



FRBR information discovery in traditional catalogues: the TELplus experience

Nuno Freire,
Rosa Galvão and
Margarida Lopes
National Library of Portugal
Lisbon, Portugal

Meeting: 135. UNIMARC

WORLD LIBRARY AND INFORMATION CONGRESS: 75TH IFLA GENERAL CONFERENCE AND COUNCIL
23-27 August 2009, Milan, Italy
<http://www.ifla.org/annual-conference/ifla75/index.htm>

“The European Library exists to open up the universe of knowledge, information and culture of all Europe’s national libraries”¹.

ABSTRACT

This paper gives an account of the experiment to develop an FRBR repository for the European Library, being carried out in the framework of the TELPlus Project. Following a brief review of the existing FRBR experiments, a number of issues for the FRBR implementation in bibliographic databases was identified that provided a foundation for the development in the TELPlus Project. The system design, the results achieved so far, as well as the final deliverables and results expected, are explained.

INTRODUCTION

The European Library (TEL)² was created in 2005 as a service of the Conference of the European National Libraries (CENL)³, aiming to offer a common access point to the distributed collections of the 48 national libraries of Europe in 32 languages. The European Library currently gives access to 150 million entries across Europe, concerning resources that can be both digital and bibliographical, and is a free service.

¹ Mission of the European Library. Available at: www.theeuropeanlibrary.org

² The European Library portal is available at: <http://www.theeuropeanlibrary.org>

³ Members of CENL are the national librarians of the Council of Europe member states. Further information about CENL is available at: <http://www.cenl.org/>

The European Library of today is the result of a number of earlier successful projects:

- Project Gabriel - Gateway and Bridge to Europe's National Libraries: It ran during 1996-2004 and established a joint web portal of European national libraries;
- TEL Project (The European Library: Gateway to Europe's Knowledge): It ran during 2001-2004, and laid down the foundations of current service. The project had the participation of 9 national libraries: Finland, Germany, Italy (Florence), Italy (Rome), Netherlands, Portugal, Slovenia, Switzerland and United Kingdom.

The success of The European Library has led to further developments and improvements. The number of national libraries members grew from the initial 9 to the current 48. Its growth was supported by European co-funded project. TEL-ME-MOR (2005-2007), had as its main goal to support, stimulate and facilitate the participation of the 10 national libraries from the New Member States, which were partners in the project; the EDL project (2006-2007) worked towards the integration of additional National Libraries and in enhancing the multilingual capabilities of The European Library portal. Both projects have contributed to the inclusion in the service of the collections of 19 national libraries.

TELplus is another project from The European Library. It is being taken forward by a consortium of 26 partners, including national libraries and research centres. TELplus aims at providing value-adding services and products for The European Library. Its primary objectives are to work towards improvements to the service architectural performance, the enhancement of multilingual access functionalities, and providing access to full-text documents originating from OCR processes. One of the lines of work of TELplus is to investigate solutions to support alternative services of searching and browsing in TEL according to the FRBR paradigm.

2. TELPLUS PROGRAMME OF WORK

The TELplus project is structured in 8 different work packages (WPs)⁴. We will be focusing in Work package 3 "Improving accessibility".

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 3 "Setting up a FRBR repository for the European Library" the one we will be focusing on. This task is led 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.

3. SETTING UP AN FRBR REPOSITORY FOR THE EUROPEAN LIBRARY

"The first [objective of FRBR] is to provide a clearly defined, structured framework for relating the data that are recorded in bibliographic records to the needs of the users of those records."⁵

⁴ Further information about TEplus programme of work available at:
<http://www.theeuropeanlibrary.org/telplus/workplan.php>

⁵ IFLA Study Group on the FRBR - *Functional Requirements for Bibliographic Records: Final Report*. K. G. Saur, 1998 (UBCIM Publications. New Series, 19).

The generic aim of this task is investigate possible improvements for search and accessibility of the contents of TEL through the implementation of the FRBR model in the search portal. Considering that the contents of TEL now amount to about 150 million catalogue entries in 32 different languages, an FRBR implementation has the potential to greatly improve the usability of the TEL portal. An FRBR service would narrow down search results by clustering similar records for the user: in the case of a well-known writer like Steinbeck, the search result for his novel “The grapes of wrath” will be clustered into a single answer, corresponding to the concept of “work”, in which the user can navigate afterwards, instead of having to browse each collection from each national library to access results.

3.1. Literature review on FRBR experiments

Implementation of the FRBR model is not easy and there are not many tools supporting it. Our first task, therefore, was to review existing literature on FRBR experiments and implementation tools⁶.

Reviewing existing literature allowed us to identify two sets of information, crucial to the development of the project:

- Identification of FRBR implementation projects in course, or already concluded;
- Identification of recurrent problems in FRBR implementation in bibliographic catalogues, mentioned in the reviewed literature.

As for the first set of information – identification of FRBR implementation projects in course, we concluded that several implementations of the FRBR model have already been tested, usually with small and well defined collections. Our implementation project differs significantly from the ones we studied in three aspects: it deals with a vast amount of information from several libraries; it deals with bibliographic records just as they are present in several flavours of MARC (MARC21, UNIMARC, etc.); and it deals with a multilingual context. The context of TEL is therefore extremely heterogeneous in data formats, bibliographic practices (which typically do not include FRBR inspired practices) and language of both materials and cataloguing.

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 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.

As for problems with FRBR implementation in bibliographic catalogues, we have identified 6 major issues:

Current MARC databases structure

FRBR makes no recommendation as to actual database structure cataloguing, but given the entity-relationship structure of the FRBR model, the logical step would be to adopt an entity relationship database structure, which would mean a radical departure from our current MARC based databases.

Presently, we have to work with the current database information structure. But the FRBR represents a very rich semantic model. It is not an easy task, to extract the FRBR semantics from the existing bibliographic catalogues.

⁶ Teixeira, Teresa [et al.] – *Report on FRBR experiments: TELplus deliverable n. D-3.6*. July 2008.

Extracting entities from the bibliographic records

It's difficult to extract and separate the four Group 1 entities from the single MARC record. Various pieces of the MARC record apply to work, expression, manifestation and item, but it has been found to be quite difficult to define explicitly that a certain field or subfield applies to a particular Group 1 entity. Also, these entities are not represented in a consistent way across libraries, and in some cases they are represented in unstructured fields, such as notes fields, which make it difficult to process automatically.

Being able to do this accurately – split the MARC record into work, expression, manifestation and item records, and then link multiple manifestation records to the correct expression records, and multiple expression records to the correct work record – is essential, however, for transforming MARC databases into a FRBR database.

Drawing the line between works

At some point a line is crossed and a new work is created. FRBR reports point out that such line of demarcation between one work and another may in fact be viewed differently from one culture to another. Consequently, the bibliographic conventions established by various cultures or national groups may differ in terms of the criteria they use for determining the boundaries between one work and another. Views probably also differ within cultures, even between individuals. And the line between works in FRBR appears to be based on one specific cultural tradition, namely the Anglo-American tradition embodied in AACR2.

The need for super-works

There has been some discussion of the need for an 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 and databases. However, the Study Group responsible for the FRBR model points out that such a need can be fulfilled by establishing relations between works.

Identifying works based on 200, 500 (UNIMARC) or 245 (MARC21) titles

Different publications may contain the same “work” in different expressions. Uniform titles are already used to identify “works” that appear under different titles. However, only a very small number of records have a uniform title entry. The experience from the Norwegian BIBSYS database shows that only 5% from a total of 4 million records has a uniform title entry. In Portugal, only 6% of the records of the National Library catalogue (from a total of 825.000) have a uniform title entry.

In 95% of the cases, therefore, the only titles usable to implement FRBR are the 200 (UNIMARC) or 245 (MARC21) fields. In many cases the title statement can be used for work title, but it is not always a good source for work identification. It may result in a large number of errors, since both title and author fields are prone to typing errors; use of abbreviations is frequent, which is enough to disable exact matching of records; and erroneous or inconsistent representation of title statements is frequent⁷.

The need to specify the nature of the relationships

Identifying the nature of the relationships between entities is crucial in an FRBR model implementation. For example, is the nature of the relationship between a person and an expression that of an author, translator, or editor? Cataloguing rules have provided for the addition of functions to the names of persons, such as “ill.”, “ed.”, which may

⁷ AALBERG, Trond – *MARC and FRBR: match or mismatch?* In: *Cataloguing 2007, Reikjavík, Iceland, 1-2 February 2007*. Reikjavík: the National and Univ. Library of Iceland, 2007

serve as an aid to identifying the nature of the relationships, and MARC records are structured to accommodate such information, but its inclusion was made optional in AACR2. In the absence of such a term/code, automatic identification of the nature of the relationship will be impossible, especially in the added entry position (7X2 in UNIMARC).

3.2. Defining a way of assessing the results

The background study described above provided the foundations for the design of the experiment to be carried out in TELplus. It was recognized that the results would have to be carefully evaluated, because the heterogeneous environment of TEL (both of languages, formats used and cataloguing practices) poses difficulties that were not addressed in previous work.

We therefore decided to select two specific collections, with very special characteristics, that will act as a “control group” during the experimental work on the prototype, thus making it easier to assess the results. These collections are built from records from the libraries participating in TEL.

It was agreed to focus on two collections. The first one is built after the general conceptual work “Bible”. This group was found relevant because of its potential wide relationships and global recognition. It stresses issues related with searching a single work with multiple expressions and manifestations in a multilingual catalogue. It is widely recognized as a very complex group of records, and very difficult to structure and present according to the FRBR model. The second collection is built with the works of the Literature Nobel Prize winners. In this case, the relevance comes from their expected endless expressions and translations into a large number of languages. This group of record is representative of the issues found on a group of works from the same author, with multiple translations in different languages.

To create these two collections, several national libraries sent us the Bible and Literature Nobel Prize records from their catalogues, in several MARC formats, encoded in ISO2709, MARCXML or MarcXchange. The number of records from each library, as well as the total, is presented below.

Library	Nobel bibliographic records	Bible bibliographic records
National Library of Portugal	3130	495
British Library	18838	34054
Royal Library of Belgium	1084	191
National Library of Latvia	1786	393
Martynas Mazvydas National Library of Lithuania	341	83
Russian State Library	20	5
National Library of the Czech Republic	2945	365
French National Library	28302	18931
German National Library	22598	1378
National Library of Spain	12444	1544
National Library of Serbia	1047	24
	92535	57463

MARC formats	Nobel bibliographic records	Bible bibliographic records
UNIMARC (4 libraries)	31793	19514
MARC21 (7 libraries)	60742	37949
	92535	57463

3.3. The FRBR system of TELplus

The system under development can be divided in two sub-systems (fig. 1). The first sub-system is responsible for extracting the FRBR entities and relations that are represented within the bibliographic records from the libraries. It is also responsible for consolidating the extracted entities, that is, to identify when two entities actually are the same real-world entity. For example, when an author name is spelled differently in different languages, or when typing errors exist on an author name or title. The final output of this system is a new repository that contains the original bibliographic records organized according to an FRBR structure.

The second sub-system is built on the FRBR structure that results from the FRBR processing. It focuses on providing a user interface to allow the TEL end user to search and browse the bibliographic information according to the FRBR paradigm.

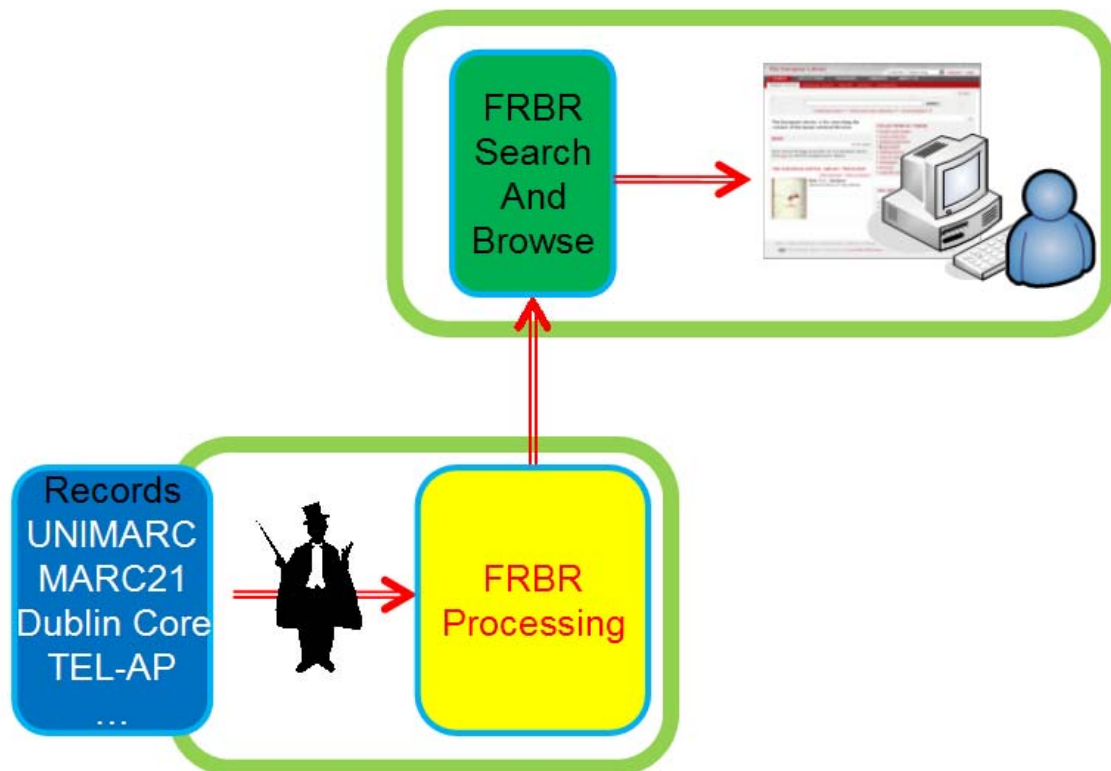


figure 1

The work is being carried out as an experiment with three stages. The experiment was designed in this way to allow the evaluation of results obtained at several stages of the FRBR processing. The three stages of the experiment and the software components are shown in figure 2.

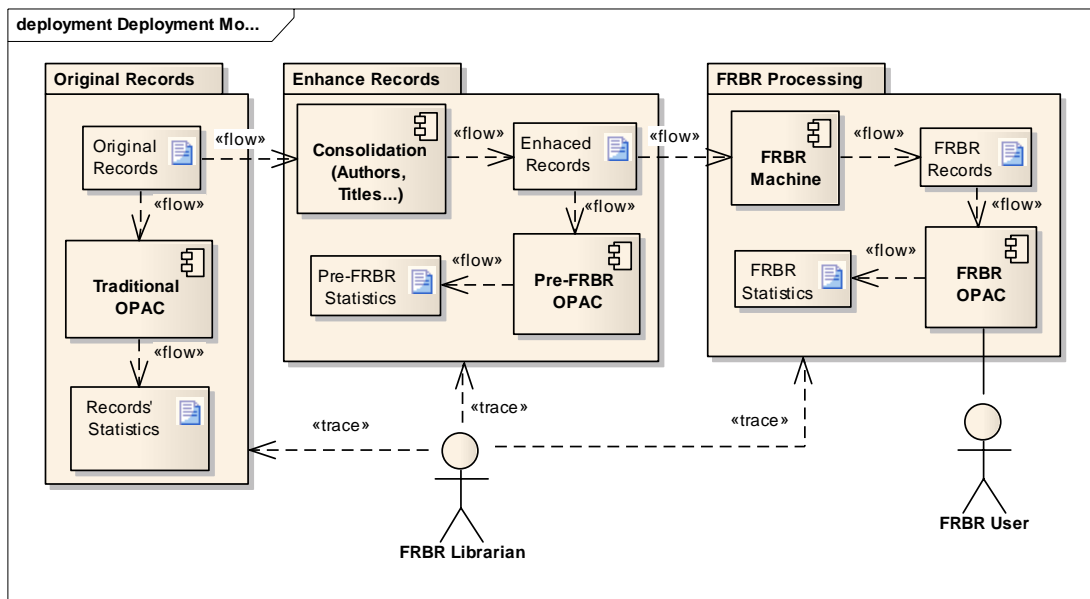


figure 2

The first stage is responsible for the evaluations of the original bibliographic records from the libraries. The FRBR system works with data encoded in XML. All Marc records that are received encoded in ISO-2709 are encoded in XML according to the MarcXchange schema. It also provides searching, access and statistics of the original records. This will allow the results of the FRBR processing to be compared against the original bibliographic information received by the system.

The second stage applies a set of techniques to the original records, with the aim of providing an initial consolidation of the entities found in the bibliographic records (mainly publication titles and author names). As in the first stage, the second stage provides searching, access and statistics about the enhanced records.

The third stage is where the final FRBR processing takes place. It extracts all FRBR entities from the original and the enhanced records, and further consolidates the entities and the existing relations between them. It is here that the definitive FRBR structure is created, and based on it the FRBR inspired user interface is provided.

The FRBR structure that results from stage three is encoded and used in the system according to FRBR in RDF⁸. This vocabulary is an expression in RDF of the concepts and relations described in the IFLA report on the FRBR. It includes RDF classes for the group 1, 2 and 3 FRBR entities and properties corresponding to the core relationships between those entities. However it does not yet describe attributes of the entities, for this reason we have extended it for TELplus by adding the properties for the 10 basic entities of FRBR.

In sum, the three stages will allow the libraries involved in this work to evaluate the results of the FRBR processing at several points, provide clues to where the FRBR processing may be improved, and check the overall quality of the resulting FRBR structure. Only the user interface on stage three is meant to be used by end-users.

Once the above experimental work is finished, the final system will be built. The experimentation will allow us to identify a set of FRBR processing techniques that can be applied to traditional MARC structured data and provide satisfactory results to the

⁸ <http://vocab.org/frbr/core>

end-user, that is results that can help the end-user in retrieving of information aided by the FRBR paradigm.

The final system will consist of just one stage, which will receive the original records and apply the FRBR processing to offer to the end-user of TEL the possibility of searching in the portal, and navigate search results, according to the FRBR paradigm, as shown in figure 3.

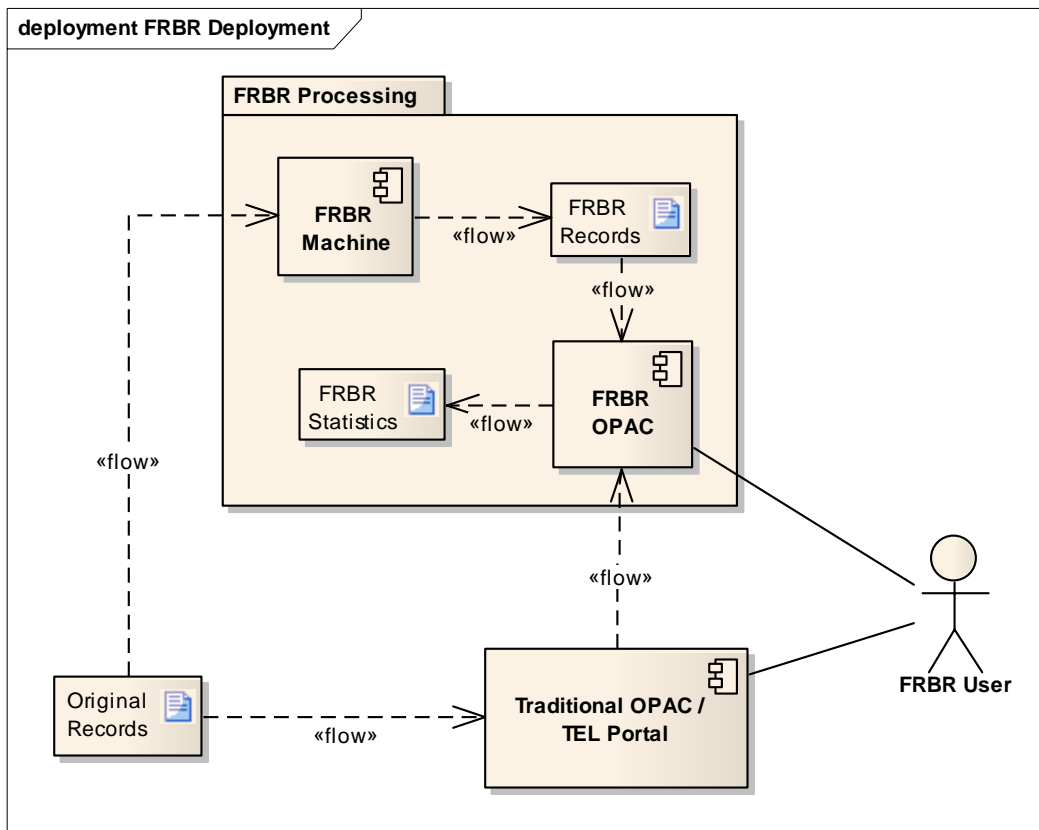


figure 3

CONCLUSION

The FRBR system is an ongoing project. The most challenging task of this project has been to create a structure for the FRBRization of bibliographic and authority records, in an information universe with very diverse characteristics, at all levels: in the information itself, in the different formats used to record it, and in the cataloguing practices used, even within the same format.

In spite of being a very complex structure, the most significant difficulties don't result from its conception or from the need to establish rules to identify FRBR entities in a record in a specific MARC format. The hardest task has been to establish rules that are valid and consistent with the several cataloguing practices, when the same MARC format is used.

This issue can only be overcome by human intervention at specific and crucial moments in the process of FRBRization. This is only natural, since it has been widely acknowledged that the implementation of the FRBR model elevates cataloguing to a higher intellectual, more complex and more rich, task. There is therefore the need to create an interface that allows intervention from the information professional, when automatic rules aren't sufficient.

In the process of FRBRization, it is clear that the information regarding the original title (corresponding to the concept of work) has to be given more importance than it gets in traditional cataloguing practices. It is also important that this element becomes an access point. It will be crucial for the c. 20%⁹ which constitutes the universe of bibliographic information that is estimated can be subjected to FRBRization (i.e., the universe that is not the “one work, one expression, one manifestation” type).

This change can be made by one of several forms. Choosing the best one is something librarians will have to address: by creating a new field in the MARC formats; by adapting and updating an already existing field, and/or by changing cataloguing rules and cataloguing practices.

In a very near future we hope to be able to give you practical results from the implementation of the FRBR system to the whole group of records given by the several libraries from TEL. Once the results are compared with the control group, and the necessary improvements are made to the procedures, this experiment will be widened to a bigger set of information.

At the end of the project, the finished system will have:

- a human user interface and a service interface (web-service) for the FRBR OPAC;
- one instance located at the Instituto Superior Técnico; and possibly one at the TEL Office too);
- the system will include a demonstrator of an FRBR OPAC; and also an online FRBR aggregation service (suitable to be integrated with the TEL portal – to be aligned with the Work Package 4).

BIBLIOGRAPHY

AALBERG, Trond – *MARC and FRBR: match or mismatch?* In: *Cataloguing 2007, Reikjavík, Iceland, 1-2 February 2007: Back to basics – and flying into the future* [online]. Reikjavík: the National and Univ. Library of Iceland, Nordic Research Library Associations Network (NFBN), Icelandic Library and Information Science Association (Upplysing), the University of Iceland – Department of Library and Information Science, 2007 [cited 20 November 2007]. Available from World Wide Web: <http://www.congress.is/cataloguing2007/>.

ANGELAKI, Georgia – *Research and innovation in The European Library; the TELplus Project. 2008*. Available from World Wide Web: http://theeuropeanlibrary.org/portal/organisation/press/documents/Paper_Sofia_Angelaki_2008.pdf

HICKEY, Thomas – *FRBR algorithms & tools* [online]. [Dublin, Ohio] : [OCLC], 2002. Available from World Wide Web: http://staff.oclc.org/~hickey/presentations/frbrAlgorithms20020620_files/frame.htm

IFLA Study Group on the Functional Requirements for Bibliographic Records - *Functional Requirements for Bibliographic Records: Final Report*. K. G. Saur, 1998 (UBCIM Publications. New Series, 19). Also available online at: www.ifla.org/VII/s13/wgfrbr/finalreport.htm.

⁹ HICKEY, Thomas – *FRBR algorithms & tools* [online]. [Dublin, Ohio] : [OCLC], 2002.

MAXWELL, Robert L. – *FRBR : a guide for the perplexed*. Chicago : ALA, 2008.

TEIXEIRA, Teresa [et al.] – *Report on FRBR experiments: TELplus deliverable n. D-3.6*. July 2008. 23 p.