

ECP-2006-DILI-510003

TELplus

Report on FRBR experiments

Deliverable number	<i>D-3.6</i>
Dissemination level	<i>Public</i>
Delivery date	<i>29 July 2008</i>
Status	<i>Final v1.4</i>
Author(s)	<i>Teresa Teixeira (IST), Margarida Lopes (BNP), Nuno Freire(BNP), José Borbinha (IST)</i>



eContentplus

This project is funded under the *eContentplus* programme,
a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable.



Contents

1	OBJECTIVES.....	3
2	INTRODUCTION TO FRBR.....	4
2.1	HISTORICAL BACKGROUND	4
2.2	FRBR AS A USER-FOCUSED APPROACH.....	4
3	FRBRIZATION.....	7
4	EFFECTIVE SERVICES.....	9
4.1	AUSTLIT: AUSTRALIAN LITERATURE GATEWAY.....	9
4.2	PERSEUS DIGITAL LIBRARY (PDL)	10
4.3	UCLA LIBRARY – FILM AND TELEVISION ARCHIVE.....	12
5	PRODUCTS & TECHNOLOGY	13
5.1	VIRTUA ILS (INTEGRATED LIBRARY SYSTEMS).....	13
5.2	AQUABROWSER.....	14
6	RESEARCH.....	15
6.1	SUPPORTING TOOLS, ALGORITHMS AND TECHNIQUES	15
6.1.1	<i>OCLC FRBR Work-Set Algorithm.....</i>	<i>15</i>
6.1.2	<i>BibSys FRBR conversion tool</i>	<i>15</i>
6.1.3	<i>MARC and the FRBR Display Tool.....</i>	<i>15</i>
6.2	OCLC FRBR PROTOTYPES	16
6.2.1	<i>OCLC FictionFinder</i>	<i>16</i>
6.2.2	<i>Curiouser.....</i>	<i>17</i>
6.2.3	<i>The FRBRization of Humphry Clinker</i>	<i>17</i>
6.3	THE NDB PROTOTYPE.....	18
6.4	BIBSYS	19
6.5	OTHER PROJECTS	21
7	REFERENCES	22

1 Objectives

This deliverable presents an overview of relevant projects, services and technology that reuse traditional bibliographic data in order to build repositories or associated OPAC systems aligned with the FRBR principles.

We are not providing an exhaustive bibliography or list of references of work related with FRBR, as many already exist. Our purpose is only to highlight what we identified as more relevant for supporting the coming up tasks in TELplus of analysis and design of a technical solution for a new FRBR search service for The European Library portal, reusing the bibliographic records of the partner libraries.

The remainder of this document is organized as follows: section 2 provides an introduction to the history and basic concepts of FRBR; section 3 presents the concept of FRBRization; section 4 presents a selection of effective services making use of FRBR concepts; section 5 lists the relevant technological products existing in the market and section 6 summarizes the main open developments and research efforts in course or held so far.

2 Introduction to FRBR

2.1 Historical background

By the late 1980s professionals in the library world, including members of the IFLA Section on Cataloguing, realized that great 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, needed a careful reflection.

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 Control, 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 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.»*

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 1998[1]. **We assume that the readers of this deliverable are familiar with this report.** Anyway, a simple synthesis is provided in the next section.

2.2 FRBR as a user-focused approach

From the very beginning, the FRBR study had the purpose to focus on the needs of users when they do search tasks, especially to:

- 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.

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 comprises the products of intellectual or artistic endeavour that are named or described in bibliographic records: Work, Expression, Manifestation and Item. The second group comprises those entities responsible for the intellectual or artistic content, the physical production and dissemination, or the custodianship of such products: person and corporate body. The third group comprises an additional set of entities that serve as the subjects of intellectual or artistic endeavour: concept, object, event and place.»

Below we reproduce the mentioned Figure 1, for our convenience:

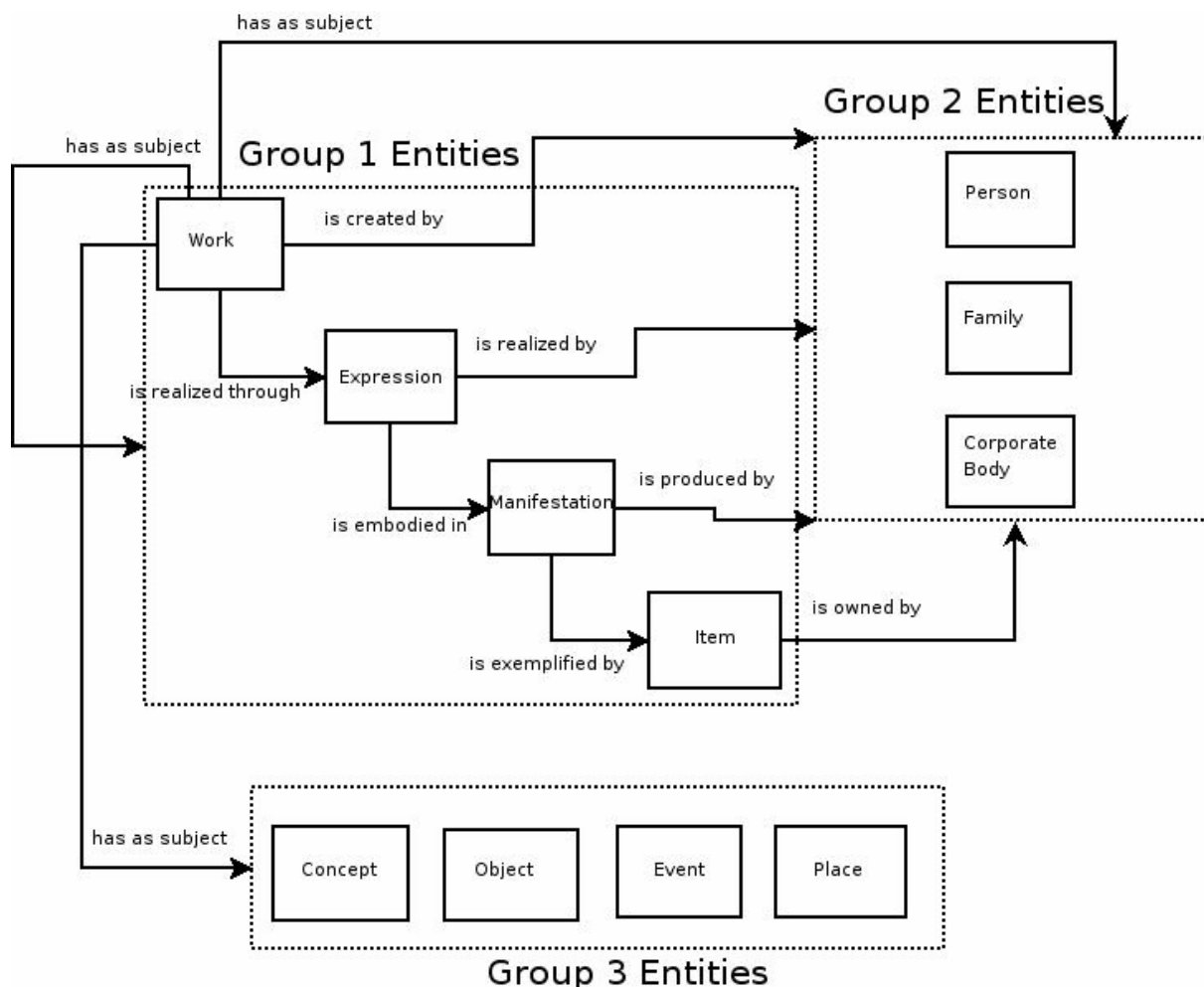


Figure 1 - The FRBR model¹

¹ Image reproduced from <http://www.miskatonic.org>

The goal of the initial FRBR study was to produce a conceptual model that would serve as the basis for relating specific attributes and relationships to the various tasks that users perform when they search information by means of bibliographic records. It is thus based in the development of a framework that identifies and clearly defines the entities of interest to users of bibliographic records, the attributes of each entity, and the types of relationships that operate between entities.

From the final report we extracted:

«The study makes no a priori assumptions about the bibliographic record itself, either in terms of content or structure. It takes a user-focused approach to analyzing data requirements insofar as it endeavours to define in a systematic way what it is that the user expects to find information about in a bibliographic record and how that information is used.» ... «The model operates at the conceptual level; it does not carry the analysis to the level that would be required for a fully developed model.».

3 FRBRization

The publication of the FRBR report started an active discussion and raised the interest for many theoretical and research activities. To follow that discussion and results, relevant references are the “FRBR Bibliography”¹, maintained by FRBR Review Group, and “The FRBR Blog”².

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 this section, we review the literature that addresses the FRBRization of existing catalogues, or in other words, OPAC systems aligned with the FRBR principles.

Yin Zhang and Athena Salaba[2] 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: The FRBR model; FRBR-related standards; FRBR applications; FRBR System development; and FRBR research. 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.

Another review of how current MARC records can be used to promote FRBRization can be found in a recent article by Martha Yee[18]. 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[3]. 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 Murtomaa[4] have detailed their efforts in analyzing MARC records to determine what attributes could best be used for automatic conversion to FRBR.

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

¹ <http://www.ifla.org/VII/s13/wgfrbr/bibliography.htm>

² <http://www.frbr.org/>

any sort of meaningful way. In FRBR literature, this problem is often referred to as the **multiple versions problem**. According to Allgood[5], «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». 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[6], «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. 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».

4 Effective Services

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.

4.1 AustLit: Australian Literature Gateway

AustLit - The Australian Literature Resource, described in [7], is a cooperative service involving eight universities and the National Library of Australia. It consists of a database designed “from scratch” according to FRBR conceptual model. It also involved the conversion of existing bibliographic data, but details about it are not available in the literature.

Unfortunately, the AustLit service is restricted, so we were not able to evaluate it independently, so we can rely only in the literature.

The AustLit data model[8] included the four FRBR group 1 entities: the Work, the Expression, the Manifestation and the Item. And it added 'event-modelling' to the bibliographic description, so Works have a creation event, Expressions have a realisation 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 desired functionality 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. The FRBR model was written with a “whole monograph” in mind. 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.

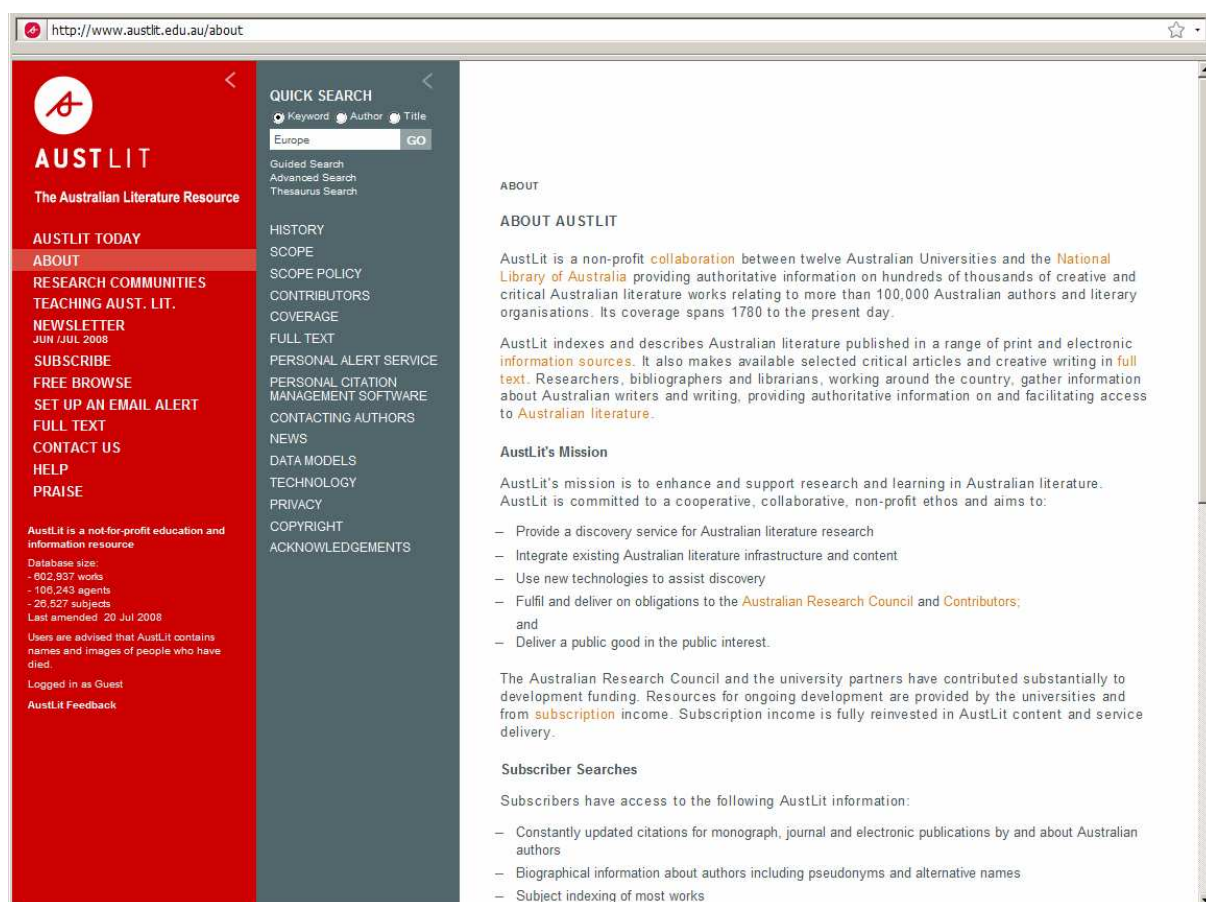


Figure 2 - The AustLit homepage

4.2 Perseus Digital Library (PDL)

The main goal of the Perseus Digital Library¹ 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 could be done using already existing metadata standards and freely available online collections.

According to Allison Babeu[9], «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». 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”,

¹ <http://www.perseus.tufts.edu/>

a larger class that encompassed twelve different Work types. A new system is under development, in a new activity started in 2008.



Figure 3 - The PERSEUS Digital Library Project

4.3 UCLA Library – Film and Television Archive

The UCLA Film and Television Archive¹ constitutes 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. To make the most of this new display, the cataloguing is done according to the new framework, though in a conventional 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 on², a preview of which can be accessed at <http://cinema.library.ucla.edu>.

Currently, one can see the organization of the information of the catalogue according to FRBR Group 1, clustered in Expressions, then in Manifestations. A screen capture of the user interface can be seen in Figure 4.



Figure 4 – User interface of the UCLA Library – Film and Television Archive

¹ <http://cinema.library.ucla.edu/>

² In FRBR blog: available at: <http://www.frbr.org/2007/01/16/midwinter-implementers#comments>

5 Products & Technology

As far as it was 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 is displayed in conformity to FRBR principles.

5.1 Virtua ILS (Integrated Library Systems)

VTLS¹ (Visionary Technology in Library Solutions) released in 2002 version 41.0 of the Virtua library system. 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 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. Figure 5 shows a screenshot from a presentation of the Virtua OPAC.

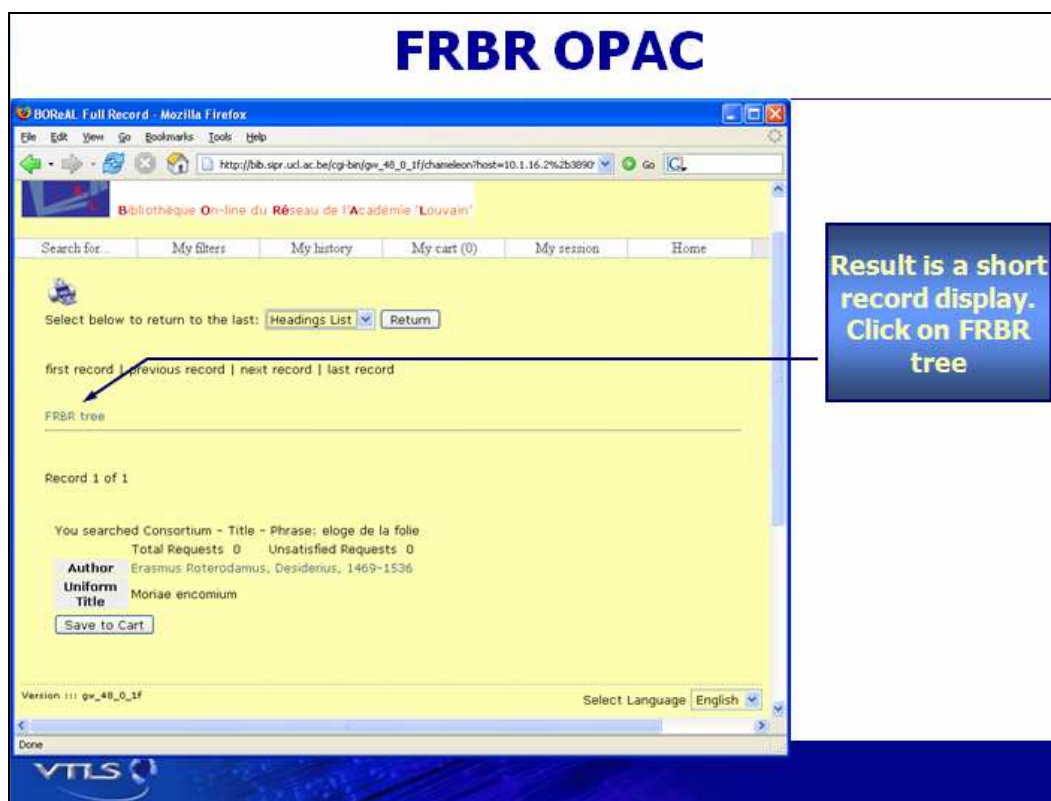


Figure 5 - Virtua OPAC²

¹ http://www.vtls.com/about_us

² Image reproduced from http://ru.is/kennarar/thorag/cataloguing2007/John_Espley.ppt

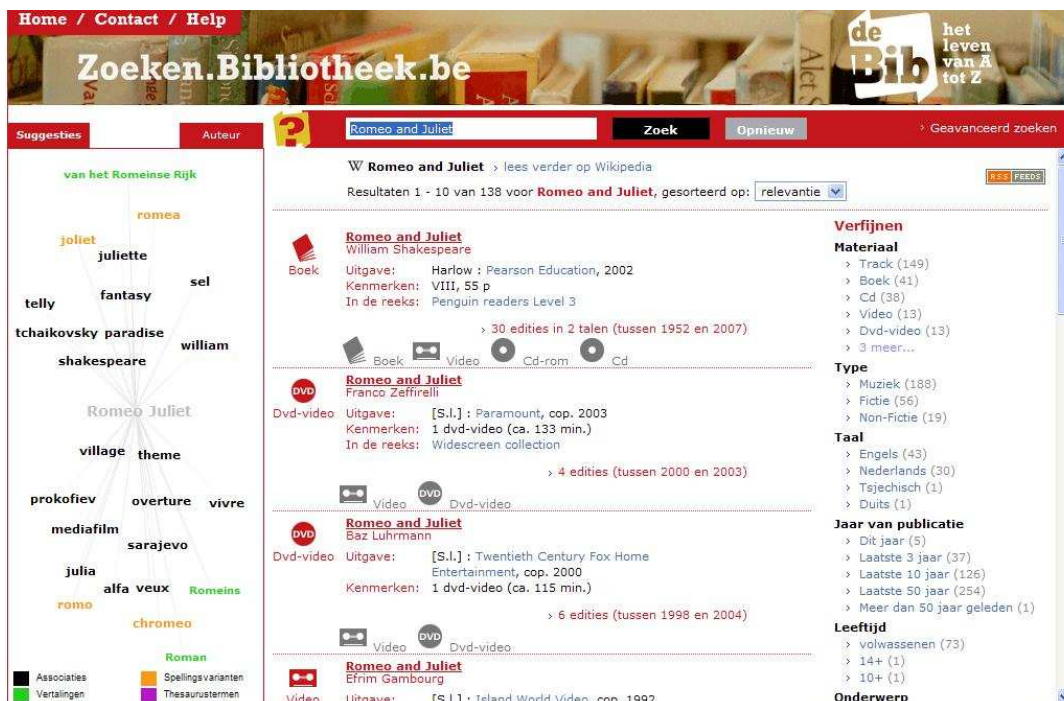
Virtua allows “flat records” and “FRBR records” to live side-by-side. The pattern followed when “splitting” records is based on the previous work by the Library of Congress (see 6.1.3). The cataloguing paradigm is still based on ISBD rules (the “Manifestation level record”), therefore it is not substantially different from any “traditional” ISBD-based record, and the cataloguing format is still basically a MARC format.

Virtua displays on the left part of the screen, the “FRBR tree”, showing the uniform title with respective Expressions, and records for Manifestations. On the right part of the screen, a complete bibliographic record for one of the Manifestations, with a link to information about Items.

5.2 AquaBrowser

AquaBrowser is a product of MediaLab¹. It can be seen when using the Flemish Central Catalogue where search results are displayed according to FRBR requisites (Figure 6).

The FRBR Work key that is produced on the AquaBrowser platform is extracted from the MARC 21 records. In AquaBrowser the key is handled as an index-entry for grouping the search results in the middle of the screen on the result list. Cataloguing is also done in a way that allows for the best FRBR display.



The screenshot shows the AquaBrowser search interface for the query "Romeo and Juliet". The interface is divided into several sections:

- Navigation:** Home / Contact / Help, Zoeken.Bibliotheek.be, de Bib het leven van A tot Z.
- Search Bar:** Search for "Romeo and Juliet" with options for "Zoek" and "Opnieuw".
- FRBR Tree (Left):** A hierarchical tree structure showing relationships between terms like "romeo", "juliet", "juliette", "sel", "tchaikovsky", "paradise", "shakespeare", "william", "Romeo Juliet", "village", "theme", "prokofiev", "overture", "vivre", "mediafilm", "sarajevo", "julia", "alfa", "veux", "Romeins", "romo", "chromeo", "Roman".
- Search Results (Center):**
 - W Romeo and Juliet > lees verder op Wikipedia
 - Resultaten 1 - 10 van 138 voor Romeo and Juliet, gesorteerd op: relevantie
 - Book: **Romeo and Juliet** by William Shakespeare. Uitgave: Harlow : Pearson Education, 2002. Kenmerken: VIII, 55 p. In de reeks: Penguin readers Level 3. > 30 edities in 2 talen (tussen 1952 en 2007).
 - Dvd-video: **Romeo and Juliet** by Franco Zeffirelli. Uitgave: [S.I.] : Paramount, cop. 2003. Kenmerken: 1 dvd-video (ca. 133 min.). In de reeks: Widescreen collection. > 4 edities (tussen 2000 en 2003).
 - Dvd-video: **Romeo and Juliet** by Baz Luhrmann. Uitgave: [S.I.] : Twentieth Century Fox Home Entertainment, cop. 2000. Kenmerken: 1 dvd-video (ca. 115 min.). > 6 edities (tussen 1998 en 2004).
 - Video: **Romeo and Juliet** by Efrim Gamburg. Uitgave: [S.I.] : Island World Video, cop. 1992.
- Refinement (Right):**
 - Verfijnen**
 - Materiaal:** Track (149), Book (41), Cd (38), Video (13), Dvd-video (13), 3 meer...
 - Type:** Muziek (188), Fictie (56), Non-Fictie (19)
 - Taal:** Engels (43), Nederlands (30), Tsjechisch (1), Duits (1)
 - Jaar van publicatie:** Dit jaar (5), Laatste 3 jaar (37), Laatste 10 jaar (126), Laatste 50 jaar (254), Meer dan 50 jaar geleden (1)
 - Leeftijd:** volwassenen (73), 14+ (1), 10+ (1)
 - Onderwerp:**

Figure 6 – The AquaBrowser search interface

¹ <http://www.aquabrowser.com/>

6 Research

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

6.1 Supporting Tools, Algorithms and Techniques

6.1.1 OCLC FRBR Work-Set Algorithm

The FRBR Work-Set Algorithm is an algorithm for clustering MARC21 bibliographic records at the FRBR 'Work' level¹. Having resources brought together under the "Works" umbrella enables users to navigate through the myriad resources available.

The algorithm is available for anyone to download and experiment with and we have identified several projects that apply it. It concentrates on bringing together clusters of records at the Work level. One reason for doing this is was that there is still some disagreement about what constitutes an Expression.

OCLC has used the algorithm to make experiments to FRBRize the WorldCat database.

6.1.2 BibSys FRBR conversion tool

Aalberg et al.[10] have researched the creation of tools for converting MARC data to the FRBR model. Their basic process involved identifying the different entities in a Marc record, selecting the MARC fields that describe each entity, and finding relationships between entities. The conversion tool they 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 with FRBR type attributes and a relationship element". The tool was applied to 4.000.000 records in the BibSys database, and they learned during this conversion process that the major issues faced were inconsistent catalogue data and scalability issues. While the authors concede that it is as yet impractical for libraries to attempt full conversion to a FRBR based data model, they believe that conversion tools such as this one can at least help create FRBRized views for catalogues.

6.1.3 MARC and the FRBR Display Tool

In 2002 the Library of Congress commissioned a study to examine the MARC 21 bibliographic and holdings formats against the FRBR. The results of the study² have been improved over the last years, and the last revision dates from 2006. It includes a proposal to

¹ <http://www.oclc.org/research/software/frbr/default.htm>

² <http://www.loc.gov/marc/marc-functional-analysis/functional-analysis.html>

map MARC 21 elements to FRBR, which also motivated the development of a FRBR Display Tool¹.

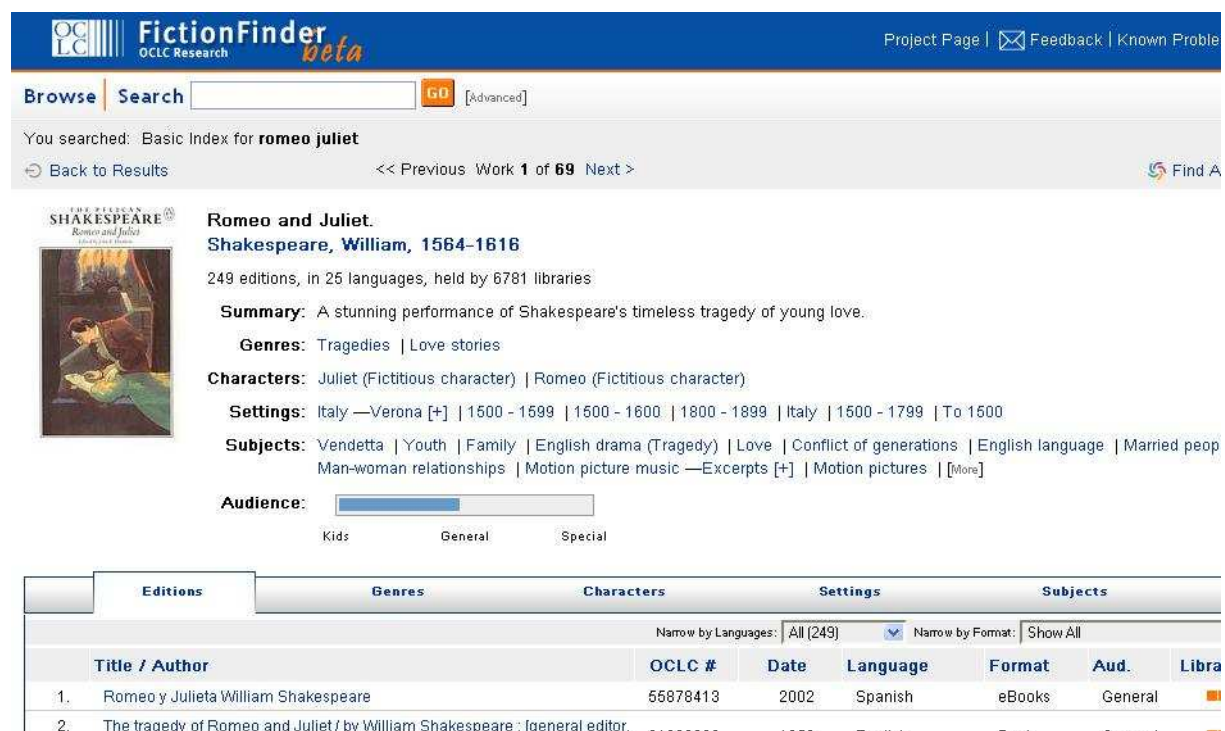
The FRBR Display Tool is coded in XSLT. It can take a group of MARC bibliographic records and cluster them into meaningful displays of Works, Manifestations, and Expressions.

6.2 OCLC FRBR prototypes

Some of the earliest experiments with FRBRization were conducted by members of OCLC Research², who in a number of articles have related information on experiments with algorithms to group bibliographic records in WorldCat into Works and Expressions[11].

6.2.1 OCLC FictionFinder

FictionFinder³ 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 provides a FRBR-inspired view of the data (Figure 7).



The screenshot shows the FictionFinder beta interface. At the top, there is a search bar with the text "You searched: Basic Index for romeo juliet". Below the search bar, there are navigation links: "Back to Results", "<< Previous", "Work 1 of 69", and "Next >". The main content area displays the title "Romeo and Juliet" by Shakespeare, William, 1564-1616. It includes a summary: "A stunning performance of Shakespeare's timeless tragedy of young love." Genres listed are "Tragedies" and "Love stories". Characters listed are "Juliet (Fictitious character)" and "Romeo (Fictitious character)". Settings include "Italy —Verona [+]", "1500 - 1599", "1500 - 1800", "1800 - 1899", "Italy", "1500 - 1799", and "To 1500". Subjects include "Vendetta", "Youth", "Family", "English drama (Tragedy)", "Love", "Conflict of generations", "English language", "Married people", "Man-woman relationships", "Motion picture music —Excerpts [+]", and "Motion pictures". An audience slider is set to "General". Below the main information, there is a table with tabs for "Editions", "Genres", "Characters", "Settings", and "Subjects". The table has columns for "Title / Author", "OCLC #", "Date", "Language", "Format", "Aud.", and "Libra".

Editions		Genres	Characters	Settings	Subjects	
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,	01609669	1950	English	Book	General	■

Figure 7 - FictionFinder

¹ <http://www.loc.gov/marc/marc-functional-analysis/tool.html>

² <http://www.oclc.org/research/projects/frbr/default.htm>

³ <http://fictionfinder.oclc.org/>

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/alices 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 cluster¹.

Pisanski and Zumer[12] analysed 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».

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.

6.2.2 Curiouser

Curiouser² is an approach to making the best use of data available about Items in WorldCat, using a user interface for exploring and selecting Works and Items. The project integrates previously developed and new techniques for display and navigation of records in a FRBR context. The Curiouser project will produce a prototype system that provides a Works-based view of Open WorldCat (OWC) content. It employs the OCLC FRBR Work-set algorithm and FictionFinder.

6.2.3 The FRBRization of *Humphry Clinker*

This work focused on the identification of Expressions. The team working in the project first manually extracted from WorldCat a set of records representing the monograph “Humphrey Clinker”[13], and analysed 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 O’Neill[13] «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 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

¹ <http://www.oclc.org/research/projects/frbr/fictionfinder.htm>

² <http://www.oclc.org/research/projects/curiouser/default.htm>

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». 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 "no matter how minor" are considered to be new Expressions, the granularity of resulting Expressions is too fine and can be almost indistinguishable from that of Manifestations. The Author concludes 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».

6.3 The NDB Prototype

The NDB Prototype¹ 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 Australia². Pisanski and Zumer[12] also analysed 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 “Super Work” where top level records were grouped together, and used form and language attributes to differentiate between numerous Works and Expressions». 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 8 shows an example for a structure of results for the “Super Work” Hamlet.

The results of this prototype have been discussed by several authors[13] stating that it is an example of the difficulties in extracting FRBR structure from MARC records, and the uncertainty on the boundaries of Expressions.

¹ <http://l101.nla.gov.au>

² <http://librariesaustralia.nla.gov.au/apps/kss>



Figure 8 - Interface of the NDB prototype

6.4 BIBSYS

Experimental FRBRization of the Norwegian BIBSYS database¹ – joint project with the participation of BIBSYS, the Norwegian University of Science and Technology (NTNU) and the National Library of Norway. Data from these two national bibliographies were analyzed following the FRBR model.

¹ <http://www.bibsys.no>



The screenshot shows the BibSys search results page for the query 'hamlet'. The interface includes a navigation bar with 'Ask - Treffliste' and a search bar. Below the search bar, there are statistics for different sources: Bibliotekbasen: 847, Emneportal: 0, ForskDok publikasjoner: 9, and ForskDok prosjekter: 0. The results are displayed in a table with columns for 'Tittel', 'Forfatter', 'År', and 'Materiale'. The table shows 8 results, with the first one being 'Hamlet' by Shakespeare, William, published in 2008. The results are sorted by year, and the total number of results is 847.

<input type="checkbox"/> Tittel	Forfatter	År	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 9 - The BibSys interface

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.¹

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 Finish 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.

¹ <http://folk.uio.no/knuthe/dok/frbr/>

- 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 (Figure 9);
- the user interface should be multilingual (English, Finnish and Norwegian) and handle different MARC formats (MARC21, FINMARC, NORMARC).

Hegna and Murtoomaa[4] 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.

6.5 Other projects

Perhaps the most significant effort in remodelling FRBR is the current effort to harmonize the CIDOC Conceptual Reference Model (CIDOC CRM) and FRBR through the creation of the ontology FRBRoo[16]. While FRBRoo is still very much in its early stages, a recent article, by Martin Doerr[15] has looked at some of its potential for libraries, with the specific example of the European Digital Library. Doerr advocates that the best strategy for creating a core ontology that will enable semantic interoperability in the library community is through the harmonization of various alternatives such as the CIDOC CRM and FRBR. These efforts started with the first European Library Automation Group (ELAG) discussions immediately after the FRBR study was published. Maja Zumer[17] explains that «The discussion in 1999 focused on detailed investigation of entities, particularly those of the first group. The conclusion was that there are, in reality, seven levels, because parts of Works, Expressions, and Manifestations should be introduced as separate intermediate levels to Work, Expression, Manifestation, and Item. In addition, there was a recommendation for an additional highest level, a so-called “top-hat” or “Super Work”, describing the original work from which other works are derived». A similar concept was advocated by Yee[18].

Library and museum communities might benefit from harmonising the two models: FRBR and CIDOC CRM. With this step, it is intended to engage in an open dialogue with all interested parties about the combined CRM and FRBRoo model, its final shaping, its application and other further work^{1,2}.

According to Patrick Le Boeuf[19] «The Intellectual rigor and demands of Martin Doerr, the main originator of the CIDOC CRM, and his expertise in modelling techniques, are very much appreciated and help us chase out every assumption that had been left implicit in the entity-relationship formalization of the model. FRBR will never look the same again.»

¹ http://cidoc.ics.forth.gr/frbr_inro.html

² http://cidoc.ics.forth.gr/frbr_graphical_representation/graphical_representation/graphical_representation.html

7 References

- [1] 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>
- [2] Salaba, A., Zhang, Y. *From a Conceptual Model to Application and System Development*, In Assis&t – The Information Society for the Information Age, Bulletin, August/September 2007. Available at: http://www.asis.org/Bulletin/Aug-07/salaba_zhang.html
- [3] 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>
- [4] Hegna, K. ,and Murtomaa, E. (2003). Data Mining MARC to find FRBR? *INTERNATIONAL CATALOGUING AND BIBLIOGRAPHIC CONTRO*. 32, 52-55.
- [5] 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.
- [6] 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.
- [7] AUSTLIT – Australian Literature Gateway [on-line electronic resource]. Available at: <http://www.austlit.edu.au>
- [8] 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>
- [9] 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>
- [10] Aalberg, T. (2006). *MARC and FRBR: Match or Mismatch?* Available at: http://ru.is/kennarar/thorag/cataloguing2007/Trond_Aalberg.ppt
- [11] 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).
- [12] Pisanski, J and M. Žumer. (2007). *Functional Requirements For Bibliographic Records: An Investigation Of Two Prototypes*. Program: Electronic Library & Information Systems, 41 (4).
- [13] 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
- [14] 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://il01.nla.gov.au/InfoOnline2007Paper.html>
- [15] Doerr, M. (2007). *Modelling Intellectual Processes: The FRBR - CRM Harmonization*. DELOS 2007. http://cidoc.ics.forth.gr/docs/doer_le_boeuf.pdf
- [16] 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. http://cidoc.ics.forth.gr/docs/frbr_oo/frbr_docs/FRBR_oo_V0.8.3.pdf

- [17] 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
- [18] 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>
- [19] Le Boeuf, P. (2005). FRBR: Hype or Cure-All? Introduction. *CATALOGING AND CLASSIFICATION QUARTERLY*. 39, 1-14.