*Frbrisation* : **towards a bright new future for national bibliographies**

**Jan Pisanski,** University of Ljubljana, Slovenia
**Maja Žumer,** University of Ljubljana, Slovenia
**Trond Aalberg,** Norwegian University of Science and Technology, Trondheim, Norway

**Meeting:**              **77. Bibliography**

**Abstract:**

*Functional Requirements for Bibliographic Records (FRBR) conceptual model is beginning to gain recognition. This paper presents FRBR including possibilities and limitations of current standards and practices in terms of expressing the FRBR model. It is claimed that national bibliographies should benefit the most from FRBR, as they contain large quantities of high quality bibliographic data on many works with multiple versions. The paper is in part based on the results of frbrisation of Slovenian national bibliography and two other bibliographic databases.*

National bibliographies are an important source of information about the intellectual production of a nation. However, the records of these bibliographies are created following rules and practices that initially were developed with the old fashioned card catalogue in mind, which limits the usability and reusability of this information for new applications in the digital environment. To increase the value of national bibliographies there is a need for information about bibliographic resources that can fully serve as machine interpretable and reusable knowledge about bibliographic resources. One of the possible solutions is the use of Functional Requirements for Bibliographic Records (FRBR) conceptual model.

**FRBR**

The FRBR conceptual model presents a general view of the bibliographic universe (FRBR, 1998). The basic elements of the FRBR model are the result of logical analysis of data in current bibliographic records. They consist of entities, their attributes and relationships between these entities. Entities of the first group are products of intellectual or artistic endeavour, entities of the second group are responsible for the content, production, dissemination or custodianship, while entities of the third group are subjects of intellectual or artistic endeavour.

**Entities of Group 1**
*Products of intellectual or artistic endeavour*

- **Work** – a distinct intellectual or artistic creation
- **Expression** – the intellectual or artistic realization of a *work* in the form alpha-numeric, musical, or choreographic notation, sound, image, object, movement, etc., or any combination of such forms.
- **Manifestation** – the physical embodiment of an *expression* of a *work*
- **Item** – a single exemplar of a *manifestation*

**Entities of Group 2**
*Entities responsible for the content, production, dissemination or custodianship of Group 1 entities*
- **Person** – an individual
- **Corporate Body** – an organization and/or a group of individuals and/or organizations

**Entities of Group 3**
*Subjects of intellectual or artistic endeavour*

When talking about FRBR in connection with national bibliographies, it has to be noted that unlike catalogues national bibliographies do not record *item*-level information.

Having a conceptual model should help users and professionals better understand bibliographic universe.  While FRBR's terminology may not be common, it represents a necessary break from the traditional usage of same appellations for different concepts (most notably *edition*). FRBR provides a framework for more user friendly displays of bibliographic data. It has the power to eliminate duplication of data in bibliographic databases. It also helps establish order and has the potential to turn linear library databases into a true networked environment by relying heavily on relations, which help provide user with context.

A common misconception about FRBR is that it is a data model or a set of cataloguing rules that specifies what implementations should be like. It is, only a theoretical model that needs to be interpreted in the cataloguing rules. Further still, each implementation will be dependent on system design. That means that unless there is a clear agreement on what FRBR represents in cultural sense, which is unlikely to happen, the model is open to different usages within different communities. As Maxwell (2008) points out we don't have any "official" literature or practical experience detailing how FRBR should be interpreted.

However, work on the new Anglo-American cataloguing rules, or rather the new standard for resource description designed for the digital world, called Resource Description and Access (RDA; http://www.rdaonline.org/) takes FRBR into consideration. This is a strong move away from the FRBR-related vicious circle (Pisanski, Žumer, 2007) and hints at appreciation of FRBR within larger library community.

An important aspect of introduction of FRBR to cataloguing rules is what Carlyle (2006) calls "cataloguer consciousness". FRBR should help cataloguers better understand why they do what they do, which in turn should lead to better service to users. While this issue is more pressing for common catalogues, it is something that must also be kept in mind when recording data for national bibliographies.

RDA has strong opposition from cataloguers, computer scientists and library science theoreticians, usually on opposite sides of the scale. Quite a lot of this opposition has to do with FRBR being the basis of RDA. In fact, the Library of Congress Working Group on the Future of Bibliographic Control at Library of Congress suggested stopping work on RDA, until, among other things, benefits of FRBR are clear (WoGroFuBiCo, 2008). Work on RDA has not stopped, however, and it is expected that RDA will be released in 2009 (Joint, 2008). This suggests that both theoretical and practical exploration of FRBR and the bibliographic universe (such as the one presented later) together with related user studies are more relevant than ever, especially since once again there are no reports on any formal user studies. First results of theoretical user studies show that FRBR on average non-librarians' mental models closely follow FRBR (Pisanski, Žumer, 2008?), while first reports of internal user studies of FRBR based catalogues also present positive results (Kilner (2005), Jepsen (2007)).

It must also be stressed that FRBR and RDA are not isolated developments in the field of cataloguing. One of the more important pieces of the puzzle is the recently published Statement of International Cataloguing Principles, built on the legacy of Paris Principles as well as FRBR (Statement, 2009).

However, even if FRBR is introduced to national bibliographies by its inclusion in the new cataloguing rules, there is still the problem of current formats not being able to fully express FRBR. Even though there are many different variations of MARC format, there is one important feature that they all have in common: there is no easy way to incorporate FRBR into the existing MARC structure. An additional problem is the vast amounts of legacy data. There appear to be two possible solutions to bringing existing data in line with FRBR. One would be to manually recatalogue the entire national bibliography, which would be both time and resource consuming and therefore not feasible under present circumstances. The second is extraction of FRBR concepts from the existing bibliographic data. This second process is referred to as *frbrisation*.

**FRBRISATION**

As it is the only serious solution that is not *apriori* deemed not feasible, *frbrisation* has to be considered a front runner for the treatment of all legacy data in light of FRBR developments. Several attempts at *frbrisation* (e.g., Hegna and Murtomaa, 2002; Hickey, O'Neill and Toves (2002), Hickey and O'Neill (2005)) show that it can be done[1]. However, results of these attempts also show that it is difficult to achieve high quality. Typical problems include insufficient or erroneously identified entities and relationships and insufficient identification of equivalent entities. Furthermore, many *frbrisation* initiatives only apply parts of the FRBR model or only process parts of the information found in a record. However, complete *frbrisation* would introduce additional problems that could further reduce the quality of results. It has to be noted that all of these problems stem from existing bibliographic data and the way it is recorded. Had the bibliographic data been recorded in a more structured manner and more consistently, there would have been fewer misidentifications in *frbrised* data.

While the basic idea of all of the *frbrisation* 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

---

[1] In fact, there are a number of *frbrised* catalogues in existence, although none of them truly follows FRBR. For more discussion, see Pisanski and Žumer (2007).

partly to local considerations in cataloguing (i.e. differences in the use of formats, cataloguing rules and standards), as well as authors' perceptions of what is feasible and what users are looking for.

On the other hand, implementation possibilities are limited by existing principles, such as the Paris Principles from 1961, cataloguing codes (such as the Anglo-American Cataloguing Rules - AACR2), and data formats (such as MARC21), as well as their actual application in practice, as cataloguers misinterpret the rules or make other mistakes. While these mistakes are usually hidden when using the "traditional" OPACs, they tend to jump out immediately when using "*frbrised*" OPAC displays.

In other words, if the quality of bibliographic records is inconsistent or bibliographic data is missing, it is difficult to automatically extract any structure, FRBR-like or other. As many FRBR prototypes are of union catalogues, consistency of data and, consequently, usability of these prototypes is further diminished. According to Buchanan (2006) error rates in the automatic detection of FRBR entities for *works* with multiple *manifestations* can rise to 50%, depending naturally on the type of resource and the quality of data.

## FRBR AND NATIONAL BIBLIOGRAPHIES

As argued by Žumer (2007, 2008), it is exactly in the case of union catalogues and national bibliographies, with their multitude of *expressions* and *manifestations* of a large number of *works*, that introduction of a conceptual model is needed and the true advantages of FRBR can properly be seen. Similarly, according to the new guidelines for national bibliographies in the digital age (2009), national bibliographies seem to be the best starting point for both *frbrisation* and original cataloguing in line with FRBR.

There are three interrelated reasons for this. Firstly, the whole range of effects of FRBR can better be observed on larger sets of bibliographic data and cumulative national bibliographies contain vast amount of data. However, unlike the next two reasons this factor alone does not put national bibliographies in a favourable position compared to union catalogues.

Next, as national bibliographies tend to be more complete than individual library catalogues, the effect of FRBR should be greater. In fact, for works of national importance, the use of FRBR should produce next to complete FRBR network[2] that is unlikely to be produced even in the largest union catalogues.

Based on these first two reasons we can safely assume that cumulative bibliographies stand to benefit the most from application of FRBR, whereas one would expect smaller units (biweekly, monthly or even yearly bibliographies) not to be as positively influenced by the use of Group 1 entities in particular.

However, there is one more important factor that works in favour of national bibliographies. As level of cataloguing in national bibliographies is generally higher than elsewhere, potential for obtaining meaningful data from *frbrisation* is greater. National bibliographies tend to include better bibliographic data than union catalogues. There is greater consistency, due to stricter rules. It is also of benefit that only one set of rules and standards is in use compared to

---

[2] Although its sheer volume makes recording of complete national intellectual production increasingly less feasible.

some union catalogues, where there may be a variety of different rules and standards followed.

It is therefore not surprising that one of the first *frbrisation* experiments by Hegna and Murtomaa (2002) was performed on Finnish and Norwegian national bibliographies. However, they did find that even under these relatively favourable circumstances, there were still problems with the correct identification of entities. They found that while elements of the FRBR model are generally present in the bibliographic records, cataloguing rules are not well suited to the model. They also found that important information is often recorded in a way that is intelligible to humans but not computers.

While high level of cataloguing, relative completeness of coverage and relatively high consistency of data might suggest that national bibliographies are as perfect as they are going to get, one has to be aware that while some order may be present in this relatively well structured data, the context is still missing. Unfortunately, much of the context could only be gained by providing explicit relationships between different bibliographic entities, which are presently not recorded or at least not in a consistent manner that would provide for easy computer manipulation.

On a slightly related note, we must also not forget that one of the objectives of FRBR was to recommend a basic level of functionality for records created by national bibliographic agencies. The International Conference on National Bibliographic Services (ICNBS), held in Copenhagen in 1998, recommended the use of components of Basic Level Record provided by the FRBR final report. However, the final implementation of these recommendations still depends on individual national bibliographic agencies. Therefore we have not looked at this issue any further.

## OUR FINDINGS

In the last section we will focus briefly on the results of *frbrisation* and analysis of MARC records from 3 different bibliographic databases: the Norwegian BIBSYS database, the Swedish Burk database and the Slovenian National Bibliography (Aalberg and Žumer, 2008). This research was conducted as part of a joint project of NTNU (Trondheim, Norway) and University of Ljubljana (Slovenia). Of the three, only Slovenian National Bibliography is a national bibliography. The suitability of Slovenian National Bibliography to the requirements for national bibliographic databases presented in FRBR final report is extensively discussed in Dimec, Žumer and Riesthuis (2005).

A tool for *frbrisation* of bibliographic records, developed at NTNU, was used for transforming MARC records into a collection of FRBR records for the entities and relationships that can be deduced from the input. The results were evaluated by manually inspecting the examples.

The tool is rule based and can therefore be adapted to particular collections, based on the MARC format used as well as other particularities, mainly related to cataloguing practice. For the three above mentioned databases three independent rule sets were developed. Based on the results, it is possible to state that national bibliographies can be *frbrised*, but not without the usual difficulties, associated with the quality of legacy data.

Our results support the theoretical claim that the cataloguing in national bibliographies is usually better than in individual catalogues, therefore *frbrised* results can potentially be better than those from catalogues. For instance, Slovenian National Bibliography consistently records all of the authors and contributors along with their appropriate relator codes. On average there are twice as many persons to be found in a Slovenian National Bibliography record as in the other two databases. This is a direct result of the much stricter approach to national bibliographies compared to the catalogues. Relator codes, consistently present in Slovenian Bibliography, were also found to be highly useful for successful determination of FRBR expressions. If no person or corporate body can be associated with an expression during the process of identification of entities, expressions can only be indentified at a categorical level which typically was the case for both the Norwegian and Swedish databases. This is in line with typical FRBR implementations which only provide sets of expression according to language, form or format. In the Slovenian database it was possible to use agents related to the expressions in the identification due to the extensive use of relator codes. This enabled individual identification of expressions as defined in FRBR. Due to the structure of MARC the data associated with group 2 entities is often insufficient for identifying specific relationships. For example, it is often impossible to establish whether a person is related to the work or the expression. The Slovenian bibliography is different from the other two because it uses UNIMARC and has a strict policy for including relator codes.

For the Slovenian bibliography some original titles were found in uniform title fields, others were found in note fields. However, the latter are relatively easy to identify due to the partly structured use of notes, as the uniform title is preceded by introductory phrase "Prevod dela:" ("Translation of work:"). However, morphological richness of Slovenian language (use of prefixes and suffixes) made some of the works more difficult to identify (e.g., when there was more than one work in a publication). In the other two databases there are specific repeatable fields for original titles that are frequently used. For all databases the number of records with a specific work-level title was low. In approximately 80% of the records the title statement is the only source for discovering what works the manifestation contains.

Being able to identify a set of entities from a record is just the first step in *frbrisation* process. The more difficult part of the process is to determine the relationships between the deduced entities. While it would help if all important relationships were recorded, it is somewhat discouraging that even some of the most obvious relationships (between a work and a person or an expression and a person) are sometimes not easy to identify, because of lack of relator codes or similar problems. As already stated, in both the Norwegian database and the Swedish database there are only a few relator codes in use, whereas in Slovenian National Bibliography there was an extensive set of relator codes. Another problem was establishing proper relationships when there were multiple persons and multiple works or expressions identified within a record. The MARC formats, as well as current cataloguing practices, do not really deal with this structural information – which in FRBR is important. However, this is not only an issue of formats and should be given appropriate consideration in cataloguing rules, including RDA.

Also, MARC records have a structure that is suited to publications that contain a single expression of a single work. The practice for cataloguing publications that contain multiple expressions is typically different between catalogues and even within one there may be a variety of solutions.

Of course, one has to be careful when drawing conclusions about the suitability of national bibliographies for *frbrisation* based on one example, especially since there may be practical variation in use of different bibliographic tools for particular national bibliographies.

**CONCLUSION**

Adopting a conceptual model that has the potential to change the nature of national bibliographies and other bibliographic databases and make them more user-friendly, brings with it plenty of complications, mostly stemming from the way existing bibliographic data is recorded. Nevertheless, the only feasible solution to making legacy data compatible with any new data is *frbrisation*.

National bibliographies are most likely the type of bibliographic databases that is the most prepared for *frbrisation* and also stands to gain the most from it. As such, they could become prime examples of the benefits of FRBR and could serve to convince the opponents of practical value of a conceptual model of the bibliographic universe.

**BIBLIOGRAPHY**

Aalberg, T. and Žumer, M. (2008). Looking for Entities in Bibliographic Records. *Digital Libraries: Universal and Ubiquitous Access to Information, 11th International Conference on Asian Digital Libraries, ICADL 2008*. 327-330.

Buchanan, G. (2006). FRBR: Enriching and Integrating Digital Libraries. *Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries*. 260-269

Carlyle, A. (2006). Understanding FRBR as a Conceptual Model: FRBR and the Bibliographic Universe. *Library Resources & Technical Services*. 50 (4), 264-273.

Dimec, Z., Žumer, M. and Riesthuis, G. (2005). Slovenian Cataloguing Practice and *Functional Requirements for Bibliographic Records*: A Comparative Analysis. *Cataloging & Classification Quarterly*. 39 (3/4), 207-227.

(FRBR) *Functional Requirements for Bibliographic Records: final report* (1998). München: KG Saur.

Hegna, K. and Murtomaa, E. (2002). *Data mining MARC to find: FRBR?* 68th IFLA Council and General Conference, August 18-24, 2002. Retrieved 21.11.2008 from: www.ifla.org/IV/ifla68/papers/053-133e.pdf

Hickey T., O'Neill, E. and Toves, J. (2002). Experiments with the IFLA Functional Requirements for Bibliographic Records (FRBR), *D-Lib Magazine*, 8 (9). Retrieved 5.1.2009 from: http://www.dlib.org/dlib/september02/hickey/09hickey.html

Hickey, T. and O'Neill, E. (2005). FRBRizing OCLC's WorldCat. *Cataloging & Classification Quarterly*, 39 (3/4), 239-251.

Jepsen, E. T. (2007). *Bibliographic relations*. Retrieved 21.6.2007 from: http://ru.is/kennarar/thorag/cataloguing2007/Erik_Thorlund_Jepsen.ppt

Joint Steering Committee for Development of RDA (2008). *RDA: Resource Description and Access*. Retrieved 19.1.2009 from: http://www.collectionscanada.gc.ca/jsc/rda.html.

Kilner, K. (2005). The AustLit Gateway and Scholarly Bibliography: A Specialist Implementation of the FRBR. *Cataloging & Classification Quarterly*, 39 (3/4), 87-102.

Maxwell, R. (2008). *FRBR: A Guide for the Perplexed*. Chicago: American Library Association.

*National bibliographies in the digital age: Guidance and new directions* (2009). In print.

Pisanski, J. and Žumer, M. (2007). Functional requirements for bibliographic records: an investigation of two prototypes. *Program*, 41 (4), 400-417.

Pisanski, J. and Žumer, M. (2008). How do non-librarians see the bibliographic universe? *Culture and identity in knowledge organisation: proceedings of the 10th International ISKO Conference, 5-8 August 2008, Montreal, Canada.* Würzburg: Ergon, 131-136.

(Statement) *Statement of International Cataloguing Principles* (2009). Retrieved 9.3.2009 from: http://www.ifla.org/VII/s13/icp/ICP-2009_en.pdf

(WoGroFuBiCo) *On the Record*. Retrieved 19.1.2009 from: http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf

Žumer, M. (2007). FRBR: The End of the Road or a New Beginning. *Bulletin of the American Society for Information Science and Technology*, 33 (6), 27-29.

Žumer. M. (2008). FRBR… *Workshop on FRBR in The European Library, Lisbon, Portugal*. Retrieved 3.3.2009 from: http://frbr.bnportugal.pt/documentos/FRBR%20Zumer.ppt