFRAME mapping Rule 2 – Extending mappings to preserve clustering of RDA Expressions of the same Work (one RDA Work with several Expressions).

RDA	BIBFRAME
rdac:C10001 - rdaw:P10078 - rdac:C10006 *Work - has expression of work – Expression*	Step 2 (when RDA *Work* has several *Expressions*) When more than one *bf:Work* instances have been created from step 1, link all *bf:Works* generated from step 1 with *bf:hasExpression* relationship Case A. (The *bf:hasExpression* property is not transitive) Relate with the *bf:hasExpression* property all possible pairs of *bf:Work* instances generated from the same RDA *Work* participating in multiple '*Work-has expression of work-Expression*' paths. Case B. (The *bf:hasExpression* property is transitive). Connect the *bf:Work* instances generated from the same RDA *Work* - participating in multiple '*Work-has expression of work-Expression*' paths – successively with the *bf:hasExpression* property.

4.3.2. Mapping of derivative relationships

RDA represents derivative relationships either between *Works* or *Expressions.* Similarly to FRBR, the representation of a derivative relationship between *Works* discloses both the existence of a derivative relationship between them, and the unawareness regarding the original set of signs used to produce the derivative one. By contrast, the representation of a derivative relationship between *Expressions* conveys the certain knowledge of both sets of signs, original and derivative ones, used during the derivation process. In BIBFRAME, derivative relationships are represented at the *bf:Work* level. Considering that the *bf:Work* class accommodates both ideas and signs, the derivative relationship involves derivation at the signs level. Hence, derivative relationships represented in RDA between *Expressions* may be mapped to the corresponding BIBFRAME properties relating *bf:Works*. This is the third mapping rule presented in Table 4-8.

Table 4-8. RDA to BIBFRAME mapping Rule 3 – Mapping derivative relationships.

RDA	BIBFRAME
rdac:C10001 - rdaw:P10148 (and all its subproperties) - rdac:C10001 *Work – is derivative (work)* (and all its subproperties) - *Work*	Ignore these relationships in the mappings.
rdac:C10006 – rdae:P20203 (and all its subproperties) - rdac:C10006 *Expression – is derivative (expression)* (and all its subproperties) - *Expression*	Map these relationships to the corresponding BIBFRAME derivation property between *bf:Work* instances (see Table 4-9).

Regarding the mapping of derivative relationships represented between RDA *Works,* these relationships are ignored (Table 4-8). In case they would not be ignored in the mapping, the *bf:Works* generated by mapping the original *Work's* paths (RDA *Work - has expression of work – Expression* paths) would all carry the derivative relationship of the original *Work.* Moreover, these *bf:Works* would be related to all *bf:Works* generated by mapping the derivative *Work's* paths (RDA *Work - has expression of work – Expression*). This would cause severe noise in the mapping. Noisy and imprecise mappings are expected in large numbers when large bibliographic families are involved, or when both original and derivative *Works* are realized in many *Expressions.* Thus, even though in RDA the *Work* instances are related, in BIBFRAME these will remain unrelated to any other member of their bibliographic family. This is a compromise that needs to be made, given that the scenario of including RDA *Work-Work* derivative relationships in the mapping will cause severe noise and false mappings, similarly to the FRBR-BIBFRAME mapping presented in Figure 4:13.

The mapping of properties used for the representation of derivative relationships from RDA to BIBFRAME is presented in Table 4-9. The BIBFRAME hierarchy of derivative relationships (see Figure 2:24) may be considered as less refined than the RDA one (see Figure 2:18 and Figure 2:19). BIBFRAME differentiates only between translation (*bf:translation*) and other types of derivation (*bf:hasDerivative*). As a result, the majority of specialized RDA properties representing derivative relationships are mapped to only one BIBFRAME property, namely, the more general *bf:hasDerivative* property. Derivative relationships are expected to be retained when mapping from RDA to BIBFRAME, but loss of the exact semantics is also anticipated.

Table 4-9. RDA to BIBFRAME mapping. Mapping of derivative relationships from RDA to BIBFRAME.

RDA	BIBFRAME
rdae:P20203 is derivative (expression)	bf:hasDerivative
rdae:P20166 is abridged as (expression)	
rdae:P20153 is adapted as (expression) rdae:P20110 is adapted as libretto (expression)	
rdae:P20211 is revised as	
rdae:P20171 is translated as	bf:translation

4.3.3. Findings

Even though in BIBFRAME there is no equivalent class to the RDA *Expression* entity, the semantics of this entity may be preserved using *bf:hasExpression* property instances to relate *bf:Works* carrying different realizations (*Expressions* in RDA terms) of the same ideational content (*Work* in RDA terms). The property may be used either as transitive or non- transitive. Moreover, the use of the *bf:hasExpression* property contributes to the formation of bibliographic families and to their exploration.

Concerning the mapping of derivative relationships from RDA to BIBFRAME, two points must be highlighted. First, derivative relationships between RDA *Works* are ignored in the mapping due to the concern that an excessive number of false relationships will be generated in BIBFRAME, especially when the mapping involves relationships between RDA *Works* with many *Expressions.* Secondly, the mapping of RDA derivative relationships is made to semantically generic BIBFRAME properties. Due to the limited number of BIBFRAME properties for the representation of derivative relationships, all RDA relationships, except for translation, are mapped to the *bf:hasDerivative* property.

4.4. Model reconciliation: BIBFRAME – RDA mapping

The thesis proceeds with the investigation of the model reconciliation method applied to the conversion of BIBFRAME to RDA. The BIBFRAME-RDA mapping is completed with four steps, *Steps (a)-(d)*. *Step (a)* executes the mapping of *bf:Work* instances to sets of RDA *Works* with their *Expressions. Step (b)* executes the mapping of derivative relationships. *Step (c)* executes the mapping of *bf:Work* properties to RDA *Work* and *Expression* properties respectively. The final *Step (d)* executes the mapping of the remaining core classes, *bf:Instance* and *bf:Item*, and inherent relationships.

4.4.1. Mapping of the *bf:Work* class

The *bf:Work* class is considered as equivalent to the following RDA triple, '*Work - has expression of work – Expression*', carrying the semantics of both RDA *Work* and *Expression* entities (Figure 4:16). There are two important parameters that need to be taken under consideration. The first parameter involves the identification of the same ideational content when examining the conversion of *bf:Work* instances to proper sets of RDA *Works* along with their *Expressions*. The second parameter involves the separation of the *bf:Work* properties to RDA *Work* properties and RDA *Expression* properties. This mapping is presented in the paragraphs 4.4.3 and 4.4.4.

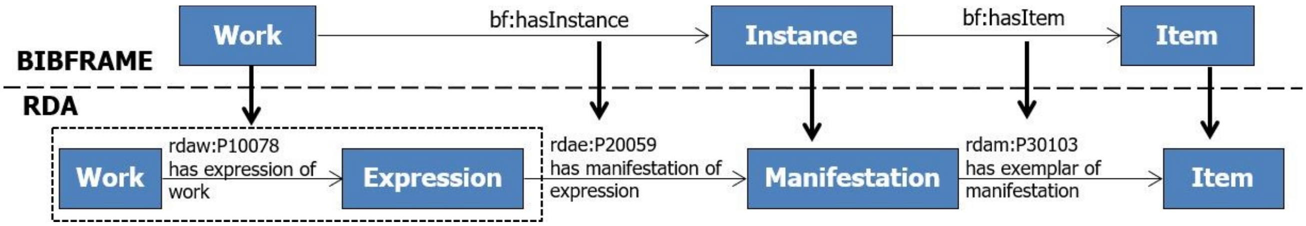


Figure 4:16. Core BIBFRAME entities and inherent relationships mapped to RDA.

Step (a) that executes the mapping of the *bf:Work* class is completed in three stages:

a1. The set of *bf:Work* instances and their corresponding relationships (Part 1 of Figure 4:17) is partitioned to subsets, each of them containing the different realizations of the same ideational content only (Part 2 of Figure 4:17). The partitions are created exploiting the existence of *bf:hasExpression, bf:translation*, and *bf:otherEdition* property instances. The *bf:hasExpression* property is used to relate *bf:Works* carrying different signs/expressions of the same sets of ideas. The *bf:translation* property is used to represent literal translations in BIBFRAME. The third property, *bf:otherEdition,* is used to relate *bf:Works* sharing the same content in different languages or content forms, e.g., edition of the same content in another media type.

a2. An RDA *Work* class is generated for each subset (Part 3 of Figure 4:17).

a3. An RDA *Expression* instance is generated for every *bf:Work* in the subset, and it is related afterwards to the RDA *Work* instance (generated in the previous step) with a *rdae:P20231 has work expressed* property instance (the inverse *rdaw:P10078 has expression of work* property is also instantiated) (Part 3 of Figure 4:17).

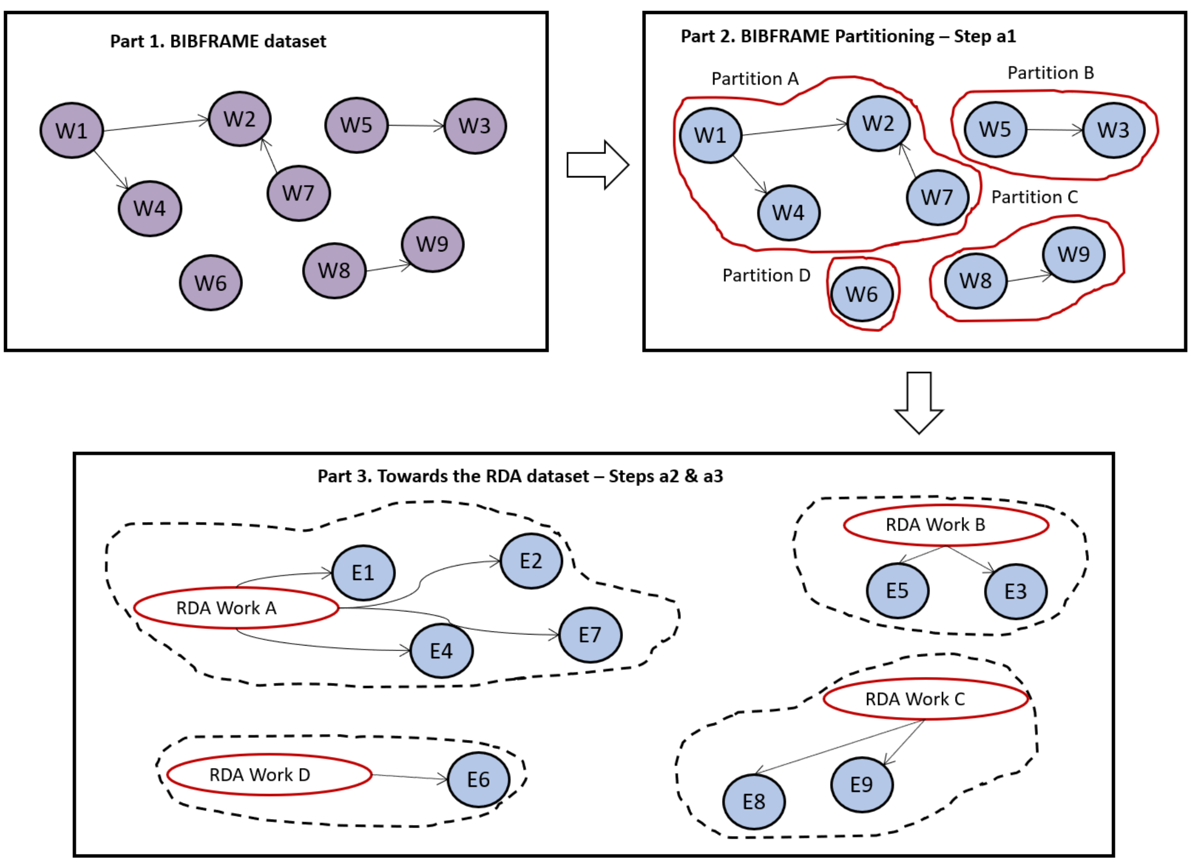


Figure 4:17. BIBFRAME to RDA mapping Step (a) - Partitioning and mapping a set of bf:Work instances to RDA Works and Expressions.

An application of the mapping *Step (a)* transforming a partition containing *bf:Works* from the Odyssey family to RDA is depicted in Figure 4:18. *Step (a1)* creates Partition A based on the existence of instances of the *bf:hasExpression*, *bf:translation*, or *bf:otherEdition* properties relating a set of *bf:Works*. In *Step (a2)*, an instance of the RDA *Work* named “The Odyssey of Homer” is generated for Partition A. In *Step (a3)*, each *bf:Work* in Partition A is mapped to an RDA *Expression* and thus four RDA *Expressions* are generated. The “Odyssey of Homer” RDA *Work*, generated in the second step, is related to the four RDA *Expressions* (identified with numbers 1, 2, 4, and 7) generated in this step with instances of the property *rdaw:P10078 has expression of work*. In Figure 4:18, each generated *Expression* carries the reference number of its corresponding *bf:Work*, e.g., the RDA *Expression* (1) is generated out of the *bf:Work* (1). In the fourth step, the mapping of *bf:translation* and *bf:otherEdition* properties generates instances of properties representing relationships at the *Expression* level, namely the *rdae:P20171 is translated as* and *rdae:P20204 is based on expression* properties. The instances of the *bf:hasExpression* property are ignored.

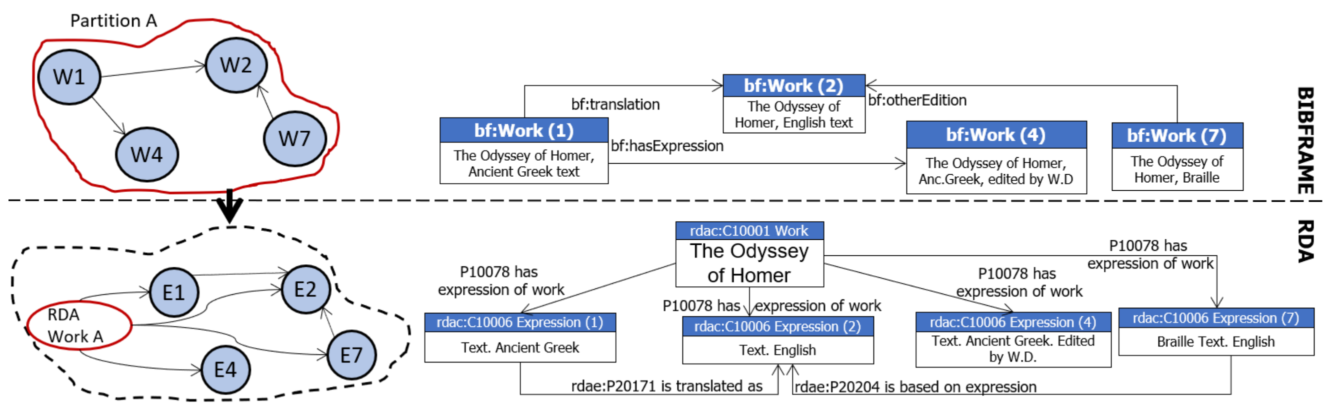


Figure 4:18. BIBFRAME to RDA mapping Step (a) - Example of mapping the Partition A that includes the bf:Works from the Odyssey family to RDA Work and Expressions.

4.4.2. Mapping derivative relationships

Step (b) of the mapping algorithm maps derivative relationships from BIBFRAME to RDA. As presented in Table 4-10, all BIBFRAME derivative relationships may be mapped to RDA. The hierarchy of derivative relationships is depicted in Figure 4:19. The high-level BIBFRAME *bf:hasDerivative property* is mapped to the higher RDA property for the representation of derivative relationships at the *Expression* level, that is, the *rdae:P20203 is derivative (expression)* property[5]. The *bf:translation* property is mapped to the *rdae:P20171 is translated as* property. Since there is no equivalent property in RDA for the representation of the ‘other editions’ cases, the *bf:otherEdition* property is mapped to an RDA generic property, the *rdae:P20203 is derivative (expression)* property, or to its inverse *rdae:P20204 is based on (expression).* It may be mapped to both properties because BIBFRAME defines the *bf:otherEdition* property as symmetric. This means that the relationship exists both ways, regardless the *bf:Works* participating in the relationship are at the domain or range side of the *bf:Work-bf:otherEdition-bf:Work* path. Therefore, the *bf:otherEdition* property may be mapped either to the *rdae:P20203 is derivative (expression)* property, or to its inverse *rdae:P20204 is based on (expression).* The mappings need to take under consideration

[5] Due to the use of the version 2.7.3 of the RDA vocabularies some slight differences may be observed in the thesis regarding RDA labels. As an example, the *rdae:P20203* property has the label ‘is derivative (expression)’ in the RDA element set (used in the thesis), while on the RDA registry its label is ‘has derivative expression’.

how the property is used in the BIBFRAME datasets that will be converted to RDA. In case a library has selected to use the *bf:otherEdition* property to relate the original *bf:Work* to its derivative *bf:Work,* then the mapping of the property will be to the *rdae:P20203 is derivative (expression)* property. If the library has selected the opposite representation, from the derivative *bf:Work* to the original one, then the *rdae:P20204 is based on (expression)* property will be used in the mapping.

It must be noted that the generic *bf:hasDerivative* property may represent many types of derivation due to the lack of more specialized properties in the BIBFRAME hierarchy of derivative relationships. Some of the derivative relationships represented with the *bf:hasDerivative* property may involve derivations of the same ideational content, e.g., cases of abridgement or revision. By contrast, other cases represented with the *bf:hasDerivative* property may involve changes of the ideational content, such as, adaptations, dramatizations, etc. The use of the *bf:hasDerivative* property does not provide any information regarding the existence of common or different ideational content among the related *bf:Works.* Therefore, this property was not considered in *Step (a)* of the mapping algorithm. Another important point is that the *bf:originalVersion* property has been excluded from the mapping (Table 4-10). Even though this property is included in the hierarchy of derivative relationships, it is suitable only for the representation of the equivalence relationship (reproduction) between *bf:Instances*, considering a) its definition "Resource is the original version of which this resource is a reproduction" (Library of Congress, 2016a), b) the BIBFRAME specifications regarding the conversion of the MARC21 534 field to BIBFRAME (Library of Congress, 2019), and c) one related clarification made in the BIBFRAME mailing list (Denenberg, 2017a).

Table 4-10. BIBFRAME to RDA mapping 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

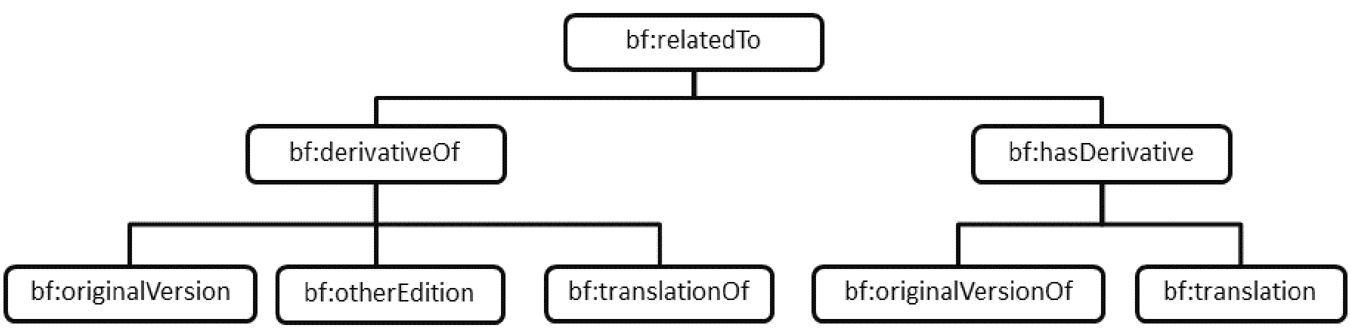


Figure 4:19. Hierarchy of derivative relationships in BIBFRAME.

Thus, in the example of Figure 4:18, all *bf:Work* level relationships are mapped to their RDA *Expression* level properties. It is worth mentioning that the *bf:hasExpression* property is ignored. The existence of common ideational content, that the *bf:hasExpession* property represents, is now represented with the *rdaw:P10078 has expression of work/rdae:P20231 has work expressed* properties instances.

4.4.3. Mapping core entities: *Work* properties

With regard to the separation of which BIBFRAME *Work* properties will be mapped to RDA *Work* properties, or to RDA *Expression* properties, the dominant characteristics that define an RDA *Work* and an RDA *Expression* are taken under consideration. Thus, in *Step (c), bf:Work* properties regarding primary contribution information and title will be mapped to RDA *Work* properties, while *bf:Work* properties regarding content type, non-primary contribution, and language information will be mapped to RDA *Expression* properties.

Titles and information about the primary contributors are considered as pertaining to RDA *Works* (Aalberg & Žumer, 2008; Hickey & O'Neill, 2009; Peponakis et al., 2011; Takhirov, 2013). BIBFRAME represents titles with instances of the *bf:Title* class. The information in this class, and specifically the whole *bf:Work-bf:title-bf:Title-bf:mainTitle-rdfs:literal* statement, is mapped to the literal value of the RDA *Work P10088 has title of work* property (Table 4-11). BIBFRAME uses classes for the representation of contributions, namely, the *bf:Contribution* class. Contributions are represented with two statements: "*bf:Work – bf:contribution – bf:Contribution – bf:agent – bf:Person",* and "*bf:Contribution – bf:role – bf:Role"*. The *bf:Contribution* class may be used to represent either primary or non-primary contributions. The information of whether the contribution is primary or not is revealed by the value of the *bf:Role* class. This class may have as a value a plain literal, e.g., author, or may take its value from a controlled vocabulary, e.g., the *aut* value for author with the URI http://id.loc.gov/vocabulary/relators/aut taken from the MARC Code List for Relators Scheme ("LC Linked Data Service: MARC Code List for Relators Scheme," n.d.). In case a cataloguing agency selects local values or values from another controlled vocabulary, different mapping rules will likely be triggered. Therefore, mapping rules must consider the policy implemented in assigning values to the *bf:Role* class. A *bf:Contribution* subclass, the *bflc:PrimaryContribution* one, may provide a more stable way for representing primary contributions in BIBFRAME, since the information that a primary contributions is being described lies in the use of the class itself (i.e., *bflc:PrimaryContribution*) and not in the value of the *bf:Role* class. The subclass has not been incorporated to the official BIBFRAME vocabulary and it remains a construct used in Library of Congress BIBFRAME-related projects.

Contrary to the BIBFRAME practice, RDA defines a separate property for each role value. This property is different when the *Agent* involved is a *Person* or a *Family.* As an example, the property for relating an RDA *Work* with its author (type *rdac:C10004 person*) is the *P10061 has author agent,* while the property for a corporate body author is the *rdaw:P10530 has author corporate body* property. Consequently, the BIBFRAME to RDA mapping regarding mapping contributions varies depending on the type of the *bf:Agent* (*bf:Person, bf:Organization,* or *bf:Family*) and its exact role (Table 4-11).

It must be once again noted that the mapping, as all mappings in this chapter, focuses on core entities, inherent relationships, and derivative relationships. Thus, a full mapping is not provided for all *bf:Work* properties, or for the entirety of values of a given vocabulary, e.g., the LC MARC Code List for Relators Scheme.

Table 4-11. BIBFRAME to RDA mapping 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	
	If bf:Agent - rdf:type - bf:Person	
	then if bflc:PrimaryContribution-bf:role-bf:Role(relators:aut)	rdac:C1001-rdawo:P10061-rdac:C1002
	if bflc:PrimaryContribution-bf:role-bf:Role(relators:pht)	rdac:C1001-rdawo:P10445-rdac:C1002
	if bflc:PrimaryContribution-bf:role-bf:Role(relators:cmp)	rdac:C1001-rdawo:P10442-rdac:C1002
	if bflc:PrimaryContribution-bf:role-bf:Role(relators:edc)	rdac:C1001-rdawo:P10444-rdac:C1002
	if bflc:PrimaryContribution-bf:role-bf:Role(relators:drt)	rdac:C1001-rdawo:P10455-rdac:C1002
	if bf:Agent-rdf:type-bf:Organization	
	then if bflc:PrimaryContribution-bf:role-bf:Role(relators:aut)	rdac:C1001–rdawo:P10530–rdac:C10011
	if bflc:PrimaryContribution-bf:role-bf:Role(relators:pht)	rdac:C1001-rdawo:P10539-rdac:C10011
	if bflc:PrimaryContribution-bf:role-bf:Role(relators:cmp)	rdac:C1001-rdawo:P10536-rdac:C10011
	if bf:Agent-rdf:type-bf:Family	
	then if bflc:PrimaryContribution-bf:role-bf:Role(relators:aut)	rdac:C1001-rdawo:P10577-rdac:C10011
	if bflc:PrimaryContribution-bf:role-bf:Role(relators:pht)	rdac:C1001-rdawo:P10586-rdac:C10011
	if bflc:PrimaryContribution-bf:role-bf:Role(relators:cmp)	rdac:C1001-rdawo:P10583-rdac:C10011

Information about content type, language and non-primary contribution is associated to the identification of RDA *Expressions* (Aalberg & Žumer, 2008; E T O'Neill, 2002; Yee, 2005). Hence, *Step (c)* proceeds with the mapping of these pieces of information from BIBFRAME to RDA (Table 4-12). BIBFRAME represents content types using different *bf:Work* subclasses. There are 11 subclasses, i.e., Text, Cartography, Audio, NotatedMusic, NotatedMovement, Dataset, StillImage, MovingImage, Object, Multimedia, and MixedMaterial. In RDA, content types are represented as values for the *rdae:P20001 has content type Expression* property. The values may be local, or they may be taken from a controlled vocabulary. In this *Step (c),* the RDA Content Type Vocabulary ("Term and Code List for RDA Content Types," 2014) is considered. In case another controlled vocabulary is selected, the mapping rules will need to alter accordingly.

BIBFRAME represents language information with the *bf:Language* class taking values from a controlled vocabulary. In the following statement *bf:Work - bf:language – bf:Language (languages:value)*, the values are from the MARC List for Languages vocabulary ("LC Linked Data Service: MARC List for Languages," 2011). This BIBFRAME triple will be mapped to the following RDA triple *rdac:C10006 Expression – rdaeo:P20006 has language of expression – languages:value*.

The *bf:Work* class may include information regarding primary and non-primary contributions. Primary contributions involve the ideational content and have already been mapped to RDA *Work* properties. Non-primary contribution, represented with the *bf:Contribution* class, involves the contribution made to the signs that the *bf:Work* carries, and will be, consequently, mapped to RDA *Expression* properties. Non-primary contribution is represented with two statements in which the *bf:Contribution* class participates: 1) *bf:Work - bf:contribution - bf: Contribution - bf:agent – bf:Agent,* and 2) *bf:Contribution-bf:role-bf:Role(relators:value)*. The first statement describes which *bf:Agent* has contributed to the signs of the *bf:Work* instance. The mapping of this statement to an RDA *Expression* property is determined by the *bf:Agent* type, i.e., if the *bf:Agent* is a Person, a Family, or a Corporate Body. The second statement describes the *bf:Agent's* role using either local values for the *bf:Role* class, or a controlled vocabulary. Similarly to the mapping of the *bflc:PrimaryContribution* class, the Library of Congress MARC Code List for Relators Scheme ("LC Linked Data Service: MARC Code List for Relators Scheme," n.d.) is used. In case another controlled vocabulary is selected, the mapping rules will need to alter accordingly. The mapping of content type, language, and non-primary contribution information from BIBFRAME to RDA is pointedly declared in Table 4-12.

Table 4-12. BIBFRAME to RDA mapping 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 –	rdac:C10006- rdaeo:P20001-
	if bf:Text	rdaco:1020
	if bf:Audio	rdaco:1012
	if bf:NotatedMusic	rdaco:1010
	if bf:Cartography	rdaco:1002
	if bf:NotatedMovement	rdaco:1009
	if bf:Dataset	rdaco:1007
	if bf:StillImage	rdaco:1014
	if bf:MovingImage	rdaco:1023
	if bf:Object	rdaco:1021
	if bf:Multimedia	rdaco:1008
	if bf:MixedMaterial	no equivalent value in rdaco vocabulary
Language	bf:Work - bf:language-bf:Language (languages:value)	rdac:C10006-rdaeo:P20006-languages:value

Non-primary contributions	bf:Work - bf:contribution - bf: Contribution - bf:agent - bf:Agent	
	If bf:Agent - rdf:type - bf:Person	
	then	
	if bf:Contribution-bf:role-bf:Role(relators:trl)	rdac:C10006-rdaeo:P20346-rdac:C1002
	if bf:Contribution-bf:role-bf:Role(relators:edt)	rdac:C10006-rdaeo:P20338-rdac:C1002
	if bf:Contribution-bf:role-bf:Role(relators:nrt)	rdac:C10006-rdaeo:P20378-rdac:C1002
	if bf:Contribution-bf:role-bf:Role(relators:prf)	rdac:C10006-rdaeo:P20351-rdac:C1002
	if bf:Contribution-bf:role-bf:Role(relators:ctb)	rdac:C10006-rdaeo:P20389-rdac:C1002
	if bf:Contribution-bf:role-bf:Role(relators:spk)	rdac:C10006-rdaeo:P20380-rdac:C1002
	if bf:Contribution-bf:role-bf:Role(relators:edc)	rdac:C10006-rdaeo:P20338-rdac:C1002
	if bf:Contribution-bf:role-bf:Role(relators:abr)	rdac:C10006-rdaeo:P20357-rdac:C1002
	if bf:Contribution-bf:role-bf:Role(relators:lbt)	rdac:C10006-rdaeo:P20053-rdac:C1002
	if bf:Agent-rdf:type-bf:Organization	
	then	
	if bf:Contribution-bf:role-bf:Role(relators:trl)	rdac:C10006-rdaeo:P20464-rdac:C1002
	if bf:Contribution-bf:role-bf:Role(relators:edt)	rdac:C10006-rdaeo:P20456-rdac:C1002
	if bf:Agent-rdf:type-bf:Family	
	then	
	if bf:Contribution-bf:role-bf:Role(relators:trl)	rdac:C10006-rdaeo:P20523-rdac:C1002
	if bf:Contribution-bf:role-bf:Role(relators:edt)	rdac:C10006-rdaeo:P20515-rdac:C1002

4.4.4. Mapping core entities and inherent relationships: *Instance* and *Item* classes

The *bf:Instance* class is used for the representation of *bf:Work* physical embodiments and may be considered as equivalent to the RDA *Manifestation* class. The *bf:Item* class is used for the individual copies of the *bf:Instances* and is mapped to the RDA *Item* class (Table 4-13). The *bf:hasInstance* property is mapped to the *rdae:P20059 has manifestation of expression* inherent relationship, while the *bf:hasItem* property is mapped to the *rdam:P30103 has exemplar of manifestation* property (Figure 4:16).

Table 4-13. BIBFRAME to RDA mapping Step (d)- Mapping the bf:Instance and bf:Item classes.

BIBFRAME	RDA
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

Mapping *bf:Instance* properties to RDA is not straightforward; BIBFRAME uses classes in comparison to the RDA that mostly uses properties. The same pieces of information may be described in BIBFRAME using classes and statements, while in RDA specialized properties are used instead. As an example, the date of publication is represented in BIBFRAME through the *bf:ProvisionActivity* class and the following path: *bf:Instance - bf:provisionActivity - bf:ProvisionActivity - bf:Instance - bf:date - xsd:integer.* This whole path will be mapped to a single RDA property, the *rdam:P30278 has date of manifestation* property. Moreover, the *bf:ProvisionActivity* class and its four subclasses, namely *bf:Publication, bf:Production, bf:Distribution,* and *bf:Manufacture,* cluster information regarding agents, dates, and places relevant to the described event. RDA describes the same events with properties for agents, dates, and places that are different for each event type. As an example, in BIBFRAME a *bf:Publication* class instance will cluster information about the *bf:Agent* (*bf:Publication-bf:agent-bf:Agent*), the exact date (*bf:Publication-bf:date-rdfs:literal*), and the place of publication (*bf:Publication-bf:place-bf:Place*). Similarly, a *bf:Distribution* class instance will use the same properties and classes to describe when and where a *bf:Agent* participated in the distribution event. In RDA, different properties are used for each case. Regarding publication, agent information is given using the *rdam:P30083 has publisher agent* property, place information is provided with the *rdam:P30088 has place of publication* property, while date with the *rdam:P30011 has date of publication* property. For distribution other properties are used; correspondingly, *rdam:P30080 has distributor agent, rdam:P30085 has place of distribution*, and *rdam:P30008 has date of distribution.*

Table 4-14 presents the mapping for some core pieces of *bf:Instance* related information. A full mapping of the *bf:Instance* properties is out of scope for this thesis. Even though, mappings are provided for a subset of *bf:Instance* properties, the difference between the BIBFRAME and RDA representations is clearly manifested. Similarly to the mapping of the *bf:Work* properties, whole BIBFRAME paths are mapped to a single RDA property.

Table 4-14. BIBFRAME to RDA mapping Step (d) - Mapping the bf:Instance properties.

Information	BIBFRAME	RDA
Carrier	bf:Instance-bf:carrier- bf:Carrier (carriers:value)	rdac:C10007 - rdam:P30001 "has carrier type" - carriers:value
Issuance	bf:Instance-bf:issuance – bf:Issuance (issuance:value)	rdac:C10007 - rdam:P30003 "has mode of issuance"- issuance:value
Media	bf:Instance-bf:media-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	bf:Instance - bf:provisionActivityStatement-rdfs:literal	rdac:C10007-rdam:P30292 "has manifestation statement" – rdfs:literal
Provision Activity	IF bf:Instance-bf:provisionActivity-bf:ProvisionActivity-rdf:type-bf:Publication then bf:Instance-bf:provisionActivity-bf:ProvisionActivity-bf:agent-bf:Agent	rdac:C10007 - rdam:P30083 "has publisher agent" - rdac:C10002 "agent"

	bf:Instance-bf:provisionActivity-bf:ProvisionActivity-bf:date-xsd:integer	rdac:C10007-rdamd:P30011 "has date of publication" - xsd:integer
	bf:Instance-bf:provisionActivity-bf:ProvisionActivity-bf:place-bf:Place	rdac:C10007 - rdam:P30088 "has place of publication" - rdac:C10009 "place"
	IF bf:Instance-bf:provisionActivity-bf:ProvisionActivity-rdf:type-bf:Distribution then	
	bf:Instance-bf:provisionActivity-bf:ProvisionActivity-bf:agent-bf:Agent	rdac:C10007 - rdam:P30080 "has distributor agent" - rdac:C10002 "agent"
	bf:Instance-bf:provisionActivity-bf:ProvisionActivity-bf:date-xsd:integer	rdac:C10007 –rdam:P30008 "has date of distribution" - xsd:integer
	bf:Instance-bf:provisionActivity-bf:ProvisionActivity-bf:place-bf:Place	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:provisionActivity-bf:ProvisionActivity-bf:agent-bf:Agent	rdac:C10007 - rdam:P30082 "has manufacturer agent" - rdac:C10002 "agent"
	bf:Instance-bf:provisionActivity-bf:ProvisionActivity-bf:date-xsd:integer	rdac:C10007-rdam:P30010 "has date of manufacture" - xsd:integer
	bf:Instance-bf:provisionActivity-bf:ProvisionActivity-bf:place-bf:Place	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:provisionActivity-bf:ProvisionActivity-bf:agent-bf:Agent	rdac:C10007 - rdam:P30081 "has producer agent of unpublished manifestation" - rdac:C10002 "agent"
	bf:Instance-bf:provisionActivity-bf:ProvisionActivity-bf:date-xsd:integer	rdac:C10007-rdam:P30009 "has date of production" - xsd:integer
	bf:Instance-bf:provisionActivity-bf:ProvisionActivity-bf:place-bf:Place	rdac:C10007 - rdam:P30086 "has place of production" - rdac:C10009 "place"

4.4.5. Findings

The BIBFRAME-RDA mapping may be considered as a challenging one. BIBFRAME defines a smaller number of core classes and derivative relationships. The main challenge is to map a set of *bf:Work* instances to proper sets of RDA *Work* and *Expression* instances. The identification of common ideational content among *bf:Works* is made using explicit representations, meaning instances of three properties all representing that the related *bf:Works* share the same ideas. These three properties are: *bf:hasExpression, bf:translation,* and *bf:otherEdition.* This approach provides a more stable approach for the mapping of the *bf:Work* class to RDA, comparing to string-comparison approaches mostly used in FRBRization projects. String comparison of information regarding authors and titles has been used in many projects (Decourselle et al., 2015; Dickey, 2008; Hickey & O'Neill, 2009; Manguinhas, Freire, & Borbinha, 2010; Sfakakis & Kapidakis, 2009) to reveal common origin. Yet, this approach is often complex and time consuming, and it depends a lot on manually typed data; literals are like to change, though, and are error-prone. In this BIBFRAME-RDA mapping, 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 developed in this thesis to assess the mappings (see paragraph 5.1 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 (Aalberg, Merčun, & Žumer, 2011; Aalberg & Žumer, 2008; Smiraglia, 2007b). 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.

With regard to the use of the *bf:hasExpression* property it must be noted that, even though, it was initially used in the RDA-BIBFRAME mapping to preserve the existence of common ideational content among the generated *bf:Works*, and to avoid the result of "orphan" *bf:Works,* in the BIBFRAME-RDA mapping it has been proved as a useful mechanism for semantic interoperability between the two models. The result is that, despite both models' different primitives and representation approaches, the same semantics may be preserved, and the interoperability may be ensured. Of course, the decision of using the *bf:hasExpression* is related to a library's cataloging policy.

The mapping of *bf:Work* properties to either RDA *Work* or *Expression* ones, was made by differentiating which *bf:Work* properties refer to ideas (and, thus, should be mapped to RDA *Work* properties), and which refer to signs (and, thus, should be mapped to RDA *Expression* properties). Information about primary contribution and titles was regarded as referring to ideas and the mapping to RDA was made using *Work* properties. Information about language, content type and other contributions was regarded as referring to signs and the mapping to RDA was made using *Expression* properties. Due to use of different primitives for the representation of the aforementioned pieces of information, information lying in specialized classes and in controlled vocabularies' values was exploited. As an example, the use of the *bf:Text* class prompted the mapping to the RDA *Expression rdae:P20001 has content type* property using the value txt taken from the RDA Content Type Vocabulary.

The mapping of derivative relationships was made to RDA *Expression* properties. There are not many properties in BIBFRAME for the representation of derivative relationships. Thus, the generic *bf:hasDerivative* property was mapped to a generic property in RDA, namely, the *rdae:P20203 is derivative (expression)* property. The *bf:translation* property to the *rdae:P20171 is translated as* property. The symmetric *bf:otherEdition* property to the generic *rdae:P20203 is derivative (expression)* property, or to its inverse *rdae:P20204 is based on (expression)* properties due to the absence of an equivalent property in RDA. The mapping of the *bf:otherEdition* property depends on the policy implemented regarding its use.

4.5. Conclusions

The main objective of this chapter has been to develop mappings by tackling each one of the heterogeneities identified in the Chapter 3. In general, BIBFRAME with its flexible definitions, the lack of specific properties for the different types of derivation, and the unification of two disjoint FRBR/RDA classes, i.e., *Work* and *Expression*, into one (the *bf:Work* class) challenged the mappings. For better mappings, a variety of approaches has been implemented.

First, the mappings considered representation approaches that are thought to be common in libraries, e.g., the realization approach for the representation of literal translations. Thus, other representation approaches enabled by the semantics of each model, as demonstrated in Chapter 3, were not considered. The selection of one representation approach over another is a cataloging policy decision that may affect mappings. The role of cataloging policies in the interoperability of data is an important finding supported by the investigation toward a BIBFRAME-EDM application profile and the three mappings developed by the thesis (FRBR-BIBFRAME, RDA-BIBFRAME, and BIBFRAME-RDA). As an example, all three mappings implement the Library of Congress approach (BIBFRAME - Bibliographic Framework Initiative, 2014; McCallum, 2017, 2018), where the *bf:Work* class is considered as equal to the union of the RDA *Work* and *Expression* entities. It should be recalled that alternative BIBFRAME representations may exist triggering different mappings. Characteristic alternative representations in BIBFRAME involve the use of the *bf:hasExpression* either with *Expression*-agnostic *bf:Work* instances, or with *bflc:Hub* instances. In the first case, the use of the *bf:hasExpression* property in an FRBR-like way presupposes the existence of *Expression*-agnostic *bf:Work* instances lacking core information contrary to the class definition. In the second case, the use of the *bf:hasExpression* property with *bflc:Hub* instances does not violate the *bf:Work* class semantics, but there is an ambiguity regarding the exact semantics of the *bflc:Hub* class and whether this class' instances may successfully accommodate the FRBR *Work* entity semantics. Thus, the thesis selected to use those BIBFRAME representations being the closest ones to the *bf:Work* class semantics. Further, the implementation of the Library of Congress approach and the utilization of instances of the *bf:hasExpression* property to cluster *bf:Works* sharing common ideational content, proved to preserve the RDA semantics of the *Work* and *Expression* entities in mappings to BIBFRAME.

Among the approaches implemented for better mappings have been the use of controlled vocabularies (all mappings), the exception of *Work* to *Work* relationships (RDA-BIBFRAME mapping), and the utilization of the *bf:hasExpression* property to represent the common ideas shared by the related *bf:Works* (RDA-BIBFRAME and BIBFRAME-RDA mapping). During the development of mappings some prerequisites and good practices have been identified. Table 4-15 accumulates the heterogeneities according to the Haslhofer & Klas classification (Bernhard Haslhofer & Klas, 2010), the thesis' approach in handling each one of them, and the findings detected during the development of mappings. The findings related to prerequisites and good practices are denoted with the "PGP" acronym.

<u>Domain conflicts</u> have been mostly identified in relation to the EDM model. For the BIBFRAME-EDM application profile, the EDM extension mechanism has been tested using typed *edm:InformationResource* class instances and

skos:Concept instances for the representation of *bf:Work* semantics. With this mechanism, BIBFRAME granularity could be added to EDM implementations. It must be noted that the addition of external BIBFRAME semantics was made using a predetermined value, i.e., "BFWork". Thus, for consistency reasons, the addition of external semantics through the use of typing properties that extend other properties' semantics must be made in conjunction with predetermined values from local or common controlled vocabularies.

Terminological mismatches were resolved by studying the definitions and the official documentation provided by each model. All three mappings presented in this chapter involve BIBFRAME. Thus, besides the definitions of the mapped models, the Library of Congress conversion specifications from MARC21 to BIBFRAME (Library of Congress, 2019) and feedback from the BIBFRAME mailing list ("BIBFRAME Mail. List," 2020) were also taken under consideration. An important note that must be made regarding definitions provided by each model, is that flexible definitions cause ambiguity, as well as varying interpretations and uses of the model's classes and properties that in the end may obstruct the development of mappings. An example of ambiguity would be the *bf:hasExpression* property enabling different representations, or the *bf:originalVersion* property which refers to reproduction despite being a sub-property of the *bf:hasDerivative* property which refers to derivation. Terminological mismatches also obstruct mappings due to increase of needed time to study each model's semantics. As an example, the common naming "Work" in FRBR and the rest of FRBR-inspired models in contrast to BIBFRAME. The meaning of "Work" in BIBFRAME is different requiring complex mappings. Another example is the *edm:isDerivativeOf* property that represents translation in EDM, whereas BIBFRAME uses the *bf:translation* property for translation and the *bf:derivativeOf* for adaptation. Adaptation is represented in EDM using the *dcterms:hasVersion* property. Existing MARC21 conversions have been used in resolving naming and terminological mismatches and in developing the mappings. MARC21 has been the de facto standard, a common language, used in libraries worldwide; thus, conversions or crosswalking tables may serve mappings in cases where models do not provide explicit definitions, or they provide flexible ones.

Abstraction level incompatibilities have been identified regarding representation constructs, core entities/classes, relationships, and bibliographic families. Even though, different representation approaches are enabled by each model, the thesis selects the one in each model that is closer to librarians' common perceptions and is frequently found in library domain implementations. A typical example is the realization approach for the representation of translation recognizing literal translations as new realizations of the same ideational content, and not as new ideational content per se. Typically, different representation approaches enabled by each model are not explicitly defined, even though they implicitly exist in the minds of the members participating in the models' editorial groups. A related example is the use of the *bf:hasExpression* property relating two *bf:Works* both carrying the same ideational content in different signs, or relating two *bf:Works* where the first one carries the ideational content only and the second one carries the same ideational content along the signs realizing it. The selection of one representation approach over another is a cataloging policy issue and it may have impact on the semantic interoperability of the dataset. Different representation approaches may cause incompatibilities even in different instances of the same model. Further, they may trigger different mappings, as it has been demonstrated in the case of the BIBFRAME-EDM application profile.

Regarding core classes, relationships and bibliographic families, a path-oriented approach has been used to develop the mappings. Paths are explicit, they present the domain and range classes of each used property, they better demonstrate each model's statements, and they actually reveal the associations between the classes and properties, already in the minds of the models' developers. The inexistence of a mapping tool that may enable path mappings was faced during the development of all mappings in this chapter. The success of the mappings of core entities without loss of information may depend on the implemented cataloging policy. As an example, the proper sets of RDA *Works* with their *Expressions* may be generated, if the cataloging policy implemented in a BIBFRAME model instance considers representing the existence of common ideational content explicitly by using the proper properties, e.g., *bf:hasExpression, bf:translation,* and *bf:otherEdition.*

Relationships may provide linkages in the bibliographic universe empowering exploration. The preservation of relationships after mappings has been a goal for the thesis' mappings. Most often, there is varying granularity between the models regarding the defined properties for the representation of bibliographic relationships. The thesis tried to preserve all relationships and their semantics. When this was not possible, the mappings were made to properties of generic nature. Even though, the semantics were lost, the relationship was pertained. The mapping of derivative relationships revealed two important issues, namely, BIBFRAME's strict approach, and the importance of *signs* in representing relationships. The dual nature of the *bf:Work* class requires the prior knowledge of both the original and the derivative signs involved in the derivation. Otherwise, the representation of the derivative relationship is not possible. As a result, the bibliographic control of *signs* (*Expression* entity/class in FRBR and RDA) is highlighted. This is a cataloging policy issue. The representation of derivative relationships at the *signs* level may enable the preservation of relationships when mapping bibliographic data modeled with either FRBR, or RDA to BIBFRAME. Contrary to FRBR and RDA and their ability to represent the progenitor *Work* for a bibliographic family, the progenitor "set of ideas" cannot be represented in BIBFRAME. The mappings use the *bf:hasExpression* property to cluster *bf:Works* realizing the same ideas. In the BIBFRAME-RDA mapping, the generated *Expressions* are clustered under the *Work* they realize using the information lying in explicit relationships, namely, the *bf:hasExpression, bf:translation,* and the *bf:otherEdition* properties.

<u>Multilateral correspondences</u> were handled using the path-oriented approach in mappings. For precise mappings, useful pieces of information were used, such as content type, carrier type, primary contribution and type of agent (Person, Corporate Body, Family), type of provision activity (publication, distribution, manufacture, production). In the cases of content and carrier type, controlled vocabularies were used triggering mappings. Thus, the cataloging of certain pieces of information using values from controlled vocabularies may enable more precise mappings. Once again, these decisions are formulated in the framework of a cataloging policy.

<u>Meta-level discrepancies</u> were overcome by using the path-oriented approach. Whole paths could be mapped to a single class or property, and vice-versa. Attribute values that may enable mappings were identified, as well as controlled vocabularies were used to increase consistency in mappings. Therefore, an important finding has been that controlled vocabularies are to be used not just for consistency, but for better mappings too. It must be noted that the selection of controlled vocabularies is a cataloging policy issue that impacts on interoperability, bearing in mind that the selection and use of different controlled vocabularies may trigger different mappings.

<u>Domain coverage incompatibilities</u> were not studied as out of scope.

By handling each type of heterogeneity, all mappings presented in this section have managed to preserve information after mappings, even though there are cases with loss of information, or loss of specificity. The following paragraphs present conclusions regarding the mapping of core entities/classes, inherent relationships and derivative relationships.

<u>Core entities/classes</u> may be preserved in all studied cases, namely the BIBFRAME-EDM application profile, the FRBR-BIBFRAME mapping, the RDA-BIBFRAME mapping and its reverse BIBFRAME-RDA mapping. In the BIBFRAME-EDM application profile, the mapping of core BIBFRAME classes depends on which EDM representation approach is used. In detail, the mapping taking the library metadata alignment report (Angjeli, Bayerische, et al., 2012) under consideration is different to the mapping that uses other EDM representation approaches. The

selection of an EDM representation that uses the *ore:Proxy* class, or typed *edm:InformationResource* class instances, or both *ore:Proxy* class and typed *edm:InformationResource* class instances prompts different mappings of the core BIBFRAME classes to EDM. It must be noted, that in the last case, when *edm:InformationResource* class instances typed as “BFWork” are used, BIBFRAME granularity is better preserved.

In the FRBR-BIBFRAME and RDA-BIBFRAME mappings, core FRBR/RDA entities may be mapped successfully to BIBFRAME. To avoid the loss of specificity, the mapping rules need to take under consideration the values of specific *Expression* and *Manifestation* attributes, namely, 1) the FRBR *form of expression* attribute along with the LC Content Types Scheme, and 2) the *form of carrier* attribute along with the Carriers Scheme. The BIBFRAME-RDA mapping of the *bf:Work* class exploits explicit representation of relationships to cluster the generated RDA *Expressions* under the RDA *Work* they realize. The mapping of *bf:Work* properties to either RDA *Work* or *Expression* ones was made exploiting specific pieces of information (e.g., titles and primary contribution for RDA *Works*) combined with values from controlled vocabularies. The mapping of *bf:Instance* properties exploited the type of agent (e.g., Person) also.

Inherent relationships have been preserved in all studied cases. In the three model reconciliation investigations, FRBR-BIBFRAME, RDA-BIBFRAME, and BIBFRAME-RDA, inherent relationships were preserved. In FRBR, or RDA, to BIBFRAME mappings, the *bf:hasExpression* property has been used to preserve the information that the generated BIBFRAME *Works* share the same ideational content.

Derivative relationships can be preserved but come with a cost in precision, especially in the cases where the mapping involves conversion of information from a model with richer representations to another model with simpler ones. FRBR, and especially RDA, identify many types of relationships in contrast with EDM and BIBFRAME. The only one case that relationships cannot be preserved are the *Work-Work* derivative relationships when mapped from FRBR or RDA to BIBFRAME.

Bibliographic families have not been preserved in all mappings. The FRBR-BIBFRAME, and the RDA-BIBFRAME mappings ignored the *Work-Work* relationships to avoid the generation of false relationships in BIBFRAME. A result of this decision was that some of the generated *bf:Works* remained unrelated to other members of their family, thus, impeding data exploration or inspection of *bf:Works* within the context of their family.

The mappings of this chapter have been assessed using a testbed. The testbed and the results of the mappings are presented in Chapter 5.

Table 4-15: Thesis' approach in tackling heterogeneities and findings. The heterogeneities are presented following the Haslhofer and Klas' categorization (Bernhard Haslhofer & Klas, 2010).

Category	Type	Heterogeneities	Thesis' approach	Findings
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	Extension mechanisms may add granularity or accommodate "external" semantics. Use of controlled vocabularies in extending semantics (PGP).
	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	Differing terminology and flexible definitions cause ambiguity/obstruct the development of mappings Model community and MARC21 conversions rules may serve mappings Lack of an updated metadata registry
Structural	Abstraction level incompatibilities	<u>Representation constructs</u>		
		Different representation approaches enabled by each model. There might be differences even between datasets using the same model.	Identification of different representation approaches enabled by each model using paths Select one approach based on librarians' common perceptions and on examples found in the library domain, e.g., realization based.	Different representation approaches: 1) exist in the minds of the members participating in the models' editorial groups – Not explicitly described. 2) may cause incompatibilities even in different instances of the same model. 3) may trigger different mappings.
		<u>Core entities/classes</u>		
		Different abstractions FRBR, RDA, LRM: four entities – WEMI BIBFRAME: three classes – WII FRBRoo: drops Manifestation / author's signs vs publisher's signs (F24 Publication Expression) EDM: almost no granularity EDM-FRBRoo: typed *edm:InformationResource* instances.	Path-oriented approach for the representation of each real-world bibliographic description case	Paths are more explicit, present domain and range classes, and represent better the statements of the model Associations between classes & properties, already in the model developers' minds, are represented with paths enabling better mappings. Lack of mapping tools. Importance of cataloging policy in mappings (PGP).
		<u>Relationships</u>		
		Translation: signs level (FRBR, LRM, RDA), concepts level (FRBRoo - derivation approach), and both concepts/signs (BIBFRAME). Aggregates - FRBRoo at signs level.	In order not to lose the relationship, mapping to more generic ones.	Strict BIBFRAME approach. Importance of signs (*Expression* entity/class). Importance of cataloging policy in mappings (PGP).
		<u>Bibliographic families</u>		
		FRBRoo: *F15 Complex Work* with other *F14 Individual Works* as members. BIBFRAME: clustering possible only if there is known connection between original and derivative signs / alternative representations with *bf:hasExpression / bflc:Hub* EDM: provider-oriented clustering.	Use of *bf:hasExpression* to cluster *bf:Works* realizing the same ideas (PGP) Use of explicit relationships in BIBFRAME-RDA mapping to check the implicit existence of a family in BIBRAME	Families are easily lost in BIBFRAME and exploration may be impeded. Importance of cataloging policy in mappings (PGP).
	Multilateral correspondences	<u>Classes</u>, e.g., *bf:Work* equals the FRBR *Work-is realized through-Expression* path *edm:ProvidedCHO* equals three disjoint FRBR entities (WEM) <u>Relationships</u>, e.g., *bf:hasDerivative* property (adaptation, summarization, transformation, etc.) edm:isDerivativeOf (translation, summarization, abstraction) dcterms:isVersionOf (versions, editions and adaptations)	Path-oriented approach Identification of other useful information to enable better mappings, e.g., content type values, type of agent (PGP).	There is critical info to be captured to enable mappings, e.g., content type, carrier type, primary contribution and type of agent (Person, Corporate Body, Family) (PGP). Values of controlled vocabularies triggered mappings. Importance of cataloging policy in mappings (PGP).
	Meta-level discrepancies	BIBFRAME uses classes where other models use properties, e.g., content type represented with attributes/properties in FRBR/RDA, and with *bf:Work* subclasses in BIBFRAME. RDA uses specific properties, while FRBRoo & LRM generic properties that can be 'typed' with specific values.	Path-oriented approach Search for attribute values that may enable mappings (PGP). Use of controlled vocabulary values for consistency & mappings (PGP).	Controlled vocabularies not just for consistency but for mappings too, e.g., roles, languages, content types, carrier types, etc. (PGP). Selection of controlled vocabularies may trigger mappings.
	Domain coverage	EDM. The providers' descriptions are really important and represented with ore:Proxy & ore:Aggregation classes.	Not studied	-

5. Mappings assessment[6]

In this chapter the creation of a testbed for the assessment of mappings is presented. The mappings were implemented using XSLT and Python languages[7]. To assess the efficiency of the mappings, the method of using gold datasets has been selected. The decisions taken for developing the gold datasets, along with the mappings' results are presented in detail. The datasets, the implemented mappings, as well as the results are presented at: http://libdata.tab.ionio.gr/models/si-mapping/si_project.html

5.1. Gold datasets

The thesis has used the findings of the investigation regarding the representation of core entities/classes, inherent relationships and other bibliographic relationships (presented in Chapter 3) to develop mapping rules (presented in Chapter 4).

To assess these mapping rules, the method of using gold datasets has been selected. The term "gold dataset" is used to describe a dataset being created under certain conditions to perform an assessment. In our case the created gold datasets are compared with the results of the implementation of the mappings described in the previous chapter. This comparison aims to assess the performance and the quality of the mappings. It is commonly thought, though, that Gold datasets are datasets created with the precaution to be valid and to include all required data for the performance of assessment tests. Therefore, it is first important to identify *what* and *how* it will be assessed.

For the purposes of this thesis, four Gold datasets were created to assess the success of mappings in terms of preserving core entities/classes, and inherent relationships. The preservation of derivative relationships is also included as a goal of the assessment study because they are among the most common relationships (G. H. Leazer & Smiraglia, 1999; Petek, 2007; Smiraglia, 1992; Smiraglia & Leazer, 1999; Tillett, 1987; Vellucci, 1995). Moreover, derivative relationships are core linking mechanisms between each bibliographic family's members (Smiraglia & Leazer, 1999; Svenonius, 2009). In detail, the four gold datasets are Gold FRBR dataset, Gold RDA dataset, and two versions of a Gold BIBFRAME dataset. The Gold FRBR dataset along with the Gold BIBFRAME, version 1 dataset were used to assess the mapping from FRBR to BIBFRAME. The assessment produced some interesting findings imposing the enhancement of both Gold datasets. Thus, both datasets were enhanced to create the Gold RDA

[6] This chapter builds upon the studies published in the following papers:

- Zapounidou S., Sfakakis M., & Papatheodorou C. (2017). Preserving Bibliographic Relationships in Mappings from FRBR to BIBFRAME 2.0. In: Kamps J., Tsakonas G., Manolopoulos Y., Iliadis L., Karydis I. (eds) Research and Advanced Technology for Digital Libraries. TPDL 2017. Lecture Notes in Computer Science, vol 10450. Springer, Cham. doi:10.1007/978-3-319-67008-9_2.
- Zapounidou S., Sfakakis M., & Papatheodorou C. (2019). Assessing the Preservation of Derivative Relationships in Mappings from FRBR to BIBFRAME. In: Garoufallou E., Sartori F., Siatri R., Zervas M. (eds) Metadata and Semantic Research. MTSR 2018. Communications in Computer and Information Science, vol 846. Springer, Cham. doi:10.1007/978-3-030-14401-2_22.
- Zapounidou S., Sfakakis M., & Papatheodorou C. (2019). Mapping Derivative Relationships from RDA to BIBFRAME 2. *Cataloging & Classification Quarterly 57* (5): 278-308. doi:10.1080/01639374.2019.1650152.
- Sfakakis M., Zapounidou S., & Papatheodorou C. (2020). Mapping derivative relationships from BIBFRAME 2.0 to RDA. *Cataloging & Classification Quarterly 58.* doi: 10.1080/01639374.2020.1821856.

[7] The mappings were implemented by Associate Professor Michalis Sfakakis, Department of Archives, Library Science and Museology, Ionian University.

and Gold BIBFRAME (version 2) datasets. These two datasets were used to assess the mapping from RDA to BIBFRAME, and vice-versa. All Gold datasets have been created with the Protégé 5.2.0. version. The Protégé resource is supported by grant GM10331601 from the National Institute of General Medical Sciences of the United States National Institutes of Health (Musen & Team, 2015). The next paragraphs present the selection of cases included in the Gold datasets, and the steps for creating each one of them.

5.1.1. Selection of cases

To ensure that the selected cases will contain derivative relationships, it was decided to focus on famous *bibliographic families* that tend to have many derivations. Eleven eminent literary works written in different languages, i.e., Ancient Greek, English, French, German, Russian, and Spanish, have been selected:

1. Cien años de soledad
2. Преступление и наказание (Crime and Punishment)
3. Don Quijote
4. Faust
5. Ἰλιὰς (Iliad)
6. Бра́тья Карама́зовы (Karamazov Brothers)
7. Madame Bovary
8. Ὀδύσσεια (Odyssey)
9. The Scarlet letter
10. Tom Sawyer
11. Wuthering Heights

Cases of derivation were identified in each *Work's* publishing history to find out that each one has been translated, abridged, adapted, and dramatized more than once. The study of the selected *bibliographic families'* publishing history used a number of literary sources (Bush, 1926; Fay, 1951; Foster, 1918; Harvitt, 1919; Nikoletseas, 2012; Remnick, 2005; Stavans, 2008; Susannah Hunnewell, 2015; Underwood, 1998), online bookstores, and digital humanities projects. Some of these digital humanities projects are: the Wikipedia page for the English translations of Homer ("English translations of Homer," 2019), the University of Michigan "Translating Homer" project (Pablo Alvarez, 2012), the Texas A&M University "Cervantes Project" ("Cervantes project: Cervantes collection," n.d.), the "Centro Virtual Cervantes" by the Instituto Cervantes ("CVC. Centro Virtual Cervantes.," n.d.), and the University of Virginia Library "Mark Twain in his times" project (Railton, 2012).

After the selection of derivation cases, the corresponding MARC21 records were collected from the Library of Congress catalog. There was a small number of records that was downloaded by the National Library of Spain. The records are stored in a Koha installation publicly available through its OPAC (http://83.212.114.162:8070). The notes in these MARC21 records were also considered, since it was not possible to collect the physical items for close inspection. The publishing history investigation revealed that some *Works' bibliographic families* have some notable derivations that further extend the bibliographic family with their own derivations. Some examples are the French translation of Don Quixote by Filleau de Saint-Martin, the English translation of Odyssey by G. Chapman and its adaptations, the English translations of Odyssey and Iliad by US scholars (R. Lattimore, S. Lombardo, R. Fitzgerald, R. Fagles) in multiple content types, the English translations of Dostoyevsky works by C. Garnett and their derivations.

Overall, the Gold datasets include cases described by the following derivative relationship types: literal translation, transformation (dramatizations, opera screenplays and librettos), adaptation (adaptation, free translation), inspiration, and other editions. The derivative relationships most frequently met in the datasets are translation, transformation, and adaptation. The “other editions” relationship involves simultaneously-published editions in other languages, regular-print reprints, and editions in another medium. In the Gold datasets the third case of “other editions” is present; derivative *Expressions* realize other textual ones in different content types, e.g., a text is read for an audiobook (spoken word) or converted to Braille alpha-numeric characters.

The Gold datasets do not include representation for other types of relationships, e.g., aggregates, whole-part, descriptive, equivalence, because they were developed to study the derivative type of bibliographic relationships. It must be noted that the collected MARC21 records include information for non-derivative relationships offering the opportunity to expand the Gold datasets with more relationships in the future. Regarding the representation of information found in the publishing history investigation and in the downloaded MARC21 records, each dataset includes instances of core entities/classes and of relationships/properties between them; subjects and other entities are not included.

5.1.2. Gold FRBR dataset

A Gold FRBR dataset was developed using the RDA element sets and vocabularies. Totally 256 records have been downloaded, the exact number of which per family is presented in Table 5-1.

Table 5-1. Works selected, and the numbers of MARC records used in the Gold FRBR dataset.

Name of Work	Number of records
Cien años de soledad	15 records
Преступление и наказание (Crime and Punishment)	29 records
Don Quijote	11 records
Faust	28 records
Ἰλιὰς (Iliad)	25 records
Бра́тья Карама́зовы (Karamazov Brothers)	21 records
Madame Bovary	32 records
Ὀδύσσεια (Odyssey)	20 records
The Scarlet letter	24 records
Tom Sawyer	31 records
Wuthering Heights	20 records

All instances in the Gold FRBR dataset, i.e., 72 *Works*, *229 Expressions*, and *257 Manifestations*, are described either implicitly or explicitly in the 256 MARC21 records (Table 5-2). The number of *Manifestation* instances is larger than the number of MARC21 records. This has happened because in one MARC21 record two *Manifestations* were described, the printed publication, and the accompanying CD which included the reading of the printed text by an actor.

Table 5-2. Occurrences of core entities in the Gold FRBR dataset.

Family	Work	Expression	Manifestation
Cien años	3	9	15
Crime&P	9	25	29
DonQuijote	4	11	11
Faust	9	29	28
Iliad	3	24	25
Karamazov	5	20	21
MmeBovary	8	27	32
Odyssey	3	17	20
ScarletLett	12	22	24
TSawyer	8	28	32
Wuthering	8	17	20
Total	72	229	257

The Gold FRBR dataset has been developed following the FRBR rules:

1. Derivative relationships have been represented using RDA properties (Table 5-3). The RDA properties used are the ones that correspond to relationships as described in FRBR (IFLA Study Group on the Functional Requirements for Bibliographic Records, 2009) and not the ones that refine FRBR relationships. All occurrences of FRBR relationships represented with RDA properties are presented per family in Table 5-4.
2. Regarding the case of literal translation, it must be noted that the realization approach has been implemented. Thus, literal translations are represented as *Expression* instances of the same *Work.* The translation relationship is represented with the *rdae:P20171 is translated as* property, only when both original and translated *Expressions* are known.
3. The representation of other derivative relationships is made according to the FRBR model specification. Some derivative relationships are represented between *Work* instance, and others between *Expression* instances using the corresponding RDA property. As an example, revisions and abridgements are represented as properties relating *Expression* instances. Adaptations and dramatizations are represented as properties relating either *Work* or *Expression* instances depending on the knowledge of the exact source texts used for the adaptation. When the source text is not known, the derivative relationship of adaptation is represented between two different *Works,* the original *Work* and the derived *Work*. When the source text is known, then the relationship is represented between two *Expressions* realizing the two different *Works,* the original *Work* and the derived one.
4. In the dataset there are many children's books with illustrations. In these cases, the illustrations were considered as integral to the intellectual realization of a *Work*. Therefore, the existence of illustrations triggered the representation of new *Expression* instances.
5. Person agents have their own URI in the dataset. If a VIAF ID (OCLC, 2018) or an LCNAF URI (Library of Congress, 2018) exists, then these person agents are related to them with *owl:sameAs* property instances. For a few person agents a corresponding VIAF ID or LCNAF URI has not been found.

Table 5-3. The derivative relationships used in the Gold FRBR dataset along with their corresponding RDA properties.

Relationship	Name of relationship in FRBR	Domain / Range	RDA property
Abridgement	has an abridgement	[E - E] (same work)	rdae:P20166 is abridged as (expression)
Revision	has a revision		rdae:P20211 is revised as
Translation	has a translation		rdae:P20171 is translated as
Adaptation	has adaptation	[E - E] (different works)	rdae:P20153 is adapted as (expression)
		[W-W]	rdaw:P10155 is adapted as (work)
Transformation	has a transformation	[W-W]	rdaw:P10016 is dramatized as (work)

Table 5-4. Occurrences of derivative relationships in the Gold FRBR dataset.

	Domain: Work – rdaw:		Domain: Expr. - rdae:				
Family	**P10155**	**P10016**	**P20171**	**P20211**	**P20166**	**P20153**	**Total**
Cien años			5				5
Crime&P	4	4	10	3		1	22
DonQuijote	1		2				3
Faust	4		2				6
Iliad	1		2	5	4	4	16
Karamazov	2	2	9	3		2	18
MmeBovary	4		2			1	7
Odyssey	1		5	3	1	4	14
ScarletLett	6	3	3				12
TSawyer	5	1			1	1	8
Wuthering	6		3				9
Total	**34**	**10**	**43**	**14**	**6**	**13**	**120**

5.1.3. Gold RDA dataset

The Gold FRBR dataset was used as a basis for the Gold RDA dataset. Yet, there were some important enhancements in the Gold RDA dataset.

1. Some MARC21 records were excluded from the gold dataset because they were published theatrical programs describing performances. Performances are out of scope in this thesis, even though the representation of performances with the new bibliographic conceptual models is not exhaustively

studied. Another reason has been the lack of guidelines regarding their representation in the target BIBFRAME model. This is further clarified in the following 5.1.4 paragraph. The total number of MARC21 records now used are 235 (Table 5-5).

2. The decision regarding illustrations considered as integral to the intellectual realization of a *Work*, thus, triggering the representation of a new *Expression* instance was revisited. After close examination of the MARC21 records and the discovery of public domain digitizations, it was made clear that the illustrations were not integral to the realization of a *Work.* Therefore, the *Expressions* including both text and illustrations were deleted, each illustration was considered as a *Work* of its own that now is aggregated in a *Manifestation* instance along with the text and other illustrations. The representation of illustrations was excluded of the Gold dataset because it needed close inspection of all cases through using either physical or digitized copies which was not possible.
3. Specialized RDA properties have been used for derivative relationships implicitly or explicitly described in the MARC21 records. As an example, there is in the dataset an adaptation of Madame Bovary as libretto. In FRBR (IFLA Study Group on the Functional Requirements for Bibliographic Records, 2009), this case is represented with the has adaptation relationship. The RDA property for this relationship is the *rdaw:P10155 is adapted as (work)*. Yet, RDA refines FRBR relationship and provides a more specialized one represented with the *rdaw:P10113 is adapted as libretto (work)* property. The properties used in the Gold RDA dataset are presented in Table 5-6.
4. Another derivative relationship was found in the dataset, for whose representation a specialized property could not be found. It is the case of “other editions” where a known textual Expression is realized using other content types, i.e., spoken word (audio book), and Braille alpha-numeric (Braille book). The relationship between the original Expression instance and the other Expression instances with the same content in spoken word or in Braille has been represented with the generic *rdae:P20203 has derivative expression* property and its inverse *rdae:P20204 is based on expression* (Table 5-6).
5. Agents (Person or Corporate Body) are identified by VIAF IDs. When there is no VIAF ID, then an LCNAF URI is selected. When no VIAF ID or LCNAF URI is found, then an agent is identified by a local URI. Previously in the Gold FRBR datasets, agents were identified using local URIs that were afterwards related with *owl:sameAs* property instances to corresponding LCNAF URIs or VIAF IDs.
6. Annotation properties have been added in *Work, Expression,* and *Manifestation* instances to provide matching values and to confirm the success of the mappings from RDA to another model. The specific structure used in these annotation properties’ values is presented in Table 5-7. The annotation values are taken from MARC21 fields’ values. The exact MARC21 fields are also referred in Table 5-7.

Due to the first two changes in the Gold RDA dataset, the numbers of core entities’ occurrences is different to the corresponding Gold FRBR dataset ones. The Gold RDA dataset includes the following entity instances: 48 *Work*s, 195 *Expressions,* and 236 *Manifestations*. The occurrences of *Work, Expression,* and *Manifestation* instances are displayed per family in Table 5-8. The occurrences of relationships in the Gold RDA dataset are presented in Table 5-9.

Table 5-5. Works selected, and the numbers of MARC records used in the Gold RDA dataset.

Name of Work	Number of records
Cien años de soledad	14 records
Преступление и наказание (Crime and Punishment)	24 records
Don Quijote	11 records
Faust	25 records
Ἰλιὰς (Iliad)	25 records
Бра́тья Карама́зовы (Karamazov Brothers)	20 records
Madame Bovary	29 records
Ὀδύσσεια (Odyssey)	19 records
The Scarlet letter	19 records
Tom Sawyer	31 records
Wuthering Heights	18 records

Table 5-6. The derivative relationships used in the Gold RDA dataset along with their corresponding RDA properties.

<table>
<tr><th>Relationship</th><th>Domain / Range</th><th>RDA property</th></tr>
<tr><td>Abridgement</td><td rowspan="4">[E - E]
(same work)</td><td>rdae:P20166 is abridged as (expression)</td></tr>
<tr><td>Other edition</td><td>rdae:P20203 is derivative (expression)</td></tr>
<tr><td>Revision</td><td>rdae:P20211 is revised as</td></tr>
<tr><td>Translation</td><td>rdae:P20171 is translated as</td></tr>
<tr><td rowspan="2">Adaptation</td><td>[E - E]
(different works)</td><td>rdae:P20110 is adapted as libretto expression
rdae:P20153 is adapted as expression</td></tr>
<tr><td>[W-W]</td><td>rdaw:P10113 is adapted as libretto work
rdaw:P10155 is adapted as work
rdaw:P10236 is adapted as opera work</td></tr>
<tr><td>Free translation</td><td>[W-W]</td><td>rdaw:P10099 is freely translated as work</td></tr>
<tr><td>Inspiration</td><td>[W-W]</td><td>rdaw:P10291 is inspiration for</td></tr>
<tr><td>Transformation</td><td>[W-W]</td><td>rdaw:P10016 is dramatized as (work)</td></tr>
</table>

Table 5-7. Annotation values' structure and examples in the Gold RDA dataset.

Entity	Annotation Structure	Example
Work	100$a Author, 100$d dates. 240 $a Title or 245$a Title.	Homer. Iliad.
Expression	240$l Language (008 Content type), 008 Language code. Contributor(s): 700 $a Name, 700$d dates 700$ e role.	English (Text), eng. Contributor(s): Fitzgerald, Robert, 1910-1985 tr.
Manifestation	Year from 008. 260$a Place:260$b Publisher, 260$c Year	1974. Garden City, N.Y., Anchor Press, 1974.

Table 5-8. Occurrences of core entities in the Gold RDA dataset.

Family	Work	Expression	Manifestation
Cien años	2	7	14
Crime&P	4	18	24
DonQuijote	4	12	11
Faust	7	25	25
Iliad	3	21	25
Karamazov	3	18	20
MmeBovary	3	22	29
Odyssey	2	15	19
ScarletLett	6	16	19
TSawyer	7	26	32
Wuthering	7	15	18
Total	48	195	236

Table 5-9. Occurrences of derivative relationships in the Gold RDA dataset.

	Domain: Work						**Domain: Expression**						
Family	**P10016**	**P10099**	**P10113**	**P10155**	**P10236**	**P10291**	**P20110**	**P20153**	**P20166**	**P20171**	**P20203**	**P20211**	**Total**
Cien años						1				4			5
Crime&P	1			2						7	1		11
DonQuijote		1								2			3
Faust				2	1					3			6
Iliad				1				1	2	10	3	2	19
Karamazov	2								1	7		2	12
MmeBovary			1	1			1		2	3			8
Odyssey				1				1	1	8	2	2	15
ScarletLett	2			1	2					2			7
TSawyer	1			4		1			4		3		13
Wuthering				5		1				3			9
Total	**6**	**1**	**1**	**17**	**3**	**3**	**1**	**2**	**10**	**49**	**9**	**6**	**108**

The Gold RDA dataset was used to assess the mapping of derivative relationships from RDA to BIBFRAME, and vice-versa. The results of both mappings are presented in 5.3 Assessment of the RDA – BIBFRAME mapping and in 5.4 Assessment of the BIBFRAME – RDA mapping.

The Gold RDA dataset is available at: http://libdata.tab.ionio.gr/models/si-mapping/resources/gold_rda_derivations_dbis_20190419.rdf . It has been uploaded in a Virtuoso RDF server and SPARQL queries can be submitted. Graph IRIs, used prefixes, and SPARQL queries are openly available at http://libdata.tab.ionio.gr/models/ds/gold-RDA-drvs_v6_20190419.html.

5.1.4. Gold BIBFRAME dataset, versions 1 and 2

There have been two versions of the Gold BIBFRAME dataset. The first version has been used in conjunction with the Gold FRBR dataset to evaluate if derivative relationships are retained in mappings from FRBR to BIBFRAME (Zapounidou, Sfakakis, & Papatheodorou, 2019a). In the first version it was tried to create original BIBFRAME descriptions being outside the limits of flat MARC21 records. The first version of the Gold BIBFRAME dataset adheres to the same principles implemented for the development of the Gold FRBR dataset, and includes:

1. one *bf:Work* class instance that represents the progenitor work of each bibliographic family in its representative language, e.g., Crime and Punishment in Russian, Iliad in Ancient Greek, Madame Bovary in French, etc.
2. *bf:translation* property instances relating all *bf:Works* with literal translations to the *bf:Work* class instance representing the progenitor, regardless there is knowledge about which exact texts were used for the literal translation. This decision exploited BIBFRAME's flexibility; it was taken to relate *bf:Works* carrying the same ideational content and to avoid rendering them orphan and unrelated to other members of their family. Other derivative relationships have been represented between *bf:Works* when there was some piece of information confirming the exact source and derived texts. Lack of such pieces of information ends in the representation of "orphan" *bf:Works* that are not related to other *bf:Works* of their family, hence, being excluded from users' possible exploration endeavors.
3. instances of the *bf:hasDerivative* property and sub-properties, namely, *bf:translation/bf:translationOf,* and *bf:otherEdition.* The *bf:originalVersionOf* property has not been used, because it refers to the reproduction relationship (Denenberg, 2017a; Library of Congress, 2016a, 2019) and may be used to relate *bf:Instance* instances only.
4. URIs where possible. Agents are identified with VIAF IDs or LCNAF URIs; roles are identified with URIs from the MARC Code List for Relators.

The first version of the Gold BIBFRAME dataset includes 230 *bf:Work* class instances and 257 *bf:Instance* class instances. Derivative relationships are set at the *bf:Work* level*.* Total number of classes' and properties' occurrences are presented in Table 5-10.

Table 5-10. Occurrences of core classes and derivative relationships in the Gold BIBFRAME dataset, first version.

Family	bf:Work	bf:Instance	bf:translation	bf:hasDerivative
Cien años	9	15	5	0
Crime&P	25	29	15	12
DonQuijote	12	11	5	1
Faust	29	28	17	4
Iliad	26	25	20	14
Karamazov	20	21	15	9
MmeBovary	27	32	13	4
Odyssey	17	20	12	9
ScarletLett	21	24	10	8
TSawyer	27	32	8	8
Wuthering	17	20	6	8
Total	**230**	**257**	**126**	**77**

This first version of the Gold BIBFRAME dataset needed to be enhanced for the reasons already presented in the Gold RDA dataset. These reasons involve:

1. the exclusion of theatrical programs from the dataset due to performances being out of scope for this thesis. Another reason has been that BIBFRAME is under development and there are no guidelines regarding the representation of performances in BIBFRAME despite the experimental LD4P outputs regarding moving images and performed music (Michelle Futornick, n.d.).
2. the representation of illustrations in aggregate publications. In the first version of the Gold BIBFRAME dataset, illustrations triggered the representation of a new *bf:Work* instance. After consulting digital copies of publications in the dataset, it was deduced that these publications were aggregating texts and illustrations. In FRBR and RDA, aggregates are represented at the *Manifestation* entity which embodies more than one *Expression* instances. In BIBFRAME 1.0, such representation was not possible due to the *bf:Work-bf:Instance* one-to-many cardinality constraint (Miller et al., 2012). Even though this constraint has been excluded from the BIBFRAME 2.0 specification, it is yet unclear if a many-to-many cardinality has been set to relate the *bf:Work* and *bf:Instance* classes. Moreover, BIBFRAME 2.0. does not provide any guideline for treatment of aggregates in its framework.

The second version is an enhancement of the first Gold BIBFRAME dataset, and it has been used to assess the preservation of derivative relationships in mappings from RDA to BIBFRAME (Zapounidou, Sfakakis, & Papatheodorou, 2019b) and vice-versa. The enhancement of the first version of the Gold BIBFRAME dataset involved the following changes:

1. the policy regarding the representation of literal translations changed. Even though, BIBFRAME intentionally uses flexible definitions that may serve different cataloguing/description policies (Schreur, 2018), the linked data mindset and the RDF language demand explicit representation of relationships between a domain class and a range class. Therefore, it was decided to use the *bf:translation* property only when both *bf:Works* including the source and derived texts (domain and range correspondingly) are known. As a result, many literal translations of the same *bf:Work* have become orphan and unrelated to the *bf:Work* – progenitor of the family.
2. The *bf:hasExpression* property has been used to cluster *bf:Works* containing the same sets of ideas realized through different sets of signs. These *bf:Works* were linked to each other successively with *bf:hasExpression* property instances. The use of the *bf:hasExpression* property as an OWL transitive one (Sean Bechhofer et al., 2004) provided a simple representation pattern.
3. The case of “other editions” was represented using the *bf:otherEdition* property. The property was used in all nine cases relating the derived *bf:Work* to the original one. This is important regarding mapping of this property to RDA.
4. Annotation properties have been added in *bf:Work,* and *bf:Instance* instances to provide explicit matching keys for entities, and for confirming the success of the mappings from BIBFRAME to another model. The specific structure used in these annotation properties’ values is presented in Table 5-11. The annotation values are taken from MARC21 fields’ values. The exact MARC21 fields are also referred in Table 5-11.

Table 5-11. Annotation values' structure and examples in the Gold BIBFRAME dataset.

Entity	Annotation Structure	Example
bf:Work	100$a Author, 100$d dates. 240 $a Title or 245$a Title. 240$l Language (008 Content type), 008 Language code. Contributor(s): 700 $a Name, 700$d dates 700$ e role.	Homer. Iliad. English (Text), eng. Contributor(s): Fitzgerald, Robert, 1910-1985 tr.
bf:Instance	Year from 008. 260$a Place:260$b Publisher, 260$c Year	1974. Garden City, N.Y., Anchor Press, 1974.

Table 5-12 presents the occurrences of core classes and derivative relationships in the second version of the Gold BIBFRAME dataset. The dataset includes 195 *bf:Works*, 236 *bf:Instances*, and 77 instances of properties representing derivative relationships. In detail, derivative relationships are represented between *bf:Works* and they include 49 instances of translation (*bf:translation*), 19 instances of various derivative relationships (*bf:hasDerivative*), and 9 instances of other editions (*bf:otherEdition*). The derivative relationships represented with the *bf:hasDerivative* refer mostly to abridgements and adaptations. They are represented with the generic *bf:hasDerivative* property due to the inexistence of particular properties for the description of derivative relationships in BIBFRAME.

Table 5-12. Occurrences of core classes and derivative relationships in the Gold BIBFRAME dataset, second version.

Family	bf:Work	bf:Instance	bf:translation	bf:hasDerivative	bf:otherEdition
Cien años	7	14	4		
Crime&P	18	24	7		1
DonQuijote	12	11	2		
Faust	25	25	3		
Iliad	21	25	10	5	3
Karamazov	18	20	7	3	
MmeBovary	22	29	3	3	
Odyssey	15	19	8	4	2
ScarletLett	16	19	2		
TSawyer	26	32		4	3
Wuthering	15	18	3		
Total	**195**	**236**	**49**	**19**	**9**

The Gold BIBFRAME dataset (second version) was used to assess the mapping of derivative relationships from RDA to BIBFRAME, and vice-versa. The results of both mappings are presented in 5.3 Assessment of the RDA – BIBFRAME mapping and in 5.4 Assessment of the BIBFRAME – RDA mapping.

The Gold BIBFRAME dataset (second version) is available at http://libdata.tab.ionio.gr/models/si-mapping/resources/gold_bf2_derivations_dbis_20190324_OnlyFamiliesC_sz20200510.rdf. It has been uploaded in a Virtuoso RDF server and SPARQL queries can be submitted. Graph IRIs, used prefixes, and SPARQL queries are openly available at http://libdata.tab.ionio.gr/models/ds/gold-BF2-drvs_v10_20190324.html .

5.1.5. Findings

The Gold datasets were developed to assess the mappings presented in Chapter 4. For their development, the same real-world cases and the same cataloging principles were used. All cases were represented, and the same principles were implemented according to each model's primitives. The representation of the selected cases was straightforward in the majority of cases. There was uncertainty mostly about BIBFRAME representations owing to the flexible definitions and the lack of documentation regarding domain and range restrictions. The lack of editing tools was also experienced during the development of the datasets.

For the better representation of relationships, a variety of resources was consulted. This was a time-consuming process but revealed important and specific relationships at the signs level, e.g., the adaptation of Odyssey by Charles Lamb was based on the Chapman's English translation. In many cases, the exact origin of a derivative text was decided based on the academic work of experts. Time or access to experts are not always available to libraries. Thus, libraries may decide when and for which parts of their collections such research will be conducted.

5.2. Assessment of the FRBR – BIBFRAME mapping

The mapping of FRBR to BIBFRAME was implemented using the XSLT language (Kay, 2017). To assess the mapping, two Gold datasets were used, an FRBR Gold Dataset and a BIBFRAME Gold Dataset (version 1). The implementation of the mapping transformed the FRBR Gold Dataset to a new BIBFRAME dataset, named BIBF1. The BIBF1 dataset was later compared to the Gold BIBFRAME dataset. The following section presents the comparison results between the BIBF1 dataset and the first version of the Gold BIBFRAME dataset.

5.2.1. Core entities/classes and inherent relationships

The occurrences of the three datasets' core entities/classes per bibliographic family are presented in Table 5-13. The 98% of *bf:Work* instances in the Gold BIBFRAME and the BIBF1 dataset are identical. The slight difference of 2% comes from the different number of FRBR *Expressions* (Gold FRBR) and of *bf:Works* (Gold BIBFRAME). The decisions regarding the representation of illustrations in both datasets caused this dissimilarity. Some illustrations were thought as integral to the text triggering the representation of a new *Expression* in the Gold FRBR dataset. When illustrations were not considered as integral but as *Expressions* of their own *Work,* they were not represented, and the *Manifestation* was considered as an aggregate. In BIBFRAME, due to the inexistence of rules regarding the handling of aggregates, all illustrated realizations were represented with new *bf:Works.* Regarding the occurrences of the *bf:Instance* class there is a 100% match between the Gold BIBFRAME dataset and the BIBF1 one.

The exact mappings are presented in the webpage 83.212.114.162/bibdata_mappings/displayMappings.php.

Table 5-13. Occurrences of core entities in the three datasets.

Family	Gold FRBR			Gold BIBFRAME		BIBF 1	
	Works	Expres.	Manifest.	Works	Instances	Works	Instances
Cien años	3	9	15	9	15	9	15
Crime&P	9	25	29	25	29	25	29
DonQuijote	4	11	11	12	11	11	11
Faust	9	29	28	29	28	29	28
Iliad	3	24	25	26	25	24	25
Karamazov Brothers	5	20	21	20	21	20	21
Madame Bovary	8	27	32	27	32	27	32
Odyssey	3	17	20	17	20	17	20
ScarletLett	12	22	24	21	24	22	24
TSawyer	8	28	32	27	32	28	32
Wuthering	8	17	20	17	20	17	20
Total	**72**	**229**	**257**	**230**	**257**	**229**	**257**

5.2.2. Derivative relationships

The comparison of results regarding the derivative relationships is presented in Table 5-14. There are three columns in the Gold FRBR dataset, Trl, LitTrl, and Deriv. The *Trl* column presents the number of instances of the property *rdae:P20171 is translated as*.

The *LitTrl* column presents the number of the literal translations found in the Gold FRBR dataset; for 43 literal translations the source text is known and the relationship is represented by the property *rdae:P20171 is translated as* (*Trl* column), while for the rest 83 literal translations, the source text is not known and no translation relationship is represented. Yet, these 83 literal translations remain *Expressions* of the same FRBR *Work,* and of their bibliographic family graph.

In the Gold BIBFRAME dataset, to avoid the danger of orphan *bf:Works* and to allow the exploration between members of the same bibliographic family, all literal translations were intentionally represented with *bf:translation* property instances regardless the source text was known or not. The number of literal translations in the Gold FRBR dataset is identical to the number of literal translations in the Gold BIBFRAME dataset. The mapping algorithm transforms relationships represented explicitly. Therefore, the 43 literal translations explicitly represented in the Gold FRBR dataset are all successfully mapped to BIBF1.

Table 5-14. Occurrence of translation and other derivations in the three datasets. The properties included in each column are presented in a note.

Family	Gold FRBR			Gold BIBFRAME		BIBF 1	
	Trl*	LitTrl*	Deriv*	Trl*	Deriv*	Trl*	Deriv*
Cien años de soledad	5	5	0	5	0	5	0
Crime & Punishment	10	15	12	15	12	10	108
Don Quijote	2	5	1	5	1	2	4
Faust	2	17	4	17	4	2	42
Iliad	2	20	14	20	14	2	33
Karamazov Brothers	9	15	9	15	9	9	53
Madame Bovary	2	13	5	13	4	2	80
Odyssey	5	12	9	12	9	5	35
Scarlet Letter	3	10	9	10	8	3	79
Tom Sawyer	0	8	8	8	8	0	128
Wuthering Heights	3	6	6	6	8	3	60
Total	**43**	**126**	**77**	**126**	**77**	**43**	**622**

*** Properties presented in each column.**
Gold FRBR, <u>Trl</u>: rdae:P20171 is translated as, <u>LitTrl</u>: no properties are included but the number of *Expressions* representing a literal translation, <u>Deriv</u>: rdae:P20211 is revised as, rdae:P20166 is abridged as (expression), rdae:P20153 is adapted as (expression), rdaw:P10155 is adapted as (work), rdaw:P10016 is dramatized as (work).
Gold BIBFRAME, <u>Trl</u>: bf:translation, <u>Deriv</u>: bf:hasDerivative
BIBF1, <u>Trl</u>: bf:translation, <u>Deriv</u>: bf:hasDerivative

The Deriv column in Gold FRBR sums the occurrences of all derivative relationships, except the *has a translation* one. The occurrences of each type of derivative relationship is presented in the Gold FRBR dataset description in Table 5-4. All types of derivative relationships, except for translation, are represented in the Gold BIBFRAME dataset with the *bf:hasDerivative* property. The total number of derivative relationships in the Gold FRBR dataset (Deriv column) is identical to the one in the Gold BIBFRAME dataset (Deriv column), 77 occurrences. Yet, after the mapping of the Gold FRBR dataset to BIBFRAME, the BIBF1 dataset presents a really large number of *bf:hasDerivative* property instances, 622. This deviation was expected because of the mapping of the 44 instances of derivative relationships represented between FRBR *Works* (Table 5-4). After the mapping to BIBFRAME, these relationships were carried to all the *bf:Works* containing realizations of the FRBR *Works* that participated in the *W-W* derivative relationships. As an example, the *Crime & Punishment* FRBR *Work* has 17 *Expressions* and has 6 other derivative FRBR *Works.* Each one of the *Crime & Punishment Work – is realized through – Expression* path is mapped to a *bf:Work* instance, 17 in total. Each one of these 17 *bf:Work* carries at least 6 *bf:hasDerivative* relationships resulting in a total of 6 x 17 = 102 occurrences! In addition, some of the 17 FRBR *Expressions* participate in *Expression-Expr*ession derivative relationships. These relationships are mapped as properties of the *bf:Works* generated by the *Work – is realized through – Expression* paths in which the *Expressions* in question participate. In this case, the generated *bf:Work* carries the 6 *bf:hasDerivative* property instances mapped from the FRBR Crime & Punishment *Work's* relationships plus the *bf:hasDerivative* property instances generated from the FRBR *Expression's* relationships. Hence, the total number of *bf:hasDerivative* property instances increases even more.

5.2.3. Findings

The FRBR-BIBFRAME mapping was successful regarding the mapping of core FRBR entities and of derivative relationships represented between *Expressions.* It must be noted, though, that, owing to the lack of specialized properties for the representation of derivative relationships in BIBFRAME, the exact semantics of the mapped FRBR relationships were lost in BIBFRAME.

Another important point is that, due to the semantics of the *bf:Work* class, all generated *bf:Works* carried the relationships of both FRBR *Work* and *Expression* entities from which they were generated. Thus, the mapping of FRBR *Work-Work* relationships produced a great number of relationships in BIBFRAME. Apparently, the majority of these relationships was wrong, and it brought "noise" to the generated BIBFRAME dataset (named as BIBF1).

5.3. Assessment of the RDA – BIBFRAME mapping

The mapping of RDA to BIBFRAME was made using the XSLT language (Kay, 2017). The implementation of the mapping transformed the RDA Gold Dataset to a new BIBFRAME dataset, named RDA2BF. The RDA2BF dataset was later compared to the Gold BIBFRAME dataset (version 2).

All three datasets, Gold RDA, Gold BIBFRAME (version 2), and RDA2BF have been uploaded in a Virtuoso RDF server and SPARQL queries can be submitted. Prefixes, graph IRIs, and the SPARQL queries used for querying each dataset, are openly available at http://libdata.tab.ionio.gr/models/si-mapping/si_project.html. In detail, the results of the RDA-BIBFRAME mapping are organized in 3 sections and are presented in 5 webpages:

- Mapping - Derivative relationships (Expression-Expression only)
    1. Display mappings
    2. RDA2BF dataset: graph IRIs, prefixes, SPARQL queries
- Test Mapping - Derivative relationships (Work-Work & Expression-Expression)
    3. Display mappings
    4. RDA2BF dataset with mapped RDA Work Relationships: graph IRIs, prefixes, SPARQL queries
- Wuthering Heights Family Example
    5. graph IRIs, prefixes, SPARQL queries, Visualizations

5.3.1. Core entities/classes and inherent relationships

Table 5-15 presents the occurrences of core entities/classes in each one of the three datasets, Gold RDA, Gold BIBFRAME (version 2), and the produced RDA2BF dataset. The RDA2BF dataset presents the exact same number of *bf:Work* and *bf:Instance* occurrences compared to the Gold BIBFRAME dataset. It is, therefore, indicated that the mapping of core RDA entities to BIBFRAME may be successful with a 100% accuracy. The clustering of mapped *bf:Work* instances that realize the same ideas was achieved using instances of the *bf:hasExpression* property. The *bf:hasExpression* property was used as an OWL transitive one relating the *bf:Work* instances successively. This mapping results in simple BIBFRAME representations and favors the preservation of RDA inherent relationships after mappings to BIBFRAME.

Table 5-15. Occurrences of core entities/classes in the three datasets.

	Gold RDA			**Gold BIBFRAME (2nd version)**		**RDA2BF**	
Family	**Work**	**Expr.**	**Manif.**	**Work**	**Inst.**	**Work**	**Inst.**
Cien años	2	7	14	7	14	7	14
Crime&P	4	18	24	18	24	18	24
DonQuijote	4	12	11	12	11	12	11
Faust	7	25	25	25	25	25	25
Iliad	3	21	25	21	25	21	25
Karamazov	3	18	20	18	20	18	20
MmeBovary	3	22	29	22	29	22	29
Odyssey	2	15	19	15	19	15	19
ScarletLett	6	16	19	16	19	16	19
TSawyer	7	26	32	26	32	26	32
Wuthering	7	15	18	15	18	15	18
Total	48	195	236	195	236	195	236

5.3.2. Derivative relationships

Table 5-16 presents the occurrences of derivative relationships in the three datasets. Derivative relationships are organized in three categories in RDA, namely derivative relationships between *Works* (WorkDeriv column), translation (Trl column), and derivative relationships between *Expressions* (ExprDeriv column).

- The RDA properties included in the WorkDeriv column are: *P10016 is dramatized as work, P10099 is freely translated as work, P10113 is adapted as libretto work, P10155 is adapted as work, P10236 is adapted as opera work,* and *P10291 is inspiration for*.
- The Trl column presents the sum of the *P20171 is translated as* property occurrences.
- The ExprDeriv column displays the sum of occurrences of the following *Expression* properties: *P20110 is adapted as libretto expression, P20153 is adapted as expression, P20166 is abridged as expression, P20203 has derivative expression*, and *P20211 is revised as*.

In the Gold BIBFRAME dataset the Trl column presents the number of the *bf:translation* property instances, while the Deriv column depicts the sum of the *bf:hasDerivative* and *bf:otherEdition* instances.

The RDA2BF dataset includes three columns. The first Trl column presents the number of *bf:translation* instances that are generated from the mapping of the 49 *P20171 is translated as* property instances of the RDA Trl column to BIBFRAME. The number of 49 *P20171 is translated as* property occurrences in RDA (RDA Trl column) are the same with the occurrences of the *bf:translation* property in the Gold BIBFRAME dataset (Trl column) and in the mapped RDA2BF dataset (Trl column). Therefore, the translation relationship is successfully preserved in RDA to BIBFRAME mappings.

The second ExprDeriv column presents the number of *bf:hasDerivative* instances that are generated from the mapping of the 28 *Expression* properties instances of the RDA ExprDeriv column to BIBFRAME. All 28 occurrences of these RDA *Expression* properties are preserved after mapping to BIBFRAME. Yet, specificity is lost because they are all mapped to the *bf:hasDerivative* property.

The third WorkExprDeriv column presents the number of *bf:hasDerivative* instances that are generated from mapping the instances of the RDA WorkDeriv and ExprDeriv columns to BIBFRAME. The *W-W* derivative relationships are ignored in the Gold BIBFRAME dataset and in the mapping. Therefore, in the Gold BIBFRAME columns there are not equivalent occurrences to the RDA WorkDeriv column properties. The scenario of including the *W-W* derivative relationships in the mapping is presented in the RDA2BF WorkExprDeriv column proving that the mapping of *W-W* derivative relationships produces redundant and erroneous derivative relationships in BIBFRAME. In detail, the 31 *W-W* derivative relationships produce 406 *bf:hasDerivative* instances. From the sum of 434 instances in the RDA2BF WorkExprDeriv column the number of the 28 *E-E* derivative relationships mapped to BIBFRAME are deducted (434-28=406). In this test mapping all *W-W* relationships were included. Even though, the Gold RDA dataset was developed for the assessment of mapping core entities, inherent relationships, and derivative relationships, it includes a small number of other types of relationships, e.g., whole-part. Redundant relationships were also produced when mapping these non-derivative relationships from RDA to BIBFRAME. Hence, the thesis recommends that all RDA *Work-Work* relationships be ignored in mappings from RDA to BIBFRAME.

Table 5-16. Occurrences of derivative relationships in the three datasets.

	Gold RDA			**Gold BIBFRAME (2nd version)**		**RDA2BF**		
Family	**WorkDeriv***	**Trl***	**ExprDeriv***	**Trl***	**Deriv***	**Trl***	**ExprDeriv***	**WorkExpr Deriv***
Cien años	1	4		4		4		10
Crime&P	3	7	1	7	1	7	1	46
DonQuijote	1	2		2		2		
Faust	3	3		3		3		22
Iliad	1	10	8	10	8	10	8	24
Karamazov	2	7	3	7	3	7	3	35
MmeBovary	2	3	3	3	3	3	3	42
Odyssey	1	8	6	8	6	8	6	19
ScarletLett	5	2		2		2		55
TSawyer	6		7		7		7	127
Wuthering	6	3		3		3		54
Total	**31**	**49**	**28**	**49**	**28**	**49**	**28**	**434**

*** Properties presented in each column.**
Gold RDA, WorkDeriv: P10016 is dramatized as work, P10099 is freely translated as work, P10113 is adapted as libretto work, P10155 is adapted as work, P10236 is adapted as opera work, and P10291 is inspiration for. Trl: P20171 is translated as. ExprDeriv: P20110 is adapted as libretto expression, P20153 is adapted as expression, P20166 is abridged as expression, P20203 has derivative expression, and P20211 is revised as.
Gold BF, Trl: bf:translation, Deriv: bf:hasDerivative and bf:otherEdition.
RDA2BF, Trl: bf:translation, ExprDeriv: bf:hasDerivative instances generated by mapping instances of the properties in Gold RDA ExprDeriv column, WorkExprDeriv: bf:hasDerivative instances generated by mapping instances of the properties in Gold RDA WorkDeriv and ExprDeriv columns.

To estimate the accuracy of mapping RDA to BIBFRAME, the RDA2BF dataset was compared to the Gold BIBFRAME dataset by dividing the "number of the mappings verified by the gold standard datasets" with the sum of the "number of mappings verified by the gold standard datasets" and the "number of mappings not verified by the gold standard datasets". The equation used is the following:

$$\text{Accuracy} = (\text{mappings verified by gold standard datasets}) \;/\; (\text{mappings verified by gold standard datasets} + \text{mappings not verified by gold standard})$$

Table 5-17 presents the results of the accuracy estimates when mapping RDA to BIBFRAME. The mappings of core entities and of the translation derivative relationship are 100% successful. Similarly, the mapping of other derivative relationships represented between RDA *Expressions* is successful in terms of preserving the relationship. The specificity of the derivation is lost, though, since all non-translation derivative relationships are mapped to the same generic *bf:hasDerivative* property. The mapping of *Work-Work* and *Expression-Expression* derivative relationships is not accurate presenting a low level of success, only 18%. This percentage will be different if other datasets are used. Yet, it must be noted that the accuracy percentage depends on the number of *Expressions* realizing the *Works* that participate in the mapped derivative relationships. In detail, the accuracy of the mapping decreases with the increase of the number of *Expressions*.

Table 5-17. Accuracy percentages of the mapping (Comparison between Gold BIBFRAME and RDA2BF core entities and derivative relationships).

Core entities		100 %
Relationships	Translation	100 %
	ExprDeriv	100 %
	WorkDeriv & ExprDeriv	28+49/434 = 18%

5.3.3. Findings

Similarly to the FRBR-BIBFRAME mapping, core RDA entities were successfully mapped to BIBFRAME. The preservation of common origin was enabled by using *bf:hasExpression* property instances to relate the generated *bf:Works* that carry different realizations of the same content. The use of the *bf:hasExpression* property as transitive resulted in simple BIBFRAME representations and preserved the clustering of all *bf:Works* realizing with different signs the same content respecting BIBFRAME model's primitives and semantics. Without the use of the *bf:hasExpression* property, all these *bf:Works* would be "orphan" and out of the context of their bibliographic family. Moreover, the exploration of the members of its bibliographic family would be impossible due to the absence of explicit links/properties relating them.

Derivative relationships represented between *Expressions* in RDA were successfully mapped to BIBFRAME, but exact semantics was lost regarding all derivative relationships except for translation. BIBFRAME provides a specialized property for translations, but all other derivative relationships may be represented with the generic *bf:hasDerivative* property. The mapping of derivative relationships represented between RDA *Works* was ignored in this mapping to avoid unnecessary "noise". A test was performed mapping all *Work-Work* relationships generating 406 *bf:hasDerivative* instances relating falsely the generated *bf:Works* in the RDA2BF dataset. The decision of ignoring *Work-Work* relationships in the RDA-BIBFRAME mapping comes with a cost; the RDA *Works* related with a *Work*-*Work* derivative relationship are mapped to *bf:Work* instances that are not related to other *bf:Works.* Thus, bibliographic families may be easily lost in BIBFRAME impeding further exploration.

5.4. Assessment of the BIBFRAME – RDA mapping

The implementation of the mapping transformed the BIBFRAME Gold dataset (version 2) to a new RDA one, called BF2RDA. The BF2RDA dataset was later compared to the Gold RDA dataset. The results of the comparison are presented in the 5.4 paragraph.

All three datasets, Gold BIBFRAME (version 2), Gold RDA, and BF2RDA have been uploaded in a Virtuoso RDF server and SPARQL queries can be submitted. Prefixes, graph IRIs, and the SPARQL queries used for querying each dataset, are openly available at http://libdata.tab.ionio.gr/models/si-mapping/si_project.html. In detail, the results of the BIBFRAME-RDA mapping are presented in 3 webpages:

1. Preprocessing and Mapping BIBFRAME to RDA
2. Display mapping (RDF XML file)
3. BF2RDA dataset: graph IRIs, prefixes, SPARQL queries

5.4.1. Core entities/classes and inherent relationships

Table 5-18 presents the occurrences of core entities/classes in each one of the three datasets, Gold BIBFRAME (second version), Gold RDA, and the produced BF2RDA dataset. The BF2RDA dataset presents the exact same number of RDA *Expressions* as the Gold RDA dataset denoting that the mapping of signs carried in *bf:Works* to RDA *Expressions* is 100% successful. Similarly, the number of the BF2RDA RDA *Manifestation* instances is the same as the Gold RDA dataset. Not surprisingly, the number of the BF2RDA *Works* is larger than the one in the Gold RDA dataset. There are 57 instances of the RDA *Work* class, contrary to the 48 instances in the Gold RDA dataset. This is due to the use of the generic *bf:hasDerivative* property relating 9 abridged *bf:Works* to their original *bf:Works.* As previously highlighted, BIBFRAME does not provide specialized properties for many derivation cases, such as abridgement, revision, adaptation, and the generic *bf:hasDerivative* property is used for their representation. In the case of the 9 *bf:Works*, even though that the existence of a derivative relationship is preserved with the use of the *bf:hasDerivative* property, the information that the abridged *bf:Works* carry the same ideational content with their original ones is lost. As a result, this representation does not provide the information needed to include abridgements in the proper subsets of the partitions generation in first step (*Step (a1)*) of the mapping algorithm, and 9 additional RDA *Works* are generated after the mapping. This difference of 9 bf:Works is less than expected. There are 10 abridgement and 6 revision cases in the Gold datasets. They are all represented with the *bf:hasDerivative* property in BIBFRAME.

Similarly to the representation of the abridgement relationship, the use of the generic *bf:hasDerivative* property for the representation of revisions has the same effect: *bf:Works* including revisions cannot be included in the proper subsets of the generated partitions during *Step (a1)* due to the loss of the needed information. Hence, the expected disparity between the Gold RDA and the BF2RDA regarding the number of *bf:Works* was a total of 16. By coincidence, all six revision cases are also translations and one abridgement case is an “other edition”. Due to the representation of translations with instances of the *bf:translation*, and of the “other edition” case with an instance of the *bf:otherEdition* property, all 7 *bf:Works* have been included in the proper subset during the execution of the *Step (a1)* of the mapping algorithm.

Table 5-18. 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**

Regarding the mapping of the *bf:Work* class to sets of RDA *Works* with their *Expressions*, it has been successful 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. All three properties indicate that the related *bf:Works* share the same intellectual content. The selection of using the *bf:hasExpression* property in a BIBFRAME model, which is a modeling/cataloging decision, can be considered as an effort to align the BIBFRAME and RDA semantics, despite their differing modeling patterns. Further, the *bf:hasExpression* property may be used for the representation of common ideational content in cases of abridgement and revision. Both cases are represented with the *bf:hasDerivative* property, which is also used for other types of derivation involving creation of new ideational content, such as, adaptation. Thus, to minimize the loss of semantics for the abridgement and revision derivative relationships, cataloging policies may choose to represent them using instance of both properties. The instance of the *bf:hasExpression* property will indicate the existence of common intellectual content among two related *bf:Works,* while the instance of the *bf:hasDerivative* property will indicate the existence of a derivative relationship between them. The difference of only 9 RDA *Works* between the Gold RDA and the BF2RDA datasets, instead of the anticipated 16, advocates for such a cataloging policy decision. This proposed modeling may ensure that the *Step (a1)* of the mapping algorithm generates the proper partitions and RDA *Work* instances.

5.4.2. Derivative relationships

Table 5-19 presents the occurrences of properties used for the representation of derivative relationships in all three datasets. The columns differentiate between translation (mapping of the *bf:translation* property) and other derivative relationships (mapping of the *bf:hasDerivative* and the *bf:otherEdition* properties). The three columns in the Gold RDA and the BF2RDA datasets present the occurrences of properties used to represent derivative relationships between RDA Works (*WorkDeriv*), translation between RDA Expressions (*Trl*), and other types of derivative relationships between RDA Expressions (*ExprDeriv*).

As mentioned, BIBFRAME Work relationships have been mapped to relationships relating RDA Expressions. In the Gold RDA dataset, there are 31 instances of *Work-Work* relationships (Gold RDA *WorkDeriv* column in Table 5-19) that cannot be represented in BIBFRAME, because the exact sets of signs used to produce the derivation are not known. These 31 derivative *bf:Works* remain unrelated to other *bf:Works* of their bibliographic families in the Gold BIBFRAME dataset. Because of this, there are no occurrences of properties relating RDA Works in the *WorkDeriv* column of the BF2RDA dataset. This suggests that, due to the absence of derivative relationships between RDA Works in the BF2RDA dataset, the exploration of *bibliographic families,* and of data in general, can be made only using the relationships at the *Expression* level.

The *Trl* columns in the Gold RDA and BF2RDA datasets present the instances of the *rdae:P20171 is translated as* property. The numbers in these two columns are the same to the number of the *bf:translation* property instances in the Gold BIBFRAME dataset. Thus, the mapping of the *bf:translation* property has been successful without any loss of its semantics.

The *ExprDeriv* columns in the Gold RDA and BF2RDA datasets present 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 (*Deriv* column). Therefore, the mapping of these two properties has also been successful. Despite the fact that the *ExprDeriv* columns in the Gold RDA and BF2RDA datasets present the same total number, it must be highlighted that different properties are instantiated in each column. There is a granularity difference between BIBFRAME and RDA. BIBFRAME represents all derivative relationships, except for translation, with the generic *bf:hasDerivative* property. RDA provides many specialized properties for the representation of derivative relationships, except for the derivative relationship represented with the *bf:otherEdition* property in BIBFRAME. Thus, even though the total of the *bf:hasDerivative* and *bf:otherEdition* instances are successfully mapped to RDA, the mapping is made to generic properties and not to the specialized ones provided by the RDA model. The consequence of this is that, despite the same number of property occurrences in the *ExprDeriv* columns of the Gold RDA and BF2RDA datasets, there is a significant loss of semantics regarding the exact nature of derivation that each property occurrence in the *ExprDeriv* column of the BF2RDA dataset represents. This result is mostly due to the BIBFRAME's less expressive semantics regarding derivative relationships, and not due to a setback of the mapping algorithm.

Table 5-19. Occurrences of derivative relationships in the three datasets (Gold BIBFRAME, Gold RDA, BF2RDA).

	Gold BIBFRAME		**Gold RDA**			**BF2RDA**		
Family	**Trl***	**Deriv***	**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	8	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	6	1	8	6	0	8	6
ScarletLett	2		5	2		0	2	
TSawyer		7	6		7	0		7
Wuthering	3		6	3		0	3	
Total	**49**	**28**	**31**	**49**	**28**	**0**	**49**	**28**

*** Properties presented in each column.**
Gold BF, Trl: bf:translation, Deriv: bf:hasDerivative and bf:otherEdition.

Gold RDA, WorkDeriv: P10016 is dramatized as work, P10099 is freely translated as work, P10113 is adapted as libretto work, P10155 is adapted as work, P10236 is adapted as opera work, and P10291 is inspiration for. Trl: P20171 is translated as. ExprDeriv: P20110 is adapted as libretto expression, P20153 is adapted as expression, P20166 is abridged as expression, P20203 has derivative expression, and P20211 is revised as.
BF2RDA, WorkDeriv: not applicable. Trl: P20171 is translated as. ExprDeriv: P20204 is based on (expression) .

Some important points need to be highlighted regarding the mapping of derivative relationships. First, BIBFRAME provides much smaller number of properties for their representation comparing to RDA. Thus, derivative relationships are successfully mapped from BIBFRAME to RDA, but the mapping is made to generic properties with a significant loss of semantics regarding the exact type of derivation represented. Secondly, due to the semantics of the *bf:Work* class, derivative relationships may be represented at the signs level and are mapped to RDA *Expression*-level properties. As a result, in cases where the exact signs used for a derivation are not known, the relationship cannot be represented and the relevant *bf:Works* remain unrelated to other *bf:Works* in their family. Further, in BIBFRAME to RDA conversions, the generated RDA dataset does not contain any *Work-Work* relationships and the exploration of bibliographic families can be made using only *Expressions* and the relationships between them. Thirdly, the representation of relationships at the signs level is recommended but it must be noted that related research will likely be needed. Given the amount of time, effort, and expertise needed for this task, a library may decide to implement this policy in certain collections of great interest or in collaboration with experts.

To estimate the accuracy of mapping BIBFRAME to RDA, the BF2RDA dataset was compared to the Gold RDA dataset by dividing the "number of the mappings verified by the gold standard datasets" with the sum of the "number of mappings verified by the gold standard datasets" and the "number of mappings not verified by the gold standard datasets". The equation used is the following:

$$\text{Accuracy} = (\text{mappings verified by gold standard datasets}) \; / \\ (\text{mappings verified by gold standard datasets} + \text{mappings not verified by gold standard})$$

Table 5-20 presents the results of the accuracy estimates when mapping BIBFRAME to RDA. The mappings of core entities is 98% successful due to 9 cases of revision. For these cases the information of common ideational content among the related *bf:Works* was lost due to the lack of a specialized property in BIBFRAME and the use of the generic *bf:hasDerivative* property. The *bf:hasDerivative* property has not been considered for the clustering of *bf:Works* in partitions due to its broad semantics incorporating both derivations that do not change the original ideas, and derivations that add new ideas to the original ones. All instances of translation present 100% accuracy. Similarly, all instances of properties denoting non-translation derivative relationships (at the signs level) were successfully mapped to properties relating RDA *Expressions.* It must be noted, though, that different properties were instantiated after the mapping (see Table 5-19). Thus, the relationships were preserved but the specificity of derivation is lost due to the generic virtue of the *bf:hasDerivative* property. Due to the semantics of the *bf:Work* class, all relationships between *bf:Works* were mapped to properties relating RDA *Expressions.* Thus, *Work-Work* derivative relationships were not mapped, and no accuracy percentages can be calculated.

Table 5-20. Accuracy percentages of the mapping (Comparison between Gold RDA and BF2RDA core entities and derivative relationships).

Core entities		98 %
Relationships	Translation	100 %
	ExprDeriv	100 %
	WorkDeriv	Not applicable (*Work* – *Work* derivative relationships cannot be mapped)

5.4.3. Findings

BIBFRAME provides flexible definitions enabling different representations. The development of the Gold BIBFRAME dataset (2nd version) was made following the same set of cataloging principles used for the development of the Gold RDA dataset. Thanks to the proper use of BIBFRAME constructs, namely the *bf:hasExpression* property, the mapping of the *bf:Work* class to sets of RDA *Works* with their *Expressions* was successful. In detail, instances of the *bf:hasExpression* property were used to relate all *bf:Works* sharing the same ideational content. Further, the property was used as an OWL transitive one and all *bf:Works* were related successively. The result has been a simple representation pattern for the clustering of many members of each bibliographic family. The use of the *bf:hasExpression* property enabled the accommodation of RDA semantics in BIBFRAME and the preservation of shared ideational content among the related *bf:Works,* clustered many of the *bf:Works* belonging to the same bibliographic family, differentiated between distinct families, and provided an interoperability mechanism between RDA and BIBFRAME. The selection of using the *bf:hasExpression* property in the Gold BIBFRAME dataset has been a modeling/cataloging decision that proved to be successful in aligning the BIBFRAME and RDA semantics, despite their differing modeling patterns.

The mapping of the *bf:Work* class to RDA exploited explicit representations of common ideational content, i.e., instances of the *bf:hasExpression, bf:translation,* and *bf:otherEdition* properties. The use of properties to trigger the mapping of the *bf:Work* class may be considered as a more sound approach comparing to string-matching of literals regarding primary contributors and titles often used in FRBRization projects and matching algorithms (Decourselle et al., 2015; Dickey, 2008; Hickey & O'Neill, 2009; Manguinhas et al., 2010; Sfakakis & Kapidakis, 2009). The implemented approach is not affected by possible inconsistencies regarding title and/or primary contribution and as a result the clustering of *bf:Works* and the mapping of the *bf:Work* class to RDA *Work* and *Expression* entities is executed successfully. Yet, inconsistencies are a real-world problem that may have an impact on the mapping of certain properties, e.g., RDA *Work* title-related properties. Thus, in the context of mapping properties, string-comparison approaches may be proven useful. It must be noted that the Gold Datasets (paragraph 5.1) developed for the assessment of the mappings do not present Title/Primary contribution inconsistencies.

Due to the use of the *bf:hasDerivative* property for representing both derivations involving the same ideational content (e.g., abridgement and revisions),and derivations involving diverse ideational content (e.g., adaptation), some of the generated RDA *Expressions* were not clustered properly and additional RDA *Works* were instantiated. Thus, for the cases of abridgement and revision that involve a derivation that does not alter the original ideational content, the use of two properties, namely *bf:hasExpression* and *bf:hasDerivative,* may be proven as a good practice. The *bf:hasExpression* property instance will represent the common ideational content among the related *bf:Works,* and the *bf:hasDerivative* property instance will represent the derivative relationship denoting which *bf:Work* is the original one, and which *bf:Work* is the derivative one. Moreover, the implementation of this good practice will deliver better mappings from BIBFRAME to RDA, correctly partitioning the *bf:Works* of the source dataset, and without generating additional RDA *Works*.

The *bf:Work* class including both ideas and signs imposed the mapping of derivative relationships to RDA *Expression* properties. All relationships were successfully mapped to RDA. Yet, due to the lack of specialized properties in BIBFRAME, the exact semantics of each derivation case was lost, and the mapping was executed to generic RDA *Expression* properties. A result of this was that in the generated BF2RDA dataset the exploration of families could be made by using *Expression*-level relationships only.

5.5. Conclusions

The main objective of this chapter has been to assess the mappings developed in Chapter 4. Three Gold datasets were created to perform the mappings and assess them. The development of the Gold datasets was time-consuming and demanded the definition and the implementation of specific rules. Moreover, the development process revealed more good practices and findings. Table 5-21 is an update of the Table 4-15 adding more actions in the thesis' approach about handling the heterogeneities found in the Haslhofer & Klas classification (Bernhard Haslhofer & Klas, 2010), and more findings detected during the assessment of mappings. For readability purposes, the updates in Table 5-21 are presented in bold font.

Domain conflicts do not exist between the mapped models, i.e., FRBR, RDA, and BIBFRAME. Yet, differing conceptualizations of the same real-world entities are common between models belonging to the same domain. The mapped models present important differences in conceptualizing real-world bibliographic description cases. The use of instances of the models representing the same real-world cases has provided an objective measure to compare the studied models. These Gold datasets have been used to assess the produced mappings. It must be noted that in the case of the Gold BIBFRAME dataset, the flexible definitions have brought uncertainty in its development process, e.g., in the representation of relationships. Even though the Protégé software was used for the development of the three Gold Datasets, the lack of more user-friendly editing tools must be highlighted.

One of the identified abstraction-level incompatibilities between the models has been the representation of relationships at a different conceptual level. To tackle this incompatibility, the thesis has tried to represent relationships at the signs level. Such information is not usually provided by the authors or publishers, nor it is ordinarily documented in bibliographic records. To find out the original signs used for the creation of derivations, a number of sources was consulted. These sources were created by experts in literature and subject librarians. Some examples are digital humanities projects, scientific articles, Wikipedia articles, etc. Expertise regarding the origins of a text or the publishing history of a *Work* may prove helpful in future library linked data projects by revealing previously unknown bibliographic relationships existing in the bibliographic universe. The explicit representation of these relationships will likely empower more exploration possibilities. Due to the amount of the needed time and resources for such research, a library may need to decide for which parts of its collection such policy would be implemented.

With regard to bibliographic families, the development of the Gold BIBFRAME dataset has shown that bibliographic families are easily fragmented in BIBFRAME due to *bf:Work* class semantics and the strict representation of relationships at the signs level only. The use of the *bf:hasExpression* property has partially preserved families as far as the clustering of *bf:Works* carrying different signs of the same ideational content is concerned. Thus, the use of the *bf:hasExpression* may be considered as an alignment between RDA and BIBFRAME enabling the representation of both intellectual content and signs in BIBFRAME despite the models' different modeling patterns. If the *bf:hasExpression* was not used, the clustering of *bf:Works* would be impossible and the majority of the *bf:Works* in the Gold BIBFRAME dataset would remain unrelated to other members of their bibliographic family. The use of the *bf:hasExpression* property did not burden the dataset with too many triples due to the thesis' approach of using the *bf:hasExpression* property as transitive for simple representations. Moreover, it was proved that the use of this property -in conjunction with the *bf:hasDerivative* property- in the representations of abridgement and revisions preserves the information about shared intellectual content between the related *bf:Works.* The implementation of this approach is a cataloging policy issue.

With regard to the mapping of the core entities/classes, all three mappings, FRBR-BIBFRAME, RDA-BIBFRAME, and BIBFRAME-RDA, preserved them fully. Only in the case of the BIBFRAME-RDA mapping, the number of generated RDA *Works* was larger comparing to the Gold RDA dataset. The representation of abridgement cases with the generic *bf:hasDerivative* property was the key factor for the generation of extra RDA *Works.* Thus, it was concluded that a better policy for the representation of abridgement and revision cases would be the use of both *bf:hasExpression* and *bf:hasDerivative* properties; the former property may represent the existence of common ideational content and enable the proper clustering of the generated RDA *Expressions* in BIBFRAME-RDA mappings, while the latter will represent the existence of a derivative relationship between *bf:Works* in BIBFRAME, and between the generated RDA *Expressions* in BIBFRAME-RDA mappings.

Even though, the mapping of properties used for the representation of derivative relationships could be described as straightforward, the mapping of derivative relationships was also challenging. During the development of the Gold BIBFRAME datasets, relationships at the ideas level (*Work-Work* relationships in FRBR and RDA) were not represented and many *bf:Work* instances became unrelated to other *bf:Works.* The implementation of the RDA-BIBFRAME mapping revealed that 1) *Work-Work* relationships must be ignored in the conversion, 2) all *Expression-Expression* derivative relationships, except for translation, were preserved in BIBFRAME but they lost in specificity, since all non-translation relationships were mapped to the *bf:hasDerivative* property. The *bf:hasDerivative* property may be used for both relating *bf:Works* sharing common ideational content (e.g., cases of revision and abridgements) and relating *bf:Works* with different ideational content (e.g., cases of adaptation, dramatization, free translation). Thus, in the BIBFRAME-RDA mapping that uses properties to cluster generated RDA *Expressions* under the *Work* they realize, the *bf:hasDerivative* property was ignored.

In the next chapter, the findings of the thesis are further discussed.

Table 5-21. Thesis' approach in tackling heterogeneities and findings – updated during the assessment of mappings (updates in bold font). The heterogeneities are presented following the Haslhofer and Klass' categorization (Bernhard Haslhofer & Klas, 2010).

Category	Type	Heterogeneities	Thesis' approach	Findings
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 **Description of the same real-world bibliographic description cases- Gold Datasets**	Extension mechanisms may add granularity or accommodate "external" semantics. Use of controlled vocabularies in extending semantics (PGP). **Use of the same real-world cases enables comparison.** **BIBFRAME's flexibility imported uncertainty regarding the representation of the cases in the Gold dataset, e.g., representation of relationships.** **Lack of editing tools**
Semantic	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	Differing terminology and flexible definitions cause ambiguity/obstruct the development of mappings Model community and MARC21 conversions rules may serve mappings Lack of an updated metadata registry
Structural	Abstraction level incompatibilities	Representation constructs		
		Different representation approaches enabled by each model. There might be differences even between datasets using the same model.	Identification of different representation approaches enabled by each model using paths Select one approach based on librarians' common perceptions and on examples found in the library domain, e.g., realization based.	Different representation approaches: 1) exist in the minds of the members participating in the models' editorial groups – Not explicitly described. 2) may cause incompatibilities even in different instances of the same model. 3) may trigger different mappings.
		Core entities/classes		
		Different abstractions FRBR, RDA, LRM: four entities – WEMI BIBFRAME: three classes – WII FRBRoo: drops Manifestation / author's signs vs publisher's signs (F24 Publication Expression) EDM: almost no granularity EDM-FRBRoo: typed *edm:InformationResource* instances.	Path-oriented approach for the representation of each real-world bibliographic description case	Paths are more explicit, present domain and range classes, and represent better the statements of the model Associations between classes & properties, already in the model developers' minds, are represented with paths enabling better mappings. Lack of mapping tools Importance of cataloging policy in mappings (PGP).
		Relationships		
		Translation: signs level (FRBR, LRM, RDA), concepts level (FRBRoo - derivation approach), and both concepts/signs (BIBFRAME). Aggregates - FRBRoo at signs level.	In order not to lose the relationship, mapping to more generic ones. **Representation of relationships at the signs level, if such info is available** (**PGP**)	Strict BIBFRAME approach Importance of signs (*Expression* entity/class) Importance of cataloging policy in mappings (PGP). **Info regarding the origins of a text or its publishing history is needed.** **Need for consulting digital humanities projects and scholarly resources** (**PGP**)**.** **Cooperation with experts may be needed in the future** (**PGP**)**.**
		Bibliographic families		
		FRBRoo: *F15 Complex Work* with other *F14 Individual Works* as members. BIBFRAME: clustering possible only if there is known connection between original and derivative signs / alternative representations with *bf:hasExpression / bflc:Hub* EDM: provider-oriented clustering.	Use of *bf:hasExpression* to cluster *bf:Works* realizing the same ideas (PGP) Use of explicit relationships in BIBFRAME-RDA mapping to check the implicit existence of a family in BIBRAME	Families are easily lost in BIBFRAME**,** and exploration may be impeded. **They may be partially preserved by using *bf:hasExpression* property instances (PGP).** ***bf:hasExpression* enables the representation of RDA semantics in BIBFRAME.** ***bf:hasExpression* as transitive provides simple representations** (**PGP**) **Use of both *bf:hasExpression* & *bf:hasDerivative* properties for the cases of abridgement and revision (PGP)** Importance of cataloging policy in mappings (PGP).
	Multilateral correspondences	Classes, e.g., *bf:Work* equals the FRBR *Work-is realized through-Expression* path *edm:ProvidedCHO* equals three disjoint FRBR entities (WEM) Relationships, e.g., *bf:hasDerivative* property (adaptation, summarization, transformation, etc.) edm:isDerivativeOf (translation, summarization, abstraction) dcterms:isVersionOf (versions, editions and adaptations)	Path-oriented approach Identification of other useful information to enable better mappings, e.g., content type values, type of agent (PGP).	There is critical info to be captured to enable mappings, e.g., content type, carrier type, primary contribution and type of agent (Person, Corporate Body, Family) (PGP). Values of controlled vocabularies triggered mappings. Importance of cataloging policy in mappings (PGP).
	Meta-level discrepancies	BIBFRAME uses classes where other models use properties, e.g., content type represented with attributes/properties in FRBR/RDA, and with *bf:Work* subclasses in BIBFRAME. RDA uses specific properties, while FRBRoo & LRM generic properties that can be 'typed' with specific values.	Path-oriented approach Search for attribute values that may enable mappings (PGP). Use of controlled vocabulary values for consistency & mappings (PGP).	Controlled vocabularies not just for consistency but for mappings too, e.g., roles, languages, content types, carrier types, etc. (PGP). Selection of controlled vocabularies may trigger mappings.
	Domain coverage	EDM. The providers' descriptions are really important and represented with ore:Proxy & ore:Aggregation classes.	Not studied	-

6. Discussion and Conclusions

This chapter summarizes the findings of the thesis and how these answer its main research question. The findings are presented and discussed, their importance for different stakeholders follows along with recommendations based on the thesis' findings. Later, the limitations of the thesis and future work are mentioned. The thesis concludes with its final statement.

6.1. Overview of the thesis

The core research question of the thesis has been "*Is semantic interoperability between conceptual bibliographic data models feasible?"*. The thesis approached the scope of contributing to the semantic interoperability of bibliographic data in the semantic web environment by posing the following four objectives:

1. study of selected conceptual bibliographic data models,
2. development of mappings,
3. assessment of mappings, and
4. identification of possible prerequisites or possible good cataloging practices for enhancing the interoperability between the library data, data exchange and data linking.

To fulfil the first objective of the thesis, granular models were selected based on the hypothesis that these may represent better common bibliographic description cases. The selected models were FRBR and its consolidation known as IFLA LRM, RDA, FRBRoo, BIBFRAME, and EDM. The selected models were studied in terms of core entities, inherent relationships, and of representing bibliographic description cases which were identified as common ones in the related literature (Bennett et al., 2003; Neill et al., 2015; Petek, 2007; Smiraglia, 1992, 1999; Smiraglia & Leazer, 1999; Tillett, 1987; Vellucci, 1995). The common bibliographic description cases identified in the literature were derivation, equivalence, and aggregates. Besides the commonality of these cases, the literature revealed the importance of bibliographic relationships in exploring bibliographic families. Exploration is a new user task added in the International Cataloguing Principles (Galeffi et al., 2017) which is fully compliant with the vision of a linked bibliographic universe. Thus, bibliographic families were included in the study. The representation of single monographs, of bibliographic relationships, and of bibliographic families in each one of the selected models revealed common ground, as well as important incompatibilities between them. These similarities and differences were organized using the Haslhofer & Klas categorization of metadata heterogeneities (Bernhard Haslhofer & Klas, 2010). The similarities support semantic interoperability and exchange of data, while the differences may impede semantic interoperability, mappings, and seamless navigation through different datasets in a unified bibliographic world. It must be noted that model-level heterogeneities were studied only, while instance-related and element-related heterogeneities were excluded from the study as out of scope (Figure 1:1). All identified semantic and structural similarities and heterogeneities are presented in Table 6-1.

To fulfil the second objective of the thesis, the development of mappings, the thesis used different approaches to tackle each type of identified heterogeneity. The thesis explored the meta-model agreement method by developing a BIBFRAME-EDM application profile, and created three mappings, 1) FRBR to BIBFRAME, 2) RDA to BIBFRAME, and 3) BIBFRAME to RDA. The thesis approach in untangling the identified structural and semantic heterogeneities is presented in Table 6-1 (column Thesis' approach).

To fulfil the third objective of the thesis, three Gold datasets have been created to perform the three mappings and to assess them. A webpage at the Database and Information Systems Research Group (part of the Laboratory on Digital Libraries and Electronic Publishing of the Ionian University, Greece) has been created to demonstrate the tools, the data and the mappings. It is available at http://libdata.tab.ionio.gr/models/si-mapping/si_project.html. The assessment of the mappings revealed that core entities/classes and inherent relationships may be preserved after mappings. There was a loss of information and specificity regarding the mapping of bibliographic relationships. Bibliographic families are more difficult to represent and to be preserved especially in BIBFRAME. The findings from the development and the assessment of mappings are analytically presented in the "Findings" column (Table 6-1).

To fulfil the fourth objective, the thesis tried to identify some prerequisites and good practices that may enable semantic interoperability between the studied models. These were identified during the development of the mappings and of the Gold datasets. They are presented in Table 6-1 either as a thesis' approach method to tackle heterogeneities (Thesis' approach column), or as a finding (Findings column). Prerequisites and good practices involve a) the representation of relationships at the signs level, b) the representation of specific pieces of information that may enable mappings, such as content type, carrier type, type of Agent, roles, etc., c) the use of common controlled vocabularies, d) cooperation with experts regarding the publishing history of works exhibiting big bibliographic families, and e) use of the *bf:hasExpression* property in BIBFRAME as transitive one. For readability reasons, both cases, prerequisites and good practices, are characterized with the "PGP" initials.

6.2. Findings of the thesis

The thesis has revealed important findings, confirming previous studies by other scholars, or opinions expressed by experts. First, some general remarks regarding the methods and tools used in the thesis are presented. Later, the findings are discussed following the Haslhofer & Klas categorization as depicted in Table 6-1.

6.2.1. General remarks for methods and tools

The thesis used a path-oriented approach in fulfilling the first and the second objectives of the thesis, the study of the models and the development of mappings. This approach facilitated the identification of 1) the modeling principles and methodology of each model, e.g., BIBFRAME uses classes where other models use properties, 2) the constraints regarding the use of each model's constructs, e.g., domain and range constraints, 3) the different representation approaches implicitly enabled by each model, and 4) similarities and heterogeneities between the models in terms of core constructs and representation approaches. All these characteristics regarding the representation of the selected bibliographic description cases would not be identified if an element-based approach would have been used instead. The correlations between classes and properties for the representation of specific bibliographic description cases -the thesis used the term "representation approach"- were expressed with paths. The use of paths enabled both the representation study and the development of mappings. Path-oriented mappings empowered successful mappings especially in cases where the models use different constructs to represent the same piece of information, e.g., BIBFRAME uses a whole path with a specific literal value to represent authorship, while RDA uses only one property. Other cases that paths empowered mappings have been those exhibiting abstraction level incompatibilities and multilateral correspondences. Further, they enabled the mapping of similar representation approaches, e.g., the mapping of the realization approach for translations from one model to another. Even though the use of paths enabled representations and mappings, the lack of tools for path-oriented mappings did not. The lack of tools that are needed for several library linked data processes has also been confirmed in OCLC (Smith-Yoshimura, 2016, 2018) and LIBER (Frosterus et al., 2020) studies, as well as in the related literature (Taniguchi, 2017b; Ullah et al., 2018; Wahid et al., 2018).

The Haslhofer & Klas categorization served as an invaluable standardization tool in the representation study - Haslhofer & Klas use the term "mapping discovery phase" (Bernhard Haslhofer & Klas, 2010) - which identified similarities and heterogeneities among the models. The results of the "mapping discovery phase" were exploited in the development of mappings. The thesis implemented a combination of approaches to tackle each type of heterogeneity. This combination of approaches may be proven successful in future semantic interoperability projects/studies.

During the development of mappings and the assessment process (second and third objectives), a significant lack of tools, also observed in (Smith-Yoshimura, 2016, 2018; Ullah et al., 2018; Wahid et al., 2018), was encountered. As an example, the inexistence of a metadata registry that includes updated definitions from several models and enables advanced search functions forced many searches in each model's documentation. The mapping process could be supported by a visual mapping tool that would ideally enable the upload of different models and include a variety of mapping functions to create 1 to 1 mappings, mappings of whole paths, and mappings using specific attribute/property values. Mapping tools already used in the library and related domains either focus on the mapping to a specific target model, e.g., the "Mapping Memory Manager - 3M" mapping tool uses the CIDOC-CRM model as target model (Marketakis et al., 2016), and the "Metadata Interoperability – MINT" tool uses the Europeana Data Model as target model (Charles, Isaac, Tzouvaras, & Hennicke, 2013), or focus on transforming legacy data from heterogeneous sources to linked data using a specific target model, e.g., the KARMA tool was used to transform data from the Smithsonian American Art Museum to the Europeana Data Model (Szekely et al., 2013).

For the assessment process, neither tools, nor gold standard bibliographic datasets were found. During the development of the Gold datasets the lack of editing tools was also experienced.

6.2.2. Semantic differences

The thesis has identified domain conflicts owing to the studied models' similar, or dissimilar domain, as well as terminological mismatches (see Table 6-1). Domain conflicts were tackled using the same real-world cases in the representation study. In the case of the most different model, the EDM model which belongs to the cultural heritage domain and exhibits a different view compared to the other models, its extension mechanism was used (instances of the *edm:InformationResource* class are further specialized using literal values modeled as instances of the *skos:Concept* class). This technique revealed that models' extension mechanisms (if they exist) may add granularity in less granular models, and they may accommodate "external" semantics without violating the internal semantics. This finding is consistent with that of the EDM-FRBRoo Application Profile Task Force (Doerr et al., 2013) who accommodated the semantics of the granular FRBRoo model to the EDM one's using the latter's extension mechanism. It also affirms the suggestion made in (Willer & Dunsire, 2013, p. 131) about the use of SKOS properties to represent same or similar semantics between metadata schemas and value vocabularies. Such extension mechanisms have been also observed in FRBRoo and IFLA-LRM regarding the use of literals for specializing properties that represent the existence of a derivative relationship without expressing the exact type of derivation, i.e., in FRBRoo the *R2.1 has type* property is used to specify the type of derivation represented with the *R2 is derivative of* property; in IFLA-LRM, the *LRM-R24 is derivation of* property may be further specialized with different literal values. Yet, both models do not provide a stable approach, i.e., a controlled value vocabulary, for populating the values of the properties that extend other properties' semantics. Both models neither identify full set of values for these properties, nor they agree on the types of derivative relationships. The existence of a common controlled vocabulary for bibliographic relationships may support consistency and preservation of bibliographic relationships' exact semantics in future mappings.

The comparison of the models by representing the same real-world cases with each model's constructs enabled the identification of each model's conceptualizations and revealed possible interoperability obstacles, such as terminological mismatches, flexible or incomplete definitions and the existence of multiple approaches in representing the same case in each model without violating its semantics. Terminological mismatches and flexible definitions, especially the BIBFRAME ones, brought uncertainty in all three studies undertaken by the thesis, representation study (Chapter 4), development of mapping (Chapter 5) and of Gold datasets (Chapter

6). This finding is consistent with related findings in (Johnston, 2005; Nillson, 2010) and also to the Professor Taniguchi's studies presenting the terminological mismatches between BIBFRAME and RDA (Taniguchi, 2017a), and the flexible BIBFRAME definitions enabling different representations (Taniguchi, 2017b, 2017a, 2018) for the same bibliographic description case. The existence of multiple representation approaches is further presented as a syntactic / abstraction-level incompatibility in the next paragraph. In the case of BIBFRAME, the MARC21 to BIBFRAME conversion rules and the BIBFRAME mailing list helped the understanding of some ambiguous definitions. MARC21, despite its scope as an exchange format, remains a commonly understood format. Thus, conversion rules from MARC21 to models, usually created by the model developers, may help the understanding of how they perceive specific properties' and classes' semantics, and the development of mappings. The existence of a model-focused community also helps, because in an active community's discussions and communications several model-related issues are resolved and clarified. Another difference related to terminology involved the use of Entity-Relationship modeling terms versus to Object Oriented one used by Semantic Web and Linked Data environments. This difference, originally observed by the W3C LLD Incubator Group Report (Baker et al., 2011) in 2011, was also affirmed later in studies conducted by Dunsire (Dunsire, 2012) and Peponakis (Peponakis, 2016). This difference is still observed in LRM (Riva, Le Boeuf, & Žumer, 2017); despite being the newest conceptual model in the bibliographic domain consolidating FRBR, FRAD, and FRSAD, Entity-Relationship modeling terms are used, namely entities, attributes, and relationships. Thus, the thesis confirms that nearly after a decade from the original observation made in (Baker et al., 2011), this incompatibility has not been resolved yet.

6.2.3. Syntactic differences

The thesis identified four types of syntactic differences, abstraction level incompatibilities, multilateral correspondences, meta-level discrepancies, and domain coverage differences (see Table 6-1).

6.2.3.1. Abstraction level incompatibilities

The <u>abstraction level incompatibilities</u> refer to models' different hierarchies for the representation of the "same real-world entities" (Bernhard Haslhofer & Klas, 2010). For readability reasons, the findings regarding them are organized in four categories:

i. representation constructs,
ii. core entities/classes,
iii. bibliographic relationships, and
iv. bibliographic families.

i. Representation constructs

The use of ambiguous definitions and wordings is not a terminology mismatch only. Ambiguity does not ensure flexibility in a model's future implementations. By contrast, it is likely to enable different representations for the same cases, further hindering interoperability, even between instances of the same model. The importance of different representation approaches has been a major finding in the thesis. For the development of mappings, focusing on the models' definitions is not enough; the representation approach(es) enabled by each model need to be identified. These representation approach(es) are not explicitly defined in most of the models, even though they implicitly exist as canons in the minds of the members participating in each model's editorial group, as well as in the documentation of each model, e.g., as domain and range restrictions. Two characteristic examples are the representation of the translation derivative relationship, and the clustering of bibliographic families' members in BIBFRAME. Translation may be either represented according to the realization approach as a new set of signs (new *Expression*) realizing the original *Work,* or according to the derivation approach as a new *Work.* In BIBFRAME, a bibliographic family's members may be clustered using different approaches, i.e., instances of the *bf:hasExpression* property, instances of the *bf:hasExpression* property with instances of *bf:Work* class lacking signs-related information (the thesis used the term *Expression*-agnostic *bf:Works*), and the use of the *bflc:Hub* class. The implementation of different representation approaches for the same bibliographic description case raises interoperability concerns among the instances of the same model. Moreover, each representation approach may trigger a different mapping to another model's constructs. The selection of approaches reflecting common perceptions among the library community, e.g., the realization-based approach for the representation of the translation case, enabled the mappings. In this context, the existence of a systematic record regarding the available representation approaches in the bibliographic and cultural heritage domains would be an important interoperability asset. This is aligned with R.Urban's vision for creating a Linked Open Data Patterns database for the Libraries, Archives, and Museums domain (LODLAM Patterns) (Urban, 2014) that could also serve as a crosswalking tool (Urban, 2013). If these canons are considered as representation constraints, then the thesis agrees with Baker, Coyle, and Petiya claiming in (Baker et al., 2014) that interoperability may be achieved as long as models share common views in terms of constraints, and proposes the adoption of more common views in the library domain. Some examples of common views, based on the thesis representation study, would be the identification of the types of bibliographic relationships, the semantic level (e.g., ideas or signs) on which they apply, etc.

ii. Core entities / classes

All studied models perceived differently the bibliographic world and what they describe. These varying perceptions impacted each model's abstractions and resulted with entities/classes having different semantics. As it has already been stated in paragraph 6.2.1, the use of the paths proved successful in expressing each model's statements and semantics, as well as domain and range restrictions.

iii. Bibliographic relationships

All studied models provide properties for the representation of bibliographic relationships; yet, there does not seem to be a consensus between the models regarding the types of bibliographic relationships, and for some relationships the semantics of the class on which these relationships apply. A distinct piece of information that has been proved important for mappings is *signs* and the relationships between them. These relationships may or may not be known by the librarians. To fill this knowledge gap, the publishing history of all bibliographic families included in the Gold Datasets, was represented in the data after consulting scholarly resources and digital humanities projects. The representation of bibliographic relationships at the signs level has been a great exploration enabler between the resources described in the Gold Datasets. Moreover, it enabled mappings especially in the case of the RDA to BIBFRAME, where a relationship at the signs level in RDA may be preserved after its mapping to BIBFRAME. It is reminded that BIBFRAME follows a strict approach enabling the representation of bibliographic relationships at the signs level only. Relationships represented between RDA *Works* cannot be mapped to BIBFRAME; the relationships are lost after the mapping and the mapped *bf:Works* become "orphan". Thus, the knowledge and the representation of bibliographic relationships at the *signs* level, when such information is available, has been identified by the thesis as a good cataloging practice. The importance of controlling information about *signs* is in agreement with Smiraglia's empirical evidence presented in (Smiraglia, 2004). In this study, Smiraglia empirically proved that the "explicit control of expressions will provide the best control over instantiation networks because it is instantiations such as translations, abridgements, and adaptations that require explicit linking". The need to cooperate with experts in order to represent a bibliographic relationship at the most specific level is in accordance with findings and views in (Creider, 2006; Rafferty, 2015; Wallheim, 2016) regarding the need for experts and collaboration to better represent the bibliographic universe in library catalogs.

iv. Bibliographic families

Bibliographic families are another enabler for exploration between bibliographic resources sharing common ideational content. FRBR and all FRBR-inspired models represent families in terms of clustering its members with relationships to the progenitor *Work.* On the contrary, *Works* in BIBFRAME and *Provided Cultural Heritage Objects* in EDM may easily become "orphan" and unrelated to other members of their bibliographic family. The *bf:hasExpression* property was used to preserve bibliographic families in BIBFRAME, but this was partially achieved. The use of the *bf:hasExpression* property enabled only the clustering of all *bf:Works* that share the same ideational content using different signs. Other *bf:Works* including other derivations (adaptations, transformations, inspirations, etc.) could not be clustered with other members of the family, due to BIBFRAME's strict approach of representing bibliographic relationships when all signs (e.g., original and derivative ones) involved in the relationship are known. The use of the *bf:hasExpression* property as a transitive one provided simpler representations and enabled the clustering of mapped RDA *Expressions* under the *Work* they realize in the BIBFRAME-RDA mapping. Recently, the Library of Congress announced its experimentations with the newly introduced *bflc:Hub* class and its use as clustering mechanism, similar to the RDA *Work* class (K. Ford, 2019a, 2019b). The thesis has expressed some reservations regarding the *bflc:Hub,* owing to the ambiguous definition of the class and to the other uses this class might serve, irrelevant ones to bibliographic families (K. Ford, 2019a, 2019b).

6.2.3.2. Other syntactic differences

Both <u>multilateral correspondences</u> and <u>meta-level discrepancies</u> were resolved using paths and controlled vocabularies. Paths enabled the one to many mappings, as well as conditional mappings using specific values. The thesis identified specific attributes/elements, such as, content type, carrier type, type of agent, etc., that may trigger better mappings preserving the semantics of the source model into the target one. The use of controlled values for populating attributes may further influence mappings. Use of common value vocabularies is likely to enable mappings, while use of uncommon or inhouse vocabularies may require new mapping rules. The thesis has provided evidence that controlled vocabularies may contribute to semantic interoperability and the preservation of semantics after mappings. There have been cases in the thesis' mappings where elements combined with certain values from controlled vocabularies triggered different mappings. This finding is in accordance with many experts suggesting the use of controlled vocabularies for consistency and ease of conversion from legacy to linked data formats (Baker et al., 2011; Edward T. O'Neill & Žumer, 2014; Suero, 2011; Wallis, 2018), as well as for better interoperability of data and easy consumption in the linked data environment (Hogan et al., 2012).

One <u>domain coverage difference</u> was identified in the thesis, the clustering of cultural heritage objects in EDM under the authority providing their metadata to Europeana. This difference was not further studied, nor tackled, as out of scope.

6.3. Importance of the findings and recommendations

The findings from this thesis may be proven useful to different user groups for a variety of reasons. The findings further support conclusions and recommendations to different stakeholders. They are all presented in Table 6-1.

6.3.1. Researchers studying semantic interoperability between models

The thesis provides a comprehensive investigation and assessment of the semantic interoperability between well-known conceptual bibliographic data models. It has focused on the interoperability between models' core modeling constructs (core entities, inherent relationships) and exploration mechanisms (bibliographic relationships and bibliographic families). It adds to the growing body of research that indicates semantic interoperability is an important issue that needs to be considered for the future integration of library linked data and the avoidance of library linked data silos. Interestingly, the methods combined for the thesis' research may support future semantic interoperability studies. The mixed methods approach implemented in this thesis consists of: 1) the Haslhofer & Klas categorization of heterogeneities, extended to include similarities too, 2) the path-oriented approach to reveal different representation approach(es) enabled by each model and to include value-based mappings using values from controlled vocabularies, and 3) Gold Standard datasets to assess the preservation of semantics after converting them from one model to another.

A natural progression of the present study is to investigate other bibliographic relationships, such as whole-part, descriptive, etc. The thesis recommends its mixed-methods approach in studying more bibliographic relationships and their corresponding representation approaches to provide full mappings. There is a chance that these studies may discover more prerequisites and good cataloging practices, or possible extension mechanisms in each model's constructs that may enable future mappings. For the provision of full mappings, extended studies that include the mapping of attributes are needed. These studies may further contribute to the identification of more attributes with values from controlled vocabularies that may enable mappings, as well as more prerequisites and good cataloging practices.

A significant finding in the thesis has been that models enable different representations for the same bibliographic description case using their constructs and without their semantics being violated. The identification and the recording of these patterns by scholars and models' development teams in a common infrastructure, similar to the LODLAM Patterns project envisioned by Professor Urban (Urban, 2013, 2014), could further contribute to the semantic interoperability in the library linked data domain.

6.3.2. Libraries - Cataloging agencies

The thesis has demonstrated that decisions taken by libraries and cataloging agencies affect the interoperability of data. These decisions involve both information selected to be recorded on regular basis, and the conceptualizations chosen to best serve their collections' description needs. Libraries can take under consideration the findings of this thesis in the formulation of cataloging policies, but also in the design of retrospective cataloging projects for the enrichment of existing bibliographic data. The findings involve (a) critical pieces of information to be included in their policies and to be systematically recorded, (b) selected controlled vocabularies, and (c) selected representation approaches for describing their collections.

Regarding points (a) and (b), the thesis has identified the following "pieces of information" as really important for mappings: content type, carrier type, primary contribution, type of agent (Person, Corporate Body, Family), and role of agent. They have been used with controlled vocabularies to trigger mappings. The thesis has used Library of Congress and RDA controlled vocabularies for the values of the aforementioned pieces of information. Thus, the cataloging of certain pieces of information using values from controlled vocabularies may enable more precise mappings. The selection of common controlled vocabularies in models' instances shall further enable their interoperability.

Regarding point (c), the thesis has identified the representation approaches for describing bibliographic description cases as a key enabler for interoperability between instances of the same model and between models also. The thesis has used the ones being most common in the library domain. As an example, the realization approach was selected for the representation of translations. Derivative relationships were preferably represented at the signs level, when such information was available. Another example, in BIBFRAME, the use of the *bf:hasExpression* property to successively relate *bf:Work* instances containing different signs for the same ideas enabled the clustering of *bf:Works* belonging to the same family without violating the semantics of the *bf:Work* class and of the *bf:hasExpression* property. Thus, representation approaches exhibiting a common perception are easier to map from one model to another. And after this finding, new issues emerge that the library community should consider. What are the representation approaches that each model enables for the representation of common bibliographic description cases?

Which ones of them disclose common perceptions among librarians? Can these common perceptions be implemented in cataloging policies so that the produced data becomes interoperable? Will future cataloging policies be oriented to representation approaches of common bibliographic description cases? Will future cataloging policies consider the RDF graph representation of the produced library data for the Semantic Web?

The support of the *explore* user task demands for different cataloging practices. The thesis' recommendations regarding libraries and cataloguing agencies are no different than experts in the field urging for less text, more URIs and structured data, and use of common vocabularies (Dunsire, Hillmann, & Phipps, 2012; Hogan et al., 2012; Wallis, 2018; Zeng, 2019). In addition, the thesis proposes that the cataloging policies will be updated for the sake of interoperability and for the support of the explore user task.

Representing bibliographic relationships is not a straightforward issue and may demand further research. The thesis has consulted many scholarly resources to identify the exact relationships between the members of the families included in the Gold Datasets. Given the huge number of publications, great amounts of time and expertise are needed to identify bibliographic relationships. And these are not available in libraries on a regular basis (Wallheim, 2016). Therefore, there must be a common agreement between National Libraries where each one will focus on the national literature, the bibliographic families that exist in it, and the bibliographic relationships between the families' members. Librarians are trained in recording the relationships, while experts may clarify the certain type of relationships. The thesis supports the establishment of collaboration between National Libraries and scholars, bibliographers, and literature departments for the description of bibliographic families, e.g., the Don Quijote family represented by the National Library of Spain, Homer's Odyssey represented by the National Library of Greece, Madame Bovary represented by the National Library of France, etc. The need of experts and the possibility of cooperation between them and the cataloging community has also been highlighted by other scholars (Creider, 2006; Rafferty, 2015; Wallheim, 2016).

The thesis' findings support the collaboration of libraries with other stakeholders and the further progression of catalogers' current mind-shift regarding library data. The output of catalogers' effort, library metadata, is no longer closed in a catalog serving only the needs of a specific community. Library data are exchanged and shared; they may be further re-purposed in new projects. The generation of library data demands the mind-shift from "act locally, think globally" to "act and think globally". Libraries need to develop collaborations, cataloging policies, and library data having interoperability in mind. This finding is consistent with that of Tallerås in (Tallerås, 2018) who urged for "new practices … to prevent new inconsistencies".

Given the "act and think globally" mentality, the thesis supports the creation of a conceptual bibliographic data models registry curated by the International Federation of Library Associations and Institutions (IFLA). The development and preservation of a metadata registry by IFLA will support research, future conversions and mappings, and interoperability of library data by providing highly-detailed information regarding different models through one central infrastructure.

6.3.3. Models' development/editorial groups

Members of the editorial groups of the studied models may take under consideration the thesis' findings in the future updates of the models. The thesis' findings involve terminology mismatches, flexible definitions, differing representation approaches, different types of relationships acknowledged by each model, and selection of controlled vocabularies. Regarding terminology, there are two recommendations. The first involves the accordance of library community terminology to the semantic web terminology. Libraries use mostly entity-relationship modeling terms, i.e., entities, attributes and relationships, while in the Semantic Web the terms classes and properties are used. This terminological discordance prevents librarians from understanding the Semantic Web and shall cause problems in future library linked data projects where librarians need to collaborate with IT staff and developers. The second involves the relabeling of properties to wordings used in other models to enable mappings. Despite the thesis focus on core entities and bibliographic relationships, there have been found attributes/properties that serve the same scope but are labelled differently. In this case, same or similar labels could enable mappings. The opposite case where same or similar labelling is used for different semantics was also noticed. Even though, this case could also be resolved through collaboration between the different models' editorial groups, it is considered more time consuming than the former one.

The issue of flexible BIBFRAME definitions must be considered by the BIBFRAME community. The thesis has provided evidence, in support of Taniguchi's findings (Taniguchi, 2013, 2017b, 2017a, 2018), that flexible BIBFRAME definitions cause ambiguity and may cause interoperability problems even between different BIBFRAME instances. Despite the need for further research regarding the semantics and the uses of the *bflc:Hub* class, the thesis urges for a better definition. Based on related findings of this thesis, it is assumed that the varying uses of the *bflc:Hub* class (enabling the collocation of various resources also, e.g., contributors, variant titles, translations, etc.) will likely cause more structural heterogeneities that need to be tackled in future mappings and conversions of data.

With regard to the representation approaches, the thesis supports the development of a semantic interoperability infrastructure similar to LODLAM Patterns envisioned by Professor R.Urban (Urban, 2014). This tool would document the representation approaches that each model enables for specific bibliographic description cases. In case it would be shared and used by different model editorial groups, it could contribute to the development of mappings. If the differing representation approaches are identified and recorded, new questions arise: can these representation approaches be uniquely identified and declared in the administrative metadata of a library dataset triggering different mappings? Will it be possible for a library to explicitly state the representation approaches it has selected for its data?

Even though, all models provide properties for the representation of bibliographic relationships there are many differences in terms of granularity. Especially, with respect to the derivation relationship, the thesis observed many differences that hindered the performed mappings. The thesis recommends the development of a vocabulary integrating the Tillett taxonomy of bibliographic relationships and the Smiraglia extensions regarding derivative relationships. Toward the development of a value vocabulary for bibliographic relationships, the first step to be made is the consensus among the library community regarding the types of bibliographic relationships existing in the bibliographic universe. If the Tillett taxonomy along with the Smiraglia extension are represented in a vocabulary, they may be used to extend the semantics of generic properties. As an example, FRBRoo provides the *R2.1 has type* property to better express the nature of the represented derivative relationship. The FRBRoo *R2.1 has type* property may have as value one of the following *E55 Type* instances, "Abridgement", "Adaptation", "Arrangement", "Imitation", "Revision", "Summary", "Transformation", and "Translation". A similar mechanism could be used for generic derivative properties in LRM, EDM, and BIBFRAME. The *LRM-R22 is a transformation of,* the *LRM-R24 is derivation of,* the *edm:isDerivativeOf*, and the *bf:hasDerivative* properties could be typed using values from a Tillett/Smiraglia vocabulary. Thus, multilateral correspondences may be avoided, no big changes need to take place regarding the introduction of many new properties in models' future updates, and the semantics of the relationship could be represented in each model and be also preserved in future mappings.

A Tillett/Smiraglia vocabulary for bibliographic relationships is recommended based on the thesis findings regarding the importance of controlled vocabularies in mappings. Their role for data consistency is unanimously acknowledged; the thesis has provided evidence that the selection of controlled vocabularies may trigger mappings. Even though mappings between controlled vocabularies may be developed, the thesis recommends the adoption of common vocabularies for assigning values to certain properties, e.g., content type and carrier type, as an easier and time-saving interoperability approach. The vocabulary editorial groups may consider the addition of new values to enable mappings. As an example, the thesis recommends the addition of a generic value for cartography (currently there are 6 values for specific cartographic types) to enable the mapping of the *bf:Cartography* class in BIBFRAME-RDA mappings.

Finally, due to the lack of tools encountered during the research, the thesis recommends the sharing of models' schemas and vocabularies in metadata registries. Existing metadata registries, such as the Open Metadata Registry ("Open Metadata Registry," 2010), are not always updated, nor include many models belonging to the same domain. Moreover, the participation of models' editorial groups in the development and the updates of metadata registries could hopefully provide more insight to software developers for advanced or new metadata registries functions than could support mapping processes, such as search of properties using filters for domain/range constraints.

6.3.4. Software developers

Even though the lack of tools is most certainly known to software developers, the thesis' thorough presentation of representations, of approaches in handling heterogeneities, and of mappings may help software developers in determining the workflow that must be supported by the mapping and data interlinking software they hopefully develop.

Based on the thesis workflow and findings, a mapping tool would ideally include an import tool, a preprocessing tool, a visual mapping editor, a convertor, and a reporting tool. The import tool would enable the importing of different model schemas/vocabularies, test data, controlled vocabularies, and patterns (if a similar to LODLAM patterns database would exist). The preprocessing tool would enable the linking of selected value vocabularies to specific properties, as well as the selection of patterns needed for the mapping. The visual mapping editor would enable the mapping with different functions, such as 1 to 1 mapping, 1 to many mappings, mappings of whole paths, pattern-to-pattern mappings, and mappings using specific attribute/property values. The convertor would implement the mapping and convert the test data. The reporting tool could provide information regarding the test data before and after the conversion, e.g., instances of classes and properties. This whole process could be stored as a project that could be further shared or edited.

Other tools that would have helped the thesis' research are user-friendly editing tools for the development of the Gold Datasets, an updated metadata registry with advanced search functions using combinations of keyword and domain/range constraints.

6.4. Limitations of the thesis

The thesis' findings have been produced studying a predefined number of models: namely, FRBR and its consolidation IFLA LRM, RDA, FRBRoo, BIBFRAME, and EDM. The semantic interoperability between these models was studied in terms of their core conceptualizations, i.e., entities/classes, inherent relationships, bibliographic relationships (derivative/translation, derivative/adaptation, equivalence, aggregates), and bibliographic families. The examination of more bibliographic relationships or of entities' attributes may reveal more incompatibilities between the models that need to be taken under consideration in future mappings.

Another limitation of the study has been that its focus on the representation of bibliographic description cases referring to monographs. The representation of other types of materials, e.g., serials, musical works, performances, was not studied.

The assessment of the mappings was performed using three Gold datasets. The use of other Gold Datasets is expected to produce different results in terms of absolute numbers and percentages.

6.5. Future work

This thesis is not the end, but it may be the starting point of new research endeavors.

The mappings provided in this thesis need to be updated to include aggregates and other bibliographic relationships. Therefore, new representation studies, updated mappings and Gold datasets will follow for each selected case. Hopefully, these studies will identify more prerequisites and good practices to be incorporated in cataloging policies for better mappings.

IFLA has published the IFLA-LRM vocabulary in July 2020. Thus, the research will continue focusing on mappings between the IFLA LRM and the BIBFRAME. The mappings will be developed following the methodology in the thesis. The thesis already provides the representations for 6 real-world bibliographic description cases (simple monographs, translation, adaptation, digitization, aggregates, and bibliographic families), as well as similarities and differences between the models. Thus, the future study will exploit the representation findings to create i) a new Gold Dataset using IFLA-LRM constructs, and ii) two mapping algorithms: IFLA LRM to BIBFRAME and vice-versa. Both mappings will be assessed using the corresponding Gold Datasets.

Lastly, the thesis has expressed reservations regarding the *bflc:Hub* class. The official definition of the class has not been published yet, and the Library of Congress still experiments with it. After Library of Congress publishes the official definition of the class, a study regarding the semantics of this class and its possible uses will follow to examine if its application may facilitate the clustering of *bf:Works* in bibliographic families.

6.6. Final statement of the thesis

Despite the expected losses of information during conversions of data, semantic interoperability is feasible, because heterogeneities can be overcome if librarians, catalogers, models' editorial groups and people involved in library linked data projects adopt a common mindset and practices, start thinking and acting at a global scale, and collaborate resolving heterogeneities of the past and preventing new ones from happening.

Table 6-1. Semantic and structural similarities/heterogeneities among the studied models, FRBR, LRM, RDA, FRBRoo, BIBFRAME, and EDM. Similarities/heterogeneities are presented following the Haslhofer and Klass' categorization (Bernhard Haslhofer & Klas, 2010). *Thesis' approach in tackling heterogeneities and findings are also presented. The thesis' findings support conclusions, suggestions, and further work.*

Category	Type		Similarities	Heterogeneities	Thesis' approach	Findings	Conclusions / Suggestions / Further Work
Semantic	Domain agreements / conflicts		Same or similar domain for bibliographic products Capture same/similar info	EDM cultural heritage domain. Different conceptualizations of real-world bibliographic description cases e.g., core entities, types of bibliographic relationships, constraints	EDM application profile with skos extension mechanism Description of the same real-world bibliographic description cases, single-volume monographs, bibliographic relationships (derivative, equivalence, and aggregates), and bibliographic families - Gold datasets	Extension mechanisms may add granularity or accommodate "external" semantics. Use of controlled vocabularies in extending semantics (PGP). Use of the same real-world cases enables comparison. BIBFRAME's flexibility imported uncertainty regarding the representation of cases in the Gold dataset, e.g., bibliographic relationships. Lack of editing tools	Extension mechanism + controlled vocabulary to represent external conceptualizations. Controlled vocabulary for bibliographic relationships incl. Tillett's taxonomy & Smiraglia's refinements. User-friendly editing tools.
Semantic	Terminological (mis)matches		FRBR, FRBRoo, LRM, RDA – WEMI BIBFRAME Item – with WEMI Item entity Many common terms, e.g., statement of responsibility	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	Differing terminology and flexible definitions cause ambiguity/obstruct the development of mappings Model community and MARC21 conversions rules may serve mappings Lack of a metadata registry	Adopt SW/LD terminology Relabel properties for easier mapping. Need for more robust definitions especially in BIBFRAME. Need for a metadata registry.
Structural	Abstraction level (in)compatibilities	Representation constructs	RDF as common syntactic language. Entities/classes with properties as attributes or as relationships. Relationships either inherent or bibliographic.	Different representation approaches enabled by each model. There might be differences even between datasets using the same model.	Identification of different representation approaches enabled by each model using paths Select one approach based on librarians' common perceptions & on examples found in the library domain, e.g., realization based.	Different representation approaches: 1) exist in the minds of the members participating in the models' editorial groups – Not explicitly described. 2) may cause incompatibilities even in different instances of the same model. 3) may trigger different mappings.	LODLAM Patterns or a similar one based on real-world bibliographic description cases Identifiable patterns imported as admin metadata in datasets enabling the triggering of different mappings.
Structural	Abstraction level (in)compatibilities	Core entities/classes	Content vs carrier	Different abstractions FRBR, RDA, LRM: four entities – WEMI BIBFRAME: three classes – WII FRBRoo: drops Manifestation / author's vs publisher's signs (F24 Publication Expression) EDM: almost no granularity EDM-FRBRoo: typed *edm:InformationResource* instances.	Path-oriented approach for the representation of each real-world bibliographic description case	Paths are more explicit, present domain and range classes, and represent better the statements of the model Associations between classes & properties, already in the model developers' minds, are represented with paths enabling better mappings. Lack of mapping tools Importance of cataloging policy in mappings (PGP).	Development of mapping tools enabling mapping of paths also.
Structural	Abstraction level (in)compatibilities	Representation of relationships	Adaptation – most abstract entity/class Equivalence/reproduction – embodiment level Aggregates-most models as signs embodied in manifestation	Translation: signs level (FRBR, LRM, RDA), concepts level (FRBRoo - derivation approach), both concepts/signs (BIBFRAME). Aggregates - FRBRoo at signs level.	In order not to lose the relationship, mapping to more generic ones. Representation of relationships at the signs level, if such info is available. (PGP)	Strict BIBFRAME approach Importance of signs (*Expression* entity/class) Importance of cataloging policy in mappings (PGP). Info regarding the origins of a text or its publishing history is needed. Need for consulting digital humanities projects & scholarly resources. (PGP) Cooperation with experts may be needed in the future. (PGP)	Consensus regarding types of relationships. Use of extension mechanisms & value vocabularies. Representation of relationships at the most specific level if there is such information available. Create cooperation networks for publishing history of well-known works. Investigate other bibliographic relationships too. Mapping of attributes too.
Structural	Abstraction level (in)compatibilities	Bibliographic families	FRBR, LRM, and RDA: clustering using progenitor *Work*	FRBRoo: *F15 Complex Work* with other *F14 Individual Works* as members. BIBFRAME: clustering possible only if there is known connection between original and derivative signs / alternative representations with *bf:hasExpression / bflc:Hub* EDM: provider-oriented clustering.	Use of *bf:hasExpression* to cluster *bf:Works* realizing the same ideas (PGP) Use of explicit relationships in BIBFRAME-RDA mapping to check the implicit existence of a family in BIBRAME	Families are easily lost in BIBFRAME, partially preserved by bf:hasExpression (PGP) *bf:hasExpression* enables the representation of RDA semantics in BIBFRAME. *bf:hasExpression* as transitive provides simple representations (PGP) Use of both *bf:hasExpression* & *bf:hasDerivative* properties for the cases of abridgement and revision (PGP) Importance of cataloging policy in mappings (PGP).	Definition of the *bflc:Hub* class. *bflc:Hub* class to be further studied regarding bibliographic families.

	Direct / Multilateral correspondences	Classes, e.g., FRBR *Manifestation / rdac:C10007 Manifestation / bf:Instance* Inherent relationships, e.g., FRBR *is embodied in/ rdae:P20059 / bf:hasInstance* Bibliographic relationships, e.g., FRBR *has a translation* /BIBFRAME *bf:translation*	Classes, e.g., *bf:Work* equals the FRBR *Work-is realized through-Expression* path / *edm:ProvidedCHO* equals three disjoint FRBR entities (WEM) Relationships, e.g., *bf:hasDerivative* (adaptation, summarization, transformation, etc.) / edm:isDerivativeOf (translation, summarization, abstraction) / dcterms:isVersionOf (versions, editions and adaptations)	Path-oriented approach Identification of other useful information to enable better mappings, e.g., content type values, type of agent (PGP)	There is critical info to be captured to enable mappings, e.g., content type, carrier type, primary contribution and type of agent (Person, Corporate Body, Family) (PGP) Values of controlled vocabularies triggered mappings. Importance of cataloging policy in mappings (PGP).	Study of mapping attributes and identifying other pieces of critical info Extension of properties for bibliographic relationships, e.g., *bf:hasDerivative* in BF to avoid multilateral correspondences, or addition of more properties for bibliographic relationships Libraries evaluate policies if they systematically record critical pieces of info that enable mappings. Update/Enrich cataloging policies.
	Meta-level matches / discrepancies	Information about the same real-world objects is captured/represented using same constructs, e.g., embodiment is captured using classes in both RDA (*rdac:C10007*) & BIBFRAME (*bf:Instance*).	BIBFRAME uses classes where other models use properties, e.g., content type represented with attributes/properties in FRBR/RDA, and with *bf:Work* subclasses in BIBFRAME. RDA uses specific properties, while FRBRoo & LRM generic properties that can be 'typed' with specific values.	Path-oriented approach Search for attribute values that may enable mappings (PGP) Use of controlled vocabulary values for consistency & mappings (PGP)	Controlled vocabularies not just for consistency but for mappings too, e.g., roles, languages, content types, carrier types, etc. (PGP) Selection of controlled vocabularies may trigger mappings.	Agreement on using common vocabularies Enrich vocabularies with new values to enable mappings, e.g., generic Cartography value.
	Domain coverage		EDM. The providers' descriptions are really important and represented with ore:Proxy & ore:Aggregation classes	Not studied	-	-

7. References

Aalberg, T., Haugen, F. B., & Husby, O. (2006). A Tool for Converting from MARC to FRBR. In J. Gonzalo, C. Thanos, M. F. Verdejo, & R. Carrasco (Eds.), *Research and Advanced Technology for Digital Libraries SE - 41* (Vol. 4172, pp. 453–456). Berlin Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/11863878_41

Aalberg, T., Merčun, T., & Žumer, M. (2011). 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)* (Vol. 7008 LNCS, pp. 128–137). https://doi.org/10.1007/978-3-642-24826-9_18

Aalberg, T., Vennesland, A., & Farrokhnia, M. (2015). A Pattern-Based Framework for Best Practice Implementation of CRM/FRBRoo. In T. Morzy, P. Valduriez, & L. Bellatreche (Eds.), *New Trends in Databases and Information Systems: ADBIS 2015 Short Papers and Workshops, BigDap, DCSA, GID, MEBIS, OAIS, SW4CH, WISARD, Poitiers, France, September 8-11, 2015. Proceedings* (Vol. 539, pp. 438–447). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-23201-0_44

Aalberg, T., & Žumer, M. (2008). 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*, 327–330. https://doi.org/10.1007/978-3-540-89533-6-36

Aalberg, T., & Žumer, M. (2013). The value of MARC data, or, challenges of frbrisation. *Journal of Documentation*, *69*(6), 851–872. https://doi.org/10.1108/JD-05-2012-0053

Alemu, G., Stevens, B., Ross, P., & Chandler, J. (2012). Linked Data for libraries: Benefits of a conceptual shift from library-specific record structures to RDF-based data models. *New Library World*, *113*(11), 549–570. https://doi.org/10.1108/03074801211282920

Angjeli, A., Baumgartner, M., Charles, V., Clayphan, R., Deliot, C., Eriksson, J. J., … Gehrke, S. (2012). *D5.2 Library domain metadata aligned with the Europeana Data Model Version 1.0*. Retrieved from http://www.europeana-libraries.eu/documents/868553/c2d8203e-1f6f-4862-b97f-589a45840eea

Angjeli, A., Bayerische, M., Chambers, S., Charles, V., Clayphan, R., Deliot, C., … Rühle, S. (2012). *D5.1 Report on the alignment of library metadata with the European Data Model (EDM) Version 2.0*. Retrieved from http://www.theeuropeanlibrary.org/confluence/download/attachments/12091395/D5.1_EDM_for_libraries_v2.0.pdf

Avram, H. (1975). *MARC: History and Implications*. Washington, D.C.

Baca, M. (Ed.). (2016). *Introduction to Metadata* (3rd ed.). Los Angeles, CA: Getty Publications. Retrieved from http://www.getty.edu/publications/intrometadata/

Baker, T., Coyle, K., Dunsire, G., Isaac, A., Murray, P., Panzer, M., … Zeng, M. (2011). *Library Linked Data Incubator Group Final Report*. Retrieved from http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/

Baker, T., Coyle, K., & Petiya, S. (2014). Multi-entity models of resource description in the Semantic Web A comparison of FRBR, RDA and BIBFRAME. *Library Hi Tech*, *32*(4), 562–582. https://doi.org/10.1108/LHT-08-2014-0081

Bekiari, C., Doerr, M., Le Bœuf, P., & Riva, P. (2015). *FRBR Object-Oriented Definition and Mapping from FRBRer, FRAD and FRSAD (version 2.4)*. Paris: International Working Group on FRBR and CIDOC CRM Harmonisation. Retrieved from http://www.cidoc-crm.org/docs/frbr_oo/frbr_docs/FRBRoo_V2.4.pdf

Bennett, R., Lavoie, B. F., & O'Neill, E. T. (2003). The Concept of a Work in WorldCat: An Application of FRBR. *Library Collections, Acquisitions, and Technical Services*, *27*(1), 45–59. Retrieved from http://dx.doi.org/10.1016/S1464-9055(02)00306-8

Berners-Lee, T. (1991). HTML tags. Retrieved July 15, 2020, from http://info.cern.ch/hypertext/WWW/MarkUp/Tags.html

Berners-Lee, T. (1998, October 14). Semantic Web roadmap - Design issues. Retrieved March 6, 2020, from https://www.w3.org/DesignIssues/Semantic.html

Berners-Lee, T. (2008). Linked Open Data. In *Linked Data Planet, New York, NY, USA, 17-18 June 2008*. New York, NY, USA. Retrieved from https://www.w3.org/2008/Talks/0617-lod-tbl/#%281%29

Berners-Lee, T. (2009). Linked Data - Design Issues. Retrieved July 5, 2020, from http://www.w3.org/DesignIssues/LinkedData.html

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*, (May), 35–43.

Biagetti, M. T. (2018). A Comparative analysis and evaluation of bibliographic ontologies. In F. Ribeiro & M. E. Cerveira (Eds.), *Challenges and Opportunities for Knowledge Organization in the Digital Age. Proceedings of the fifteenth International ISKO conference, Porto, July 9-11 2018.* (pp. 499–510). Baden-Baden: Ergon Verlag. https://doi.org/10.5771/9783956504211-499

BIBFRAME - Bibliographic Framework Initiative. (2014). *BIBFRAME Profiles: Introduction and Specification (Draft - 5 May 2014)*. Retrieved from http://www.loc.gov/bibframe/docs/bibframe-profiles.html

BIBFRAME mailing list. (2020). Retrieved March 22, 2020, from https://listserv.loc.gov/cgi-bin/wa?A0=BIBFRAME

BIBFRAME ontology - hasExpression. (2016). Retrieved January 30, 2019, from http://id.loc.gov/ontologies/bibframe.html#p_hasExpression

Bibframe2Schema.org Community Group. (2018, December 4). Retrieved March 1, 2020, from https://www.w3.org/community/bibframe2schema/

Biswas, S., & Rath, D. S. (2014). From Maunsell to Lubetzky: A journey back in search of the root of FRBR among the cataloguing codes of anglo-american origin. *Annals of Library and Information Studies*, *61*(1), 7–14.

Borgman, C. L. (1997). From Acting Locally to Thinking Globally: A Brief History of Library Automation. *The Library Quarterly: Information, Community, Policy*, *67*(3), 215–249. Retrieved from http://www.jstor.com/stable/40039721

Bowen, J. (2010). Moving library metadata toward linked data: Opportunities provided by the extensible catalog. In *Proceedings of the International Conference on Dublin Core and Metadata Applications* (pp. 44–59).

Bray, T., & Sperberg-McQueen, C. M. (1996, November 14). Extensible Markup Language (XML): W3C Working Draft 14-Nov-96. Retrieved July 6, 2020, from https://www.w3.org/TR/WD-xml-961114

Bruce D'Arcus, & Frédérick Giasson. (2016). Bibliographic Ontology Specification. Retrieved July 18, 2019, from http://www.dublincore.org/specifications/bibo/

Bush, J. N. D. (1926). English Translations of Homer. *PMLA*, *41*(2), 335. https://doi.org/10.2307/457438

Cagnazzo, L. F. (2017). *Linked Data: Implementation, Use, and Perceptions across European National Libraries*. University of Strathclyde. Retrieved from https://local.cis.strath.ac.uk/wp/extras/msctheses/papers/strath_cis_publication_2702.pdf

Calhoun, K. (2006). *The changing nature of the catalog and its integration with other discovery tools.* Washington, DC: Library of Congress. Retrieved from http://www.loc.gov/catdir/calhoun-report-final.pdf

Carlyle, A. (1996). Ordering author and work records: An evaluation of collocation in online catalog displays.

Journal of the American Society for Information Science. https://doi.org/10.1002/(SICI)1097-4571(199607)47:7<538::AID-ASI6>3.0.CO;2-V

Carlyle, A., Ranger, S., & Summerlin, J. (2008). Making the pieces fit: Little women, works, and the pursuit of quality. *Cataloging and Classification Quarterly*, *46*(1), 35–63. https://doi.org/10.1080/01639370802182992

Cervantes project: Cervantes collection. (n.d.). Retrieved September 9, 2019, from http://cervantes.tamu.edu

Charles, V., Isaac, A., & Manguinhas, H. (2017). EDM roadmap. Retrieved July 16, 2019, from https://docs.google.com/document/d/1omFFf4KsNZAAnaOUvXOvxRc1nmrKcLHxrrB58KwLY2g/edit#

Charles, V., Isaac, A., Tzouvaras, V., & Hennicke, S. (2013). Mapping cross-domain metadata to the Europeana Data Model (EDM). *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8092 LNCS*, 484–485. https://doi.org/10.1007/978-3-642-40501-3_68

Cole, T. W., Han, M. J., Weathers, W. F., & Joyner, E. (2013). Library Marc Records Into Linked Open Data: Challenges and Opportunities. *Journal of Library Metadata*, *13*(2–3), 163–196. https://doi.org/10.1080/19386389.2013.826074

Coyle, K. (2006). Mass Digitization of Books. *Journal of Academic Librarianship*, *32*(6), 641–645. https://doi.org/10.1016/j.acalib.2006.08.002

Coyle, K. (2010a). FRBR, the Domain Model. *Library Technology Reports*, *46*(2), 20–26.

Coyle, K. (2010b). Metadata models of the world wide web. *Library Technology Reports*, *46*(2), 12–19. https://doi.org/10.5860/ltr

Coyle, K. (2016). *FRBR, before and after : a look at our bibliographic models*. Chicago: ALA editions. Retrieved from https://www.kcoyle.net/beforeAndAfter/978-0-8389-1364-2.pdf

Coyle, K. (2017). Creating the Catalog, Before and After FRBR. In *Encuentro di Catalogacion y Metadatos, Universidad Nacional autonoma de Mexico; September 12, 2017*. Mexico: Universidad Nacional autonoma de Mexico. Retrieved from http://kcoyle.net/mexico.html

Coyle, K., Fallgren, N., Folsom, S., Godby, J., Hearn, S., & Jones, E. (2017). *PCC SCS / LDAC Task Group on the Work Entity: Preliminary White Paper*. Retrieved from https://www.loc.gov/aba/pcc/documents/PoCo-2017/WorkEntitity Preliminary White Paper-2017-09-27.pdf

Creider, L. S. (2006). Cataloging, Reception, and the Boundaries of a "Work." *Cataloging & Classification Quarterly*, *42*(2), 3–19. https://doi.org/http://dx.doi.org/10.1300/J104v42n02_02

Cutter, C. A. (1904). *Rules for a dictionary catalog* (4th ed., r). Washington, D.C.: Government Printing Office. Retrieved from https://archive.org/details/rulesforadictio06cuttgoog/page/n5

CVC. Centro Virtual Cervantes. (n.d.). Retrieved September 9, 2019, from https://cvc.cervantes.es/portada.htm

datos.bne.es 2.0. (2019). Retrieved August 10, 2019, from http://www.bne.es/en/Inicio/Perfiles/Bibliotecarios/DatosEnlazados/datos2-0/

Davis, I., Newman, R., & D'Arcus, B. (2005). Expression of Core FRBR Concepts in RDF. Retrieved July 18, 2019, from http://vocab.org/frbr/core

DCMI: Dublin Core™ Element Set, v 1.0: Reference Description. (1998, September 1). Retrieved June 22, 2020, from https://www.dublincore.org/specifications/dublin-core/dces/1998-09-01/

Decourselle, J., Duchateau, F., & Lumineau, N. (2015). A Survey of FRBRization Techniques. In Sarantos Kapidakis, C. Mazurek, & M. Werla (Eds.), *Research and Advanced Technology for Digital Libraries. TPDL 2015. Lecture Notes in Computer Science* (Vol. 9316, pp. 185–196). Springer, Cham.

https://doi.org/10.1007/978-3-319-24592-8 14

Deliot, C. (2014). Publishing the British National Bibliography as Linked Open Data. *Catalogue & Index*, (174), 13–18. Retrieved from http://www.bl.uk/bibliographic/pdfs/publishing_bnb_as_lod.pdf

Delsey, T., Dullabahn, B., & Heaney, M. (1999). *The Logical Structure of the Anglo-American Cataloguing Rules - Part II*. Retrieved from http://www.rda-jsc.org/archivedsite/docs/aacr2.pdf

Denenberg, R. (2017a). Re: bf:originalVersionOf [Electronic mailing list message]. Retrieved March 22, 2020, from https://listserv.loc.gov/cgi-bin/wa?A2=ind1702&L=BIBFRAME&P=34509

Denenberg, R. (2017b). Re: Relation 1:1 work/instance in BIBFRAME 2.0 ? Retrieved August 12, 2020, from https://listserv.loc.gov/cgi-bin/wa?A2=bibframe;5756bef2.1704

Denton, W. (2007). FRBR and the History of Cataloging. In A. G. Taylor (Ed.), *Understanding FRBR: What It Is and How It Will Affect Our Retrieval Tools* (pp. 35–58). Westport, Connecticut: Libraries Unlimited. Retrieved from http://hdl.handle.net/10315/1250

Deutche National Bibliothek. (2016). *The Linked Data Service of the German National Library: Modelling of bibliographic data*. Retrieved from http://www.dnb.de/SharedDocs/Downloads/EN/DNB/service/linkedDataModellierungTiteldaten.pdf?__blob=publicationFile

Dickey, T. J. (2008). FRBRization of a Library Catalog: Better Collocation of Records, Leading to Enhanced Search, Retrieval, and Display. *Information Technology and Libraries*, *27*(1), 23. https://doi.org/10.6017/ital.v27i1.3260

Digital Public Library of America. (n.d.). Retrieved July 18, 2019, from https://dp.la/

Dodds, L., & Davis, I. (2012). *Linked data patterns: A pattern catalogue for modelling, publishing, and consuming Linked Data*. Retrieved from https://patterns.dataincubator.org/book/linked-data-patterns.pdf

Doerr, M., Gradmann, S., Le Boeuf, P., Aalberg, T., Bailly, R., & Olensky, M. (2013). *Final Report on EDM – FRBRoo Application Profile Task Force*. *Europeana Network*. Retrieved from http://pro.europeana.eu/files/Europeana_Professional/EuropeanaTech/EuropeanaTech_taskforces/EDM_FRBRoo/TaskfoApplication Profile EDM-FRBRoo.pdf

Dublin Core Metadata Initiative. (2012). *DCMI metadata terms*. Retrieved from http://dublincore.org/documents/dcmi-terms

Duchateau, F., Lumineau, N., & Aalberg, T. (2018). Impact of open and linked data on bibliographic catalogs. *Ingenierie Des Systemes d'Information*, *23*(3–4), 57–93. https://doi.org/10.3166/ISI.23.3-4.57-93

Dunsire, G. (2007). RDA and library systems. *BiD: Textos Universitaris de Biblioteconomia i Documentació*, *19*.

Dunsire, G. (2012). Representing the FR family in the Semantic Web. *Cataloging and Classification Quarterly*, *50*(5–7), 724–741. https://doi.org/10.1080/01639374.2012.679881

Dunsire, G. (2015). FRBRer model. IFLA. Retrieved from http://iflastandards.info/ns/fr/frbr/frbrer/

Dunsire, G. (2019). The IFLA Library Reference Model and RDA. In *prepared for Segundo Coloquio sobre RDA en América Latina y el Caribe, Biblioteca Nacional de Chile, Santiago, 21 October 2019 (postponed)*. Santiago de Chile, Chile. Retrieved from http://www.rda-rsc.org/sites/all/files/Dunsire IFLA LRM and RDA.pdf

Dunsire, G., Hillmann, D., & Phipps, J. (2012). Reconsidering Universal Bibliographic Control in Light of the Semantic Web. *Journal of Library Metadata*, *12*(2–3), 164–176. https://doi.org/10.1080/19386389.2012.699831

English translations of Homer. (2019). In *Wikipedia*. Retrieved from https://en.wikipedia.org/wiki/English_translations_of_Homer

Europeana. (2017). *Definition of the Europeana Data Model v5.2.8*. Den Haag. Retrieved from https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation//EDM_Definition_v5.2.8_102017.pdf

Europeana Libraries | Europeana Pro. (n.d.). Retrieved July 16, 2019, from https://pro.europeana.eu/project/europeana-libraries

Europeana Pro: Linked Open Data. (2019). Retrieved August 10, 2019, from https://pro.europeana.eu/page/linked-open-data

Falcone, A., Greben, J., & Lorimer, N. (n.d.). Linked Data for Production (LD4P). https://doi.org/10.1145/3184558.3186201

Fattahi, R. (1996). Super records: An approach towards the description of works appearing in various manifestations. *Library Review*, *45*(4), 19–29. https://doi.org/10.1108/EUM0000000004129

Fattahi, R. (1997). AACR2 and Catalogue Production Technology. In J. Weihs (Ed.), *International Conference on the Principles and Future Development of AACR Toronto, Canada, October 23-25, 1997* (pp. 17–43). Ottawa; London; Chicago: Canadian Library Association; Library Association Publishing; American Library Association. Retrieved from http://epe.lac-bac.gc.ca/100/200/300/jsc_aacr/aacr_cat/r-aacr2.pdf

Fay, H. C. (1951). Chapman' s Materials for His Translation of Homer. *The Review of English Studies*, *2*(6), 121–128. Retrieved from http://www.jstor.org/stable/511025

Ford, K. (2010). ID.LOC.GOV, 1 ½ Years: Review, Changes, Future Plans, MADS/RDF. In *2010 Digital Library Federation (DLF) Fall Forum, 1-3 November, 2010 - Palo Alto, California, USA.* Retrieved from https://id.loc.gov/static/presentations/kefo_dlf_id.pdf

Ford, K. (2019a). Concerning Identities : For Things, but not the easy Things [Contains Hubs Part I]. In *3rd European BIBFRAME Workshop*. Stockholm: National Library of Sweden. Retrieved from https://www.kb.se/download/18.d0e4d5b16cd18f600eaea/1569322333140/2019-kefo-identity.pdf

Ford, K. (2019b). Concerning Relationships : Hubs Part II. In *3rd European BIBFRAME Workshop* (pp. 17–18). Stockholm: National Library of Sweden. Retrieved from https://www.kb.se/download/18.d0e4d5b16cd18f600eafb/1569324367132/2019-kefo-relationships.pdf

Ford, P. (2006). Homer In The French Renaissance. *Renaissance Quarterly*, *59*(1), 1–28. Retrieved from http://www.jstor.org/stable/10.1353/ren.2008.0159

Foster, F. M. K. (1918). *English Translations from the Greek: A Bibliographical Survey*. New York: Columbia University Press.

Freire, N., Borbinha, J., & Calado, P. (2007). Identification of FRBR works within bibliographic databases: An experiment with UNIMARC and duplicate detection techniques. In D.-L. Goh, T. Cao, I. Sølvberg, & E. Rasmussen (Eds.), *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 4822 LNCS, pp. 267–276). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-77094-7_36

Frias, J. A., & Rios Hilario, A. B. (2002). Visibility And Invisibility Relationships In Bibliographic Library Catalogue. In M. J. López-Huertas & F. J. Munoz-Férnandez (Eds.), *Challenges in Knowledge Representation and Organization for the 21st Century. Integration of Knowledge across Boundaries. Proceedings of the Seventh International ISKO Conference, 10-13 July 2002 Granada, Spain* (pp. 264–270). Granada, Spain: ERGON VERLAG. Retrieved from https://www.ergon-verlag.de/isko_ko/downloads/aikovol08200239.pdf

Fritz, D. (2016a). *RDA and WGA treatment of aggregates Abstract*. *RSC Aggregates Working Group*.

Fritz, D. (2016b). *RSC/AggregatesWG/1: Summary of responses*. *RSC Aggregates Working Group*. Retrieved from http://www.rda-rsc.org/sites/all/files/RSC-AggregatesWG-1-AggregatesWG-response-rev.pdf

Frosterus, M., Dadvar, M., Hansson, D., Lappalainen, M., & Zapounidou, S. (2020). *Linked Open Data: impressions & challenges among Europe's research libraries*. Retrieved from https://doi.org/10.5281/zenodo.3647843

Futornick, M., & Younes, R. (2017). bibliotek-o. Retrieved July 18, 2019, from https://wiki.duraspace.org/display/LD4P/bibliotek-o

Galeffi, A., Bertolini, M. V., Bothmann, R. L., Rodríguez, E. E., & McGarry, D. (2017). *Statement of International Cataloguing Principles (ICP)*. *IFLA Cataloguing Section, IFLA Meetings of Experts on an International Cataloguing Code,*. Den Haag. Retrieved from http://www.ifla.org/publications/statement-of-international-cataloguing-principles

Gilliland, A. J. (2016). Setting the stage. In Martha Baca (Ed.), *Introduction to Metadata* (3rd ed.). Los Angeles, CA: Getty Publications. Retrieved from http://www.getty.edu/publications/intrometadata/setting-the-stage/

Glennan, K., & James, K. (2018). Representative Expressions. In *RDA Toolkit Redesign Update and Preview, Preconference to ALA Midwinter meeting in Denver, Colorado*. Denver, CO. Retrieved from http://www.rda-rsc.org/sites/all/files/Representative expressions.pdf

Godby, C. J. (2013). *The Relationship between BIBFRAME and OCLC's Linked-Data Model of Bibliographic Description: A Working Paper*. Dublin, Ohio: OCLC Research. Retrieved from https://www.oclc.org/content/dam/research/publications/library/2013/2013-05.pdf

Godby, C. J., & Denenberg, R. (2015). *Common Ground: Exploring Compatibilities Between the Linked Data Models of the Library of Congress and OCLC*. Retrieved from http://www.oclc.org/content/dam/research/publications/2015/ oclcresearch-loc-linked-data-2015.pdf

Godby, C. J., & Vizine-Goetz, D. (2017). BIBFRAME and OCLC Works: Defining models and discovering evidence. In *ALA Annual 2017: Library of Congress BIBFRAME Update Session. 26 June 2017*. Chicago.

Godby, C. J., Wang, S., & Mixter, J. K. (2015). *Library Linked Data in the Cloud: OCLC's Experiments with New Models of Resource Description*. *Synthesis Lectures on the Semantic Web: Theory and Technology 9*. Morgan & Claypool Publishers. https://doi.org/10.2200/S00620ED1V01Y201412WBE012

Goossens, P., & Mazur-Rzesos, E. (1982). Hierarchical relationships in bibliographic descriptions: Problem analysis. In A. H. Helal & J. W. Weiss (Eds.), *Hierarchical Relationships in Bibliographic Descriptions. INTERMARC Software-Subgroup Seminar 4, Library Systems Seminar (March 25-March 27, 1981)* (pp. 13–128). Essen: Essen University Library. https://doi.org/10.1086/601371

Gorman, M., & Oddy, P. (1998). The Anglo-American Cataloguing Rules Second Edition: their history and principles. In J. Weihs (Ed.), *The principles and future of AACR : proceedings of the International Conference on the Principles and Future Development of AACR Toronto, Canada, October 23-25, 1997*. Ottawa: Canadian Library Association ; London : Library Association Publishing ; Chicago : American Library Association. Retrieved from https://epe.lac-bac.gc.ca/100/200/300/jsc_aacr/aacr_sec/r-aacr2e.pdf

Gorman, M., Winkler, P. W., & American Library Association. (1978). *Anglo-American cataloguing rules*. Chicago: ALA.

Gueguen, G., Glendon, I., Charles, V., Dean, R., Bahneman, G., Hardesty, J., … Stitzlein, H. (2017). Introduction to the DPLA Metadata Application Profile, version 5.0, 56. Retrieved from http://dp.la/info/map

Hahn, T. B. (2006). Impacts of mass digitization projects on libraries and information policy. *Bulletin of the American Society for Information Science and Technology*, *33*(1), 6.

Hallo, M., Luján-Mora, S., Mate, A., & Trujillo, J. (2016). Current state of Linked Data in digital libraries. *Journal of Information Science*, *42*(2), 117–127. https://doi.org/10.1177/0165551515594729

Harvitt, H. J. . (1919). Hugues Salel , Poet and Translator. *Modern Philology*, *16*(11), 595–605. Retrieved from http://www.jstor.org/stable/432908

Haslhofer, B., & Isaac, A. (2011). Data.europeana.eu the europeana linked open data pilot. In *Proceedings of the International Conference on Dublin Core and Metadata Applications*. Retrieved from https://dcpapers.dublincore.org/pubs/article/view/3625

Haslhofer, Bernhard, & Klas, W. (2010). A Survey of Techniques for Achieving Metadata Interoperability. *ACM Comput. Surv.*, *42*(2), 7:1--7:37. https://doi.org/10.1145/1667062.1667064

Hickey, T. B., & O'Neill, E. T. (2009). FRBRizing OCLC ' s WorldCat. *Cataloging & Classification Quarterly*, *39*(3–4), 239–251. https://doi.org/10.1300/J104v39n03_15

Hillmann, D., Coyle, K., Phipps, J., & Dunsire, G. (2010). RDA Vocabularies : Process , Outcome, Use. *D-Lib Magazine*, *16*(1/2). Retrieved from http://dlib.org/dlib/january10/hillmann/01hillmann.html

Hogan, A., Umbrich, J., Harth, A., Cyganiak, R., Polleres, A., & Decker, S. (2012). An empirical survey of Linked Data conformance. *Journal of Web Semantics*, *14*, 14–44. https://doi.org/10.1016/j.websem.2012.02.001

Home - schema.org. (2019). Retrieved July 18, 2019, from https://schema.org/

Hopkinson, A. (1984). International access to bibliographic data: MARC and MARC-related activities. *Journal of Documentation*, *40*(1), 13–24. https://doi.org/10.1108/eb026754

ICOM-CIDOC. (2019). CIDOC CRM Family models and collaborations: FRBRoo. Retrieved July 13, 2019, from http://www.cidoc-crm.org/frbroo/

IFLA -- Statement of International Cataloguing Principles (ICP) 2016. (2019). Retrieved September 5, 2020, from https://www.ifla.org/publications/node/11015

IFLA FRBR Working group on Aggregates. (2009). *Differences between the works-of-works (Tillett) and the manifestation-of-expressions (O'Neill-Žumer) definitions of aggregates* (FRBR Working group on Aggregates). Milan. Retrieved from http://www.ifla.org/files/assets/cataloguing/frbrrg/aggregates-definition-differences.pdf

IFLA PRESSoo Review Group, & Bœuf, P. Le. (2017). *Definition of PRESSoo: A conceptual model for Bibliographic Information Pertaining to Serials and other Continuing Resources (version 1.3)*. Den Haag. Retrieved from https://www.ifla.org/files/assets/cataloguing/PRESSoo/pressoo_v1-3.pdf

IFLA Study Group on the Functional Requirements for Bibliographic Records. (1998). *Functional Requirements for Bibliographic Records Final Report* (Vol. 19). München: IFLA. Retrieved from http://www.ifla.org/files/assets/cataloguing/frbr/frbr.pdf

IFLA Study Group on the Functional Requirements for Bibliographic Records. (2009). *Functional Requirements for Bibliographic Records Final Report*. The Hague: IFLA. Retrieved from https://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf

IFLA Working Group on Content Designators. (1977). *UNIMARC: Universal MARC format*. London: IFLA International Office for UBC. Retrieved from file://catalog.hathitrust.org/Record/000130834

IFLA Working Group on Content Designators. (1980). *UNIMARC: universal MARC format* (2d. ed. re). London: IFLA International Office for UBC. Retrieved from file://catalog.hathitrust.org/Record/000130827

IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR). (2009). *Functional Requirements for Authority Data: A conceptual model*. (G. E. Patton, Ed.). München: K.G.Saur.

IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR). (2013).

Functional Requirements for Authority Data: A Conceptual model. The Hague. Retrieved from http://www.ifla.org/files/assets/cataloguing/frad/frad_2013.pdf

IFLA Working Group on the Expression Entity. (2007). *Functional Requirements for Bibliographic Records Chapter 3: Entities Changes approved to the FRBR text Clean Version*. Retrieved from http://www.ifla.org/files/assets/cataloguing/frbr/amend-1998-1-clean.pdf

IFLA Working Group on the Functional Requirements for Subject Authority Records. (2010). *Functional Requirements for Subject Authority Data (FRSAD): A Conceptual Model*. The Hague. Retrieved from http://www.ifla.org/files/assets/classification-and-indexing/functional-requirements-for-subject-authority-data/frsad-final-report.pdf

International Federation of Library Associations. (1974). *ISBD (M): International standard bibliographic description for monographic publications*. London: IFLA Committee on Cataloguing.

International Federation of Library Associations and Institutions. ISBD Review Group., & International Federation of Library Associations and Institutions. Cataloguing Section. Standing Committee. (2011). *ISBD : International Standard Bibliographic Description* (Consolidat). Berlin; München: De Gruyter Saur. Retrieved from https://www.ifla.org/files/assets/cataloguing/isbd/isbd-cons_20110321.pdf

International Federation of Library Associations and Institutions. (2009). *Statement of International Cataloguing Principles*. *Library* (Vol. 90). Den Haag. https://doi.org/10.1063/1.2958287

Isaac, A. (2013). *Europeana Data Model Primer*. The Hague. Retrieved from https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf

Isaac, A., Waites, W., Young, J., & Zeng, M. (2011). *Library Linked Data Incubator Group : Datasets , Value Vocabularies , and Metadata Element Sets*. Retrieved from https://www.w3.org/2005/Incubator/lld/XGR-lld-vocabdataset-20111025/

ISBD Review Group, & Galeffi, A. (2015). Superseded ISBDs. Retrieved July 7, 2019, from https://www.ifla.org/isbd-rg/superseded-isbd-s

Jett, J., Cole, T. W., Page, K. R., & Downie, J. S. (2016). Enhancing Scholarly Use of Digital Libraries : A Comparative Survey and Review of Bibliographic Metadata Ontologies. *ACM/IEEE-CS on Joint Conference on Digital Libraries*, 35–44.

Johnston, P. (2005). XML, RDF, and DCAPs. Retrieved April 3, 2020, from http://www.ukoln.ac.uk/metadata/dcmi/dc-elem-prop/

Joint Steering Committee for Development of RDA. (2009). RDA — Resource Description and Access : Prospectus. Retrieved January 10, 2019, from http://www.rda-jsc.org/archivedsite/rdaprospectus.html

Joint Steering Committee for Revision of AACR. (2006). *RDA / ONIX Framework for Resource Categorization* (Vol. 5JSC serie). Retrieved from http://www.rda-jsc.org/archivedsite/docs/5chair10.pdf

Jones, R., MacGillivray, M., Murray-Rust, P., Pitman, J., Sefton, P., O'steen, B., & Waites, W. (2011). Open Bibliography for science, technology, and medicine. *Journal of Cheminformatics*, *3*(10), 1–10. https://doi.org/10.1186/1758-2946-3-47

Kay, M. (2017). *XSL Transformations (XSLT) Version 3.0: W3C Recommendation 8 June 2017*. Retrieved from https://www.w3.org/TR/2017/REC-xslt-30-20170608/

Knapp, J. F. (1968). Design considerations for the MARC magnetic tape formats. *Library Resources and Technial Services*, *12*(3), 275–85. Retrieved from http://downloads.alcts.ala.org/lrts/lrtsv12no3.pdf

Kondylakis, H., Doerr, M., & Plexousakis, D. (2006). *Mapping Language for Information Integration. Technical report 385*. Heraklion, Crete, Greece: Institute of Computer Science, FORTH-ICS. Retrieved from http://www.ics.forth.gr/tech-reports/2006/2006.TR385_Mapping_Language_Information_Integration.pdf

Koster, L. (2012). Old silos , new silos , no silos: From redundancy to aggregation or distribution? In *SWIB 2012*. Köln, Germany. Retrieved from https://www.slideshare.net/lukask/old-silos-new-silos-no-silos

Kovari, J., Folsom, S., & Younes, R. (2017). Towards a BIBFRAME Implementation: The bibliotek-o Framework. In *International Conference on Dublin Core and Metadata Applications* (pp. 52–61). Washington, D.C. Retrieved from http://dcpapers.dublincore.org/pubs/article/view/3854/2039

Kruk, S. R. (2004). *MarcOnt initiative: bibliographic description and related tools utilising Semantic Web technologies*. Galway. Retrieved from https://web.archive.org/web/20071117023028/http:/www.marcont.org/marcont/pdf/sk_marcont_project.pdf

Kruk, S. R., Dabrowski, M., & Synak, M. (2009). Bibliographic Ontology. *Semantic Digital Libraries*, 103–122. Retrieved from http://dx.doi.org/10.1007/978-3-540-85434-0_8

Kruk, S. R., & Zimmermann, K. (2005). MarcOnt - Integration Ontology for Bibliographic Description Formats. In *International Conference on Dublin Core and Metadata Applications 2005* (pp. 231–234). Madrid. Retrieved from http://dcpapers.dublincore.org/pubs/article/download/829/825

Lapôtre, R. (2017). Library metadata on the web: The example of data.bnf.fr. *JLIS.It*. https://doi.org/10.4403/jlis.it-12402

Lassila, O., & Swick, R. R. (1999). *Resource Description Framework (RDF) Model and Syntax Specification*. Retrieved from https://www.w3.org/TR/PR-rdf-syntax/

LC Linked Data Service: Authorities and Vocabularies - Carriers Scheme. (2014). Retrieved November 16, 2019, from http://id.loc.gov/vocabulary/carriers.html

LC Linked Data Service: Authorities and Vocabularies - Content Types Scheme. (2014). Retrieved May 26, 2019, from http://id.loc.gov/vocabulary/contentTypes.html

LC Linked Data Service: BIBFRAME Ontology LC Extension. (2017). Retrieved January 22, 2020, from http://id.loc.gov/ontologies/bflc.html

LC Linked Data Service: MARC Code List for Relators Scheme. (n.d.). Retrieved August 3, 2019, from http://id.loc.gov/vocabulary/relators.html

LC Linked Data Service: MARC List for Languages. (2011). Retrieved August 3, 2019, from http://id.loc.gov/vocabulary/languages.html

Le Boeuf, P. (2013). Transforming FRBR into FRBRoo. *Ciclo "Biblioteche, Libri, Documenti: Dall'informazione Alla Conoscenza" Organizzato Dalla Prof.Ssa Maria Teresa Biagetti*. Rome: Rome, Università La Sapienza. Retrieved from http://old.cidoc-crm.org/docs/Transforming FRBR into FRBRoo.pdf

Leazer, G.-H. (1993). *A conceptual plan for the description and control of bibliographic works*. Columbia University.

Leazer, G. H., & Smiraglia, R. P. (1999). Bibliographic families in the library catalog: A qualitative analysis and grounded theory. *Library Resources and Technical Services*, *43*(4), 191–207.

Library of Congress' Network Development and MARC Standards Office. (2020, June 8). MARC 21 XML Schema. Retrieved June 26, 2020, from https://www.loc.gov/standards/marcxml/

Library of Congress. (n.d.-a). Chronicling America. Retrieved July 18, 2019, from https://chroniclingamerica.loc.gov/

Library of Congress. (n.d.-b). MARC standards. Retrieved from http://www.loc.gov/marc/

Library of Congress. (2011a). *A Bibliographic Framework for the Digital Age*. Retrieved from http://www.loc.gov/bibframe/news/framework-103111.html

Library of Congress. (2011b). Transforming our Bibliographic Framework: A Statement from the Library of Congress (May 13, 2011). Retrieved July 8, 2019, from http://www.loc.gov/bibframe/news/framework-051311.html

Library of Congress. (2016a). BIBFRAME 2.0 Vocabulary List View - originalVersion property. Retrieved March 21, 2020, from http://id.loc.gov/ontologies/bibframe.html#p_originalVersion

Library of Congress. (2016b). Overview of the BIBFRAME 2.0 Model. Retrieved April 14, 2018, from https://www.loc.gov/bibframe/docs/bibframe2-model.html

Library of Congress. (2017). *BIBFRAME 2.0 RDF Conventions*. Washington, D.C.

Library of Congress. (2018). Library of Congress Name Authority File (NAF). Retrieved June 9, 2018, from http://id.loc.gov/authorities/names.html

Library of Congress. (2019). MARC 21 to BIBFRAME 2.0 Conversion Specifications (BIBFRAME - Bibliographic Framework Initiative, Library of Congress). Retrieved March 22, 2019, from https://www.loc.gov/bibframe/mtbf/

Library of Congress. (2020, July 15). EAD: Encoded Archival Description. Retrieved August 11, 2020, from https://www.loc.gov/ead/

Library of Congress Network Development and MARC Standards Office. (2002). *Displays for Multiple Versions from MARC 21 and FRBR*. *Functional analysis of the MARC 21 bibliographic and holdings formats*. Washington, DC. Retrieved from http://www.loc.gov/marc/marc-functional-analysis/multiple-versions.html

Library of Congress Network Development and MARC Standards Office. (2019, January 17). Metadata Encoding and Transmission Standard (METS) Official Web Site | Library of Congress. Retrieved June 26, 2020, from http://www.loc.gov/standards/mets/

Library of Congress Standards. (2020). Metadata Object Description Schema (MODS): Official Web Site. Retrieved February 25, 2020, from http://www.loc.gov/standards/mods/

Library of Congress Working Group on the Future of Bibliographic Control. (2008). *On the record*. Washington, D.C. Retrieved from https://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf

Lopatin, L. (2006). Library digitization projects, issues and guidelines: A survey of the literature. *Library Hi Tech*, *24*(2), 273–289. https://doi.org/10.1108/07378830610669637

Lovins, D., & Hillmann, D. (2017). Broken-world vocabularies. *D-Lib Magazine*, *23*(3–4), 1–8. https://doi.org/10.1045/march2017-lovins

Lubetzky, S. (1953). *Cataloging Rules and Principles: A Critique of the A.L.A. Rules for Entry and a Proposed Design for Their Revision*. Washington: Library of Congress.

Lubetzky, S. (1969). *Principles of Cataloging. Final Report. Phase I: Descriptive Cataloging.* Los Angeles, CA. Retrieved from https://files.eric.ed.gov/fulltext/ED031273.pdf

Lubetzky, S. (1986). Principles of descriptive cataloging. In M. Carpenter & E. Svenonius (Eds.), *Foundations of cataloging: a sourcebook* (pp. 104–112). Michigan: Libraries Unlimited.

Malmsten, M. (2008). Making a library catalogue part of the semantic web. In *Proceedings of the International Conference on Dublin Core and Metadata Applications*. Retrieved from https://dcpapers.dublincore.org/pubs/article/view/927

Manguinhas, H. M. Á., Freire, N. M. A., & Borbinha, J. L. B. (2010). FRBRization of MARC records in multiple catalogs. *Proceedings of the 10th Annual Joint Conference on Digital Libraries - JCDL '10*, 225. https://doi.org/10.1145/1816123.1816157

Marketakis, Y., Minadakis, N., Kondylakis, H., Konsolaki, K., Samaritakis, G., Theodoridou, M., … Doerr, M.

(2016). X3ML mapping framework for information integration in cultural heritage and beyond. *International Journal on Digital Libraries*, (June). https://doi.org/10.1007/s00799-016-0179-1

McCallum, S. (2017). BIBFRAME Development. *JLIS.It, Italian Journal of Library, Archives & Information Science*, *8*(3), 71–85. https://doi.org/10.4403/jlis.it-12415

McCallum, S. (2018). BIBFRAME Introduction. In *European Bibframe Workshop*. Florence. Retrieved from http://www.casalini.it/EBW2018/web_content/2018/presentations/McCallum_1.pdf

McCallum, S. (2019). BIBFRAME expansion and access. In *3rd European BIBFRAME Workshop*. Stockholm: National Library of Sweden. Retrieved from https://www.kb.se/download/18.d0e4d5b16cd18f600eabd/1569242059929/stockholm2019-EWS-BFdevelopment-mccallum.pdf

Mercun, T., Zumer, M., & Aalberg, T. (2017). Presenting Bibliographic Families Using Information Visualization: Evaluation of FRBR-Based Prototype and Hierarchical Visualizations. *Journal of the Association for Information Science and Technology*, *68*(2), 392–411. https://doi.org/10.1002/jasist.23659

Michelle Futornick. (n.d.). LD4P Outputs - LD4P public website - DuraSpace Wiki. Retrieved September 12, 2019, from https://wiki.duraspace.org/display/LD4P/LD4P+Outputs

Miller, E., Ogbuji, U., Mueller, V., & MacDougall, K. (2012). *Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services*. *Library of Congress*. Washington, DC: Library of Congress. Retrieved from http://www.loc.gov/marc/transition/pdf/marcld-report-11-21-2012.pdf

Musen, M. A., & Team, P. (2015). The Protégé Project: A Look Back and a Look Forward, *1*(4), 4–12. https://doi.org/10.1145/2757001.2757003

National Library of Medicine. (2015). Experimentation with BIBFRAME at the National Library of Medicine. Retrieved July 7, 2015, from https://github.com/fallgrennj/BIBFRAME-NLM

National Library of Sweden. (2019). Data Model. Retrieved August 10, 2019, from https://id.kb.se/doc/model

Neill, E. O., Žumer, M., & Mixter, J. (2015). FRBR Aggregates: Their Types and Frequency in Library Collections. *Library Resources & Technical Services*, *59*(3), 120–129. https://doi.org/http://dx.doi.org/10.5860/lrts.59n3.120

Nikoletseas, M. M. (2012). *The Iliad - twenty centuries of translation*. Charleston, SC: Michael Nikoletseas.

Nillson, M. (2010). *From interoperability to harmonization in metadata standardization: designing an evolvable framework for metadata harmonization*. KTH, Stockholm. Retrieved from http://kmr.nada.kth.se/papers/SemanticWeb/FromInteropToHarm-MikaelsThesis.pdf

Noruzi, A. (2012). FRBR and Tillett's Taxonomy of Bibliographic Relationships. *Knowledge Organization*, *39*(6), 409–416.

O'Neill, E., Cato, A., Goossens, P., Kuhagen, J., Tillett, B., Van Nuys, C., & Žumer, M. (2011). *Final Report of the Working Group on Aggregates*. Oslo. Retrieved from http://www.ifla.org/files/assets/cataloguing/frbrrg/AggregatesFinalReport.pdf

O'Neill, E T. (2002). FRBR: Functional Requirements for Bibliographic Records. Application of the entity-relationship model to Humphry Clinker. *Library Resources and Technical Services*, *46*(4), 150–159. https://doi.org/http://dx.doi.org/10.5860/lrts.46n4.150

O'Neill, Edward T., & Žumer, M. (2014). Round Table on the Role of Controlled Vocabularies in the Semantic Web: Discussion Notes. *Cataloging and Classification Quarterly*, *52*(1), 123–128. https://doi.org/10.1080/01639374.2013.860339

OCLC. (2015). *OCLC prints last library catalog cards*. *OCLC News Releases 2015 - US*. Dublin, OH. Retrieved from https://cdm15003.contentdm.oclc.org/digital/collection/p15003coll6/id/386

OCLC. (2018). VIAF: The Virtual International Authority File. Retrieved June 9, 2018, from https://viaf.org/

Open Metadata Registry. (2010). Retrieved August 30, 2020, from http://metadataregistry.org/

Pablo Alvarez. (2012). Translating Homer. Retrieved September 9, 2019, from https://www.lib.umich.edu/online-exhibits/exhibits/show/translating-homer--from-papyri

Panizzi, A. (1841). Rules for the Compilation of the Catalogue. In *The Catalogue of Printed Books in the British Museum, vol. 1* (p. [v]–ix). London: British Museum.

Panizzi, A. (1985). Mr. Panizzi to the Rigth Hon, the Earl of Ellesmere. - British Museum, January 29, 1848. In M. Carpenter & E. Svenonius (Eds.), *Foundations of Cataloging: A Sourcebook* (pp. 15–47). Littleton, Colorado: Libraries Unlimited, Inc.

Park, H., & Kipp, M. (2019). Library Linked Data Models: Library Data in the Semantic Web. *Cataloging & Classification Quarterly*, *0*(0), 1–17. https://doi.org/10.1080/01639374.2019.1641171

Park, H., & Kipp, M. E. I. (2015). Evaluation of Mappings from MARC to Linked Data. *Advances in Classification Research Online*, *25*(1). https://doi.org/10.7152/acro.v25i1.14908

Park, J.-R., Andrew Brenza, & Richards, L. (2020). BIBFRAME Linked Data: A Conceptual Study on the Prevailing Content Standards and Data Model. *Intech*. https://doi.org/10.5772/intechopen.91849

Patrício, H. S., Cordeiro, M. I., & Ramos, P. N. (2020). From the web of bibliographic data to the web of bibliographic meaning : structuring , interlinking and validating ontologies on the semantic web. *International Journal of Metadata, Semantics and Ontologies*, *14*(2), 124–134. Retrieved from http://dx.doi.org/10.1504/IJMSO.2020.108318

Peponakis, M. (2012). Conceptualizations of the Cataloging Object: A Critique on Current Perceptions of FRBR Group 1 Entities. *Cataloging & Classification Quarterly*, *50*(5–7), 587–602. https://doi.org/10.1080/01639374.2012.681275

Peponakis, M. (2016). In the name of the name: RDF literals, ER attributes, and the potential to rethink the structures and visualizations of catalogs. *Information Technology and Libraries*, *35*(2), 19–38. https://doi.org/10.6017/ital.v35i2.8749

Peponakis, M., Sfakakis, M., & Kapidakis, S. (2011). FRBRization: using UNIMARC link fields to identify Works. *World Library and Information Congress : 77th IFLA General Conference and Assembly*, 1–12.

Peroni, S., & Shotton, D. (2012). FaBiO and CiTO: ontologies for describing bibliographic resources and citations. *Journal of Web Semantics*, *17*, 33–43. https://doi.org/10.1016/j.websem.2012.08.001

Petek, M. (2007). Derivative bibliographic relationships in the Slovenian online catalogue COBIB. *Journal of Documentation*, *63*(3), 398–423. https://doi.org/10.1108/00220410710743315

Pettee, J. (1936). The Development of Authorship Entry and the Formulation of Authorship Rules as Found in the Anglo-American Code. *The Library Quarterly*, *6*(1). https://doi.org/10.1086/613854

Picco, P., & Ortiz Repiso, V. (2012). The contribution of FRBR to the identification of bibliographic relationships: The new RDA-based ways of representing relationships in catalogs. *Cataloging and Classification Quarterly*, *50*(5–7), 622–640. https://doi.org/10.1080/01639374.2012.680847

Powell, C. K. (2008). OPAC integration in the era of mass digitization: The MBooks experience. *Library Hi Tech*, *26*(1), 24–32. https://doi.org/10.1108/07378830810857771

Rafferty, P. (2015). FRBR , Information , and Intertextuality. *Library Trends*, *63*(3), 487–511.

Railton, S. (2012). Mark Twain in His Times. Retrieved September 9, 2019, from http://twain.lib.virginia.edu/index2.html

Ranganathan, S. R. (1955). *Heading and canons: Comparative study of five catalogue codes*. Madras: S. Viswanathan.

Rasmussen Pennington, D., & Cagnazzo, L. (2019). Connecting the silos: Implementations and perceptions of linked data across European libraries. *Journal of Documentation*, *75*(3), 643–666. https://doi.org/10.1108/JD-07-2018-0117

RDA Registry | Data using the RDA vocabularies. (2017). Retrieved January 27, 2019, from https://www.rdaregistry.info/rgData/rdaCuries.html

RDA Toolkit. (2016, October 12). 3R Project: Kickoff Announcement. Retrieved March 1, 2020, from https://www.rdatoolkit.org/3Rproject/announcement

RDA Toolkit. (2019, October 15). Completion of 3R Project. Retrieved March 1, 2020, from https://www.rdatoolkit.org/node/202

Reed-Scott, J. (1999). *Cataloging of Resources Digitized for Preservation. SPEC Kit 249 [and] SPEC Flyer 249*. Washington, DC. Retrieved from https://files.eric.ed.gov/fulltext/ED437954.pdf

Reese, T. (2013). MarcEdit Development. Retrieved July 11, 2020, from http://marcedit.reeset.net/

Register, R., Cohn, K., Hawkins, L., Henderson, H., Reynolds, R., Shadle, S. C., … Yue, P. W. (2009). Metadata in a digital age: New models of creation, discovery, and use. *Serials Librarian*, *56*(1–4), 7–24. https://doi.org/10.1080/03615260802672445

Remnick, D. (2005). The Translation Wars. *The New Yorker*. Retrieved from https://www.newyorker.com/magazine/2005/11/07/the-translation-wars

Riva, P., Bœuf, P. Le, & Žumer, M. (2017a). *IFLA Library Reference Model: A Conceptual Model for Bibliographic Information*. Den Haag. Retrieved from https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017_rev201712.pdf

Riva, P., Bœuf, P. Le, & Žumer, M. (2017b). *Transition Mappings : User Tasks, Entities, Attributes, and Relationships in FRBR, FRAD, and FRSAD mapped to their equivalents in the FRBR-Library Reference Model*. Den Haag. Retrieved from https://www.ifla.org/files/assets/cataloguing/frbr-lrm/transitionmappings201708.pdf

Riva, P., Le Boeuf, P., & Žumer, M. (2017). *IFLA Library Reference Model: A Conceptual Model for Bibliographic Information*. Den Haag: IFLA. Retrieved from https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017.pdf

Riva, P., & Oliver, C. (2012). Evaluation of RDA as an implementation of FRBR and FRAD. *Cataloging and Classification Quarterly*, *50*(5–7), 564–586. https://doi.org/10.1080/01639374.2012.680848

Riva, P., & Žumer, M. (2018). FRBRoo , the IFLA Library Reference Model , and now LRMoo : a circle of development. In *IFLA WLIC 2018 Kuala Lumpur*. IFLA. Retrieved from http://library.ifla.org/2130/

Salaba, A., Merčun, T., & Aalberg, T. (2018). Complexity of Work Families and Entity-Based Visualization Displays. *Cataloging & Classification Quarterly*, *0*(0), 1–26. https://doi.org/10.1080/01639374.2018.1529008

Schreiber, G., Raimond, Y., Manola, F., Miller, E., & McBride, B. (2014, June 14). RDF 1.1 Primer: W3C Working Group Note 24 June 2014. Retrieved March 6, 2020, from https://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624/

Schreur, P. (2018). The Evolution of BIBFRAME: from MARC Surrogate to Web Conformant Data Model. In *IFLA WLIC 2018* (pp. 1–10). Kuala Lumpur. Retrieved from http://library.ifla.org/2202/1/141-schreur-en.pdf

Sean Bechhofer, Frank van Harmelen, Jim Hendler, Ian Horrocks, Deborah L. McGuinness, Peter F. Patel-Schneider, & Lynn Andrea Stein. (2004). OWL Web Ontology Language Reference: W3C Recommendation 10 February 2004. Retrieved July 25, 2019, from https://www.w3.org/TR/owl-ref/

Seeman, D., & Goddard, L. (2014). Preparing the Way: Creating Future Compatible Cataloging Data in a Transitional Environment. *Cataloging & Classification Quarterly*, (September 2015), 1–10.

https://doi.org/10.1080/01639374.2014.946573

Sfakakis, M., & Kapidakis, S. (2009). Eliminating query failures in a work-centric library meta-search environment. *Library Hi Tech*, *27*(2), 286–307. https://doi.org/10.1108/07378830910968236

Shotton, D., & Peroni, S. (2019). FRBR-aligned Bibliographic Ontology (FaBiO), version 2.1. Retrieved from https://sparontologies.github.io/fabio/current/fabio.html

Simon, A., Di Mascio, A., Michel, V., & Peyrard, S. (2014). We grew up together: data. bnf. fr from the BnF and Logilab perspectives. In *IFLA 2014*. Retrieved from http://ifla2014-satdata.bnf.fr/pdf/iflalld2014_submission_Simon_DiMascio_Michel_Peyrard.pdf

Smiraglia, R. P. (1992). *Authority control and the extent of derivative bibliographic relationships*. The University of Chicago, Chicago.

Smiraglia, R. P. (1999). Derivative Bibliographic Relationships among Theological Works. *Proceedings of the ASIS Annual Meeting*, *36*, 497–506.

Smiraglia, R. P. (2004). Authority Control of Works: Cataloging's Chimera? *Cataloging & Classification Quarterly*, *38*(3–4), 291–308. https://doi.org/http://dx.doi.org/10.1300/J104v38n03_22

Smiraglia, R. P. (2005). Instantiation : Toward a Theory. In L. Vaughan (Ed.), *Data, information and knowledge in a networked world; Annual conference of the Canadian Association for Information Science. London, Ontario. June 2-4 2005.* (pp. 1–8). London, Ontario. Retrieved from http://www.cais-acsi.ca/ojs/index.php/cais/article/view/226

Smiraglia, R. P. (2007a). Bibliographic families and superworks. In A. G. Taylor (Ed.), *Understanding FRBR: what it is and How it will Affect our Retrieval Tools* (pp. 73–86). Westport, Connecticut: Libraries Unlimited.

Smiraglia, R. P. (2007b). The "Works" Phenomenon and Best Selling Books. *Cataloging & Classification Quarterly*, *44*(3–4), 179–195. https://doi.org/10.1300/j104v44n03_02

Smiraglia, R. P. (2009). Defining Bibliographic ' Works :' Naïve classification for Terminology Generation. In C. N. Peña (Ed.), *Memoria del I Simposio Internacional sobre Organización del Conocimiento: Bibliotecología y Terminología* (pp. 7–17). México: Universidad Nacional Autónoma de México. Retrieved from http://iibi.unam.mx/voutssasmt/documentos/organizacion_del_conocimiento_corto.pdf

Smiraglia, R. P., & Leazer, G. H. (1999). Derivative bibliographic relationships: The work relationship in a global bibliographic database. *Journal of the American Society for Information Science*, *50*(6), 493–504. https://doi.org/10.1002/(SICI)1097-4571(1999)50:6<493::AID-ASI4>3.0.CO;2-U

Smith-Yoshimura, K. (2016). Analysis of international linked data survey for implementers. *D-Lib Magazine*, *22*(7–8), 1–11. https://doi.org/10.1045/july2016-smith-yoshimura

Smith-Yoshimura, K. (2018). Analysis of 2018 International Linked Data Survey for Implementers. *Code4Lib Journal*, *42*. Retrieved from https://journal.code4lib.org/articles/13867

Spalding, C. S., American Library Association, Library of Congress, Library Association, American Library Association. Division of Cataloging and Classification, & Library of Congress. Descriptive Cataloging Division. (1967). *Anglo-American cataloguing rules*. London: Library Association.

Spellman, R., & Holley, R. (2011). An Overview of the Google Books Project and Other Digitization Initiatives. *Journal of Library and Information Science*, *37*(1), 18–30.

Sridhar, M. S. (2004). OPAC vs card catalogue: A comparative study of user behaviour. *Electronic Library*, *22*(2), 175–183. https://doi.org/10.1108/02640470410533443

Statement of Principles adopted by The International Conference on Cataloguing Principles Paris, October 1961. (1961). Retrieved from https://www.ifla.org/files/assets/cataloguing/IMEICC/IMEICC1/statement_principles_paris_1961.pdf

Stavans, I. (2008). One Master, Many Cervantes: Don Quixote in translation. *Humanities: The Magazine of the National Endowment for Humanities*, *29*(5). Retrieved from http://www.neh.gov/humanities/2008/septemberoctober/feature/one-master-many-cervantes

Suero, D. V. (2011). *Library Linked Data Incubator Group : Use Cases*. Retrieved from https://www.w3.org/2005/Incubator/lld/XGR-lld-usecase-20111025/

Suominen, O., & Hyvönen, N. (2017). From MARC silos to Linked Data silos? *O-Bib. Das Offene Bibliotheksjournal*, *4*(2), 1–13. https://doi.org/10.5282/O-BIB/2017H2S1-13

Susannah Hunnewell. (2015). Richard Pevear and Larissa Volokhonsky, The Art of Translation No. 4. *Paris Review*, (213). Retrieved from https://www.theparisreview.org/interviews/6385/richard-pevear-and-larissa-volokhonsky-the-art-of-translation-no-4-richard-pevear-and-larissa-volokhonsky

Svenonius, E. (2009). *The Intellectual Foundation of Information Organization*. Cambridge, Mass.; London: MIT Press.

Svensson, L. G. (2013). Are current bibliographic models suitable for integration with the web? *Information Standards Quarterly*, *25*(4), 7–13. Retrieved from http://dx.doi.org/10.3789/isqv25no4.2013.02

Szekely, P., Knoblock, C. A., Yang, F., Zhu, X., Fink, E. E., Allen, R., & Goodlander, G. (2013). Connecting the Smithsonian American Art Museum to the Linked Data Cloud. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7882 LNCS*, 593–607. https://doi.org/10.1007/978-3-642-38288-8-40

Takhirov, N. (2013). *Extracting Knowledge for Cultural Heritage Knowledge Base Population*. Norwegian University of Science and Technology. Retrieved from http://hdl.handle.net/11250/253516

Tallerås, K. (2017). Quality of Linked Bibliographic Data: The Models, Vocabularies, and Links of Data Sets Published by Four National Libraries. *Journal of Library Metadata*, *17*(2), 126–155. https://doi.org/10.1080/19386389.2017.1355166

Tallerås, K. (2018). *Metadata Structures of the Bibliographic Universe : Transformation , Interoperability , Conceptualizations , and Quality*. OsloMet – Oslo Metropolitan University. Retrieved from https://oda-hioa.archive.knowledgearc.net/handle/10642/6622

Taniguchi, S. (2002). A conceptual model giving primacy to expression-level bibliographic entity in cataloging. *Journal of Documentation*, *58*(4), 363–382. https://doi.org/10.1108/00220410210431109

Taniguchi, S. (2012). Viewing RDA from FRBR and FRAD: Does RDA represent a different conceptual model? *Cataloging and Classification Quarterly*, *50*(8), 929–943. https://doi.org/10.1080/01639374.2012.712631

Taniguchi, S. (2013). Aggregate and Component Entities in RDA: Model and Description. *Cataloging & Classification Quarterly*, *51*(5), 580–599. https://doi.org/10.1080/01639374.2013.763316

Taniguchi, S. (2017a). Examining BIBFRAME 2.0 from the Viewpoint of RDA Metadata Schema. *Cataloging & Classification Quarterly*, *55*(6), 387–412. https://doi.org/10.1080/01639374.2017.1322161

Taniguchi, S. (2017b). Is BIBFRAME 2.0 a Suitable Schema for Exchanging and Sharing Diverse Descriptive Metadata about Bibliographic Resources? *Cataloging & Classification Quarterly*, *0*(0), 1–22. https://doi.org/10.1080/01639374.2017.1382643

Taniguchi, S. (2018). Mapping and Merging of IFLA Library Reference Model and BIBFRAME 2.0. *Cataloging & Classification Quarterly*, *56*(5–6), 427–454. https://doi.org/10.1080/01639374.2018.1501457

Term and Code List for RDA Content Types. (2014). Retrieved August 3, 2019, from https://www.loc.gov/standards/valuelist/rdacontent.html

The British Library. (2019). The British National Bibliography as Linked Open Data. Retrieved August 10, 2019, from https://bnb.data.bl.uk/

The Library 100. (2019). Retrieved September 9, 2019, from https://www.oclc.org/en/worldcat/library100.html

Tillett, B. (1987). *Bibliographic Relationships: Toward a Conceptual Structure of Bibliographic Information used in Cataloging*. University of California, Los Angeles. Retrieved from http://proquest.umi.com/pqdweb?did=753729371&Fmt=7&clientId=43732&RQT=309&VName=PQD

Tillett, B. (1991). A taxonomy of bibliographic relationships. *Library Resources & Technical Services*, *35*(2), 150–158.

Tillett, B. (1994). IFLA Study on the Functional Requirements of Bibliographic Records : Theoretical and Practical Foundations. In *60th IFLA General Conference - Conference Proceedings - August 21-27, 1994* (pp. 1–5). Retrieved from https://archive.ifla.org/IV/ifla60/60-tilb.htm

Tillett, B. (2004). *What is FRBR?: A conceptual model for the bibliographic universe*. Washington, D.C. Retrieved from www.loc.gov/cds/downloads/FRBR.PDF

Ullah, I., Khusro, S., Ullah, A., & Naeem, M. (2018). An Overview of the Current State of Linked and Open Data in Cataloging. *Information Technology and Libraries*, *37*(4), 47–80. https://doi.org/10.6017/ital.v37i4.10432

Underwood, S. (1998). *English Translators of Homer*. *English Translators of Homer*. Plymouth, UK: Northcote House Publishers Ltd. https://doi.org/10.2307/j.ctv5rf17h

Urban, R. J. (2013). Representation patterns for cultural heritage resources. *Proceedings of the ASIST Annual Meeting*, *50*(1). https://doi.org/10.1002/meet.14505001123

Urban, R. J. (2014). LODLAM Patterns. Retrieved March 1, 2020, from https://web.archive.org/web/20160812081154/http://chi.cci.fsu.edu/person/rurban/projects/lodlam-patterns/

Vellucci, S. L. (1995). *Bibliographic relationships among musical bibliographic entities : A conceptual analysis of music represented in a library catalog with a taxonomy of the relationships discovered*. Columbia University.

Verona, E. (1959). Literary Unit Versus Bibliographical Unit. *Libri*, *9*(1–4), 79–104. https://doi.org/10.1515/libr.1959.9.1-4.79

W3C. (2015). Linked Data. Retrieved September 2, 2020, from https://www.w3.org/standards/semanticweb/data

W3C Library Linked Data Incubator Group. (2012). Retrieved from https://www.w3.org/2005/Incubator/lld/

Wahid, N., Warraich, N. F., & Tahira, M. (2018). Mapping the cataloguing practices in information environment: a review of linked data challenges. *Information and Learning Science*, *119*(9–10), 586–596. https://doi.org/10.1108/ILS-10-2017-0106

Wallheim, H. (2016). From complex reality to formal description: Bibliographic relationships and problems of operationalization in RDA. *Cataloging and Classification Quarterly*, *54*(7), 483–503. https://doi.org/10.1080/01639374.2016.1200169

Wallis, R. (2018). MARC and beyond: our three Linked Data choices. In *IFLA WLIC 2018 – Kuala Lumpur, Malaysia – Transform Libraries, Transform Societies in Session 113 - Information Technology*. Kuala Lumpur, Malaysia: IFLA. Retrieved from http://library.ifla.org/id/eprint/2124

Weihs, J., & Howarth, L. C. (2008). Uniform titles from AACR to RDA. *Cataloging and Classification Quarterly*, *46*(4), 362–384. https://doi.org/10.1080/01639370802322853

Wiesenmüller, H. (2012). *Some comments on the Final Report of the FRBR Working Group on Aggregates*. Stuttgart. Retrieved from http://www.mendeley.com/profiles/heidrun-wiesenmuller/

Willer, M., & Dunsire, G. (2013). *Bibliographic Information Organization in the Semantic Web*. *Bibliographic Information Organization in the Semantic Web*. Oxford, UK: Chandos Publishing.

Wilson, P. (1968). *Two kinds of power: An essay on bibliographical control*. Berkeley, Los Angeles, London: Univ of California Press.

Working Group on FRBR/CRM Dialogue, Bekiari, C., Doerr, M., Le Boeuf, P., & Riva, P. (2016). *Definition of FRBROO: A Conceptual Model for Bibliographic Information in Object-Oriented Formalism*. Den Haag. Retrieved from http://www.ifla.org/files/assets/cataloguing/FRBRoo/frbroo_v_2.4.pdf

Yee, M. M. (1993). *Moving image works and manifestations*. University of California. https://doi.org/10.16953/deusbed.74839

Yee, M. M. (1994). What is a work? part 1: The user and the objects of the catalog. *Cataloging and Classification Quarterly*. https://doi.org/10.1300/J104v19n01_03

Yee, M. M. (1995). What is a work? part 4: Cataloging theorists and a definition abstract. *Cataloging and Classification Quarterly*, *20*(2), 3–24. https://doi.org/10.1300/J104v20n02_02

Yee, M. M. (2005). FRBRization: A method for turning online public finding lists into online public catalogs. *Information Technology and Libraries*, *24*(2), 77–95. https://doi.org/10.6017/ital.v24i2.3368

Zapounidou, S., Sfakakis, M., & Papatheodorou, C. (2019a). Assessing the preservation of derivative relationships in mappings from FRBR to BIBFRAME. In E. Garoufallou, F. Sartori, R. Siatri, & M. Zervas (Eds.), *Metadata and Semantic Research. MTSR 2018. Communications in Computer and Information Science, vol 846.* (pp. 230–241). Springer, Cham. https://doi.org/https://doi.org/10.1007/978-3-030-14401-2_22

Zapounidou, S., Sfakakis, M., & Papatheodorou, C. (2019b). Mapping derivative relationships from RDA to BIBFRAME 2. *Cataloging & Classification Quarterly*, *57*(5), 278–308. https://doi.org/10.1080/01639374.2019.1650152

Zeng, M. L. (2019). Interoperability. *Knowledge Organization*, *46*(2), 122–146. https://doi.org/10.5771/0943-7444-2019-2-122.Abstract

Zeng, M. L., & Qin, J. (2016). *Metadata* (2nd ed.). Atlanta, GA, USA: ALA Neal-Schuman.

Žumer, M., & O'Neill, E. T. (2012). Modeling Aggregates in FRBR. *Cataloging & Classification Quarterly*, *50*(5–7), 456–472. https://doi.org/10.1080/01639374.2012.679547

Appendix A. List of publications

Zapounidou, S., Sfakakis, M., & Papatheodorou, C. (2013). Highlights of library data models in the era of Linked Open Data. In: Garoufallou E., Greenberg J. (eds) Metadata and Semantics Research. MTSR 2013. *Communications in Computer and Information Science, vol 390*, 396-407. Springer, Cham. doi:10.1007/978-3-319-03437-9_38.

Zapounidou S., Sfakakis M., & Papatheodorou C. (2014). Integrating library and cultural heritage data models: the BIBFRAME - EDM case. In: *Proceedings of the 18th Panhellenic Conference on Informatics. PCI 2014. ACM*. doi:10.1145/2645791.2645805.

Zapounidou S., Sfakakis M., & Papatheodorou C. (2014). Library Data Integration: Towards BIBFRAME Mapping to EDM. In: Closs S., Studer R., Garoufallou E., Sicilia MA. (eds) *Metadata and Semantics Research. MTSR 2014. Communications in Computer and Information Science, vol 478,* 262-273. Springer, Cham. doi:10.1007/978-3-319-13674-5_25.

Zapounidou, S., Sfakakis, M., & Papatheodorou, C. (2017). Representing and integrating bibliographic information into the semantic web: A comparison of four conceptual models. *Journal of Information Science, 43*(4), 525-553. doi:10.1177/0165551516650410.

Zapounidou S. (2017) Studying Conceptual Models for Publishing Library Data to the Semantic Web. In: Kamps J., Tsakonas G., Manolopoulos Y., Iliadis L., Karydis I. (eds) *Research and Advanced Technology for Digital Libraries. TPDL 2017. Lecture Notes in Computer Science, vol 10450,* 652-655. Springer, Cham. doi:10.1007/978-3-319-67008-9_63.

Zapounidou S., Sfakakis M., & Papatheodorou C. (2017). Preserving Bibliographic Relationships in Mappings from FRBR to BIBFRAME 2.0. In: Kamps J., Tsakonas G., Manolopoulos Y., Iliadis L., Karydis I. (eds) *Research and Advanced Technology for Digital Libraries. TPDL 2017. Lecture Notes in Computer Science, vol 10450*, 15-26. Springer, Cham. doi:10.1007/978-3-319-67008-9_2.

Zapounidou S., Sfakakis M., & Papatheodorou C. (2017). The representation of bibliographic families in library data models and their preservation in mappings: the case of the mapping from FRBR to BIBFRAME. *Europeana Tech Insight 8*. https://pro.europeana.eu/page/issue-8-tpdl#the-representation-of-bibliographic-families-in-library-data-models-and-their-preservation-in-mappings-the-case-of-the-mapping-from-frbr-to-bibframe

Zapounidou S., Sfakakis M., & Papatheodorou C. (2019). Assessing the Preservation of Derivative Relationships in Mappings from FRBR to BIBFRAME. In: Garoufallou E., Sartori F., Siatri R., Zervas M. (eds) *Metadata and Semantic Research. MTSR 2018. Communications in Computer and Information Science, vol 846*, 230-241. Springer, Cham. doi:10.1007/978-3-030-14401-2_22.

Zapounidou, S., Sfakakis, M., & Papatheodorou, C. (2019). Mapping Derivative Relationships from RDA to BIBFRAME 2. *Cataloging & Classification Quarterly*, *57*(5), 278-308. doi:10.1080/01639374.2019.1650152.

Zapounidou, S., Sfakakis, M., & Papatheodorou, C. (2020). Semantic Interoperability between Bibliographic Models. In Frosterus, M. (2020). *LIBER 2020 - Workshop: LIBER Linked Open Data Working Group – LOD Publication for Libraries*. doi:10.5281/zenodo.3904123

Sfakakis, M., Zapounidou, S., & Papatheodorou, C. (2020). Mapping derivative relationships from BIBFRAME 2.0 to RDA. *Cataloging & Classification Quarterly, 58* (7), 603-631. doi: 10.1080/01639374.2020.1821856.