

ISCTE

Instituto Universitário de Lisboa

Departamento de Ciências e Tecnologias da Informação

Contributions to FRBRization

**The library catalogue in a new digital environment:
how FRBR can help produce better displays in
OPACs**

**Maria Teresa Martins Dias Teixeira
(Licenciada)**

Dissertação para a obtenção do Grau de

Mestre em

Estudos de Informação e Bibliotecas Digitais

Orientador: Professor Doutor Pedro Faria Lopes

Co-Orientador: Professor Doutor José Luís Borbinha

January, 2010

Resumo

As bibliotecas enfrentam, neste momento, enormes desafios. Os utilizadores trabalham num ambiente de rede, onde a oferta de informação e os meios para a pesquisa são diversos e amigáveis. Os catálogos das bibliotecas, vulgarmente designados por OPACs, tornaram-se, pelo contrário, sistemas pouco intuitivos.

Este trabalho tem o objectivo de mostrar como os FRBR, os Requisitos Funcionais dos Registos Bibliográficos, podem contribuir favoravelmente para novas visualizações nos catálogos das bibliotecas com um impacto muito positivo no acesso à informação. Os catálogos actuais apresentam longas listas de registos bibliográficos sem qualquer organização e aparente relação entre estes. Estas visualizações são muito confusas para os utilizadores.

O estudo sobre os FRBR, produzido pela IFLA, a Federação Internacional das Associações de Bibliotecas, recomenda um novo modelo conceptual para o universo bibliográfico orientado para os utilizadores.

As bibliotecas esperam poder vir a usar os FRBR para transformar as longas listas de registos em novos formatos de visualização onde a informação é agrupada de acordo com determinadas categorias ou *clusters*, indo desta maneira ao encontro das necessidades dos seus utilizadores.

Temos uma herança de milhões de registos num formato MARC, isto é, num formato legível por computador, que não é compatível com o modelo FRBR. A solução pode estar, para já, na FRBRização do catálogo, isto é, no desenvolvimento de uma nova interface de pesquisa que permita a visualização de acordo com o novo modelo sem alterar o catálogo. A investigação levada a cabo baseou-se no levantamento de experiências e implementações mais relevantes e foi integrado no projecto TELplus, o projecto da biblioteca digital europeia, a Europeana, com o qual este trabalho de tese mantém até um certo ponto uma estreita relação. O trabalho relata como, com base no conhecimento adquirido, se fez a FRBRização de uma amostra de registos extraídos da PORBASE, o Catálogo Colectivo Nacional, e se avançou no desenvolvimento de um protótipo de pesquisa.

Palavras chave: OPAC, catálogo, pesquisa de informação, FRBR, IFLA, FRBRização, utilizadores, PORBASE, Projecto TELplus, Europeana.

Abstract

Libraries face new challenges. There is a new networked environment with access to more user-friendly search tools than those that exist in libraries, the OPACs, the Online Public Access Catalogues. These became less intuitive and difficult to use.

This work has the goal of showing how FRBR, the Functional Requisites for Bibliographic Records, can help libraries with a better display of search results on catalogues enhancing better conditions for information access. Current catalogues provide long alphabetical lists of records. These displays, with no evident organization, are becoming more and more confusing to the end users. FRBR is a study produced under the auspices of the International Federation of Library Associations (IFLA), that recommends a new conceptual model of the bibliographic universe with a strong user focus.

Libraries can make use of FRBR to display bibliographic records grouping them into relevant categories or clusters, providing a better understanding of what is being retrieved.

Though the implementation of FRBR may bring relevant advantages to the world of libraries in the future, there is a legacy of millions of records in an old format, MARC, Machine Readable Cataloguing, that cannot be forgotten. Converting all these records to a new FRBR system would be a daunting task. The solution may be in the design of a new search interface that may display the information according to the new model without changing the catalogue. Researching what is being done all over the world in with FRBR applications as related to different types of collections and information environments and a brief description of some of the issues that have been identified in each area, was accomplished and integrated in TELplus, a project created to provide value-adding services and products for The European Library, with which this thesis is connected to a certain point.

Further, this work gives a detailed report of how, with this knowledge, a FRBRization of a sample of records extracted from PORBASE, the national union catalogue, was implemented, and a development of a searching tool prototype took place.

Keywords: OPAC, catalogue, information searching, FRBR, IFLA, FRBRization, PORBASE, TELplus project, European Library.

Acknowledgements

This has been an arduous project for me, but in the end, very rewarding. Nothing that comes without sacrifice is memorable or valuable. To my daughter, Susana Costa, I give my thanks and love, for always being there when I needed an ear to hear me and a shoulder to lean on. Her strength and wise counsel were always timely, to be tapped and used when needed. Also, I would be remiss to not thank my two teachers and counselors, Doutor Pedro Faria Lopes, who helped me to design and implement the body of my thesis, who guided me and provided valuable advice in its genesis and implementation, and Doutor José Luís Borbinha, whose experience in the field was invaluable in creating a concrete implementation rather than just a theoretical one. FRBR is here to stay, and those of us who believe in its value and importance must insure that it does not fall prey to the search engines that are not specific enough to do the job the client user requires. To that end this FRBR thesis addresses many of the concerns of the client user, while keeping in mind that porting the obsolete systems into the modern integrations of FRBR will require a major effort that is both relevant and useful.

“The model developed for this study represents an initial attempt to establish a logical framework to assist in the understanding and further development of conventions for bibliographic description. It is intended to provide a base for common understanding and further dialogue, but it does not presume to be the last word on the issues it addresses.” (FRBR Report, 1998)

Table of Contents:

1	Introduction.....	7
1.1	Objectives and methodology.....	11
1.2	Introduction to the subject: What is FRBR?.....	11
1.2.1	Historical background:.....	11
1.2.2	Explaining the model.....	14
1.2.3	Methodology.....	16
1.3	Structure of this work.....	17
2	State of the art of FRBR.....	20
2.1	The impact of Research on FRBR Development:.....	20
2.2	Major implementations of FRBR.....	25
2.2.1	Worldcat.....	25
2.2.2	Austlit.....	27
2.2.3	Perseus Digital Library (PDL).....	29
2.2.4	UCLA Library.....	31
2.3	Products & Technology.....	33
2.3.1	Virtua ILS (Integrated Library Systems).....	33
2.3.2	Aquabrowser.....	35
2.4	Research: Supporting Tools, Algorithms and Techniques.....	36
2.4.1	OCLC FRBR Work-Set Algorithm.....	37
2.4.2	FRBR Display Tool.....	38
2.4.3	FRBR Floater.....	38
2.4.4	BibSys FRBR conversion tool.....	39
2.4.5	OCLC FRBR prototypes.....	40
2.4.5.1	OCLC FictionFinder.....	40
2.4.5.2	The FRBRization of Humphry Clinker.....	42
2.4.5.3	The NDB Prototype.....	43
2.4.5.4	BYSYS.....	45
2.4.6	An important lesson learned.....	47
3	Detailed description of the problem.....	49
4	How and why FRBR can help solve the problem.....	56
4.1	Introduction and rationale for the research.....	56
4.2	Methodology.....	59
4.3	Study of the model.....	60
4.4	Finding the set of rules to FRBRization.....	72
5	Results and discussion.....	88
6	Conclusions and future work.....	94
	Bibliography and References.....	99

1 Introduction

This thesis is the result of the work that has been developed to give support to two sub tasks of TELplus¹, a project of the Conference of the European National Libraries (CENL)² that is being taken by a consortium of 26 partners, national libraries and research centres that built a pool of research and innovation to provide value-adding services and products for The European Library (TEL)³.

This research project is composed of eight work packages (WPs) to increase access to digital content and making the European Library gateway usable. The main work areas are: Optical Character Recognition (OCR) of scanned texts; full text search with the use of semantic web techniques; building a centrally managed infrastructure of distributed repositories with the use of the Open Archives Initiative-Protocol for Metadata Harvesting (OAI-PMH); setting up a metadata repository based on the Functional Requirements for Bibliographic Records (FRBR) concept model; building a services infrastructure and investigating user profiles for providing personalised services.

We will concentrate on Work Package 3 (WP3), which aims at setting up a repository and the development of solutions to support alternative services of searching and browsing in The European Library based on the Functional Requirements for Bibliographic Records (FRBR) concept model. “The objective of WP3, whose leader is Bibliothèque Nationale de France, is “to improve accessibility in three ways: by improving full-text searching; by aggregating related resources; and investigating automatic vocabulary mappings”. It is composed of several tasks (five in total), being task

¹ <http://www.theeuropeanlibrary.org/portal/organisation/cooperation/telplus/>

² <http://www.cenl.org/>

³ <http://search.theeuropeanlibrary.org/portal/en/index.html>

3 “Setting up a FRBR repository for the European Library” the one we will be focusing on. This task is lead by Instituto Superior Técnico of the Technical University of Lisbon, and has the participation of the National Library of Portugal and The European Library Office. Several other national libraries also contribute to this FRBR task with bibliographic and authority data, and as advisors.” (Freire, 2009)

TELplus project was funded by the eContentplus programme of the European Commission with the purpose of enhancing digital content in Europe and started on October 2007 and will last twenty seven months. When this project is accomplished The European Library (TEL) is ready to offer a common access point to the distributed collections of the 48 national libraries of Europe in 32 languages. The mission of the European Library is described as follows: “The European Library exists to open up the Universe of knowledge, information and cultures of all Europe’s national libraries”⁴ The European Library gives access to 150 million entries across Europe, covering bibliographic and digital resources and is a free service.

“In conclusion it should be mentioned that the important research and development being realised in the context of the TELplus project such as the REPOX System⁵, along with the service infrastructure and the mass digitisation of content, will not only help innovate The European Library portal but will also feed into the development of a ground-breaking new service – Europeana⁶. Europeana is the cross-domain pan European service offering access to the digital holdings of museums, archives, audio-visual archives and libraries” (Angelaki, 2008).

Work Package 3 is a multifaceted work package combining innovative research for the European Library in different areas of work. The main goal of task 3 is to research and support the development of better searching tools to improve access to TEL digital contents through the implementation of the FRBR model in the Gateway. Using harvested bibliographic records from partner libraries, some experiments will take place to help in the development of searching and browsing tools according to the FRBR model. This model will inspire us in the organization of those harvested re-

⁴ Mission of the European Library. Available at: www.theeuropeanlibrary.org

⁵ Veja Freire, N., Galvão, R., Lopes, M.(2009) - FRBR information discovery in traditional catalogues: the TELplus experience. Available at: <http://www.ifla.org/files/hq/papers/ifla75/135-freire-en.pdf>

⁶ <http://www.europeana.eu/portal/index.html>

records into different categories or clusters like translations, editions or different media type variations of the same work. This user focused perspective has a great impact on usability as it can narrow down search results by clustering similar records for the user.

The work group in charge of task 3 is focused in the development of a search interface designed according to the FRBR model to be used in the context of that repository, TEL-The European Library's Gateway. Instituto Superior Técnico de Portugal (IST) and Biblioteca Nacional de Portugal (BNP) are partners in this project.

This thesis gave support to Sub Task 3.3.1: Review of the state of FRBR experiments, D3.6 – Report on FRBR experiments (Teixeira, 2008) and Sub Task 3.3.2: Develop a system that, using all the bibliographic records available in The European Library (TEL), can produce a FRBR metadata repository.

The first part of the research project consisted of a state of the art study of FRBR and it was completed and delivered on the 29th July, 2008. It has been updated to be included in this thesis.

With that study, once familiar with what had been done in this field and having learnt with others practical experiments and results, we were able to contribute with our own experiments to the development of a search interface – a prototype - that might suit this new way of organizing knowledge for display purposes on *Online Public Access Catalogues*.

Besides our own experiments, “A relevant project for our work was Aalberg's BibSys FRBR conversion tool. This tool aids in the process of extracting the FRBR entities and relations from bibliographic data in Machine Readable Catalogue (MARC) based formats. This tool was developed and applied to 4 million records in the BibSys database – the union catalogue of Norway. Although there are some issues to solve related with inconsistent catalogue data and scalability, this tool has proved to be valuable in creating FRBRized views for catalogues» (Freire, 2009).

The latter step, the prototype FRBR repository, as well as the conformity and usability tests will not be ready in time to include in this Master Thesis work. The prototype was planned to be ready in June 2009, but its development has been delayed.

This work is strictly about how to FRBRize a catalogue. In the course of this project, old Machine Readable Catalogue (MARC) records are analyzed to identify the FRBR

entities and to organize them into groups or clusters that correspond to our users' interests. The results that helped us to build the rules to accomplish a FRBRized catalogue will be discussed. We also have identified some issues.

So, due to the Master Thesis schedule, the thesis ends with a set of rules to guide the development of a FRBRized searching interface, and before the development of the prototype search interface takes place. This will be soon implemented, tested and validated after user surveys done with end-users and library professionals.

FRBR model has become an important topic, many debates and workshops take place everywhere in the world and new rules for bibliographic description have already been written influenced by this new conceptual model. These new rules, Resources Description and Access (RDA) are being discussed and translated. It will take a certain time before they start being used. Library systems need also to be redesigned to be in harmony with this new way of describing and organizing information. It is not easy to make the shift from the old system to the new one. There is a legacy of millions of records that are in accordance with the old system. FRBRization is the automatic way we have to convert an old MARC catalogue into a display inspired in the new model. This is an important job to accomplish, considering the millions of records that have been produced in the old MARC format, and though they may have to be kept in the old format, they may be displayed according to new FRBR inspired visualizations and fit this way end users' interests. FRBR is focused on end users and FRBRization is a means, maybe the only means available for the time being, to offer users a better display of the information.

When the retrieved bibliographic record sets are displayed into categories or clusters the users understand more easily what is being shown to them.

In short, in the context of this thesis the process of FRBRizing a catalogue is analyzed and the issues that FRBRization raises are discussed. A set of rules to help a FRBRized way of displaying information has been built.

FRBR has become so important in the library environment that it has given birth to a set of new words like FRBRize and FRBRization, and these are adopted and used frequently in this work.

1.1 Objectives and methodology

The main objective of this work is to give a contribution to the FRBRization of an Online Public Access Catalogue (OPAC), enhancing better displays to make searching and browsing easier to the catalogues' users.

We started by analyzing our old Machine Readable Catalogues (MARC) in order to try to identify if FRBR entities were present in the bibliographic records. "The first question we asked ourselves was: Can we find the FRBR structure in the existing MARC records? If we look at a single record, we see that there is information about the work, the expression and the manifestation..."(Hegna and Murtomaa, 2002).

1.2 Introduction to the subject: What is FRBR?

FRBR is an acronym for *Functional Requirements for Bibliographic Records* the title of a study⁷ undertaken by the International Federation of Library Associations and Institutions (IFLA). So, FRBR and IFLA will be two acronyms used frequently in the course of this study.

This study produced a new model of the bibliographic universe. This model includes not only documents and objects that exist in libraries, archives and museums, but also persons, bodies and families that have any kind of relationship with those documents or the objects as authors, owners or producers. Concepts are also included, as they are necessary to describe all these entities. The model shows us how these entities relate to each other and the ways users of libraries interact with databases to get what they need. This model is expected to help organize the burst of information that is characteristic of the late XX and XXI centuries.

1.2.1 Historical background:

Some FRBR roots come from the Paris Principles, dating from 1961. IFLA, forty years ago, decided to analyse the cataloguing theory and practice in a meeting held in Paris in 1961. A set of new rules were produced and therefore known as the Paris Principles.

⁷ IFLA Study Group on the Functional Requirements for Bibliographic Records, *Functional Requirements for Bibliographic Records, Final Report*, UBCIM Publications, New Series, vol. 19 (München: K. G. Saur, 1988); also available at http://www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf

Later, in 1969, there was another International Meeting of Cataloguing Experts in Copenhagen with the purpose of establishing international standards for the form and content of bibliographic descriptions. In 1971, the International Standard for Monographic Description (ISBD) was published. Soon other rules of the same ISBD family followed this one. These standards and the Paris Principles were the core of rules that were followed internationally and served as the bibliographic core for new or translated cataloguing rules adopted national and internationally.

By the late 1980s professionals in the library world, including members of the IFLA Section on Cataloguing, realized that important changes were happening in the library environment. The way information was being organized, especially with the use of automated systems, new formats, electronic publishing, networked access and the new web resources, recommended a careful reflection. The catalogue had lost the structure and the hierarchical organization of the former days.

We can read now in the introduction of the FRBR Report, that there were dramatic changes happening that required a rethinking of the way information was being organized:

- The development of automated systems for the creation and processing of bibliographic data;
- Shared cataloguing and the need to avoid duplicate efforts;
- Economic pressure to simplify the cataloguing process and to do it according to a “minimal level”.
- The emergence of new formats, electronic publishing and networked access to information.

These concerns led to another meeting in Stockholm, known as The Stockholm Seminar on Bibliographic Records, held on 1990 and sponsored by IFLA Universal Bibliographic Control and International MARC (UBCIM) Programme and the IFLA Division of Bibliographic Control⁸. This was the chosen moment to debate all those issues.

The participants in the Seminar knew the economic realities faced by libraries and the need to reduce the cost of cataloguing, but they also acknowledged the importance of

⁸ <http://www.ifla.org/en/about-the-division-of-bibliographic-control>

meeting user needs considering the different kinds of resources they were now looking for and the various contexts within which bibliographic records were used. One of the nine resolutions approved in that Seminar led to a study to define the functional requirements for bibliographic records (FRBR). The terms of reference written for that study were: "The purpose of the study is to delineate in clearly defined terms the functions performed by the bibliographic record with respect to various media, various applications, and various user needs. The study is to cover the full range of functions for the bibliographic record in its widest sense – i.e. record that encompasses not only descriptive elements, but access points (name, title, subject, etc., other "organizing" elements (classification, etc.) and annotations." (FRBR Report, 1998)⁹

The study group was also charged to recommend a basic level of functionality and basic data requirements for records created by national bibliographic agencies.

The result of that process was the creation of the IFLA Study Group on the Functional Requirements for Bibliographic Records, which published its report "Functional Requirements for Bibliographic Requirements" (FRBR) in 1980. A simple synthesis is provided below.

The FRBR report had two main goals. One was the provision of a structured framework for relating the data that are given in the bibliographic records to the needs of the users of those records. The second goal was to recommend a basic level of functionality for records created by national bibliographic agencies.

This report would:

- Determine in full the functions of the bibliographic record and identify its primary users;
- Develop a framework identifying the full range of relationships that exist between the entities;
- Identify the functions the bibliographic record is supposed to perform for each entity;

⁹ IFLA Study Group on the Functional Requirements for Bibliographic Records (1988). Functional requirements for bibliographic records: final report.

- Identify the key attributes of each entity or relationship that are needed to allow them to perform the functions mentioned in the previous statement;
- Recommend a basic level of functionality that could be required of bibliographic records produced by national bibliographic agencies.

The objectives referred above imply that the model to be used for the study was entity-relationship. Entity-relationship (E/R) was chosen because it is a general model that can be used for any domain and has been developed in the computer science profession for a very practical purpose, to assist in the design of real databases.

The entity-relationship model has great potential both to be a basis for data creation operations such as cataloguing and, from the users' perspective, to allow much better search functionality and greater clarity in information display.

A bibliographic record is a set of data that describe the entities existing in libraries catalogues and national bibliographies. There are many applications where bibliographic records are used for a variety of purposes.

For the purposes of this study, these are the generic tasks that are performed by users when they make searches in those different applications:

- **find** entities that correspond to the users' search criteria;
- **identify** the entity (confirm that the entity found is the entity the user was looking for);
- **select** an entity from the resulting group appropriate to the user's needs;
- **obtain** the selected entity.

1.2.2 Explaining the model

According to the FRBR report "The entities that have been defined for this study represent the key objects of interest to users of bibliographic data. The entities have been divided into three groups and are shown in Figure 1. The first group (Group 1 Entities) comprises the products of intellectual or artistic endeavour that are named or described in bibliographic records: Work, Expression, Manifestation and Item. The second group (Group 2 Entities) comprises those entities responsible for the intellectual or artistic content, the physical production and dissemination, or the custodian-

ship of such products: person and corporate body. The third group (Group 3 Entities) comprises an additional set of entities that serve as the subjects of intellectual or artistic endeavour: concept, object, event and place.”

Entities are the key objects of interest to users of bibliographic data. The entities have been defined into three groups:

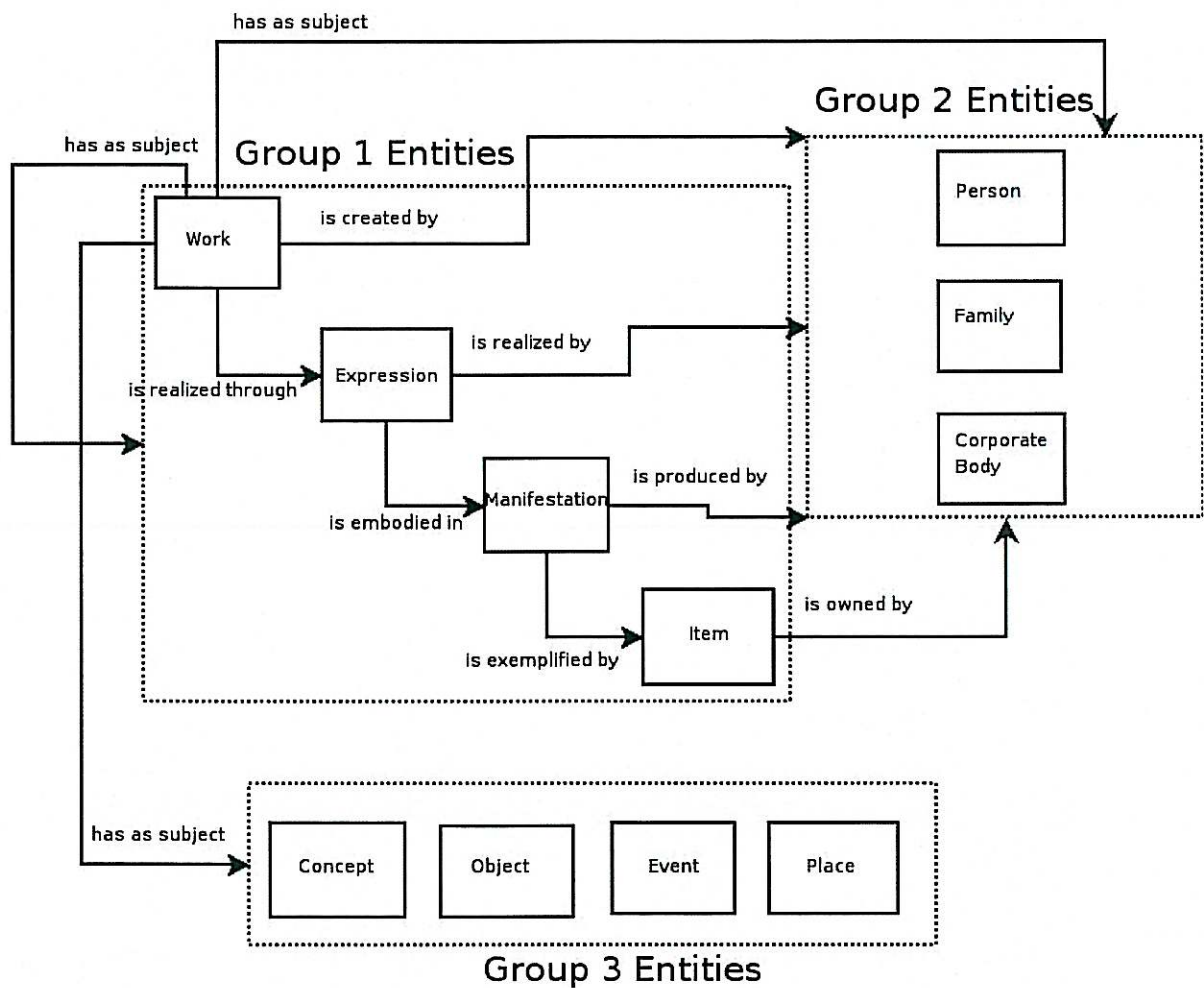


Figure 1 - The FRBR model¹⁰

¹⁰ Image reproduced from <http://www.miskatonic.org> and integrated in the TELplus project (TEIXEIRA, 2008).

The model in a simple way:

- An entity is something that can be distinctly identified. For example, persons, corporate bodies, events, concepts, works and publications may be entities.
- Entities are grouped into “entity sets”. Entities sets may have “subsets”. The entity set “person” may have subsets as “illustrator”, “photographer”, and so on.
- In the model, individual entities can also have “roles”. A role is the function an entity performs in a relationship. In a bibliographic database, the subsets referred above, “illustrator”, “photographer” should be considered “roles” rather than subsets, since a given person may perform any of those roles.
- A relationship is an association among two or more entities.
- As with entities, in the model, relationships are grouped in relationship sets. For example, in a given database the relationship set “translator – translation” might include many specific instances of this relationship.
- Entities and relationships are defined by attributes.
- Individual attributes or combination of attributes are the characteristics that uniquely identify entities. For example, one attribute of the entity set “person” might be “name”, because one of the characteristics of persons is that they have names. Another could be “date”.

1.2.3 Methodology

The methodology is based on an entity analysis technique that is used in the development of conceptual models for relational database systems.

The first measure is to isolate the key objects that are of interest to users. These objects are defined at a high level, as high as possible. This means that the analysis focuses attention not on data but on the “things” the data describe. Each of the entities defined for the model serves as a focal point for a cluster of data. At a high level an

entity diagram is able to define the relationships between different types of entities. The next step is to identify characteristics and attributes of each entity.

The entity-relationship structure derived from the analysis of entities, attributes and relationships was used as the framework for assessing the relevance of each attribute and relationships to the tasks performed by users of bibliographic records.

According to its drafters, FRBR is based on the "entity-relationship analyses technique" introduced by Peter Chen in the mid 1970's". Now widely used in database design, the model divides a given data universe (e.g. the data required to run a business) into specific entities linked by specific relationships.

1.3 Structure of this work

This document is organized as follows:

Chapter 1:

Introduction: This chapter starts explaining the association of this work with task 3 of Work Package 3 of the TELplus project, a research project for setting up a repository based on the Functional Requirements for Bibliographic Records (FRBR) concept model. It also designates the partners in this project. Furthermore, it introduces to the reader the main goal of this task 3, which is the research and the development of better searching tools to improve access to TEL digital contents through the implementation of the FRBR model in the Gateway.

Objectives and methodology of the master work: The main objective of this work is to give a contribution to the FRBRization of an Online Public Access Catalogue (OPAC), recommending the clustering of information according to the FRBR model to provide better displays in order to make searching and browsing easier to the catalogues' users.

Old MARC records are analyzed to identify where FRBR entities are present and then grouped according to those entities. These rules become the keys to FRBRization, with which the development team is able to build a new display.

Introduction to the subject: In this chapter we say what FRBR is about. We explain briefly that this is a new model of the bibliographic universe based on Entities and

Relationships. We further discuss how the model shows ways these entities interact with each other and ways users of libraries interact with databases to obtain what they need (Figure 2 - The FRBR model).

Chapter 2:

A state of the art of FRBR and FRBRization: a “state of the art” was accomplished and integrated in this study and this is the basis to move forward towards what will be the solution of a problem that is the loss of structure in the organization of the catalogued information nowadays. This is reflected in the way information is displayed. There has been a lot of research around this topic. The most relevant are discussed in this study. There are already a few experimental implementations to inspire this work. We do a brief overview of FRBR applications and implementations according to different types of collections and information environments with a description and discussion of some of the issues that have been identified for each area.

Chapter 3:

Detailed description of the problem: Current catalogue displays provide long alphabetical lists of bibliographic records with no evident organization. These results are confusing to the end user. Technological advances caused losses of bibliographic structure, absence of hierarchical display and no use of relationships. The flat structure of current databases cannot provide a hierarchical display of search results. MARC was originally designed as a flat-file system, with all information about a book or other format item stored within a single bibliographic record divided into fields or subfields. The old printed catalogues, highly structured, were thus transformed into lists of unrelated records. This is a problem that is likely to get worse in the foreseeable future if we consider the speed with which all the information is growing.

Chapter 4:

How and why FRBR can help solve the problem : The FRBR model has potential to transform long lists of alphabetical records into highly structured OPAC displays

according to the FRBR entities. Benefits include easier searching, clustering, and use of bibliographic relationships. Thus the end user will have a better experience when browsing and searching online. We explained how we worked to make the transition of the FRBR possible by manually extracting the FRBR structure from existing data, using a sample of bibliographic records exported from PORBASE.

Chapter 5:

Results and discussion: Having learned with what other researchers did as far as keys to FRBRization are concerned, we have built our own keys and applied them to a sample of bibliographic records. Tables with keys for FRBRization are given and also the results of our experiments that we analyze and discuss in detail. We summarize the main issues and obstacles to FRBRization.

Chapter 6:

Conclusions and further work: We discuss the overall benefits of FRBRization despite all the issues and the advantages of transforming our databases into entity-relationship databases following a FRBR model. This will improve access to library catalogues and meet library's users needs.

2 State of the art of FRBR

IFLA, and particularly the FRBR review group, played an important role in the promotion of FRBR research. The group created the FRBR bibliography with more than 500 entries and updated it regularly. Visiting and analyzing a major part of these 500 entries was a great challenge. Some books have already been published dealing exclusively with FRBR: (Smiraglia, 2002; LeBoeuf 2005; Taylor, 2007; Maxwell, 2008; Zhang and Salaba, 2009).

2.1 The impact of Research on FRBR Development:

FRBR began as a conceptual entity-relationship model for improving the structure of bibliographic records and has become a world-wide debated topic with implications not just for cataloguing but for many aspects of libraries and librarianship.

The FRBR report¹¹ stressed that “The model operates at the conceptual level; it does not carry the analysis to the level that would be required for a fully developed data model...The model developed for this study represents an initial attempt to establish a logical framework to assist in the understanding and further dialogue, but it does not presume to be the last word on the issues it addresses. Certain aspects of the model merit more detailed analysis and there are dimensions of the model that could be extended.” O’Neill draws the conclusion that “FRBR is not a fully developed model but rather a model that requires continuing refinement, interpretation, and development.”(O’Neill, 2002)

Barbara Tillet had already explained: “The universe [for the model] is characterized in terms of the entities in it and the relationships that hold among them. As such, the conceptual schema is not restricted by the capabilities of any particular database system and is independent of any particular record definition.” (Tillet, 1994)

¹¹ FRBR Report, 2008, p. 4,5

“On this basis, FRBR is highly theoretical, and as mentioned above, meant to be system neutral.” (Maxwell, 2008) Taking into account the fact that FRBR is open to many interpretations and implementations, according to what Zhang and Salaba explained (Zhang and Salaba, 2009), we will do an overview of the most relevant published work on this topic and experimental work.

In 2005, Patrick Le Boeuf published a book with the title “Functional Requirements for Bibliographic Records (FRBR): Hype or Cure-All?” It was considered “Very valuable for revealing significant differences in the way the FRBR concepts of work, expression, manifestation, and item are currently defined in such non-AACR2-using communities as those of rights managers, bibliographers, rare book catalogers, managers of oral tradition resources, and system designers.” (Yee, 2005)

A great number of advantages have been identified for the end user and these include “easier searching, focused results, clustering at the work level, understanding and using bibliographic relationships, and better navigation of the catalog for end users.” (Noerr et al. 1998)

Though many advantages have been identified, FRBR is expected to be more beneficial to certain types of resources, for instance, collections of fictional works, music collections, serial collections and other aggregate works that have been published in a variety of ways or published in different editions, by different publishers in different mediums. In short, FRBR has more advantages for those works that have many expressions and manifestations. FRBR is not so relevant for works with one expression and some manifestations and this group seems to represent a major part of our bibliographic universe. This idea seems to be common sense. However, the advantages are so great when a work has many expressions and manifestations, that it is worth thinking in terms of FRBR.

“The application of FRBR to WorldCat, the world’s largest union catalog, demonstrates several potential benefits in library catalogs. First, the sample data suggests that the task of applying FRBR may not be as burdensome as *a priori* estimates might suggest: FRBR can be applied non-trivially to only a small percentage of works in WorldCat. At a maximum, 20% of the works would be candidates (i.e., works with two or more manifestations); in practice, however, the percentage is likely to be lower. Analysis suggests that concentrating on relatively large works, in particular those

works whose content has been augmented, revised, or consists of collections of other works (a relatively small portion of the catalog) might be sufficient to capture the lion's share of benefits potentially available from implementing FRBR." (Bennett, Lavoie and O'Neill, 2003)

We realize, however, that FRBR provides great opportunities for creating retrieval systems that better support the end user information seeking.

As FRBR is a theoretical model and the report makes no recommendations, we can say that current efforts are exploratory. These efforts include:

- Theoretical discussions and recommendations based on some experiments,
- The creation of prototype systems for research and fully functional systems,
- Development of algorithms and support software as part of the implementation process.

Yin Zhang and Athena Salaba (Yin Zhang and Athena Salaba , 2007) of the University of Kent provided a broad overview of FRBR as a result of a study they did together with other FRBR experts, where they tried to identify the most critical issues and challenges that FRBR had to face. They identified critical issues in five areas:

1. FRBR model;
2. FRBR-related standards;
3. FRBR applications;
4. FRBR System development;
5. FRBR research.

"After a decade of discussion, exploration, and development, a better understanding of FRBR is still needed. The full potential of FRBR and its impact on library communities and beyond remain to be seen. Additionally, a clear direction for future FRBR research, application, and implementation still needs to be established." (Zhang and Salaba, 2009)

This study provides a good overview of current system developments and tools to create FRBR-based systems. It also confirmed that a number of these FRBR-based

systems are based on the FRBRization of existing data, rather than on creating new data from scratch.

FRBRization means the conversion from traditional single entity model to the four entity FRBR model.

An important line of work has been the development of processes to apply FRBR as an implementation model for existing catalogues, which is often referred to as FRBRization. In section 2.1 of this work, we review the literature that addresses the FRBRization of existing catalogues, or in other words, OPAC systems aligned with the FRBR principles.

Another review of how current MARC records can be used to promote FRBRization can be found in a recent article by Martha Yee (Yee, 2005). Yee believes that the most productive line of FRBR research may well be to investigate how the attributes in already existing MARC bibliographic, authority and holdings records may be used to enable FRBRization of OPACs.

These studies are relevant to show us that much of the information needed to FRBRize catalogues is already present in MARC data. Actually most OPACs are not true catalogues, but instead serve as online finding lists. Building new FRBR-based displays, by both correcting legacy catalogue data or developing adequate library systems that will be able to manage a "FRBR view" will likely be significant challenges.

The challenge of FRBRizing legacy data and the reality of current catalogue systems were also addressed by Maja Zumer (Zumer, 2005). Zumer stresses that, «to make the transition to FRBR possible, it is necessary to extract the FRBR structure from existing data».

Following this trend, Hegna and Murtooma have detailed their efforts in analyzing MARC records to determine what attributes could best be used for automatic conversion to FRBR. They tried to identify the FRBR structure in the existing MARC records and they came to the positive conclusion that elements of the FRBR model are, to some extent present, in the MARC record, because a bibliographic record may describe both the work and the manifestation, contain traces of the expression and contain some relations in the added entries, notes and subject descriptions. But, paradoxically, the conclusion may be also negative, because the cataloguing rules fit the card catalogue and printed bibliography, and do not fit the FRBR model.

The FRBR model has potential to improve not just the quality of the records but also the user experience when browsing and searching online. One of the major issues many users find when searching a library catalogue is the display of multiple occurrences of a Work. It occurs through multiple records for all of its different Manifestations and also through multiple records for each of those Manifestations' different formats, which typically are not clustered in any sort of meaningful way. In FRBR literature, this problem is often referred to as the multiple versions problem. According to Allgood, "users today have no patience for confusing OPAC displays with multiple hits for equivalent resources", so he lists a number of items that will be available in a "FRBR-aware catalogue", including the ability to "index and retrieve elements or attributes present in both the authority file (i.e., Works and Expressions) and in the bibliographic/holdings file where Manifestation and Item resides". (Allgood, 2007) Rather than changing cataloguing rules, Allgood, observes that FRBR's greater influence may be upon how library management systems designers develop OPACs to cluster the Manifestation-level descriptions into Work and Expression-level displays for users."

After trying to summarize and draw conclusions from these FRBR-related documents, it appears that, according to Edward T. O'Neill, "FRBR has been widely accepted and that some projects succeed in a far better organization of content". But there are some problems that will keep librarians busy for some time: extracting and separating the four Group 1 entities from the single MARC bibliographic record, or identifying, in large databases, those bibliographic records that are instances of the same Work or the same Expression. Thus, it may be taken for granted that there will be some difficulties converting MARC bibliographic databases with respect to FRBR Group 1 entities. Several pieces of the MARC bibliographic record apply to work, expression, manifestation and to item sometimes, but it has been very difficult to define clearly which field or subfield applies to a particular FRBR Group. In spite of all these difficulties, Edward T. O'Neill says that "The doubters have been few and their reservations have generally been limited to particular aspects of the model rather than the rejection of the model as a whole". (O'Neill, 2002) However, in a later paper (Bennett, Lavoie and O'Neill, 2003), it is assumed that more research needs to be done "A key area for further work is the need to transform the conceptual definitions of the FRBR entities into clear, implementation guidelines".

2.2 Major implementations of FRBR

This section describes the first major implementations of FRBR. These systems are either databases designed from the start according to the model, or are based on the result of a FRBRization process.

FRBR-based Library Catalogue Systems are mainly traditional systems converted to new ones conforming to the new FRBR requisites. Some of them are the result of the application of tools created to FRBRize those catalogues or the way the OPACs display the information. FRBR is a conceptual model, rather than a data model, which is why a number of researchers have implemented vastly different FRBRized systems. It is as open to interpretation as it is broadly accepted. Barbara Tillet explained that this entity/relationship model was chosen because it is a general model that can be used for any domain or universe: "... It is perceived as being more easily understood, more stable, and easier to design than a schema conditioned by assumptions pertaining to what constitutes a bibliographic record or by storage and efficiency consideration." (Tillet, 1994)

Following is a brief overview of FRBR implementations related to different types of collections and information environments with a concise description and discussion of the main issues that have been identified for each area.

2.2.1 Worldcat

Among the FRBR-based systems, some are full-scale working systems such as OCLC's WorldCat.org.

WorldCat is the world's largest network of library content and services. WorldCat libraries are dedicated to providing access to their resources on the Web, where most people start their search for information.¹²

Until March 2009, WorldCat contained over 1.2 billion items from. WorldCat uses the OCLC Work-Set Algorithm to FRBRize the collection clustering records by work, expressions and manifestations. The project uses FRBR concepts as work, expression,

¹² <http://www.worldcat.org/whatis/default.jsp>

and manifestation to show how records can be grouped to make OPAC displays more meaningful to the end user.

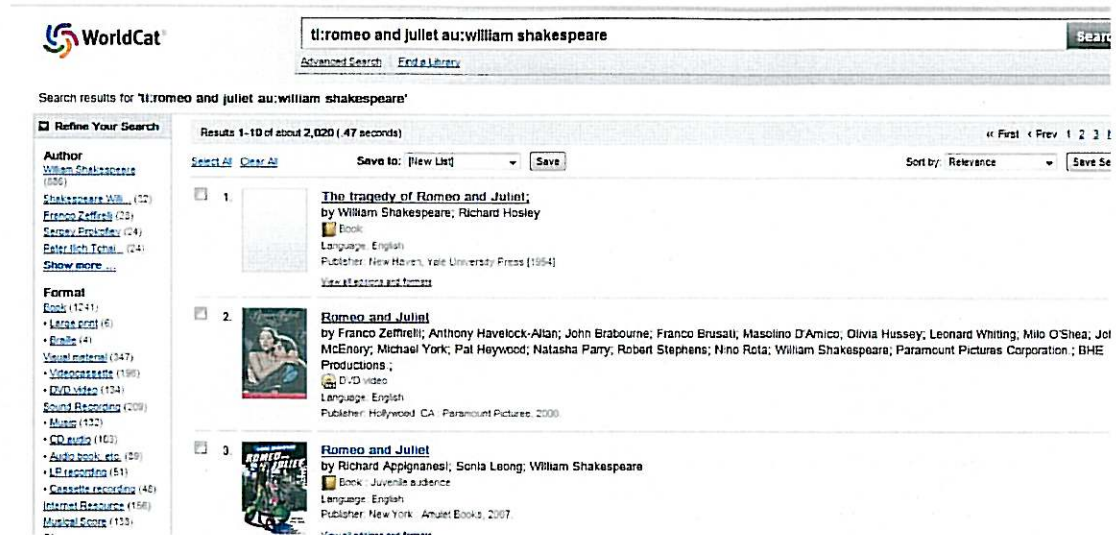


Figure 2 – WorlCat catalogue

The number of editions on the screen is reduced. The screen in Figure 2.1 displays the result of a search for FRBR. We found two works. Below the short description of each work we can read the following text in blue (hiperlink) “View all editions and formats”. If we click here we will have a list all editions and formats available.

Compared to normal online catalogues, this display is useful as it offers users links to collocate and view all editions and formats with facet-based refinement options by specific author, format, year, and language on the left side of the screen.

In WorldCat, the user does not find long lists of bibliographic records with no apparent organization. Records, in this catalogue, are clustered into work, expressions and manifestations and the user can browse the catalogue following a hierarchy according to the FRBR model from work to expression, then to manifestation, then to item. The user is searching the work in the catalogue, but it is the item that he wants to obtain. He or she may want to borrow it or buy it. These two options are given to the user in this catalogue. The library or the bookstore are connected to this catalogue.

The screen in Figure 2.1 shows the result of a search for the work William Shakespeare’s Romeo and Juliet. The first hit is relevant, but it is not a work in the real sense

(the idea, the product of intellectual activity), because it has a specific format, publisher and year attached to it. However, compared to online catalogues this display is useful and has a link to collocate and view all editions and formats together with faceted-based refinement options by author, format, year and language.

WorldCat is one of the best available bibliographic resources. Some students have been using it lately and their satisfaction is great. As it is one of the greatest databases, they find easily what they are looking for, easily get related resources there or on the web. Navigating in this catalogue is a good experience and is paradigmatic of what an OPAC should be. Here the experience how FRBR model can improve the functionality of search and retrieval tools is really good, though there are some issues that need to be addressed. "More research needs to be done, however, to examine the issues and challenges associated with the implementation of the FRBR model in library catalogs. A key area for further work is the need to transform the conceptual definitions of the FRBR entities into clear, implementation guidelines." (Bennett, Lavoie and O'Neill, 2003) The authors recognize that the implementation guidelines would solve part of the problem, as the identification of FRBR entities are ambiguous.

2.2.2 Austlit

AustLit¹³ – The Australian Literature Resource is a cooperative service involving eight universities and the National Library of Australia. It consists of a database designed "from scratch" according to the FRBR conceptual model. It also involves the conversion of existing bibliographic data, but details about it are not available in the literature.

Unfortunately, the AustLit access is not free. It is only available to students and teachers of those universities. So, I could not evaluate this service, I cannot but rely in the literature.

The AustLit data model included the following FRBR group 1 entities: the Work, the Expression and the Manifestation. Item was not included in the catalogue. Item level information is stored in the traditional library. And it added 'event-modelling' to the

¹³ Available at: www.austlit.edu.au

bibliographic description, so Works have a creation event, Expressions have a realization event, and Manifestations have an embodiment event. In the AustLit model, Works, Expressions and Manifestations all have attributes, and Creation, Realisation and Embodiment events all have attributes. The model was also augmented by the integration of the entity of "Super Work" for collecting a number of Works together.

After specifying the above functionality, a FRBR model with a higher entity, the Super Work, it was decided to build a FRBR database, as it was considered that there were no commercial systems that could support the data models or the complex relationship concepts required. All AustLit entities, including events and attributes, are topics, and relationships between those entities are also topics: The AustLit Gateway includes more than 4 million topics. In a year the system was designed and built, and 400.000 records were migrated from twelve different databases to the new one.

Regarding the conversion of data, the major problems had to do with the interpretation of FRBR and the practical implementation, and not with the models chosen. "FRBR was written with a "whole monograph" in mind, this particular application had to expand the model to fit the needs of many other types of works." (Ayes, Fitch and Kilner, 2003) Therefore only a small part of records in AustLit fit this model, as most documents in the AustLit database are individual non-monograph items (individual poems, reviews and articles), representing complex groups of documents such as poem sequences and author series.

"FRBR is somewhat print oriented. Many of its examples are print or recorded sound resources." (Taylor, 2007) There is a new world when we think of art objects, digital objects, archives and other resources of information.

The main audience of the FRBR report is the traditional library environment. FRBR does not offer any guidelines for particular applications or rules for describing resources in a variety of domains and settings.

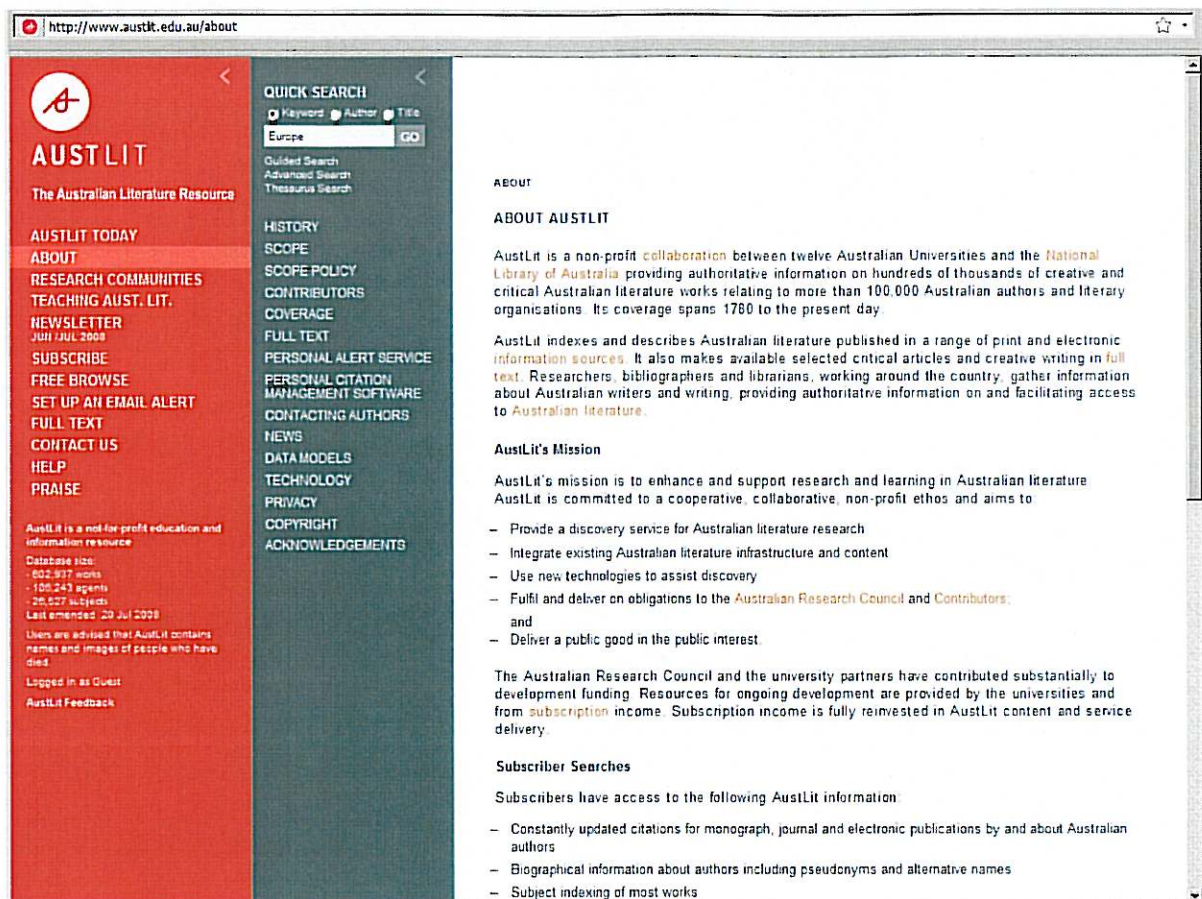


Figure 3 – The AustLit homepage

2.2.3 Perseus Digital Library (PDL)

The main goal of the Perseus Digital Library¹⁴, at Tufts University, Massachusetts, was the creation of a FRBRized catalogue which was able to include the Perseus digital collection and other digital collections of the classical domain from selected libraries. Perseus also became an important FRBR case study since it was able to demonstrate what it could be done using already existing metadata standards and freely available online collections.

«We are seeking to support the four user tasks of the FRBR model, or how to “Search, Identify, Select and Obtain”, rather than to create a FRBR catalogue» (Babeu, 2008).

14 <http://www.perseus.tufts.edu/hopper/>

This project provides FRBRized search through various Manifestations by means of standard Work identifiers to identify the Works contained therein.


The Perseus “FRBR Catalogue 1.0” was an experiment to explore the possibilities of FRBR. The FRBR model was augmented with event modelling, in order to allow for the addition of temporal attributes to the data. The research also introduced the concept of the “Super Work”, a larger class that encompassed twelve different Work types. A new system is under development, “FRBR-Inspired Catalog 2.0”, in a new activity started in 2008. What began as the catalogue for the Perseus collection online has now grown to include a much wider collection of texts. The purpose was to create a FRBRized catalogue of their collection and other selected digital collections from million book libraries (Google Initiative) for the classical domain, Perseus Digital Library also expected to become a case study and to demonstrate what work can be done using existing metadata standards and freely available online collections. Metadata Object Description Schema (MODS)¹⁵ and Metadata Authority Description Standard (MADS)¹⁶ are being used to promote metadata interoperability and records available in a XML format. One of the great challenges of this catalogue is connected with metadata and determining how thousands of XML records that are being created will be stored, indexed, and linked to each other. The Perseus collection has a great number of fragmentary works. They are waiting for a FRBR schema to emerge to reshape data according to that new model. Another benefit of applying FRBR to these types of materials is the emphasis not only on work-to-work relationships, but also on linkages through part-whole and whole-part relationships

The hierarchical MODS/MADS catalogue effectively separates FRBR levels into manageable segments. These segments in turn provide easily updatable and reusable building blocks for further cataloguing and networked catalog reuse. When a new translation is added, the standard identifier for the work is located and specify it together with some publication information.

¹⁵ <http://www.loc.gov/standards/mods/>

¹⁶ <http://www.loc.gov/standards/mads/>

http://www.perseus.tufts.edu/



The Perseus Digital Library

Gregory Crane, Editor-in-Chief, Tufts University

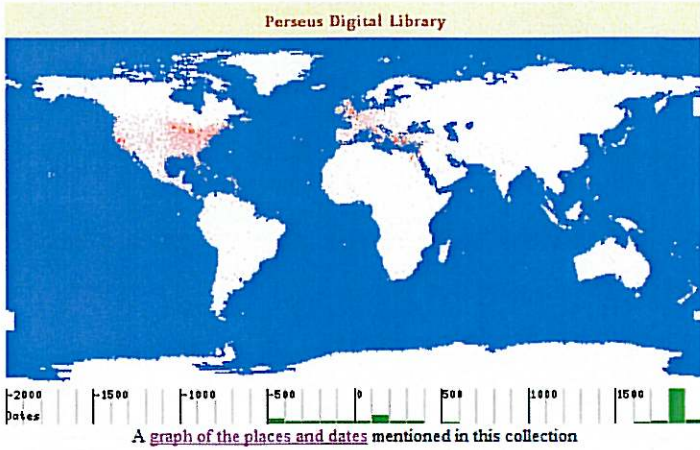
Configure Display · Help Library · Tools · Publications · Collaborations · Support Perseus

PERSEUS IS CHANGING! PLEASE VISIT PERSEUS 4.0 FOR THE LATEST VERSION

Perseus Contents

- [Classics](#)
Greek, Latin,
Archaeology
- [Papyri](#)
Duke Data Bank
- [English](#)
[Renaissance](#)
Shakespeare,
Marlowe, ...
- [London](#)
Bolles Collection
- [California](#)
[Upper Midwest](#)
[Chesapeake](#)
Library of
Congress
- [Tufts History](#)
Since 1852
- [Boyle Papers](#)
History of
Science

Home site:
[Somerville, MA](#)
Mirror sites:
[Berlin, Germany](#)
[Chicago, IL](#)



Announcements

- **Perseus 4.0 released** -- a new implementation of the Perseus Digital Library.

Perseus 4.0, a new Java-based version of the Perseus Digital Library, is available for testing. It contains a faster, more manageable back-end and a more modern look and feel. Many features of Perseus are now available as XML services -- for example, developers can extract [well-formed XML fragments](#) of primary sources with full TEI-conformant markup in order to create their own front ends. [Read more...](#)

- View [all recent announcements...](#)

Exhibits

About Perseus

Perseus is an evolving digital library, engineering interactions through time, space, and language. Our primary goal is to bring a wide range of source materials to as large an audience as possible. We anticipate that greater accessibility to the sources for the study of the humanities will strengthen the quality of questions, lead to new avenues of research, and connect more people through the connection of ideas.

Figure 4 - The PERSEUS Digital Library Project

2.2.4 UCLA Library

The UCLA Film and Television Archive¹⁷ is the second largest collections of media materials in the United States, containing more than 220,000 motion picture and television materials. The OPAC, built in 2007, displays many of the principles of FRBR.

¹⁷ <http://cinema.library.ucla.edu/>

To make the most of this new display, the cataloguing is done according to the new framework, though in a traditional system. In this system, the authority records are Work records, a bibliographic record is an Expression record, and a holdings record is a Manifestation record. A new FRBR based user interface is being worked on, a preview of which can be accessed at the UCLA Library Catalogue already made available.

Currently, one can see the organization of the information of the catalogue according to FRBR Group 1, clustered in Expressions, then in Manifestations. It supports hierarchical browsing, although it requires a lot of steps to go through the hierarchy and interpretation of the results. Two screen captures of the user interface can be seen in Figure .

The screenshot shows the UCLA Library Catalogue interface for the Film & Television Archive. The search results are as follows:

#	Title search	Uniform Title	Full Title
[1]	Blue velvet (Motion picture : 1986)	Blue velvet (Motion picture : 1986)	Blue velvet / De Laurentis Entertainment Group... executive producer, Richard Roth... writer-director, David Lynch.
	<i>Title has multiple holdings</i>		
[2]	Blue velvet (Motion picture : 1986)	Hollywood stars (Television program) Unidentified issue, no 9	Hollywood stars Unidentified issue, no 9.
	<i>Title has multiple holdings</i>		
[3]	Blue velvet (Motion picture : 1986)-trailer		Trailers Collection no. 306 reel no. 149.

Figure 5 – User interface of the UCLA Library – Film and Television Archive

UCLA LIBRARY <i>Catalog</i>		<i>Film & Television Archive</i>		Contact the Archive Research and Study Center		
UCLA Library Collections	New Search	Back to List	Back to Titles	Search History	Help	Exit/Logout

Database Name: Film and Television Archive
 Search Request: Title = blue velvet
 Search Results: Displaying 1 of 5 entries

Blue velvet / De Laurentiis Entertainment Group ; executive producer, Richard Roth ; writer-director, David Lynch. 1986.

Cast: Kyle MacLachlan, Isabella Rossellini, Dennis Hopper, Laura Dern, Hope Lange, Dean Stockwell, Jack Nance, Brad Dourif.

Credits: Director of photography, Frederick Elmes; sound design, Alan Splet; production design, Patricia Norris; edited by Duwayne Dunham; music composed by Angelo Badalamenti.
 Credits supplied from videodisc sleeve.
 Malin gives running time as 120 min.

Genre(s)/Form(s): Features.

Credits heading(s): Lynch, David, 1946- direction

BBID (expression): 8795

Database: Film and Television Archive

Location: NON-CIRCULATING RESEARCH AND STUDY CENTER COPY

Inventory Number: VD201 M

Collection: MP Motion Picture Collection

Format: 2 videodiscs of 2 (laser optical CLV) (121 min.) : sd., Dolby stereo., col. ; 12 in.

Notes: ADDITIONAL PHYSICAL DESCRIPTION: "Wide screen edition." Letterbox format to accommodate wide screen on video. Closed captioned. NOTE Home Video. VD201 M

HLDID (manifestation): 14407

Location: Non-circulating SRLF research copy

Figure 6 – User interface of the UCLA Library – Film and Television Archive

2.3 Products & Technology

As far as it is possible to determine, only two commercial solutions with FRBR support are currently available on the market. These are mainly user interfaces using FRBR principles. This means that the data is kept in conventional bibliographic databases, but are displayed in conformity to FRBR principles.

2.3.1 Virtua ILS (Integrated Library Systems)

VTLS¹⁸ (Visionary Technology in Library Solutions) released in 2002 version 41.0 of the Virtua library system¹⁹. For the very first time, a vendor made it possible for any library to create its own "FRBR catalogue." MARC records can be "split" into the 4

¹⁸ <http://www.vtls.com/products/virtua>

¹⁹ http://www.vtls.com/media/en-US/brochures/vtls_virtua.pdf

levels of the FRBR Group 1 of entities, and any cataloguer can decide to account for bibliographic families rather than isolated documents, thanks to the FRBR structure.

These FRBR records can coexist with old MARC records. The software is “FRBR aware” and automatically displays formats according which record is being accessed. VTLS also offers a software as service (SaaS) for the display according to the FRBR model. Libraries have their catalogues in traditional formats and when they link to this service they have a FRBRized way of displaying their records. “This implementation allows libraries to take advantage of the FRBR model without massive local efforts.” (Zhang and Salaba, 2009) VTLS visited Portugal to participate in the FRBR Workshop and we could see a demonstration of this system. The available documents of this presentation referred the Université Catholique de Louvain as a user of this system. There is no more information available. Figure shows a screenshot from a presentation of the Virtua OPAC.

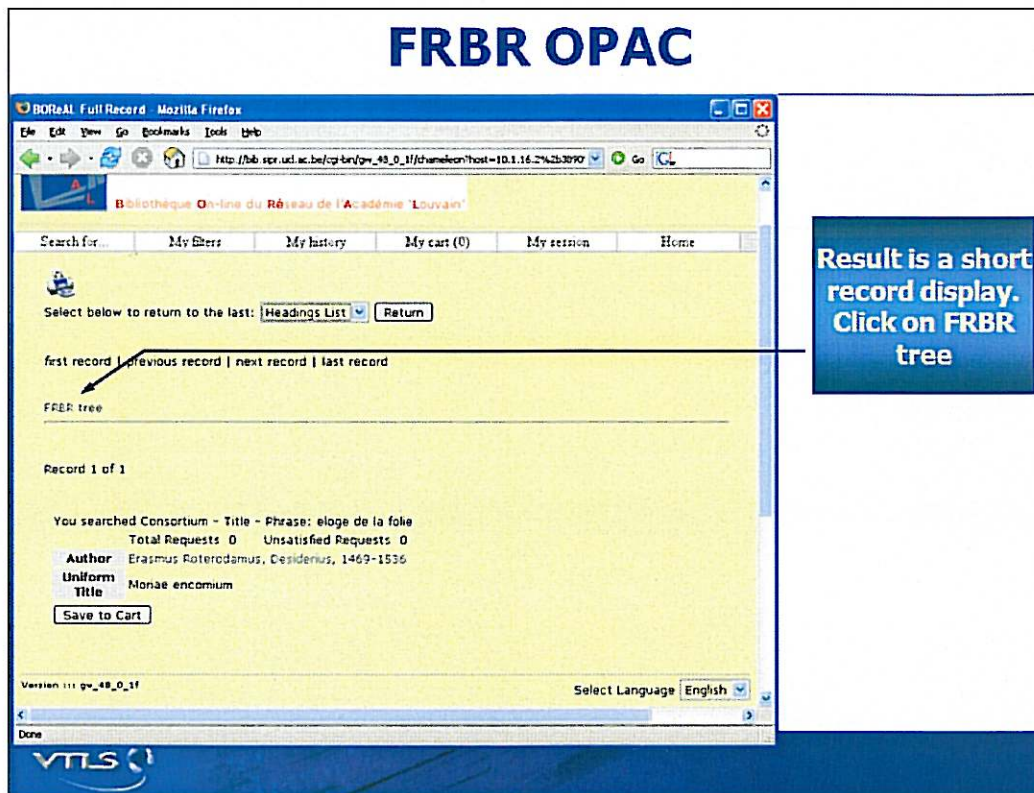


Figure 7 - Virtua OPAC

2.3.2 Aquabrowser

AquaBrowser²⁰ is a product of MediaLab, a private company. Medialab was a little company in Amsterdam (Netherlands) that is bought by Bowker²¹. It can be seen when using the Flemish Central Catalogue where search results are displayed according to FRBR requisites.

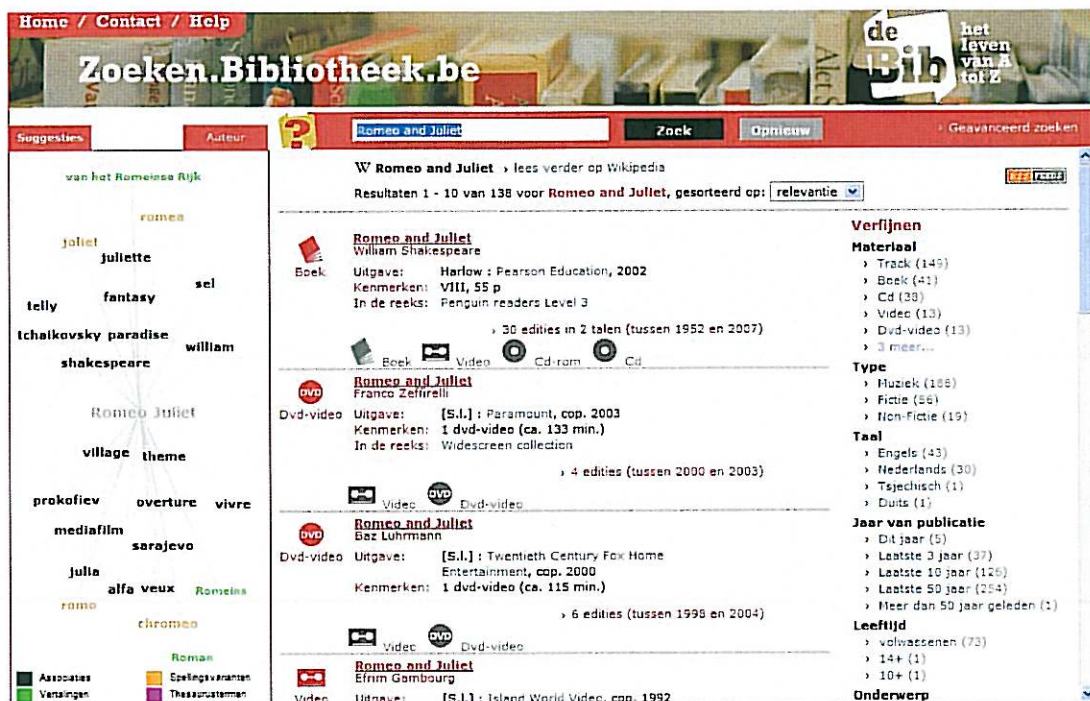


Figure 8 – The AquaBrowser search interface

The FRBR Work key that is produced on the AquaBrowser platform is extracted from the MARC 21²² records, another MARC format. In AquaBrowser the key to FRBRize the catalogue²³ is handled as an index-entry for grouping the search results in the

²⁰ <http://www.aquabrowser.com/>

²¹ <http://www.bowker.com>

²² <http://www.loc.gov/marc/bibliographic/>

²³ <http://www.slideshare.net/rcallewaert/frbr-in-open-vlacc-rosemie-callewaert-503758>

middle of the screen on the result list. The algorithm looks first for a uniform title to combine this string with the main author. If there isn't a uniform title the system looks for an original title. If there is an original title (for a translated work) this title is also combined with the main author. If there isn't an original title the system takes the publication title and combine the publication title with the main author. Cataloguing is also done in a way that allows for the best FRBR display. (Callewaert, 2008)



Figure 9 – FRBR model present on AquaBrowser search interface

2.4 Research: Supporting Tools, Algorithms and Techniques

This section presents some relevant research activities that lead to the development of tools for FRBRization, or prototypes of FRBR OPACs.

Software tools and algorithms were developed for automatic processing, it is therefore important to know how if they can achieve fully what we expect them to do. However, it is not easy to find evaluation work.

2.4.1 OCLC FRBR Work-Set Algorithm

The FRBR Work-Set Algorithm²⁴ was developed by a OCLC research group to examine the issues associated with the automatic FRBRization process. It is an algorithm for clustering MARC21 bibliographic records at the FRBR ‘Work’ level. It uses both authority records and bibliographic records. Having resources brought together under the work entity, that enables users to navigate through the resources available. The work-set algorithm is an automated method of identifying all manifestations of a work that exists in a bibliographic database. It does not attempt to identify expressions or to associate manifestations with expressions. It is based on the simple concept that all manifestations of a given work share a common author and title; therefore a combination of the author and of the title should be enough to identify all manifestations.

“The research work-set algorithm generates an author/title key for each bibliographic record. These keys can then be used to bring work-sets together. The current algorithm ignores format so that the generated work-sets are sometimes at a higher level than a FRBR work.” (Hickey and Toves, 2005)

The algorithm is available for anyone to download and experiment with and we have identified several projects that apply it. It is not a software that is ready to use, but the algorithm can be implemented using a programming language.

The algorithm has been used in many FRBR system development projects, including OCLC FictionFinder and Worldcat.org and is considered one of the more successful “FRBRization” applications. More recently, Libraries Australia’s FRBR prototype system and Kent State University’s FRBR project have adapted this algorithm for their respective FRBR implementations.

“The OCLC FRBR Work-Set Algorithm has been commonly used in many FRBR system development project; however, there have not been investigations into its effectiveness.” (Carlyle, Ranger, and Sumlnerlin, 2008)

²⁴ <http://oclc.org/research/activities/past/orprojects/frbralgorithm/default.htm>

2.4.2 FRBR Display Tool

In 2002, the Library of Congress, Network Development and MARC Standards Office²⁵, commissioned a study to examine the MARC 21 bibliographic and holdings formats against the FRBR. The results of the study²⁶ have been improved over the last years, and the last revision dates from 2006. It includes a proposal to map MARC 21 elements to FRBR, which also motivated the development of the FRBR Display Tool.

The FRBR Display Tool is coded in XSLT. It can take a group of MARC bibliographic records and cluster them into meaningful displays of Works, Manifestations, and Expressions.

The FRBR Display Tool (version 2.0) can also be downloaded freely. It helps analyze MARC data through the work, expression, manifestation, and item entities of the FRBR model and creates a display in conformity with the model.

This tool allows libraries to experiment FRBR without changing catalogue practices and provides end users with better collocating retrieved records by grouping them according to FRBR entities.

Like other tools that use MARC records, the effectiveness of the FRBR Display Tool depends, to a great extent, on the quality of the data. Errors and inconsistent cataloguing practices can be obstacles. As referred in the project Website, the tool has some limitations: it works best with name and title fields and broader searches and there are other issues that should be analyzed before being used.

2.4.3 FRBR Floater

FRBR Floater²⁷ is an innovative new service that enables users to view, in an easy-to-read OPAC window, the various editions and formats owned by the library of any title searched. The user may then simply browse the list and select the one item that is

²⁵ <http://loc.gov/marc/marc-functional-analysis/tool.html>

²⁶ <http://www.loc.gov/marc/marc-functional-analysis/functional-analysis.html>

²⁷ <http://montesanoassociates.com/apps-msafrbr.htm>

most appropriate. It was developed by Monte Sano Associates and if someone wants to use it must pay for it. It is an on-line service.

“Libraries need not re-catalog their collections or manipulate their bibliographic databases, because we use a sophisticated algorithm, based on the new international FRBR standard, to harvest the needed data from your existing MARC catalog records. As a result, libraries and their users can enjoy the benefits of FRBR without the time and expense of database analysis and re-design.” This is what we can read in their homepage.

2.4.4 BibSys FRBR conversion tool

Trond Aalberg from the Norwegian University of Science and Technology, Trondheim, Norway, has researched the creation of tools for converting MARC data to the FRBR model. His basic process involved identifying the different entities in a MARC record, selecting the fields that describe each entity, and finding relationships between entities. The conversion tool he designed has as its input MARC records in the MarcXchange XML format, and outputs “a record for each entity in a format that extends the MarcXchange format with FRBR type attributes and a relationship element.” (Aalberg, 2006) The tool was applied to 4.000.000 records in the BibSys database, and the author has learned during this conversion process that the major issues faced were inconsistent cataloguing data and scalability issues. While the author concedes that it is as yet impractical for libraries to attempt full conversion to a FRBR based data model, he believes that conversion tools such as this one can at least help create FRBRized views for catalogues.

It is recognized that this conversion tool for BYBSIS can be applied to convert MARC catalogues to FRBR entities at all levels. However there were not enough tests. The tools have however been used to convert 4.000.000 records in the BIBSYS bibliographic database “The developers did caution that, at this stage, a full conversion to a FRBR-based data model may not be realistic or desirable for most libraries due to the limited knowledge about the application aspect of the model and the lack of standardized formats for FRBR.” (Aalberg, Haugen, and Husby, 2006)

2.4.5 OCLC FRBR prototypes

Some of the earliest experiments with FRBRization were conducted by members of OCLC Research group²⁸, who in a great number of articles have related all these experiments with algorithms to group bibliographic records in WorldCat into Works and Expressions (Bennett, Lavoie and O'Neill 2003).

In the next sections the most important OCLC FRBR prototypes are described:

2.4.5.1 OCLC FictionFinder

FictionFinder²⁹ is a prototype that reflects the organization of FRBR according to FRBR Group 1 (work, expression, and manifestation) but giving a focus on work. It brings together expressions and manifestations of the same work for 2.5 million bibliographic records for fiction. The records are clustered at the work level using the OCLC FRBR Work-Set Algorithm.

The Expression-level grouping is by language and the groupings make use of existing bibliographic, authority, and holdings information.

It provides a FRBR-inspired view of the data. (

Figure) Here we were searching the work "Romeo and Juliet" and we obtained a result clustered on the work "Romeo and Juliet", with 249 expressions.

²⁸ <http://oclc.org/research/projects/frbr/default.htm>

²⁹ <http://fictionfinder.oclc.org/>

Browse Search GO [Advanced]

You searched: Basic Index for romeo juliet

Back to Results

<< Previous Work 1 of 69 Next >

Find A



Romeo and Juliet.
Shakespeare, William, 1564-1616

249 editions, in 25 languages, held by 6781 libraries

Summary: A stunning performance of Shakespeare's timeless tragedy of young love.

Genres: Tragedies | Love stories

Characters: Juliet (Fictitious character) | Romeo (Fictitious character)

Settings: Italy —Verona [+] | 1500 - 1599 | 1500 - 1600 | 1800 - 1899 | Italy | 1600 - 1799 | To 1600

Subjects: Vendetta | Youth | Family | English drama (Tragedy) | Love | Conflict of generations | English language | Married people | Man-woman relationships | Motion picture music —Excerpts [+] | Motion pictures | [More]

Audience: Kids General Special

Editions		Genres	Characters	Settings	Subjects	
		Narrow by Languages: All (249)		Narrow by Format: Show All		
Title / Author	OCLC #	Date	Language	Format	Aud.	Libra
1. Romeo y Julieta William Shakespeare	55878413	2002	Spanish	eBooks	General	
2. The tragedy of Romeo and Juliet / by William Shakespeare ; [general editor,	01600668	1959	English	Book	General	

Figure 10 – FictionFinder

Records are clustered into works using the FRBR Work-Set Algorithm. The algorithm collects manifestation records into groups based on author and title information from bibliographic and authority records. Author names and titles are normalized according to the NACO Authority File Comparison Rules to construct a work key (e.g., "carroll, lewis\1832 1898/Alice's adventures in wonderland" is the key for Lewis Carroll's Alice in Wonderland). All records with the same key are grouped together in a work set or cluster.

Pisanski and Zumer analyzed this prototype and explained that it does not follow the FRBR model, partially due to issues with both the FRBR model and cataloguing data, but also emphasized that: «It has to be pointed out that FRBR is neither a standard nor a data model. In other words, FRBR in no way implies what the implementation should be like». (Pisanski and Zumer, 2007)

The OCLC Fiction Finder prototype uses normalized title/author as the key for clustering records, and works were ranked according to libraries that owned them. One major issue the authors had with the FictionFinder, was that it focused too exclusively

on the concept of the work with no easy ability to sort results by Manifestation level data, such as a specific publisher or illustrator name.

2.4.5.2 The FRBRization of Humphry Clinker

The FRBRization of the work *The Expedition of Humphry Clinker*, or *Humphrey Clinker*³⁰, is a case study in the application of FRBR to an old catalogue based on an entity-relationship model. “Humphry Clinker was chosen for several reasons, it has been previously studied, it is widely held and it is a work of mid-level complexity.” (O’Neill, 2002) O’Neill also declared that the goal of this study was focused on:

- examining the benefits and drawbacks associated with creating an entity-relationship model for a work
- better understanding the relationship between bibliographic records and the bibliographic objects they represent,
- determining if information available in bibliographic records is sufficient to reliably identify the FRBR entities,
- developing a data set that can be used to compare and evaluate FRBRization algorithms.

The team working in the project first manually extracted from WorldCat a set of records representing the monograph “Humphrey Clinker”, and analyzed the ability of their algorithm to discover expressions within this set of records. They learned that their algorithm was able to identify 28 expressions in the set versus the 41 located through manual inspection, and was able to pull out 10 of 11 identified manifestations. Ultimately, they decided that due to the difficulty of algorithmically identifying expressions, their future work for the time being would concentrate on the identification of works rather than expressions.

According to the researcher “The goal of this study was to go beyond organizing bibliographic records to organizing the bibliographic objects represented by bibliographic records. This effort was focused on: examining the benefits and drawbacks associated

³⁰ <http://www.oclc.org/research/activities/past/orprojects/frbr/clinker/default.htm>

with creating an entity relationship model for a work, a better understanding of the relationship between bibliographic records and the bibliographic objects they represent, determining if information available in bibliographic records is sufficient to reliably identify the FRBR entities, and develop a data set that can be used to compare and evaluate FRBRization algorithms". (O'Neill, Edward T. 2002) The test bed was WorldCat, where all possible Humphrey Clinker records were searched.

This case study proved that the FRBR notion of work is a valuable concept, providing the means to aggregate bibliographic units and simplifying database organization and retrieval. Works can reliably be identified from bibliographic records. Identifying expressions, however, is far more problematic. When any modifications to a work are considered to be new expressions, the granularity of resulting expressions is too fine and can be almost indistinguishable from that of manifestations. Like work, expression is abstract. The Authors conclude this study saying: "The irony is that the FRBR model provides minimal benefits to the small works that can be reliably FRBRized, but fails on the large and complex works where it is most needed." (O'Neill, Edward T. 2002)

We can draw the conclusion that FRBR exposes errors and cataloguing inconsistencies that were hidden. The study of 179 bibliographic records in OCLC's WorldCat for different manifestations of Tobias Smollet's *The Expedition of Humphry Clinker* proved that inconsistencies in the bibliographic records were a serious impediment to identifying expressions. "The FRBRization challenge is to find an algorithm that is effective with less than perfect data." (O'Neill, 2007)

2.4.5.3 The NDB Prototype

The NDB Prototype³¹ is a simple demonstration of searching MARC bibliographic records. The database being searched is a copy of the Australian National Bibliographic Database (NBD) from March 2008. It contains 16 million bibliographic records with holdings information for Australian libraries. The same data (more up to date) is also publicly accessible through Libraries Australia. Pisanski and Zumer also

³¹ <http://librariesaustralia.nla.gov.au/apps/kss>

analyzed this prototype and explained that it “was not limited to just books but included movies and other materials and it grouped FRBR data at various levels, included a new entity called “Superwork” where top level records were grouped together, and used form and language attributes to differentiate between numerous works and expressions.” (Pisanski and Zumer, 2007) But the authors further explain that “algorithms for eliciting FRBR structure” will only work as well as the bibliographic records on which they are based.

The demonstrator extracts topics and relationships from records retrieved from a simple full text search to present search results. Related records are grouped into an FRBR-like structure. Figure 6 shows an example for a structure of results for the “Superwork” Hamlet.

The results of this prototype have been discussed by several authors stating that it is an example of the difficulties in extracting FRBR structure from MARC records, and the uncertainty on the boundaries of expressions.

BIBSYS Ask - Treffliste Logg på

[Søkekilder] [Søk] [Treffliste] [Dokument] [Samlekurv] [Søkehistorikk] [Min side] [?] [] []

Dine valgte søkekilder er endret

Bibliotekbasen: **847** Emneportal: **0** ForskDok publikasjoner: **9** ForskDok prosjekter: **0**

Resultat av søket: *hamlet*

Viser: 1 - 10 1 2 3 4 5 6 7 8 9 10 Totalt antall: 847

<input type="checkbox"/> Tittel	Forfatter	Ar	Materiale
<input type="checkbox"/> Hamlet	Shakespeare, William	2008	
<input type="checkbox"/> Hamlet / edited and with an introduction by Harold Bloom ; volume editor Brett Foster	Bloom, Harold Foster, Brett	2008	
<input type="checkbox"/> Hamlet : av William Shakespeare / programredaksjon: Hanne Tømta ... [et al.] ; Mine Seren (red.)	Shakespeare, William Tømta, Hanne 1968- Seren, Mine 1971-	2008	
<input type="checkbox"/> Hamlet eller Hamilton? : litteraturvetenskapens problem och möjligheter	Forslid, Torbjörn Ohlsson, Anders	2007	
<input type="checkbox"/> Hamlet. fantasy-overture	Čajkovskij, P.	2007	
<input type="checkbox"/> William Shakespeare's Hamlet / en film av Kenneth Branagh	Branagh, Kenneth Christie, Julie Crystal, Billy..m.fl..	2007	
<input type="checkbox"/> Hamlet : William Shakespeare	Sundvor, Yngve Shakespeare, William	2007	
<input type="checkbox"/> Hamlet : av William Shakespeare : studiemateriell : for musikk, dans og drama	Shakespeare, William	2007	

Figure 12 - Interface of the NDB prototype

This project was created to investigate the possibilities and techniques for applying the FRBR model to old MARC catalogues. Records of the BYSIS bibliographic catalogue have been successfully FRBRized through the development of an XML-based tool that extracts relevant FRBR entity information from MARC records. The project also produced a software tool that can automatically interpret and extract existing MARC record information based on the FRBR model and can therefore be used to FRBRize a MARC-based bibliographic catalogue.

The project had the following goals:

- **Analyze hit lists:** Investigating the possibility to construct an algorithm for analyzing a set of hits from an OPAC search in the light of FRBR. That is, trying to identify Works, Expressions, and Manifestations on the basis of the MARC data of the records in the hit list; identify problem areas in this regard; the investigation starts with hit lists based on author search and then moves on to title and subject search.; the algorithm must be able to cope with different

MARC formats; the test data of the project is gathered from both the Finnish national bibliography and from the Norwegian national bibliography.

- **Identify a Work level:** based on the results from the previous step, try to recognize the problems which arise when trying to identify a Work level in the catalogue as a whole.

The results show that though the information in the MARC records holds attributes relevant for identifying the Work, Expression and Manifestation entities, the accuracy and formal syntax are too simple to be properly handled by programs. Some of the results may be used to present better hit lists in OPACs.

- **OPAC design:** using the possible results from the previous steps and on the basis of IFLA's Guidelines for OPAC displays, try to design a new kind of user interface. The user interface should be multilingual (English, Finnish and Norwegian) and handle different MARC formats (MARC21, FINMARC, NORMARC).

Hegna and Murtomaa have detailed their efforts in analyzing MARC records to determine what attributes could best be used for automatic conversion to FRBR. They determined that even though MARC records held attributes that could be used to identify Work, Expression and Manifestation entities, inconsistencies in cataloguing and other errors would prove a great difficulty to automated processing. (Hegna and Murtomaa, 2003)

2.4.6 An important lesson learned

After ten years of exploration and development, a better understanding of FRBR is still needed. We have found two main types of problems. In the first place, the lack of guidelines to help implementation of the FRBR model. Many authors have agreed that the model is too theoretical and open to interpretations. FRBR entities are difficult to identify, expression entity being the most difficult of all. The frontier between work and expression is subjective and is dependent on different cultural backgrounds. In the second place, errors and inconsistencies in cataloguing make things worse. Automatic conversion of old MARC catalogues relies on perfect data. That is why "in national

bibliographies the results are better than in the individual catalogues, therefore FRBRized results can potentially be better than those from catalogues.” (Pisanski, Zumer and Aalberg, 2009) An example of this is the Slovenian National Bibliography that consistently enter the name of the authors and contributors with relator codes (relator codes indicate the role the contributors have in the work) and these are highly useful to identify expressions of a work. That is why some authors conclude that the FRBRization challenge is to find an algorithm that is effective with less perfect data.

3 Detailed description of the problem

The first card catalogues had a filing organization for the works of prolific authors that helped the users find everything in that library that belonged to a particular author or that was related with it.

The organization of that card catalogue was as follows:

- Author
 - titles of work filed alphabetically behind guide card for author's name
 - each work might have a group of cards (different editions, translations, etc.)

Within a group of cards for a work you often found the following arrangement:

- original publication
- other editions, other formats
- translations (alphabetical by language)
- works in which this work counts as subject (reviews, criticisms, etc.)
- related works:
- movies based on the original book
- songs from musical based on the book
- adaptations
- commented texts, etc.

This was the scenario we had in the library twenty years ago when we had the manual card catalogues. We could browse the card files and navigate from a work to its re-

spective translations or adaptations and see related works. Though it was a card file you could find structure there and there were card files (SEE ALSO) that guided you to other related works or subjects.

Nowadays, on-line catalogues display long lists of records with no structure at all. Users get confused with these lists with the absence of a hierarchical display. Automation failed to produce the structure of the card catalogue.

Here it is a long list of 186 records when a user searches the work “Os Maias” by Eça de Queirós in PORBASE catalogue. The user has to scroll down this long list without a good understanding of what is being shown. Then one realizes that what was sent to the screen is everything existing in that catalogue and that corresponded to the search formula, but one could see that there was no evident previous work done there by the library to organize that information or to help the user exploring those resources.

1. Os Maias : episódios da vida romântica / Eça de Queiroz. Porto : Lello & Irmão, 1950.

2. Os Maias / Eça de Queirós. Lisboa : Círculo de Leitores, 1956.

3. Os Maias : episódios da vida romântica / Eça de Queirós ; introd. Esther de Lemos. 2a ed. Lisboa : Ulisseia, 1986.

4. Os Maias ; Novelas / Eça de Queirós. Moscovo : Khulozestvennaja Literatura, 1956.

5. Os Maias / adaptação teatral do original de Eça de Queirós por José Bruno Carneiro ; pref. J. Almeida Pavão ; estudo de Carlos Reis. Lisboa : Imp. Naz.-Casa da Moeda, 1964.

6. Os Maias : episódios da vida romântica / Eça de Queirós. 3a ed. Mem Martins : Europa-América, 1987.

7. Os Maias / Eça de Queiroz. 6a ed. Lisboa : Círculo de Leitores, 1962.

Figure 13 – PORBASE search result list

How good it would be if the user had received the information structured into clusters, allowing the navigation according to his interests, first the work, then the expres-

sions and then the manifestations or simply follow a link to related works or to other works of the same author or still linking to other catalogues. A system like this would create an environment of networked resources where navigation would be easy and discovery of resources enhanced.

“Whatever enabled exploitative power was efficacious; whatever obfuscated exploitative power, and this was most of the bibliographical apparatus, was not efficacious.”³³ (Wilson, 1968)

Actually library automation brought some problems in the way information was displayed in the OPACs and created, using the words of Patrick Wilson, “a bibliographical apparatus that was not efficacious.” The structure of printed catalogues was hard to implement and automation brought instead long lists of unstructured and unrelated bibliographic data. “Technological advance has brought with it a steady deterioration in the integrity of bibliographic structures since the time of Panizzi and, with it, an undermining of bibliographic objectives.” (Svenious 2000, 64)

In a current’s online catalogue, PORBASE, as we have just seen an example, when the user looks for a work and obtains hundreds of bibliographic records, what does he do with a long list with no hierarchical structure? When he browse all those records he is confused. Which record to select?

If the user had a FRBRized set of records he would have a small list that would fit any computer screen (no need to scroll a long list of unstructured records) and with an organization that would be clear to him, more or less, the work on the top, then a cluster on expressions (editions, translations, etc.) and he would browse the second group, till he found the manifestation he was looking for and would hurry to the library that hold this item available to borrow it.

When searching a public online catalogue users find multi occurrences of a work, not only through multiple records for all its different manifestations, but also multiple records for each of those manifestation different formats, without being clustered in a meaningful way to the user. “Libraries’ earliest experience with the proliferation of copies of resources in different physical formats was with the reproduction of printed materials, first in micro formats, then in digital formats. Library cataloguing rules re-

³³ <http://www.universityofcalifornia.edu/senate/inmemoriam/PatrickWilson.htm>

quired each new iteration in a different format to have its own way entry in the catalogue. Although seemingly efficient in allowing virtual “cloning” of catalogue information from one version to another, in the end this practice proved to have a very negative impact on the usability of the catalogue, causing an increase in catalogue entries for what to many users is essentially the same resource.” (Coyle and Hillmann, 2007)

This issue could be solved if one moved from manifestation level cataloguing to expression level cataloguing. This is a challenge for a FRBR catalogue and a reason why new cataloguing rules based on the new model are needed.

Cataloguing has been performed at the manifestation level, because we all have learnt that we should catalogue the book in hand. Jennifer Bowen presented a work³⁴ to show us the attempts of the Anglo American Cataloguing Rules³⁵ (AACR)’s Format Variation Working Group (FVWG) to determine if expression level cataloguing is practically possible to achieve and even suggested that this could be done using collocation tools. The FVWG not only decided that libraries would continue the practice of cataloguing manifestations, but that they also “reaffirmed the need to provide access to expressions within catalogues, but recommended that this be achieved instead through an exploration of expression-based collocation, rather than expression-based cataloguing” (Bowen, 2005).

Julian Everett Allgood, a cataloguer at New York University Libraries, wrote an article called “Serials and Multiple Versions, or the Inexorable Trend Toward Work-Level Displays,” in the new *Library Resources & Technical Services* (July 2007, 51:3). You can read in the abstract: “The proliferation of multiple versions for bibliographic works presents numerous challenges to the catalogue and, by extension, to the cataloguer. Fifteen years after the Multiple Versions Forum held in Airlie, Virginia, online public access catalog (OPAC) users continue to grapple with confusing displays representing numerous serial manifestations (i.e., versions) resulting from the Anglo-American Cataloguing Rules’ (AACR2) cardinal principle (Rule 0.24). Two initiatives offer hope for more coherent OPAC displays in light of a renewed focus upon user needs: the ongoing revision of AACR2, and the International Federation of

³⁴ Available at <http://www.libraries.psu.edu/tas/jca/ccda/docs/rda0506.pdf>

³⁵ <http://www.aacr2.org/>

Library Associations and Institutions' Functional Requirements for Bibliographic Records (FRBR) model. A third potential tool for improving OPAC displays exists within a series of standards that have developed to parallel library needs, and today offer a robust communications medium: the MARC 21 authority, bibliographic, and holdings formats." This paper summarizes the challenges posed by multiple versions and presents an analysis of current and emerging solutions.

Karen Koyle's article "Future considerations: The Functional Library Systems Record" offers us an assessment of the issues involved in how FRBR might improve library records and catalog systems and give us some examples of how FRBRized catalogs may look like." Coyle says that, "Unfortunately, the bibliographic view of what is functional and the system views of functional are not being discussed in concert. Bringing these two reform movements together would be a better formula for success than either of them would have its own" (Coyle 2004). Indeed many analyzes of FRBR as a model often include little examination of how such models might be implemented in an actual library system. Another essential point Coyle makes is that FRBR shifts the focus away from the physical descriptions of individual publications to a concentration on their textual content and the relationships between them. The FRBR model, Coyle contends will move us "toward a view of a universe of inter-linked publications where users eventually will not need to be concerned with differences in formats or the vagaries of nearly identical printings of the same works."

Many library catalogues today do not support all the functions of the catalogue as recommended over a century ago by Cutter and, in the last decade by Lubetzky. Catalogues fail to collocate all versions of the same work and the decision to ignore bibliographic relationships between different works and among expressions and manifestations of the same work. Lubetzky said: "An adequate catalogue, concerned about the actual needs of a reader, must be designed to tell one not only whether the particular book he or she seeks is in the library but also what other editions of the work and what other works of the author the library has. That was the object of Panizzi's rules."³⁶

³⁶ Seymour Lubetzky, "The Fundamentals of Bibliographic Cataloguing and AACR2," in Svenonius and McGarry, p. 370.

Patrick LeBoeuf explains that "the steady deterioration that Elaine Svenonius denounces consists in a loss of bibliographic structure, the absence of a hierarchical display of entries, the lack of sense for the context in which each new translation, edition, and adaptation of a work inevitably takes place. The wonderful syndetic structure of printed catalogues has yielded to databases that are barely more than collections of unrelated monads." (LeBoeuf, 2005)

Even in the Report of the Library of Congress Working Group on the Future of Bibliographic Control, January 2008, "On the Record", one can read "Users are making new demands on metadata. Thanks to rapid innovations in Web technology and to the ubiquity and utility of Web search engines, most users now conduct their research in multiple discovery environments: search engines, online booksellers, course management systems, specialized databases, library catalogues, and more."

Library users got used to work in this networked environment, where they have access to friendly web search engines and became more demanding. Library catalogues cannot be competitive with Google and other search engines. Librarians and their professional communities and associations, like IFLA, realized that an effort was needed to be done to make the catalogue become a more useful tool and more user-friendly. Librarians realized that economic pressures made libraries simplify the cataloguing practices and do more "minimal level" to keep pace with the continuing growth of publishing activity and that this caused great damages to the catalogue and that this would make more difficult to respond more effectively to an increasingly broad range of user needs. It was this changing environment and new user challenges that formed the backdrop to the 1990 Stockholm seminar on bibliographic records, sponsored by IFLA. In this seminar it was recognized the need of a standard for a "basic" or "core" level. There were nine resolutions adopted in this seminar, one of which led to the FRBR model approved in 1998. "The model developed for this study represents an initial attempt to establish a logical framework to assist in the understanding and further development of conventions for bibliographic description. It is intended to provide a base for common understanding and further dialogue, but it does not presume to be the last word on the issues it addresses." (FRBR Report, 1998)

Yee affirms that OPACs are not truly catalogues, but instead serve as online finding lists. The future is dark, she adds "The rest of the world has become enamored of Google. Google cannot carry out the objectives of the catalog either. But if our choice

is between OPACs, which are expensive but cannot carry out the objectives of the catalog, and Google, which is cheap and cannot carry out the objectives of the catalog, I know what the choice is likely to be.” (Yee, 2005)

How can FRBR contribute to a better catalogue?

The lesson that we have already learned and that we cannot forget is that FRBR model is expected to be more beneficial when applied to big catalogues and to certain types of resources than to others and within a collection those that may benefit the most are those consisting of works expressed in different ways, different editions, editions by different publishers, translations and still those published in different mediums. Among these are fiction, music, serials and aggregate works with many expressions and manifestations. There is no great advantage for works with one expression and a few manifestations.

However, FRBR is librarians’ last hope. “The FRBR model could therefore be seen as a (last chance?) attempt at restoring both bibliographic structures and bibliographic objectives, at regaining the Paradise we have lost by computerising our catalogues, and at giving the correct answer to the Sphinx who keeps facing us with the Digital Era Dilemma: “Bibliographic Control, or Chaos?”” (LeBoeuf, 2005).

The current catalogues are the problem. What can be done?

The solution to this problem could be the combination of three things: “We need to accept the principles and the entity-attribute-relationship model espoused by the Functional Requirements for Bibliographic Records (FRBR). We need to construct rules for creating cataloguing and other metadata based on this model. And we need to design systems that will display our metadata using the model conceptualized in FRBR.” (Taylor, 2007)

This thesis evolves according to those principles and entity-attribute-relationship model adopted by FRBR and gives contributions to the development of a FRBRized search interface. In the context of this work, rules for creating cataloguing and other metadata or the design of new systems will not be explored.

The goal of this thesis is to give a contribution to the process of FRBRization and to the development of a search interface, considering that it is the only way now available to display information that is stored in the old MARC format in a structured way that is meaningful to the end user.

4 How and why FRBR can help solve the problem

4.1 Introduction and rationale for the research

Current catalogues are a legacy that concerns us. For about fifty years that libraries have been creating records in MARC³⁷ format. These records created in MARC cannot be forgotten, and there are some experiments done to find a way to convert their visualizations into new FRBR displays. That is what is called FRBRization. More precisely, FRBRization means the conversion of an old catalogue to a new way of displaying the information, by means of identifying and using the entities present in the MARC format to cluster the information according to the FRBR entities (Group 1): work, expression and manifestation.

We tried to learn with some authors that had already tried to FRBRize catalogues or portions of catalogues. FRBR was applied to small samples, then to more relevant sets of records and finally to entire catalogues. An example of this is the application of FRBR to WorldCat, which proved to be a good testbed, the world's largest union catalogue, where it was demonstrated in the context of several experiments that library catalogues may benefit from this application. One of the first O'Neill's experiments was the application of FRBR model to Humphry Clinker (O'Neill, 2002) to determine benefits and drawbacks with the creation of such an entity-relationship model. One of the conclusions of this experiment it was that while the identification of works and manifestations was easy, identification of expressions was problematic.

In many ways, FRBR exposes errors and inconsistent cataloguing practices that before were unnoticed. For instance, in studying the 179 bibliographic records in

³⁷ MARC is an acronym that stands for Machine Readable Cataloguing

OCLC's WorldCat for different manifestations of Tobias Smollett's *The Expedition of Humphrey Clinker*, O'Neill found that inconsistencies in the bibliographic records were a serious impediment to identifying expressions. There were different names for Smollett as a main entry, and a number of bibliographic records did not include the *Expedition of Humphry Clinker* either in the title statement or as uniform title. The variation in title and main entry makes it extremely difficult to algorithmically identify all the manifestations of a work. Many manifestations were identified by the algorithm complemented with manual work and with an intensive manual review.

"OCLC is investigating how best to implement IFLA's Functional Requirements for Bibliographic Records (FRBR). As part of that work, we have undertaken a series of experiments with algorithms to group existing bibliographic records into works and expressions. Working with both subsets of records and the whole WorldCat database, the algorithm we developed achieved reasonable success identifying all manifestations of a work." (Hickey, O'Neill and Toves, 2002)

Though the results were good, the algorithm made some errors on complex works. This experiment suggested further study to understand the magnitude and consequences of the errors. Another experiment explores the concept of work in WorldCat, using the hierarchy of bibliographic entities defined in FRBR model. A methodology is described for constructing a sample of works by applying the model to randomly selected WorldCat records. This sample is used to estimate the number of works in WorldCat and describe their characteristics. The conclusion was that FRBR has advantages if applied to a small part of the catalogue, those works with two or more manifestations (20% of the works in the catalogue would be candidates). "A random sample of 1.000 bibliographic records was selected from WorldCat... The list of candidate records for each work in the sample was then reviewed manually, and these records were supplemented by ad hoc manual searching using OCLC's First Search to investigate other variations in authors or titles ..." (Bennett, Lavoie and O'Neill, 2003)

Applying FRBR to catalogues may be easier if we have algorithms available to help us in the process of FRBRizing. Hickey published the FRBR Work-Set algorithm he created and used in all these experiments (Hickey, Toves, 2005)

Martha Yee (Yee, 2005) made an important review of how current MARC records can be used to promote FRBRization. Yee thinks that the most productive line of FRBR research may be to scrutinize how the attributes in already existing MARC bibliographic, authority and holdings records may be used to enable FRBRization of OPACS. In order to support her contention that much of the information needed to FRBRize catalogues is already present in MARC record, Yee gives a comprehensive list of MARC fields that can serve as identifiers for works and expressions in both bibliographic and authority records. The solution to the problem, according to Yee's opinion, is to "find and educate system designers who can grasp the fact that the complexity of our records is a direct result of the inherent complexity of the bibliographic universe." (Yee, 2005)

Maja Zumer also worked on these issues «to make the transition of the FRBR possible, it is necessary to extract the FRBR structure from existing data.» (Zumer, 2007) Zumer acknowledges that this will be a difficult task, considering the errors and inconsistent cataloguing practices in legacy bibliographic data and stresses the fact that a great deal of important information within the catalogue exists only in an unstructured format, such as notes fields, which makes automated processing rather difficult.

Hegna and Murtooma analyzed MARC records to determine what attributes could be best used for automatic conversion to FRBR. They determined that even though MARC records held attributes that could be used to identify work, expression and manifestation entities, inconsistencies in cataloguing and other errors would prove a great hindrance to automated processing (Hegna and Murtooma, 2002).

Further research has been done by Aalberg related with the creation of tools for converting MARC data to a FRBR model for a project between the Norwegian University of Science and Technology, the Norwegian bibliographic database BIBSYS and the National Library of Norway (Mönch and Aalberg 2003, Aalberg, et al. 2006). The FRBRizer tool developed by Trond Aalberg uses XSLT Technology, supports FRBR extraction and allows the specification of rules for MARC based formats, mapping the information used for FRBR extraction and defining key attributes or relations used for aggregation.

These several attempts at FRBRization show that it can be done. However, results of these attempts show that it is difficult to achieve high quality. Typical problems are in-

clude insufficient or erroneously identified entities and relationships and insufficient identification of equivalent entities. Furthermore, many fibrisation initiatives only apply parts of the FRBR model or only process parts of the information found in a record. However, complete FRBRization would introduce additional problems that could further reduce the quality of results. It has to be noted that these problems come from the existing MARC records. If the records had been created in a more structured way and more consistently, there would have been fewer misidentifications in fibrised data.

While the basic idea of all fibrisation algorithms is generally the same, that is to map MARC (sub)fields to FRBR entities, the actual products may vary in size and detail, owing to partly to local considerations in cataloguing (differences in the use of cataloguing rules and standards). While these mistakes are usually hidden when using the traditional catalogues, they jump out immediately when using FRBR.

Having learned from these implementations, we felt we wanted to do the same experience, manually FRBRize a sample of records extracted from PORBASE catalogue. Having done this experience and identified the main problems, we felt we could then recommend what to do to FRBRize a MARC based catalogue.

FRBRization is the way to go, though all the problems identified by different authors, because it helps better manage resources in a digital environment, making the most of the relationships among bibliographic records for better navigation in a resource sharing environment where collections tend to become larger.

4.2 Methodology

My study includes the following steps:

- Analyze the FRBR model to learn about entities, attributes and relationships. Understanding how the model works, it is essential to cluster bibliographic records according to entities, attributes or relationships that are present in the MARC format.
- Find the set of rules that support FRBRization and these are generally based on values assigned to User Tasks (see FRBR report, p. 83).

- Finally find keys to FRBRization.
- Select a work out of PORBASE, the Portuguese Union Catalogue, with a great number of editions that could be a good basis for a manual FRBRization, considering that we were not going to use any algorithm. “Os Maias” by Eça de Queirós was the work suggested for me to start with.
- Identification of the records that should be included in group entities

4.3 Study of the model

In the first step, we started by studying FRBR entities, attributes and relationships (see FRBR Report, 1999): Analyze FRBR model to know more about entities, attributes and relationships.

Understanding how the model works, is essential to cluster bibliographic records according to **entities, attributes or relationships** that are present in the MARC format.

FRBR Entities: These entities are divided into three groups:

Group 1 (products of intellectual or artistic nature):

- Work
- Expression
- Manifestation
- Item

Group 1 Entities and Primary Relationships

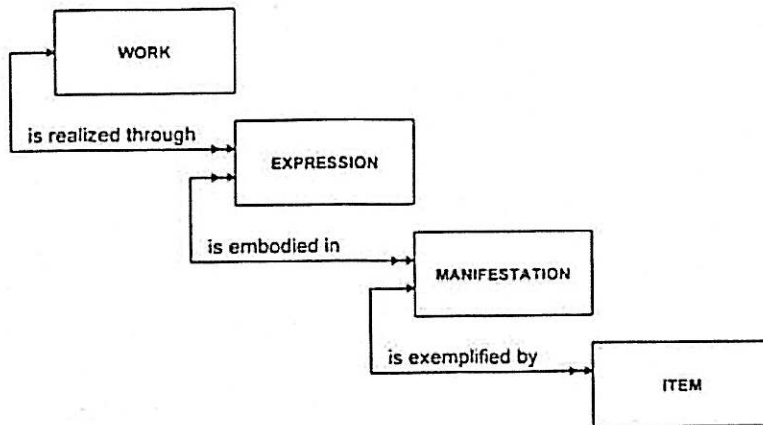


Figure 14 – Group 1 Entities and Primary relationships

(source: FRBR Report)

Group 2 (responsible for content, production, or custodianship of Group 1 entities):

- Person
- Corporate body

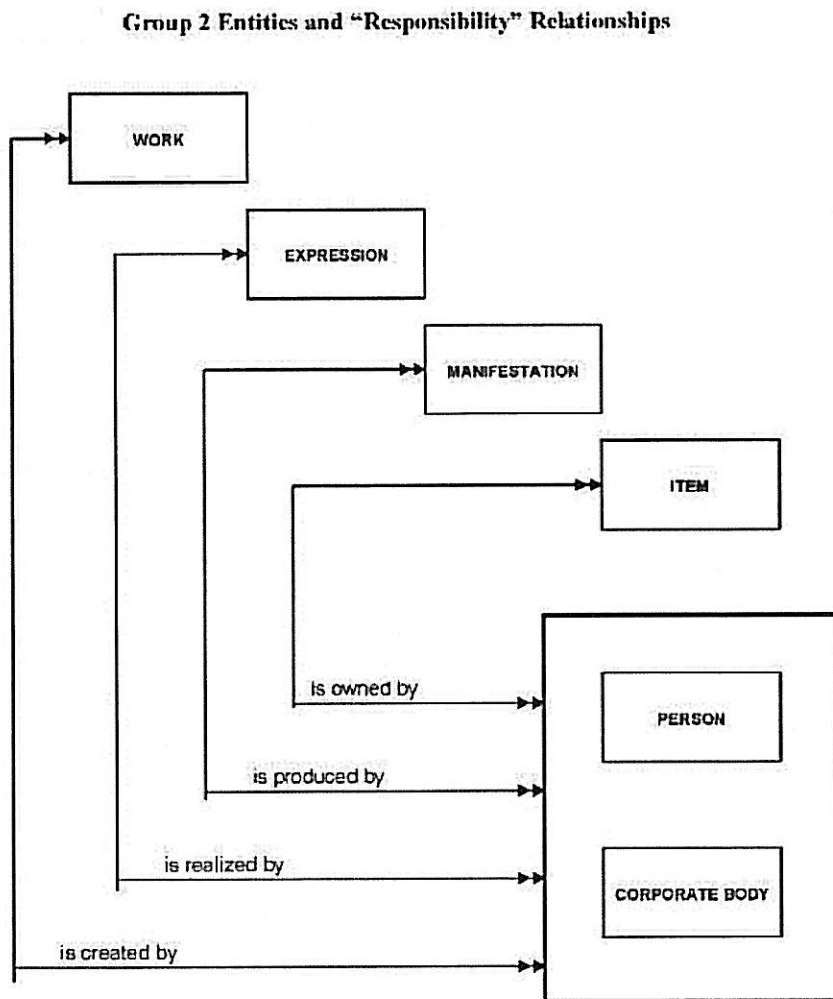


Figure 15 – Group2 Entities and Responsibility relationships

(source: FRBR Report)

Group 3 (may serve as subjects of Group 1 entities)

- Group 1 and 2 entities
- Concept
- Object
- Event
- Place

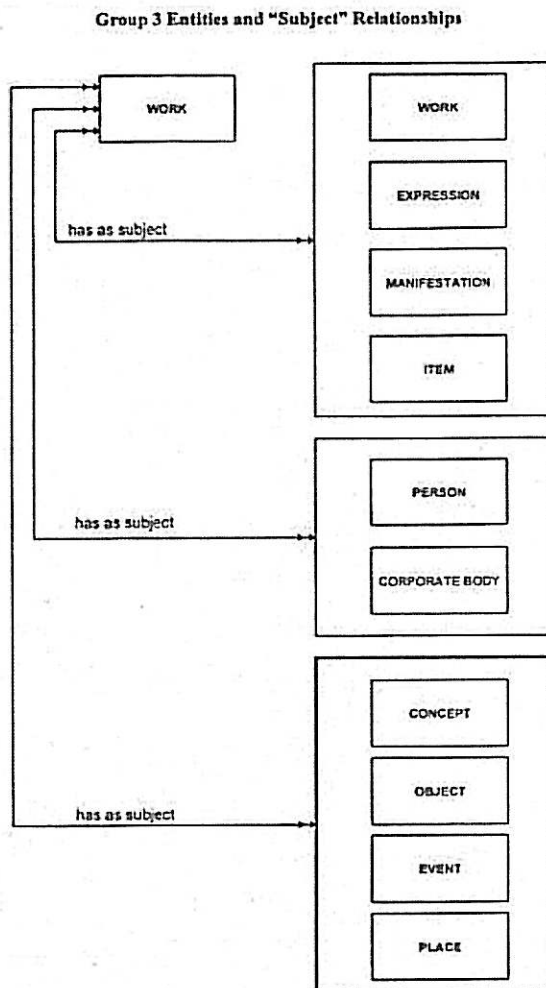


Figure 16 – Group 3 Entities (may serve as subjects of Group 1 entities)

(source: FRBR Report)

Some definitions and examples of these entities, according to Taylor (Taylor, 2007):

- **Work**

- Distinct intellectual or artistic creation
- Abstract entity with no single material object one can point to
- Recognized through individual expressions of the work
- Revisions, updates, abridgements, enlargements, translations, musical arrangements, and dubbed or subtitled versions of a film are considered to be expressions of the same work
- Paraphrases, rewritings, adaptations from one literary or art form to another, abstracts, digests, and summaries are considered to be new works
- Examples of works:
 - Shakespeare's *Romeo and Juliet*
 - Annie Proulx's *Brokeback Mountain*
 - Mozart's *The Magic Flute*
 - Vincent van Gogh's *Iris*
 - Michelangelo's *David*

- **Expression**

- Realization of a work in alpha-numeric, musical, or choreographic notation, sound, image, object movement, etc.; or any combination of such forms
- A new expression excludes aspects of physical form (e.g. typeface) that do not change intellectual or artistic realization of the work
- A change in form (e.g., change from written word to spoken word or addition of artistic content) or a change in intellectual approach (e.g.,

translation from one language to another or a revised edition) results in a *new expression*

- e.g., Shakespeare's *Romeo and Juliet*
 - e1 - original English text
 - e2 - the text illustrated with scenes from the movie
 - e3 - version translated into Portuguese
- e.g., for a work of Franz Schubert:
 - e1 - the composer's score
 - e2 - a performance by Amadeus Quartet
 - e3 - a performance by the Cleveland Quartet
- **Manifestation**
 - Physical embodiment of an expression of a work
 - When production involves changes in the physical form (or format), it results in a new manifestation
 - Changes in physical form include changes in display characteristics (e.g., font size, page layout), changes in physical medium (e.g., change from paper to microfilm), or changes in container (e.g., change from videocassette to DVD [but with no additions, modifications, etc.])
 - Changes in production signaled by a change in publisher, etc., also result in new manifestations
 - e.g., Shakespeare's *Romeo and Juliet*
 - e1 – original text document
 - m1 – New York : Applause, c2001 publication
 - m2 – archival copy of m1
 - m3 – Heinemann (London) 1926 publication
 - e.g., The New York Times
 - e1 – paper (vs. Web) version

- m1 – print-on-paper format
- m2 – microfilm format

- Manifestations are what we traditionally catalogue, describe, and provide subject analysis for.

- **Item**

- A single exemplar of a manifestation: although in some instances an “item” may consist of more than one physical object (e.g., a two-volume monograph, or a three-disk recording)
- Normally the same as the manifestation itself
- Variations external to the intent of the producer of the manifestation can occur in individual items (e.g., damaged copy, copy autographed by author, copy bound by a library, etc.)

- Items are what we collect, house, and provide physical and/or intellectual access to.

FRBR Attributes

Attributes are properties or characteristics. All entities are defined by a set of attributes that distinguishes them from each other within the FRBR model. For instance, Work has twelve attributes. The first is “title of the work” (FRBR Report, 4.2.1, p.33)³⁸. As an identifying characteristic of a work, in the Anglo-American tradition, “title” is equal to “uniform title”. FRBR defines it as “the word, phrase, or group of characters naming the work.” A second attribute of work is “form of work” (FRBR Report, 4.2.2, p. 33), this is the class to which the work belongs. “Date of the work” is the third attribute and is the date the work was originally created (the first date associated with the work). FRBR recognizes that different works may share the same title, so a fourth attribute is “other distinguishing characteristics”. Other attributes are: “intended termination”, “intended audience”, “context for the work”. Three attributes for musical works “medium of performance”, “numeric designation”, and “key”, and two for cartographic works “coordinates” and “equinox”.

³⁸ <http://www.ifla.org/en/publications/functional-requirements-for-bibliographic-records>

In this conceptual model of the bibliographic universe entities are specified and their attributes as well as the relationships between entities. FRBR associates user tasks to a number of entity attributes, thus indicating the value of these attributes in supporting a specific user task. These associations could be used to generate a list of attributes that should be mandatory for a bibliographic description in order to support a number of fundamental catalog functions. Attributes could be further divided into those attributes that are either desirable or support secondary functions and attributes that are optional.

- So, in short, the logical attributes of a *work* defined for this study are the following (FRBR Report, p. 32-33):

- title of the work
- form of work
- date of the work
- other distinguishing characteristic
- intended termination
- intended audience
- context for the work
- medium of performance (musical work)
- numeric designation (musical work)
- key (musical work)
- coordinates (cartographic work)
- equinox (cartographic work)

- The logical attributes of an *expression* defined for this study are the following: (FRBR Report, p. 35):

- title of the expression
- form of expression

- date of expression
 - language of expression
 - other distinguishing characteristic
 - extensibility of expression
 - revisability of expression
 - extent of the expression
 - summarization of content
 - context for the expression
 - critical response to the expression
 - use restrictions on the expression
 - sequencing pattern (serial)
 - expected regularity of issue (serial)
 - expected frequency of issue (serial)
 - type of score (musical notation)
 - medium of performance (musical notation or recorded sound)
 - scale (cartographic image/object)
 - projection (cartographic image/object)
 - presentation technique (cartographic image/object)
 - representation of relief (cartographic image/object)
 - geodetic, grid, and vertical measurement (cartographic image/object)
 - recording technique (remote sensing image)
 - special characteristic (remote sensing image)
 - technique (graphic or projected image)
- The logical attributes of a *manifestation* defined for this study are the following (FRBR Report, p. 40-41):

- title of the manifestation
- statement of responsibility
- edition/issue designation
- place of publication/distribution
- publisher/distributor
- date of publication/distribution
- fabricator/manufacturer
- series statement
- form of carrier
- extent of the carrier
- physical medium
- capture mode
- dimensions of the carrier
- manifestation identifier
- source for acquisition/access authorization
- terms of availability
- access restrictions on the manifestation
- typeface (printed book)
- type size (printed book)
- foliation (hand-printed book)
- collation (hand-printed book)
- publication status (serial)
- numbering (serial)
- playing speed (sound recording)
- groove width (sound recording)

- kind of cutting (sound recording)
 - tape configuration (sound recording)
 - kind of sound (sound recording)
 - special reproduction characteristic (sound recording)
 - colour (image)
 - reduction ratio (microform)
 - polarity (microform or visual projection)
 - generation (microform or visual projection)
 - presentation format (visual projection)
 - reduction ratio (microform)
 - system requirements (electronic resource)
 - file characteristics (electronic resource)
 - mode of access (remote access electronic resource)
 - access address (remote access electronic resource)
- The logical attributes of an *item* defined for this study are the following (FRBR Report, p. 48):
 - item identifier
 - fingerprint
 - provenance of the item
 - marks/inscriptions
 - exhibition history
 - condition of the item
 - treatment history
 - scheduled treatment

- access restrictions on the item

FRBR Relationships

Fundamental to an understanding of FRBR is an understanding of bibliographic relationships and these can exist between every FRBR entity. They can exist between entities in a single work or between different works. For instance, creator relationship can exist between persons and works, that is, between group 1 and group 2 entities. Relationships can still exist within the context of Group 2. For example a relationship exists between a corporate body that was replaced by another. Relationships exist between works and their subjects and we can still find relationships between a broader term and its more specific terms.

- Bibliographic relationships have been at the heart of cataloguing theory for more than a century.
- Relationship considerations follow Cutter's famous objects of the catalogues:
 1. Enable the user to find a book of which either the author, title, or subject is known;
 2. Show the user what the library has by a given author, on a given subject or in a given kind of literature;
 3. Assist in the choice of a book as to its edition or character.
- To accomplish any of the user tasks, the user must perform a relationship query.
- Bibliographic relationships between FRBR Group 1 entities began to be examined seriously more recently, when Barbara Tillet wrote a dissertation titled "Bibliographic Relationships" (Tillet, 1980). Tillet's research resulted in a "taxonomy" of bibliographic relationships. A taxonomy is a classification scheme showing relationships between the members of a group of terms. She advanced that the bibliographic universe can be classified into seven different types of relationships:

- Equivalence relationships (with the same intellectual content, a copy for example);
- Derivative relationships (translations);
- Descriptive relationships (work that describes another);
- Whole-part relationships (between resources and their parts);
- Accompanying relationships (supplement and complements);
- Sequential relationships (serials multipart);
- Shared-characteristic relationships (common author, title, subject).

4.4 Finding the set of rules to FRBRization

A fundamental task consists in finding the set of rules that support FRBRization and these are generally based on values assigned to User Tasks (see FRBR report, p. 83).

The FRBR model is to some extent present in the MARC record. A bibliographic record describes the work and the document (manifestation). Even data describing the expression might be found in the record. The information can be obtained from a semantic or a syntactic point of view.

The identification of a set of entities from a MARC record is the first step in the FRBRization process. The most difficult part of the process is to determine the relationships between the entities, because they may have not been added in the record, even those between a work and a person or an expression and a person. In these cases, relator codes (relator codes show us which roles authors have in works or expressions) are missing or are seldom used. Another problem is in the establishment of clear relationships when there are multiple persons and multiple works or expressions identified within a record. The MARC formats and the current cataloguing practices do not deal with this structural information, which is important in FRBR.

MARC was originally designed as flat-file system. Particularly difficult is the extraction and separation of Group 1 entities from the single MARC record. Various fields or subfields apply to work, expression, manifestation and sometimes item.

Let us see how the FRBR report can help us. (FRBR Report, 1998, p. 88 – 96)

- **Identifying the work:**

as expected, the FRBR relationship between a work and the creator is ranked of high value in the identification of a Work (FRBR table 6.1, p.88).

	Basic level	Should include these specific data elements	Value
<i>Identify a work</i>	Title of the work	Title heading for the work	high
	Relationships between the persons and or corporate bodies responsible for the work	Name headings for persons and/or corporate bodies with principal responsibility for the work	high

Table 1 – Attributes and relationships of work

- **Identifying an expression:**

If one seeks a particular expression, the focus will be in a relationship that is “responsible for content” or a “form of expression”, or “language of expression”, or “use restrictions on the expression”. These attributes have high values in the identification of an expression (FRBR table 6.2, p. 92);

	Basic level	Should include these specific data elements	Value
<i>Identify an expression of a work</i>	Relationships between the persons and or corporate bodies responsible for the expression	Name headings for persons and/or corporate bodies with principal responsibility for the expression	high
	Form of expression	Note on form of expression (note1)	high
	Language of expression (note2)	Addition to uniform title – language Note on language	high
	Other distinguishing characteristic	Addition to uniform title – other distinguishing characteristic Note on other distinguishing characteristic	high

Table 2 – Attributes and relationships of expression

Note 1: A note on form of expression is considered a basic requirement only if the form of expression cannot be inferred from other data in the record.

Note 2: Language of expression is considered a basic requirement only if the linguistic content of the expression is significant.

- **Identifying a manifestation:**

One seeks the name of the publisher and the date of publication (FRBR table 6.6, p. 93); FRBR is written as if it were possible to obtain a manifestation, but what the user actually wants to obtain is a single instance of a manifestation, that is, an item.

	Basic level	Should include these specific data elements	Value
<i>Identify a manifestation</i>	Title of the manifestation	Title proper	high
	Statement of responsibility	Statements of responsibility identifying the individuals and/or groups with principal responsibility for the content First statement of responsibility relating to the additional edition statement	high
	Edition/issue designation	Edition statement Additional edition statement	high
	Publisher/distributor	Name of publisher, distributor, etc.	high
	Date of publication/distribution	Date of publication/distribution	high

Table 3 – Attributes and relationships of manifestation

Selecting an item is relevant to identify: “call number”, as an item identifier, “URL” access to remote-access digital versions, “access restriction” is another attribute.

Mapping MARC fields to FRBR entity identifiers and finding keys to FRBRization (Question tags mean that, though FRBR rank these attributes, we have no means to identify those attributes due to cataloguing practices or inconsistencies):

FRBR attribute	FRBR value	Fields/Subfields	MARC Tags (by priority)
title of work	high	Title fields: uniform titles or original titles	500\$a\$k 200\$a 300\$a, 304\$a
relation to person responsible	high	Main entry	700\$a\$b
intended termination	high	?	?
form of work	moderate	?	?

Table 4 – Identifying a work

To get the work we had to look for titles in the fields of title statement (MARC tag 200), considering that uniform titles (MARC tag 500) have not been applied these last years. Cataloguing best practices advise the use of the uniform titles. Should this be used, it would be very easy to identify the work. The relation to person responsible of the work (MARC tag 700), we found in the correct field. Intended termination of work is of moderate value for us, but it could be relevant in case of electronic documents. Form of work would be convenient to use, though FRBR, considers it of moderate value in identifying a work, because change of literary form may represent a new work. So, in this study the title of the work and the relation to person responsible will be used as attributes identifying works.

FRBR attribute	FRBR value	Fields/Subfields	MARC Tags (by priority)
relation to person responsible	high	added entries (for translators...)	702\$a\$b\$4
form of expression	high	?	?
date of expression	moderate	?	?
title of expression	moderate	?	?
language of expression	high	language of publication	101

Table 5 – Identifying an expression

Language of an expression is of high value in identifying different expressions of a work, but one has to look also for the entity responsible for expression, the translator. The translator is found in field 702. If we are in the presence of a translation, the code language will be found in the field 101. Title of expression is not covered by actual cataloguing rules, but it is usually the same title of the manifestation. Date of the expression is the same date of the first manifestation published.

FRBR attribute	FRBR value	Fields/Subfields	MARC Tags (by priority)
title	high	title field	200\$a
statement of responsibility	high	statement responsibility field	200\$f
edition	high	edition field	205\$a
publisher	high	publisher field	210\$c
date	high	date field	210\$d

form of carrier	high	physical description	215\$a
extent of carrier	moderate	physical description	215\$a
physical medium	moderate	physical description	215\$c
manifestation identifier	high	ISBN	010

Table 6 – Identifying a manifestation

The manifestation title is found in field 200\$a, and is the original or the translated title. This is an important attribute to identify the manifestation as well the statement of responsibility that we find in subfield (200\$f). Edition and date of publication are redundant sometimes. Publisher (210\$c) and date (210\$d) are important to distinguish different manifestations. Extent of carrier can also help separate different manifestations. Manifestation identifier, ISBN numbers are given in 010 field and is good to differentiate one manifestation from another.

FRBRizing step by step

“Os Maias” by Eça de Queirós, was the work chosen to be studied. PORBASE, the Portuguese Union Catalogue, was our testbed. This work has a great number of editions that could be a good basis for a manual FRBRization and a good option to be explored in the context of this work as a first experiment.

The first search formula was word in the title = “Os Maias” to see what the catalogue had with this title. The result obtained was a long list of 191 bibliographic records with no apparent organization. I had to browse the list, at least 20 pages like this one, to try to understand what had been retrieved by the computer:

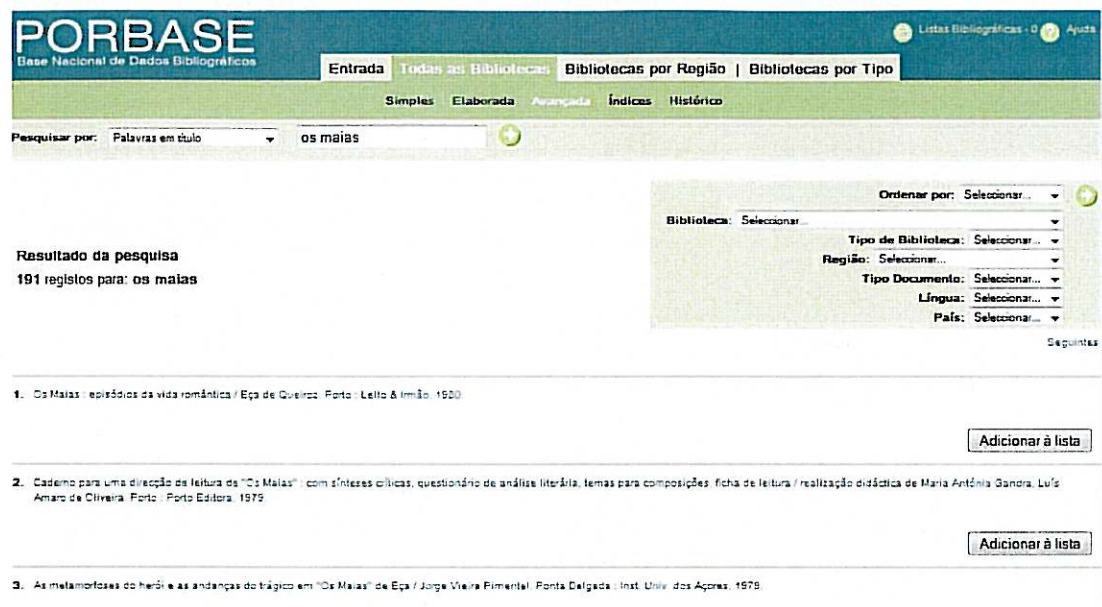


Figure 17 – PORBASE: search result screen

Search result:

The 191 records were displayed in a long list with no structure at all. This list could, on a second step, be organized according to an alphabetic order of title or publication date. No more options were given to organize such a long list of results.

Looking at the list we recognize the work “Os Maias” represented by a great number of expressions, a great number of manifestations, and a relevant number of new works about the work “Os Maias”. We also retrieved what we call “noise”, records that do not correspond to the search criteria. These last had that word in the title, but with a different meaning.

This sample of records proved to good enough to work with, because there was a great number of expressions and manifestations and, at the same time, it was not too big to make a manual FRBRization an impossible task. In a second search formula, we only had to try to remove the “noise”, that is to say, the group of records that were not related with the work “Os Maias”

So, an extraction of a sample of bibliographic records from PORBASE catalogue was prepared so that the first manual FRBRzitation test might be accomplished. To avoid

“noise”, those useless records that would add volume to our sample, an efficient retrieval effort was done, by using the following search formula:

The search formula for extracting the bibliographic family was:

Title of the work = “Os Maias” and Author = “Queirós, Eça de” or Subject Author-Title = Queirós, Eça de – Os Maias or Subject “Queirós, Eça de”

Translating this formula to MARC language:

title		200\$a\$c\$d\$e	5XX\$a	
author		700\$a\$b		
OR				
	604 = \$aQueirós,\$bEça de,\$f1845-1900\$t-Os Maias			
OR				
	675\$a"821.134.3 Queirós, Eça de" (a subfield with this string, eventually longer)			
OR				
	675\$a, where 675\$a"869.0-1" and "Queirós, E" (every subfield \$a may contain simultaneously these two expressions)			

Table 7 – Search formula to retrieve “Os Maias” from PORBASE

The results of this extraction were 114 bibliographic records. Then they were exported to an Excel sheet, where records were ordered by a record identifier, in this case, the BCN (Bibliographic Control Number).

There was no automatic tool to support this exercise, so I had to look at every record and try to identify manually which FRBR entities were present there. I did that by looking at the ISBD description, and always had to confirm analyzing the MARC for-

mat and look for more information. The items were not available to me, so some interpretation errors could have been made.

NCB	Expressão	Obra	Super-obra	Notas
975		975		Porto: Lello & Irmão, imp. 1980
10373		10373		Lisboa: Círculo de Leitores, imp. 1986
15288		15288		2ª ed., Lisboa: Uíssaia, D.L. 1986
15291		15291		trad. Servo-Croata, 1985
18252		18252		estudo, introd. J. Tomas Ferreira, Europa-América, D.L. 1985
27673		27673		trad. Italiano, Unvers, 1978
35840		35840		Mem Martins: Europa-América, D.L. 1987
43556		43556		8ª ed., Lisboa: Círculo de Leitores, imp. 1982
43557		43557		9ª ed., Lisboa: Círculo de Leitores, imp. 1982
57622		57622		Lisboa: Círculo de Leitores, 1975
85742		85742		Mem Martins: Europa-América [D.L. 1982]
108567		108567		3ª ed., Lisboa: Uíssaia, D.L. 1988
123473		123473		Lisboa: Círculo de Leitores, imp. 1988
194143		194143		Porto Editora, 1989
227690		227690		trad. Polaco, cop. 1988
249512		249512		introd. Esther Lemos, Uíssaia, 1992
258716		258716		Porto Editora, 1991
297260		297260		4ª ed., Porto: Lello & Irmão, [1903-1920]
297468		297468		Livr. Chardron, 192-
322921		322921		obras aumentadas (Exaço do texto e notas), texto reproduzido em Braille, Livros do Brasil, 1991
342456		342456		Exaço do texto e notas de Helena Cidade Moura, Livros do Brasil, D.L. 1992
343695		343695		introd. Esther de Lemos, 5ª ed., Uíssaia, 1992
408791		408791		trad. Alemão, Piper, 1956
414130		414130		[Lisboa] Círculo de Leitores, D.L. 1992
414830		414830		Exaço dos textos e notas de Helena Cidade Moura, Livros do Brasil, 1970
423810		423810		[Lisboa] Círculo de Leitores, D.L. 1993
450691		450691		trad. Italiano, Gherardo Casini, imp. 1956
454986		454986	454986	1ª ed. Livr. Chardron, Cas Editora Lugan & Genélix Sucessores 1988)
460547		460547		Porto Editora, 1994
498373		498373		4ª ed., Mem Martins: Europa-América, D.L. 1994
509014		509014		6ª ed., Livr. Chardron, de Lello & Irmão 1923

Figure 18 – Excel sheet with a part of the extraction of bibliographic records

So, now I was expected to be able to separate the work, the expressions and the manifestations, according to the FRBR model, and create the respective groups or clusters.

The identification of the work, expressions and manifestations was manually done according to:

				FRBR value
work	700\$a\$b	relation to person responsible	Queirós, Eça de	high
	200\$a	title of work	Os Maias	high
	?	Intended termination		high
	?	form of work		moderate
expression	700\$a\$b 702\$a\$b	relation to person responsible	translators,	high
	?	form of expression		high
	?	date of expression		moderate
	200\$a	title of expression		moderate
	101\$a	language of expression		high
manifestation	200\$a	title	“Os Maias”, other transla- tions of the title	high
	200\$f\$g	statement of respon- sibility	author, translators, illustrators...	high
	205\$a	edition	several ed.	high
	210\$c	publisher	several pub.	high
	210\$d	date	several dates	high
	215\$a	extent of carrier	number of pages	moderate
	010\$a	manifestation iden- tifier	ISBN	high

Table 8 – Keys to identify FRBR entities in the bibliographic record “Os Maias”

- **Identifying the work:**

The relevant fields from a bibliographic record for a work are shown here:

200\$aOs Maias\$fEça de Queirós

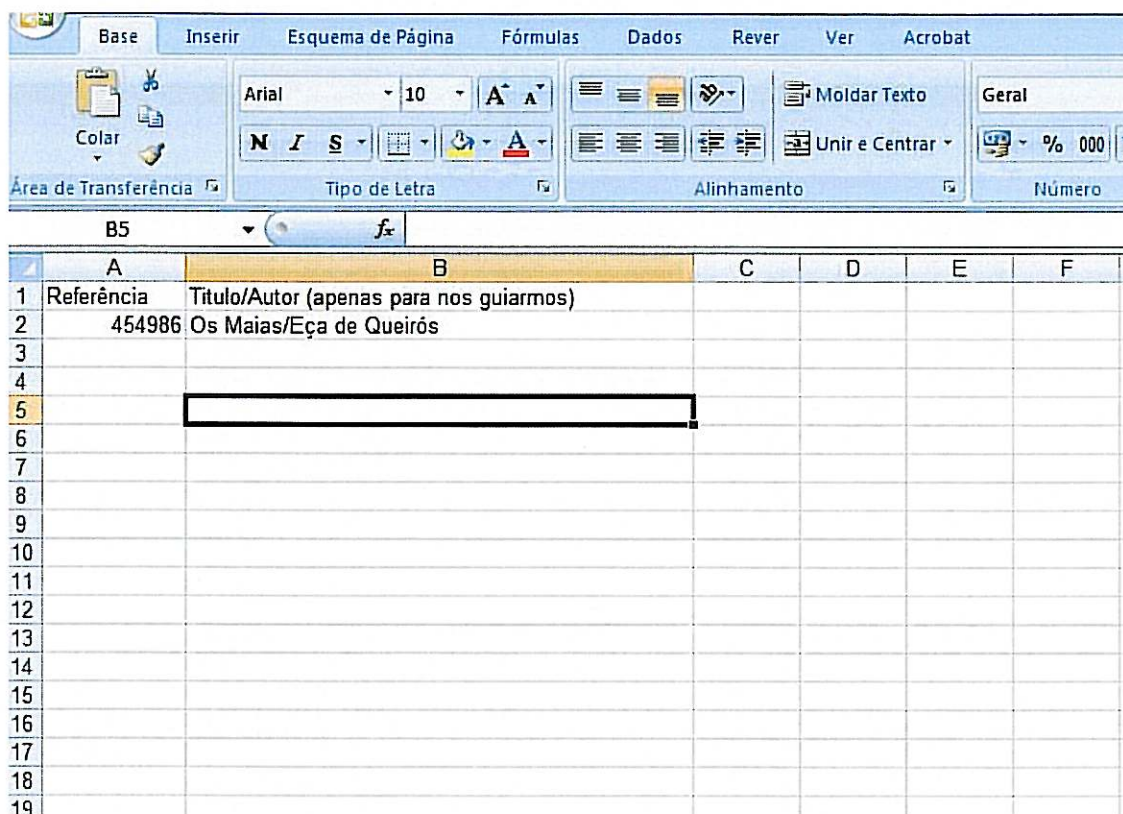


Figure 19 – Excel sheet with results for “work entity”

- **Identifying an expression (among others)**

The relevant fields from a bibliographic record for a typical expression are shown here (in red):

101\$aIt (language of expression)

200\$a Maia (title of expression)

304\$a Tít. orig.: Os Maias

702\$aMandillo\$bEnrico\$4340 (translator)

- Identifying a manifestation (among others)

And the relevant fields from a bibliographic record for a typical manifestation are shown here (in blue):

200\$aOs Maias\$fEça de Queirós

210\$aLisboa\$cCírculo de Leitores\$dimp.1986

215\$423 p.\$d

700\$aQueirós\$bEça de

We have grouped Expressions (in red) and Manifestations (in blue):

	B	C	D	E	F	G	H	I	J	K	L	M	N	O
32	454986	454986		Porto Editora, 1989										
33	194307	194307		2005aO leitor e a verdade oculta\$eensaio sobre os Maias										Família
34	195857	195857		2005aOs Maias [de]Eça de Queirós\$por José Tomaz Ferreira										Família
35	218366	218366		2005aA obra "Os Maias" de Eça de Queiroz\$bTexto policopiado										Família
36	227679	227679		2005aOs Maias em análise\$eantologia comentada										Família
37	227690	454986		trad. Polaco, . cop. 1988										
38	244225	244225		2005aLeitura dos Maias\$fSemana de Estudos Queirozianos										Família
39	247242	247242		2005aEça e os Maias\$eem anos depois\$eactas...										Família
40	15288	454986		Os Maias/introd. Esther Lemos. 4ª ed. Ulisseia, 1988										
41	250797	250797		2005a"Os Maias" de Eça de Queirós\$bVisual gráfico\$eexposição										Família
42	254708	254708		2005aUm duplo centenário\$e"Os Maias" (1888) de Eça de Queiroz\$fJoão Medina										Família
43	256100	256100		2005aO Clube do Eça apresenta vinte personagens de "Os Maias"\$fAlves de										Família
44	454986	454986		Porto Editora, 1991										
45	281268	281268		2005aSerá o Alencar dos "Maias" um retrato do BulhãoPato?										Família
46	454986	454986		4ª ed.; Porto. Lello & Irmão, [1903-1920]										
47	454986	454986		Liv. Chardron, 192-										
48	303107	454986		2005aThe Maias\$fEça de Queiroz\$gtranslated by...										
49	322921	454986		obras aumentadas (fixação do texto e notas). texto reproduzido em Braille: Livros do Brasil, 1991										
50	342456	454986		fixação do texto e notas Helena Cidade Moura. Livros do Brasil, D.L. 1992										
51	15288	454986		introd. Esther de Lemos, 5ª ed., Ulisseia, 1992										
52	408791	454986		trad. Alemão, Piper, 1986										
53	57622	454986		[Lisboa] Círculo de Leitores. D.L. 1992										
54	342456	454986		fixação dos textos e notas de Helena Cidade Moura. Livros do Brasil, 1970										1
55	57622	454986		[Lisboa] Círculo de Leitores. D.L. 1993										
56	450691	454986		trad. Italiano, Gherardo Casini, imp. 1956										
57	454986	454986		1ª ed. Liv. Chardron, Cas Editora Lugan & Geneloux Sucessores, 1888)										Obra
58	460547	454986		Porto Editora, 1994										
59	467155	467155		2005aMara Eduarda e Carlos da Maia\$bVisual gráfico\$gAlberto de										Família
60	85742	454986		4ª ed. Mem Martins: Europa-Aménica. D.L. 1994										
61	454986	454986		6ª ed., Liv. Chardron, de Lelo & Irmão, 1923										
62	454986	454986		7ª ed., Liv. Chardron de Lelo & Irmão, 1924										
63	454986	454986		introd. Esther de Lemos, 7ª ed., Ulisseia, imp. 1994										

Figure 20 – Excel sheet with results for “expression and manifestation entities”

We faced an unexpected problem: Several works, building a bibliographic family and the need of a Super Work to better collocate the Work.

The Works and the bibliographic family around a Work or the need of a Super Work:

	A	B	C	D	E	F	G	H	I	J
1	Referência	Título/Autor (apenas para nos guiarmos)								
2	454986	Os Maias/Eça de Queirós. 1ª ed.								
3	2068	Caderno para uma direcção de leitura de "Os Maias" \$realiz. Luís Ama								
4	2120	As metamorfoses do herói e as andanças do trágico em "Os Maias"								
5	4291	Introdução à leitura d'Os Maias \$Carlos Reis, 2ª. Ed.								
6	9200	Caderno auxiliar de leitura de Frei Luís de Sousa, Viagens... Os Maias								
7	27468	Os Maias/adaptação teatral do original de Eça de Queirós ... \$gestudo de Carlos Reis								
8	38694	O tempo da invenção criadora no romance os Maias, de Eça...								
9	38714	Quelques aspects de la critique sociale dans os Maias								
10	84984	Para a compreensão dos Maias como um todo orgânico								
11	106512	As máscaras do desengano \$... \$Isabel Pires de Lima								
12	106907	Polémicas de Eça de Queirós....								
13	124431	Os Maias, cent ans après								
14	190565	Os Maias de Eça de Queirós \$catálogo de da exposição \$forg.								
15	194307	O leitor e a verdade oculta \$ensaio sobre os Maias								
16	195857	Os Maias [de] Eça de Queirós \$por José Tomaz Ferreira								
17	218366	A obra "Os Maias" de Eça de Queiroz \$b Texto policopiado								
18	227679	Os Maias em análise \$antologia comentada								
19	244225	Leitura dos Maias \$Semana de Estudos Queirosianos								
20	247242	Eça e os Maias \$cem anos depois \$actas...								
21	250797	"Os Maias" de Eça de Queirós \$b Visual gráfico \$e exposição								
22	254708	Um duplo centenário \$e "Os Maias" (1988) de Eça de Queiroz \$f João Medina								
23	256100	O Clube do Eça apresenta vinte personagens de "Os Maias" \$f Alves de								
24	281268	200 \$a Será o Alencar dos "Maias" um retrato do Bulhão Pato?								
25	467155	Maria Eduarda e Carlos da Maia \$b Visual gráfico \$g Alberto de								
26	1060466	Satyras e epistola 200 \$e a propósito dos Maias, s. n., 1889, tem título uniforme, 500 \$a Sátiras e epistola, 701 B. Pato								
27										
28										
29										

Figure 21 – Excel sheet with results for “bibliographic family”

The manual grouping of records were then converted to HTML lists of bibliographic control numbers (BCN), to try the first FRBRized visualizations. These bibliographic control numbers represented clusters according to FRBR entities: work, expression and manifestations. These clustering were accomplished manually:

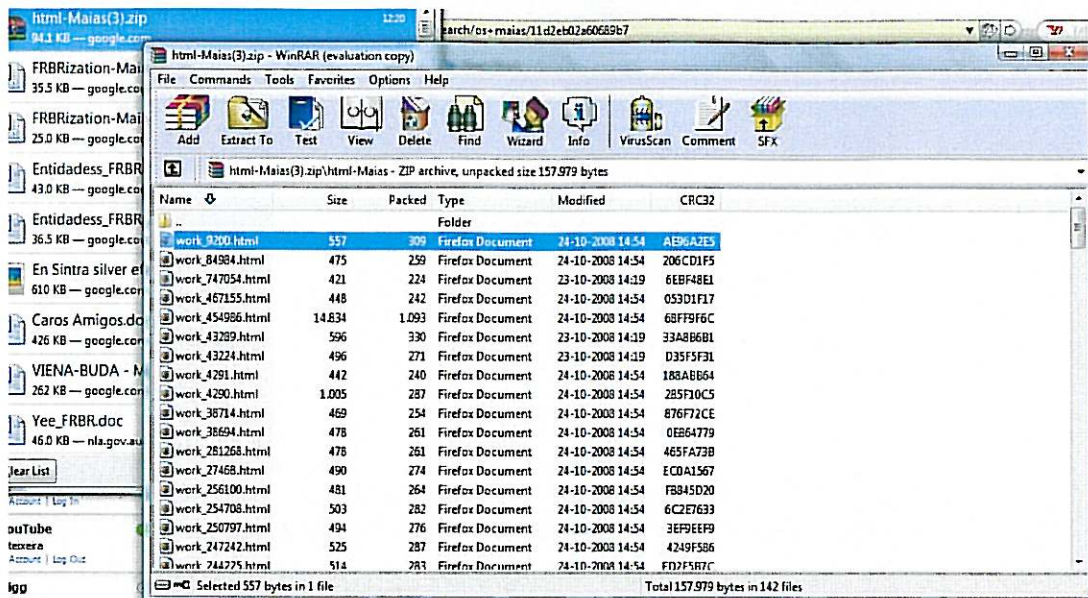


Figure 22 – HTML list of clusters

Clicking in a bibliographic control number (BCN), we could see the clustering:

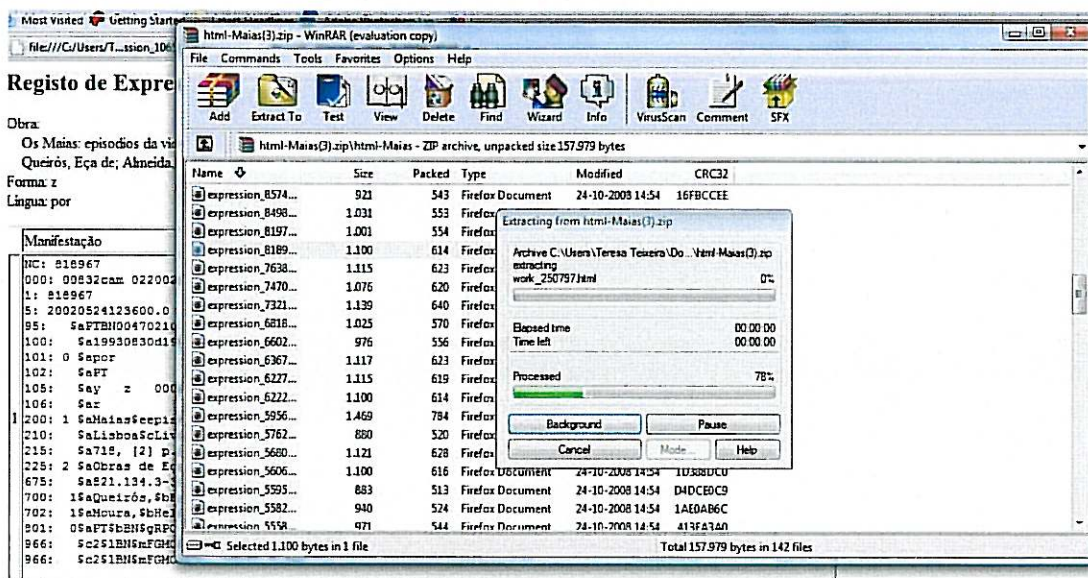


Figure 23 – HTML list of clusters, opening a cluster

Clustering in work, expressions and manifestations was achieved, as you can see, was first visualized like this: a list of works and expressions followed by clusterings in expressions and manifestation:

- Os maíns: episódios da vida romântica
Queirós, Eça de; Marques, Bernardo; Moura, Helena Cidade
- Os Maíns
Queirós, Eça de
- Os Maíns: texto integral
Queirós, Eça de
- Os Maíns: episódios da vida romântica
Queirós, Eça de

	Expressão	Manifestações
1	<ul style="list-style-type: none"> • Os Maíns: episódios da vida romântica Queirós, Eça de Forma: a Língua: por	1 (ver)
2	<ul style="list-style-type: none"> • Os Maíns Queirós, Eça de • Os Maíns: episódios da vida romântica Queirós, Eça de; Almeida, Fialho de Forma: z Língua: por	2 (ver)
3	<ul style="list-style-type: none"> • Os Maíns: episódios da vida romântica Queirós, Eça de; Lemos, Ester de Forma: a Língua: por	1 (ver)
4	<ul style="list-style-type: none"> • Os Maíns ; Queirós, Eça de 	1 (ver)

Figure 24 – the first visualization of the “expression entity”

An example of an expression with two manifestations:

Registo de Expressão:

Obra:
Os Maíns: episódios da vida romântica
Queirós, Eça de; Almeida, Fialho de
Forma: z
Língua: por

Manifestação
NC: 10373 000: 00614cam 02200217 04500 1: 10373 5: 19981111090900.0 95: \$aFTEH00010652 100: \$a19860701d1986 k y0pora0103 ba 101: 0 \$apcr 102: \$aFT 200: 1 \$aMaíns: episódios da vida romântica 210: \$aLisboa: ScCírculo de Leitores, Sdimp. 1986 215: \$a423 p. \$d25 cm 225: 2 \$aRomances completos de Eça de Queiroz 675: \$a021.134.3-31*18*5vBHSzpor\$3833768 700: 1\$aQueirós, \$bEça de, \$f1845-1900\$310528 801: 0\$aFISbBHSgrFC 966: \$1BNSMFCMONS=L. 36640 v. 998: \$aFSE19 - 00284
NC: 454986 000: 00E26cam 2200277 450 1: 454986 5: 20030121143100.0 95: \$aFTEN00657400 100: \$a19940906d1888 k y0pory0103 ba 101: 0 \$apcr 102: \$aFT

Figure 25 – the first visualization of the “manifestation entity”

5 Results and discussion

The first step of this manual FRBRization was to identify cluster types that could be used to group records representing different editions of works.

Cluster types (work, expressions and manifestations) were identified based on relationships among items and results of other earlier research projects (Hegna, Murtomaa, 2002), (Hickey, 2002), (Bennett, Lavoie, O'Neill, 2003) and (Yee,2005), Aalberg, 2007).

Relationship-based cluster types included in this study included:

- Eds. with amplifications, introductions and prefaces
- Large print, editions
- Editions in collections of 2 or more works
- Other language editions
- Nonbook format editions
- Braille editions
- Performances
- Adaptations

The second step was to analyze the records provided by PORBASE in order to implement the cluster types identified. Thus, cluster types were built based on the identification of FRBR entities in certain MARC fields and sub-fields, using FRBRization keys, through a record-by-record analysis. Keys proved that they were not enough and so contents of fields like title (other title fields), author, edition, publication, physical description, notes and language were rigorously analyzed.

The FRBRization of “Os Maias” made us able to identify within a sample of bibliographic records, the “Work”, the “Expressions” and the “Manifestations”, and the results we obtained following these rules:

Work: was identified with a combination of the title (200\$a), in the absence of a uniform title (500\$a), with the main author (700\$a\$b), and the earliest published edition, date of publication (200\$d).

Expression: expressions were grouped combining edition fields (205\$a), code language (101\$a), role of secondary authors (702\$a\$b\$4).

Manifestations: manifestations were grouped according to publication place, publisher and date (210\$a\$c\$d), sometimes we looked in physical descriptions for details that might help us.

I came to the conclusion that FRBRization, due to inconsistent cataloguing practices, might be a difficult task, though not impossible to accomplish. I had rather say that it is possible to accomplish, and it is advantageous too, but an automatic tool to help in the first selection of the bibliographic records should be made available, then FRBRization could be refined by manual selection. A full automatic FRBRization of “Os Maias” could be problematic to obtain due to:

- It was difficult to extract and separate the four Group 1 entities from the bibliographic records. Some pieces apply to work, expression and manifestation. But it was difficult to define explicitly that a certain field or subfield applies to a particular entity.
- Impossible to select the work due to the missing uniform title. We used the title statement to identify the work. The date helped too, though the publication date is of high value to the manifestation. To the work is the intended termination date that is considered of high value.
- Impossible to distinguish automatically the work from all the expressions. This is most problematic point of FRBRization. The frontier between a work and an expression is not clear and when an expression becomes another work is still more unclear. This required manual work and much study. In FRBR, variants that incorporate revisions or updates are regarded as the same work. The line

is crossed and a new work is found when there are paraphrases, rewritings, adaptations for students, children, musical variation of a theme, adaptation to the theatre or cinema, as well as adaptations of one literary form to another. The frontier between works in FRBR look to be based on cultural aspects, namely the Anglo-American tradition embodied in the Anglo American Cataloguing Rules (AACR2). This makes this frontier depend on the subjectivity of the cataloguer.

- It was difficult to identify the instances of the same work or expressions and the manifestations of a single expression.
- Relationships were curiously the easiest task for me. Portuguese cataloguing practices apply the relator codes and so it was immediate to see who the translator, illustrator or editor was.

The results that we obtained were the following:

	Work (Super Work?)	Expressions	Manifestations	Other Works	No valid records
“Os Maias”	1	50	90	24	2

Figure 26 – results of frbrisation: creation of clusters

In this sample, we found an expression corresponding to single manifestation and others where expressions represent clusters of a relevant number of manifestations.

From the sample of 26 new works we found amplifications, adaptations, performances and descriptive work, essays and others.

FRBRization, though manually, was reasonably accomplished. This means that although the difficulties it was possible to identify the four FRBR entities in the single

entity MARC record and make clusters according to work, expressions and manifestations.

These are the main issues identified and the following lessons were learned:

- It is not easy to split a MARC bibliographic record, based on a single entity, into four different entities (the FRBR model). Different pieces of the MARC bibliographic record apply to work, expression, manifestation and item and it was never defined explicitly that a certain field or subfield applies to a particular FRBR Group or entity.
- One of my biggest difficulties, and that is a major FRBR issue, was to identify the entity “expression” and understand the frontier between an expression and a new work. The line between works in FRBR appears to be based on one specific cultural tradition, namely the Anglo-American tradition embodied in AACR2. Definition of expression needs a further clarification.

FRBRization is a difficult task to accomplish due to inconsistent cataloguing practices or to budgetary constraints requiring a less detailed cataloguing, for example, uniform title, important to identify the work, is no longer registered, or the role of a secondary author is seldom used and that could help identify an expression. Is the nature of the relationship between a person and an expression that of author (e.g. producer of a revised edition) being recorded? Translator? Editor? Cataloguing rules have long provided for the addition of designation of functions to headings for persons, such as “ill.” (illustrator), “ed.” (editor), “tr.” (translator), and “comp.” (compiler), which might have served as an aid to identifying the nature of the relationship. This addition was however made optional in AACR2, and the Library of Congress instructed its cataloguers not to apply the option. In the absence of such a term (or code) in a MARC bibliographic record, automated identification of the nature of the relationship will in many cases be impossible.

- FRBRization is difficult to accomplish manually. One needs an algorithm to support the conversion of a MARC catalogue to a FRBRized one, but inconsistencies in cataloguing rules make it difficult too. For example, title (and subtitle) and author fields are prone to errors, use of abbreviations is frequent, which is enough to disable exact matching of records. So, automatic FRBRization is a start and it can be done automatically and end with a manual intervention to improve the results.
- We found a group of new works. These have overcome that frontier between a work and a new expression and are what we call new works: adaptations, criticisms, reviews, annotated editions, dramatizations and so on. There has been some discussion of the need for a FRBR Group 1 entity at a higher level than work. The purpose of such an entity would be to allow better collocation between related works in catalogues.
- The group of new works introduces us to a new concept, what we call the "bibliographic families" - groups of works that share common intellectual content. As for the types, or patterns, observed in bibliographic families, the majority have simultaneous and successive editions, as well as translations. The term of superwork is sometimes used similarly with that of bibliographic family, to mean all works and their subsets that descend from a common intellectual origin. Identifiers are needed to construct derivative relationships and to collocate the subsets of a superwork. However, there is no need to create special identifiers for superworks; the work identifiers, which they already possess (a uniform title), suffice for the purpose of structuring bibliographic relationships." The superwork, is roughly equivalent to the work entity defined by FRBR. That is, it represents the abstract intellectual conception, as the creator might have imagined it, before its realization as text (or music or art, etc.) The uniform-title heading was the collocating device preferred by cataloguers through most of the twentieth century to cause all elements of bibliographic family to file together in the catalog.

Later, a second approach to FRBRization was started with two much larger samples of records relevant to a wide public, "The Bible" and "Literature Nobel Prizes". This

FRBRization could not be done manually due to the large amount of records and so an interface was built (Figure 27) to assist in the process of grouping FRBR entities. The librarians involved in this process made suggestions to improve the functionalities of this interface, so that it might become more efficient.

It is exactly here that this master thesis ends. I refer this second approach, because this one was possible with what was learned with the first experiment, the FRBRization of “Os Maias”. The TEL project team is still developing the interface. It is expected that, based on critiques and suggestions, the new interface to support the FRBRization process will be soon available.

6 Conclusions and future work

Although there are still many issues that need to be solved, FRBR has made positive progresses. A lot of research and experimental work have been done. Problems are identified and there is common sense about the advantage of using this new model for a better display of information, though the different cultural backgrounds of cultural environments, cataloguing rules and practices.

New cataloguing rules aligned with the FRBR model will help solve most of the problems in the future. With the evolution of new technologies, it is expected that FRBR will help develop more effective information systems to support user information needs and information seeking in the Web.

FRBR is libraries' great hope to organize information in a way that corresponds to users' interests. There was a great change taking place concerning users' behaviours. Library end-users became expert using search engines and other Internet resources and were demanding new capabilities of online systems. Ranking, organizing and clustering were features they have got used in Google and other search engines against library long lists of unstructured and apparently unrelated records. Actually, there are bibliographic sites, like Amazon, LibraryThing and others, where bibliographic information is organized in a way that it makes clear to the user if a given work is interesting or not to get: there are reviews, rankings and so on. Why not give the same features to end-users in the library?

Networked environment, new technologies, new digital libraries and databases provide a wide world access to resources to everybody. The library long lists of results with no implicit organization or ranking are too confusing for them. So, users are now more demanding and the tendency is to abandon library searching tools and move to the Internet environment where they have friendly searching interfaces and a better access to the documents they are seeking.

Libraries aware of this new trend are trying to evolve to give the end-users a new searching interface with better displays focused on their needs.

Technological advances, like library automation, and cataloguing practices using minimum levels of description are the main responsible for the loss of hierarchical structure of the library catalogue. Something was lost of great value in the library catalogue.

Almost two decades ago, these problems were discussed in a conference, the Seminar on Bibliographic Records, held in Stockholm in 1990. One of the resolutions was that a study should be carried out to define the functional requirements for bibliographic records in relation with to the variety of user needs and the variety of media. In, 1991, a group was formed by IFLA.

This group worked for six years and then issued the Functional Requirements for Bibliographic Records (FRBR), that has introduced us to a new conceptual model of the bibliographic universe with a strong user focus.

The library community is aware of the problems related with the loss of hierarchical structure in the OPAC displays. FRBR is a model that may provide better searching, retrieval and display by collocating records based on relationships among the entities of the model. The entities represent intellectual or artistic products that are described in bibliographic records: work, expression, manifestation and item (Group 1). They may represent those that are responsible for the intellectual content; persons or corporate bodies (Group 2). They may still represent subjects of works, such as concept, object and place (Group 3). The clustering of information according to these entities will enhance a better navigation in the search results.

This new model for the bibliographic universe is the last hope to library catalogue and an answer to the functionalities the users are demanding. The library community is exploring its potential. A relevant number of experiments, applications and implementations based or inspired on FRBR are taking place all over the world. FRBR is changing the library environment: cataloguing rules and library systems will have to change too to be in conformity with this new model.

There is a legacy of millions of records that it is not easy to convert to the new model, but that legacy cannot be abandoned. So, there is a process that is being studied and tested, that is what we call FRBRization, a process to convert the old MARC format

catalogue into a new visualization of the records according to FRBR entities. The old catalogue is based on a single entity and the new model is based on four entities, this means that a lot of work has to be done to accomplish that transformation, like the need to split the MARC record into four entities. FRBRization experiments are being made all over the world and the process is developed based on the identification of FRBR entities that are present in the bibliographic records of the old catalogue, clustering search results according to those entities, and making OPAC displays look in conformity with the new model.

The FRBRization process is not an easy task, especially because inconsistent cataloguing practices makes this process difficult to be automatically achieved. Manual work has to be done to improve the results. FRBR exposes errors and inconsistencies that were previously largely unnoticed and are impediments to identifying FRBR entities with accuracy. However, research groups or individuals have developed algorithms that may help the process of FRBRization, though some manual work has to be done to improve results.

In the course of this work, our own experiment of FRBRization was explained in detail. On a first approach to FRBRization, we started with a relevant sample of bibliographic records, one that might be a guarantee of a good number of editions, translations, adaptations and so on. We chose the work "Os Maias" by Eça de Queirós and retrieved the data from PORBASE, the Portuguese Union Catalogue. The records were exported to an Excel sheet. It was with the help of this tool that we achieved the manual grouping into clusters. We have identified cluster types that could be used to group bibliographic records according to FRBR entities: work, expression and manifestation. Cluster types were identified based on relationships among items. The manual FRBRization was done, according to keys we have built and learned to use based on the identification of the entities on the old MARC format. Earlier research projects were a good school and we followed their work. This manual FRBRization gave us some knowledge how an automatic tool like an algorithm might be built and gave us the experience on how the whole process was done. This study builds on earlier research projects to deepen our understanding of FRBR and what can be done to convert millions of old records into new visualizations that display a new organization of information and create impact in the end users.

Results from this study and identification of the main FRBR issues that need to be addressed:

- The abstract nature of the FRBR model requires interpreting and experimenting efforts in the future.
- Definition of certain entities, particularly the entity expression. This entity has caused a fair amount of confusion.
- The need of a new entity the “Superwork”. The superwork is mainly to be used as a collocating device designed to link related works such as all Hamlet works, including texts, criticisms, adaptations, performances, movies, that are considered “new works”.
- More research needed to explore the success of clustering on other types of works using automatic FRBRization methods.
- New cataloguing rules that are convergent to the model: these have been developed and have just been published, called Resource Description and Access (RDA). How this is going to change library cataloguing we do not know yet, but great changes are expected.
- New FRBR-based record structures for new FRBR implementations: this is what we all expect, new automated systems inspired in the new model.
- New software to support the FRBRization process: this is a process that has to be guaranteed as the only way to convert the old legacy into new FRBR displays.
- New user studies to confirm that the new systems or OPAC displays benefit end users.
- Need to validate the FRBR model against real situations and different cultural communities to make sure the model is valid.
- Need to harmonize the library FRBized data with other data coming from other types of information services, for instance, archives and museums.

People who have grown up with the rich resources available in the Internet, are inclined to believe that Google can satisfy most of their information needs. Google is

friendly, but is not sufficient and library has a better expertise to organize information. So, Google and a library system are not concurrent but may complement each other. Searches on a range of file formats, such as PDF, for example, have hyperlinks that give immediate access to full text of documents on Google side, but if a user needs authority and quality of content he will make use of the library catalogue.

So, the libraries need to change the way their information is displayed. FRBR can actually be the solution to the problem. The so-called FRBR-tree that we find in Virtua system by VTLS, reminds us of the catalogue of the “old days”, and gives us an accurate graphic representation of the context for each new member of a “bibliographic family”.

FRBR has the merit to have focused everybody’s attention in the search interface and in the users’ interests. There is a lot of work being done to improve the library catalogue. The future of the catalogue, as a discovery system for users, has received great attention lately in library literature and will have developments soon. We found in the available literature reports with lists of requirements for library catalogues to become competitive in a digital environment. These reports also call for experimentation with FRBR. The need to promote the catalogue to the “network level” is clear. We should think of the catalogue as a tool able to promote connections between users and relevant resources, drawing users back from other discovery environments such as Google to the catalogue. This needs further investigation and experimentation, but it is a good start to see some authors thinking of the catalogues connecting with search engines. Connecting resources with end-users is an idea to foster and investigate in the future to come.

FRBR relationships will be very useful in this new digital networked environment. It will make browsing and navigation even easier through clustered bibliographic records, enhancing better access to information and promoting the linking of the end-users to the resources. These environments pose great new challenges.

Bibliography and References

- Aalberg, T. (2006). A Tool for Converting Bibliographic Records. Available: www.ercim.org/Ercim_News/enw66/aalberg.html
- Aalberg, Trond, Frank B. Haugen, and Ole Husby. (2006). A Tool for Converting MARC to FRBR. In *Lecture Notes in Computer Science: Vol. 4172. Research and Advanced Technology for Digital Libraries*.
- Aalberg, T. (2006). MARC and FRBR: *Match or Mismatch?* Available at: http://ru.is/kennarar/thorag/cataloguing2007/Trond_Aalberg.ppt
- Allgood, J. E. (2007). *Serials and Multiple Versions, or the Inexorable Trend Toward Work-Level Displays*. *Library Resources & Technical Services*, 51(3), pg. 160-178.
- Angelaki, Georgia (2008). *Research and Innovation in the European Library: the TELplus Project*. The Netherlands: The European Library Office.
- Anglo-American Cataloguing Rules, Second Edition, 2202 Revision, prepared under the direction of the Joint Steering Committee for Revision of AACR (Chicago: American Library Association, 2002).
- AUSTLIT – *Australian Literature Gateway* [on-line electronic resource]. Available at: <http://www.austlit.edu.au>
- Ayres, M.-L., Kilner, K., Fitch, K., and Scarvell, A. (2002). *Report on the Successful AustLit: Australian Literature Gateway Implementation of the FRBR and INDECS Event Models, and Implications for Other FRBR Implementations*. Available at: <http://www.ifla.org>. Available at: <http://www.ifla.org/IV/ifla68/papers/054-133e.pdf>
- Babeu, Alison. (2008). *Building a "FRBR-Inspired" Catalog: The Perseus Digital Library Experience*. Perseus Digital Library. Available at: <http://www.perseus.tufts.edu/~ababeu/PerseusFRBRExperiment.pdf>
- Bennett, R., B. Lavoie, B., and E. O'Neill. (2003). The Concept of a Work in WorldCat. *Library Collections, Acquisitions and Technical Services*, 27(1).
- Bowen, Jennifer. (2005). *FRBR: Coming soon to your library?* Accessed Apr 7, 2009. Available at: www.ala.org/alcts/alctsconted/presentations.htm.
- Calhoun, Karen. (2006). *The Changing Nature of the Catalogue and its Integration with other Discovery Tools*. Final Report prepared for the Library of Congress. Available at: <http://www.loc.gov/catdir/calhoun-report-final.pdf>
- Callewaert, R. (2008). FRBR in the Flemish Central Catalog. ALA. Available at: <http://www.slideshare.net/rcallewaert/frbr-in-open-vlacc-rosemie-callewaert-503758>.

- Carlyle, Allyson. (2006). "Understanding FRBR as a Conceptual Model: FRBR and the Bibliographic Universe," *Library Resources & Technical Services* 50, no. 4 (October 2006): 264-273.
- Coyle, Karen. (2004). *Future Considerations: the Functional Library System Record*, *Library Hi Tech* 22, no.2.
- Dellit, A. and T. Boston. (2007). *Relevance Ranking Of Results From MARC-Based Catalogues: From Guidelines To Implementation Exploiting Structured Metadata*. Paper presented at Information Online 2007, <http://1101.nla.gov.au/InfoOnline2007Paper.html>
- Dempsey, Lorcan. (2006). *The Library Catalogue in the New Discovery Environment: some thoughts. Ariadne*. Available at: <http://www.ariadne.ac.uk/issue48/dempsey/>
- Doerr, M. (2007). *Modelling Intellectual Processes: The FRBR - CRM Harmonization*. DELOS 2007. Available at: http://cidoc.ics.forth.gr/docs/doi_le_boeuf.pdf
- Doerr, M. and P. LeBoeuf. (2007). *FRBR: Object-Oriented Definition and Mapping to the FRBR-ER*. Version 0.8.3. Draft. International Working Group on FRBR and CIDOC CRM Harmonisation, available at: http://cidoc.ics.forth.gr/docs/frbr_oo/frbr_docs/FRBR_oo_V0.8.3.pdf
- FRBR Report, see, IFLA Study Group on the Functional Requirements for Bibliographic Records. (1988). *Functional requirements for bibliographic records: final report*. UBCIM publications, new ser., V. 19. München: K. G. Saur. Available at <http://www.ifla.org/VII/s13/frbr/index.htm>
- FRBR Work-Set Algorithm. (2005). Accessed Dec.2, 2009. Available at: <http://www.oclc.org/research/activities/past/orprojects/frbralgorithm/algorithm.htm>
- Freire, N., Galvão, R., Lopes, M. (2009) - *FRBR information discovery in traditional catalogues: the TELplus experience*. Available at: <http://www.ifla.org/files/hq/papers/ifla75/135-freire-en.pdf>
- IFLA Study Group on the Functional Requirements for Bibliographic Records. (1988). *Functional requirements for bibliographic records: final report*. UBCIM publications, new ser., V. 19. München: K. G. Saur. Available at <http://www.ifla.org/VII/s13/frbr/index.htm>
- Hegna, K., and Murtomaa, E. (2003). *Data Mining MARC to find FRBR?* *International Bibliographic Control*. 32, 52-55. Also available at: <http://folk.uio.no/knuthe/dok/frbr/datamining.pdf>. (Accessed Decemcer, 2nd., 2009).
- Hickey T. B., O'Neill, Toven, J.. (2002). *Experiments with the IFLA Functional Requirements for Bibliographic Records (FRBR)*
- Hickey T. B., Toven, J.(2005). *FRBR Work-Set Algorithm*. OCLC, April.
- Le Bœuf, P. (2005). *Functional requirements for bibliographic records (FRBR): Hype or cure-all?* Binghamton, NY: Haworth Information Press.
- Lubetzky, Seymour. (1960). *Code of Cataloguing Rules: Author and Title Entry*. An Unfinished Draft for a New Edition of Cataloguing Rules Prepared for the Catalog Code Revision Committee (Chicago: ALA, 1960).
- Lubetzky, Seymour. (2001). *Writings on the Classical Art of Cataloguing* (Englewood, Colo.: Libraries Unlimited).
- Maxwell, R. L. (2008). *FRBR: A guide for the perplexed*. Chicago: American Library Association.
- Noerr, P., Goossens, P., Matei, D., Otten, P., Peruginelli, S., and Witt, M. (1998) *User benefits from a new bibliographic model: follow-up of the IFLA functional requirements*

- study. 64th IFLA general conference. Available online at: <http://www.ifla.org/IV/ifla64/084-126e.htm>
- O'Neill, A. (2007). *The Impact of Research on the Development of FRBR*. In Understanding FRBR: what it is and how it will affect our retrieval tools. Westport, Conn: Libraries Unlimited.
- O'Neill, Edward T. (2002). *FRBR: Functional Requirements for Bibliographic Records: Application of the Entity-Relationship Model to Humphry Clinker*. OCLC., available at: http://www.oclc.org/research/publications/archive/2002/oneill_frbr22.pdf
- Panizzi, Antonio (1841). "Rules for Compilation of the Catalogue," in British Museum, The Catalogue of Printed Books in the British Museum (London, 1841).
- Pisanski, J and M. Žumer. (2007). *Functional Requirements For Bibliographic Records: An Investigation Of Two Prototypes*. Program: Electronic Library & Information Systems, 41 (4).
- Salaba, A., Zhang, Y. *From a Conceptual Model to Application and System Development*, In Assist& – The Information Society for the Information Age, Bulletin, August/September 2007. Available at: http://www.asis.org/Bulletin/Aug-07/salaba_zhang.html
- Smiraglia, R.P. (2001) *The "nature" of a work: implications for the organization of knowledge*. Lanham: Scarecrow, p. 15
- Smiraglia, R. P. (2002). *Works as entities for information retrieval*. Binghamton, N.Y.: Haworth Information Press.
- Svenious, Elaine. (2002). *The Intellectual Foundation of Information Organization*. Cambridge Press:MIT Press.
- Taylor, A. G. (2007). Understanding FRBR: What it is and how it will affect our retrieval tools. Westport, Conn: Libraries Unlimited.
- Teixeira, T. [et al.] – Report on FRBR experiments: TELplus deliverable n. D-3.6. July 2008. 23 p.
- Tillet, B.B. (1987). Bibliographic relationships: toward a conceptual structure of bibliographic information used in cataloging. Ph.D. dissertation: University of California, Los Angeles
- Tillet, B.B. (1991). A taxonomy of bibliographic relationships. *Library resources and technical services* 35(2):150-158
- Tillet, B.B. (1994). IFLA Study on the Functional Requirements of Bibliographic Records: Theoretical and Practical Foundations. Speech delivered at the 1994 IFLA Conference held in Havana, Cuba.
- Yee, M. M. (2005). Frbrization: a Method for Turning Online Public Finding List into Online *Public Catalogue*. *Information Technology and Librarians*. 24 (3), 77-95. Available at: <http://repositories.cdlib.org/postprints/715>
- Zhang, Y., & Salaba, A. (2009). *Implementing FRBR in libraries: Key issues and future directions*. New York: Neal-Schuman.
- Zumer, Maja. (2007) FRBR: The End of the Road or a New Beginning. *ASIS&T Bulletin*, August/September, 2007. Available at: <http://www.asis.org/Bulletin/Aug-07/Zumer.pdf>
- Zumer, M. (2005). *Implementation of FRBR: European Research Initiative*. *CATALOGING AND CLASSIFICATION QUARTERLY*. 39, 229-238. Available at: Digital Object Identifier: 10.13.00/J104v39n03_14