

Žumer, Maja

Do we need to change? Do we want to change? : the future of bibliographic information systems

ProInflow. 2011, vol. 3, iss. 2, pp. 138-149

ISSN 1804-2406

Stable URL (handle): <https://hdl.handle.net/11222.digilib/133726>

Access Date: 23. 02. 2024

Version: 20220831

Terms of use: Digital Library of the Faculty of Arts, Masaryk University provides access to digitized documents strictly for personal use, unless otherwise specified.

Maja Žumer

DO WE NEED TO CHANGE? DO WE WANT TO CHANGE? THE FUTURE OF BIBLIOGRAPHIC INFORMATION SYSTEMS

Zajímavosti z oboru

Abstrakt:

Poprvé ve své dlouhé historii čelí knihovny konkurenci. Existuje mnoho dalších poskytovatelů informací a uživatelé si vyhledávají informace kdekoli. I přes to, že knihovní katalogy mají výhodu v tom, že jsou kontrolované, konzistentní a bohaté na informace, zdá se, že nevyužívají svého plného potenciálu. Proto je nutné se v této oblasti posunout dále. Functional Requirements for Bibliographic Records (FRBR) představuje nové paradigma, které by nejen mohlo nabídnout mnohem intuitivnější poskytování bibliografických informací, ale též využívat nástrojů a služeb sémantického webu. V článku jsou představeny některé současné výzkumné aktivity, které otvírají cestu k lepším bibliografickým informačním systémům, které by měly být předmětem dalšího zkoumání.

Klíčová slova: *Functional Requirements for Bibliographic Records (FRBR), knihovní katalogy, sémantický web, bibliografická data, informační systémy*

Abstract:

For the first time in their long history libraries are facing competition. There are many different information providers and users find information elsewhere. The clear advantages of the library catalogue, such as authority control, consistency and the wealth of information are obviously not utilised to their potential. A step further is therefore urgently needed. Functional Requirements for Bibliographic Records (FRBR) provides a new paradigm which could not only enable more intuitive presentation of bibliographic information, but also open this information using Semantic Web tools and services and therefore promote exchange and reuse across domains. Several current research activities are presented. They all pave the way to better bibliographic information systems, which should be developed without further delay.

Keywords: *Functional Requirements for Bibliographic Records (FRBR), library catalogues, Semantic Web, bibliographic data, information systems*

1 Introduction

For the first time in their long history libraries are facing competition. There are many different information providers and users find information elsewhere. We see reports that some users are actively avoiding searching a library catalogue even when they want to borrow a book. Users prefer simple, intuitive tools such as Google and Amazon. What was repeatedly reported by researchers in the 1980's and 1990's, such as Borgman (1986, 1996) is becoming even more obvious. The message is clear: even library users do not find the catalogue attractive and easy to use. The clear advantages of the library catalogue, such as authority control, consistency and the wealth of information are obviously not utilised to their potential.

Modern bibliographic information systems should provide users with a more efficient way of accessing and using information and libraries should expose users to the broader context that is absent from the catalogues today. By doing this in an open manner, libraries could have a major influence on the development of Semantic Web.

Not only has the library community not been active enough in developing their tools, most library catalogues are still based on the model that worked very well for card catalogues but has been obsolete ever since computers were first introduced. But things are moving forward. Functional Requirements for Bibliographic Records (FRBR) (1998), the conceptual model of the bibliographic universe, has been developed. The FRBR family now includes Functional Requirements for Authority Data (FRAD) (2010) and Functional requirements for Subject Authority Data (FRSAD) (2011), which extend FRBR into authority data and the subject relationship. As indicated by the results of our studies (Pisanski and Žumer, 2010a, 2010b, 2010c), FRBR is intuitive and could (or rather, should) be used as the foundation for new bibliographic information systems. Recently some cataloguing codes based on these documents, as well as the International Cataloguing Principles, have been developed, perhaps most importantly RDA, which is currently being tested. Another important area is the ongoing development of namespaces by IFLA Namespaces Task Group.

The library community should not hesitate any more: the new paradigm has to be accepted and implemented as soon as possible. Research, both basic and applied, needs to pave the way.

2 Why FRBR?

FRBR is a revolutionary development. While librarians have been analysing their activities and a body of theoretical knowledge has been accumulated, this is the first attempt at a fully developed conceptual model of the bibliographic universe.

The *FRBR-model* was published by the *International Federation of Library Associations and Institutions* in 1998 and is generally considered to be an important contribution to our understanding of the entities and relationships that are of interest to end users of bibliographic information. The model is general and covers a broad range of intellectual and artistic products and is for this reason potentially able to serve as a foundation for semantic interoperability across a broad range of metadata resources related to intellectual and artistic endeavour. The innovative contribution of the FRBR-model is the introduction of concepts that reflects intellectual and artistic endeavour at various levels of abstraction, the actor entities involved in the creation and ownership of these entities, the relationships that may exist between the various entities in the model, and the attributes needed to identify and describe the entities. In the context of FRBR the catalogue is not seen as a sequence of self-contained bibliographic records but rather as a network of connected entities, enabling users to perform seamlessly the tasks of finding, identifying, selecting and obtaining information.

The set of entities used to model intellectual and artistic products is the core of the model and consists of the entities *work*, *expression*, *manifestation* and *item*. The *work entity* represents intellectual and artistic creation at the highest level of abstraction independent of any specific form, language or medium. In our discourse we often identify creations such as the play “Hamlet” by William Shakespeare as a distinct unit independent of the various translations or adaptations of this work and the work entity generally captures this abstraction. A work entity is in a sense the conceptual “content” that is shared between all different but still comparable externalizations of this particular play. The *expression entity* is another abstraction that is needed to enable the modelling of the intricate real-world relationships that may exist between intellectual creations. An expression entity identifies a distinct externalization of a work in a specific language or form. A work such as the play “Hamlet” may exist in different translations and adaptations and each of these can be identified as a unique derivation of the same work. The next level the model introduces is the *manifestation entity*. Manifestations reflect the way content is published in the shape of produced publications. A particular Slovenian translation of “Hamlet” published as a book identified by a specific ISBN is an instance of the manifestation entity. The same expression can be published in different manifestations and a single manifestation may contain more than one expressions. The *item entity* is the last entity in the abstraction hierarchy and reflects the actual copies that exist of a particular manifestation e.g. on the shelves of different libraries.

In addition to the entities representing products at different levels of abstraction the FRBR-model emphasises the actors that are involved and the different relationships they have to the product entities. Actors (persons and corporate bodies) may be the

creators of works, the ones responsible for the realization of an expression, the producers or publishers of manifestations or owners of items.

FRBR is explicitly focused on users of bibliographic information and their information needs. It is therefore surprising that, while the model is based on the insight and experience of experts, no user studies were performed during the development of FRBR due to organisational and time constraints (Madison, 2005). This fact has been often commented on and many have called for a user verification of FRBR (Library of Congress, Working Group on the Future of Bibliographic Control, 2008) and user verification was identified as one of the most important research topics in a Delphi study by Zhang and Salaba (2009). The decision to respond to that call led to our study performed in 2009 and published in 2010 (Pisanski and Žumer (2010a, 2010b, 2010c).

A study of non-librarians' mental models of the bibliographic universe was undertaken in order to understand whether these mental models resemble the FRBR conceptual model (for further details of the study, see Pisanski and Žumer (2010a, 2010b)). As it was the first study of its kind, it aimed at capturing broad patterns and focused only on FRBR Group 1 entities and only used examples for textual works. It was a rather small-scale study with 30 participants from different backgrounds.

None of the participants had any knowledge of FRBR or any aspect of bibliographic information systems beyond occasionally using a library catalogue. Additionally, no part of the study actually referred to catalogues or libraries. In other words, we avoided the influence of any existing systems for recording bibliographic information as much as possible. The design of the study therefore provided for capturing highly abstract and “pure” mental models of the bibliographic universe and not its surrogates.

The study consisted of three parts: card sorting, concept mapping and comparison task, all three of which are used as mental model elicitation techniques. Card sorting required of participants to sort cards with simple textual descriptions of instances of FRBR entities into at least three groups according to the level of abstractness. Ideally card sorting would lead to four clear categories, corresponding to FRBR's Group 1 entities (work, expression, manifestation, item).

For concept mapping a question ‘What comes out of what?’ was asked, using the same set of cards as in card sorting. It was also explained to the participants that connections between individual cards were sought. What we expected to elicit here were mental models that resembled an application of FRBR as a directed graph connecting cards, essentially a derivation chain flowing from works to items.

For the comparison task interviews focusing on similarity and substitutability of two real-life objects in a pair (two books or a book and a DVD of a movie) were conducted, followed by ranking of these pairs according to their perceived substitutability.

We then analysed the results with the help of cluster analysis for the first and second tasks, consensus map for the second task and simple statistics based on rankings for the third task. From the interviews we also gathered anecdotal evidence.

At least the results of the second and third task show that on average mental models of the bibliographic universe are FRBR-like in terms of Group 1 entities. Concept mapping found that the most frequent connections were generally the ones that would have been established based on FRBR. Also, in this task the mental models that were the most alike were those that were FRBR-like. Even clearer were the results of ranking of pairs according to the substitutability of the items in a pair in their task. Although no individual mental model was exactly like FRBR, most individual mental models were close to FRBR. In fact, if one disregards the deliberately introduced borderline case, 7 participants had exact FRBR groupings and 5 more only had some further groupings within the entities, essentially further dissecting FRBR entities.

On the other hand, the results of the first task were influenced by how closely the participants followed the given criterion. While some participants had trouble understanding the difference between sorting based on things described as compared to the descriptions themselves (e.g., as evidenced by descriptions of categories that referred to vagueness of descriptions), some sorts were clearly based on another criterion or even a combination of criteria. However, in both the first and second task a close connection between the work and its original expression (regardless of the language) was detected. This would suggest that a special place for original expression might be needed in any conceptual model of the bibliographic universe. Other than this distinction, no alternative model of the bibliographic universe was found.

Proximity to FRBR was investigated for all three tasks. While mental models were generally FRBR-like, they were not exactly the same. Also, during different tasks individual participants' mental models varied. In fact, very few participants had stable mental models in regard to their proximity to FRBR. Generally speaking, the more concrete their task was and the more they thought about the bibliographic universe, the more FRBR-like participants' mental models were.

We are now continuing the study. For this phase we took some of the graphs (derivation chains) resulting from the second task and presented them to participants (students of different disciplines) together with the simple descriptions of entities represented as nodes. The participants had to choose the graph which best

represented their view of the connections between entities listed. The preliminary results show that the graph representing FRBR was predominantly chosen and we are still analysing the results in more detail.

The results therefore confirm that FRBR could and should be used as the foundation for more user-focused bibliographic information systems.

3 Research topics paving the way

There are several key issues that we need to focus on, always keeping the purpose of cataloguing in mind.

3.1 Harmonisation and development of the model

The unification of models FRBR, FRAD, and FRSAD (the FRBR family) is urgent. While all build on the same foundation, different modelling decisions have resulted in incompatible solutions. In addition, the model has to be developed further (e.g. aggregates) and verified. For example, we still lack research-based evidence on exactly what attributes and relationships different user groups require. Newer initiatives (FRBR, ISBD, RDA) just seem to copy attributes from one another and current cataloguing practice without much analysis.

A small study by Leskovec (2005) confirms that the attributes and relationships recorded in current catalogues do not always correspond to user needs. She analysed user requests in a public library and found that most users search for expressions, groups of expressions (e.g. any edition of a work in a particular language) and sometimes even works in general. Some users search for manifestations (i.e. particular editions) when they are particularly interested in the first or latest edition or when they are looking for publications with additional materials, such as illustrations or commentaries. However, while catalogue records describe manifestations in detail, information about respective work(s) and expression(s) is not always evident and many important relationships and attributes are not recorded (e.g., whether a text is integral or abridged, information about sequels, etc.).

Our research group is currently extending this study to a broader sample with the goal to verify which attributes and relationships are needed to support the user tasks.

FRBR is a conceptual model of the bibliographic universe and defines all entities and relationships, but focuses particularly on Group 1. FRAD expands the model in the area of authority data for Group 2 entities and *works*; FRSAD deals with the subject

relationship. FRBR and FRSAD both start from the end-user perspective and the analysis is based on the tasks that users perform during the interaction with bibliographic systems. FRAD, on the other hand, focuses primarily on librarians performing authority control and models the cataloguing process. The two approaches necessarily resulted in differences which will have to be analysed and resolved during the harmonisation process, which will be undertaken by the FRBR Review Group in the near future.

The FRBR model is very general and in some parts rather vague. It is not intended to be directly implemented as a data model. But when the implementations are developed, questions arise and there is a need for stricter definitions and specifications. The FRBR Review Group responds to such issues by establishing working groups. The Expression Working Group is an example: they prepared an amendment to the definition in expression, which is more operational and specifies the conditions under which a cataloguer makes an informed decision.

Aggregates (all composites of individually created dependent/independent works) are another area where problems were identified. A working group has been established to: *“Explore the treatment of aggregates in the FRBR model. Common aggregates to be considered include: (1) Collections, selections, and anthologies, (2) Augmentations (original text augmented with illustrations, notes, introductions, etc.), (3) Monographic series, (4) Serials, (5) Multi-part monographs, and (6) Integrating resources.”*

Aggregates are mentioned several times in FRBR, but they are not treated systematically or consistently and the FRBR report does not provide sufficient guidance.

The work is not finished yet, but a working group report is expected by the next IFLA conference in August 2011.

3.2 Frbrisation

Libraries have been creating bibliographic data for centuries. Most of this data is now in the form of MARC records in integrated library systems. The data, created according to FRBR is (will be) very different. The huge amounts of legacy data will somehow have to coexist with born-FRBR data. The best solution for that is frbrisation, the automatic extraction of FRBR entities, attributes and relationships from legacy data. There have been several attempts at frbrisation (Hegna & Murtomaa, 2002; Hickey & O'Neill, 2005; Pisanski et al, 2009). The results were generally found to be relatively satisfactory if the data was complete and consistent (which is not always the case), but frbrisation is not trivial and it requires much

customisation to allow for the differences in individual cataloguing rules and practices, which often even change with time.

To enable exchange and reuse of bibliographic data, we need global identifiers for all entities. While some are already in place and relatively broadly used (for example ISBN for manifestation), other are not used much or missing completely. There is even no general agreement what FRBR entity individual identifiers identify (Pisanski et al, 2010). All stakeholders: libraries, publishers and rights management organisations will have to cooperate in the development and maintenance of this very important infrastructure.

3.3 Presentation

FRBR relationships require a different manner of presentation than the flat lists we are used to now. Research in presentation options is needed; we are currently looking into visualisation as a possible solution.

There has been a small number of FRBR implementations, but so far nobody has really dealt with the problem of results display in the context of FRBR. Until now, research on the field has mainly focused on the frbrisation of bibliographic records and ignored the question of how to transfer the concept into user interface area and create a view which would show FRBR structure and relations within the results display.

Traditional bibliographic information systems offer only a limited and linear results list where relations and links between records are rarely presented or pointed out. Due to this limited option for real interaction, a substantial financial input is often lost as libraries are not able to make a good presentation of what they really hold, thus consuming users' time and causing their frustration over the system. Visualization of information may present a good solution for the above discussed problems as many of visualization techniques enable better interactivity, results overview and network relations - the main advantage of FRBR.

We have already developed some visualisation scenarios and the next step will be the user testing of prototypes. We expect to see how users interact with such systems, whether they find them intuitive and which visualisation scenario(s) they prefer.

3.4 Bibliographic data and Semantic Web

No paper, claiming to be discussing the current situation of the information infrastructure can avoid the mention of Semantic Web. Information about cultural objects is very much the focus of interest on the Web and in recent years we see an increasing demand to disseminate and reuse this information beyond the library domain and across domains. Semantic web technologies can be used to expose and

interpret the meaning of the data, open access enables third parties to develop innovative new services for existing data, and new knowledge can be created by linking related and complementary data from different sources.

Libraries have for decades created metadata records describing the products of intellectual and artistic endeavour such as literature, music and other forms of expression. The use of this rich metadata has traditionally been limited to library services, but could and should be reused and integrated with other sources in innovative services that enable users to learn about, discover, annotate and discuss our cultural heritage.

One step towards the goal of giving data semantic meaning is the Linked data initiative. Technically, Linked data refers to data published on the Web in such a way that it is machine-readable, its meaning is explicitly defined, it is linked to other external data sets, and can in turn be linked to from external data sets. One of architectural prerequisites of linked data initiative is to use URIs (Uniform Resource Identifiers) as identifiers for things, since URIs identify any kind of object or concept. Several knowledge information systems (controlled vocabularies) have been published as Linked data and there are also some experiments in exposing bibliographic data in the same way.

The activities in this area also include the investigation of possible formats for FRBR-based data and harmonisation with other domains. The best example of the latter is the ongoing work of FRBR/CRM Harmonisation Working Group, a joint effort of IFLA and ICOM CIDOC (International Council of Museums – International Committee for Documentation). Since libraries and museums share users and types of materials, it is important that a common view of cultural heritage information be developed for the benefit of the users. The goal is to bring together (harmonize) the library model (FRBR) and museum model (CRM: Conceptual Reference Model). In preparing an object-oriented version of FRBR, additional goals are to check FRBR's internal consistency, enable interoperability and integration, to extend the scope of both conceptual models, and to open the road toward future applications. The most current version of the object-oriented view of FRBR was published as "FRBR object-oriented" or FRBRoo v. 1.0.1 in 2010. The harmonized model is now being further developed to include the full FRBR family, i.e. also FRAD and FRSAD.

4 Conclusion

The title of the paper begins with two questions. I will start with the second: "Do we want to change?" Probably not. In general, libraries and librarians are reluctant to change. Without making any value judgment about this (assumed?) characteristic of librarians one must note that librarians have been collecting information resources

and developing information tools over decades and even centuries. Therefore, it is unfeasible to often change directions regarding the way things are done, and a conservative approach is in many cases not only expected, but even necessary. On the other hand, resisting all change may stop development and prevent the library from offering better service to the user.

The answer to the first, “do we – libraries and librarians – have to change” is obvious. Yes, the change is urgent. Research is paving the way and it should result in standards, rules, guidance and recommendations, but also sound justification for change.

Libraries should develop their bibliographic information systems to better serve their users and thus reaffirm their position. They may be running out of time for experimentation and comfortable, slow-paced approach to changes. Users find catalogues hard to use and not intuitive enough. Additionally, most of them are not willing to invest their time into learning the details of any system, even if it could help them achieve better results. They have been exposed to simple and intuitive interfaces of other domains and expect that from libraries. Libraries create valuable metadata, but they have not adapted their systems to the requirements of the end-users. The goal should be clear: to provide a solution which users will find superior in comparison with the competition. This is the only scenario for libraries to maintain their position in the information sector.

References

1. *FRBRoo; Object-oriented definition and mapping to FRBRer (version 1.0.1, January 2010)* (2010). http://www.cidoc-crm.org/docs/frbr_oo/frbr_docs/FRBRoo_V1.0.1.pdf
2. Hegna, K. and Murtomaa, E. (2002). *Data mining MARC to find: FRBR?*. <http://folk.uio.no/knuthe/dok/frbr/datamining.pdf>
3. Hickey, T. and O'Neill, E. (2005). FRBRizing OCLC's WorldCat. *Cataloging & Classification Quarterly*, 39 (3/4), 239-251.
4. International Federation of Library Associations and Institutions. Study Group on the Functional Requirements for Bibliographic Records (1998). *Functional Requirements for Bibliographic Records: final report*. Munich, Germany: KG Saur
5. International Federation of Library Associations and Institutions. Working Group on Functional Requirements and Numbering of Authority Records. (2009). *Functional Requirements for Authority Data: a conceptual model*. Munich, Germany: KG Saur.

6. International Federation of Library Associations and Institutions. Working Group on Functional Requirements for Subject Authority Records. (2010). *Functional Requirements for Subject Authority Data (FRSAD): a conceptual model*. June 2010. <http://www.ifla.org/files/classification-and-indexing/functional-requirements-for-subject-authority-data/frsad-final-report.pdf>
7. Leskovec, M. (2005). *Delo, izrazna oblika, pojavná oblika : kaj uporabniki res iščejo? (Work, expression, manifestation: what are user really looking for)*. BS thesis. Ljubljana: Univerza v Ljubljani, Filozofska fakulteta.
8. Library of Congress. Working Group on the Future of Bibliographic Control. (2008). *On the record: report of the Library of Congress Working Group on the Future of Bibliographic Control*. <http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf>
9. Madison, O. (2005). The origins of the IFLA study on Functional Requirements for Bibliographic Records. *Cataloging & Classification Quarterly*. 39 (3/4). 15-37.
10. Pisanski, J. and Žumer, M. (2010a). Mental models of the bibliographic universe. Part 1: Mental models of descriptions. *Journal of Documentation*, 66 (5), 643-667
11. Pisanski, J. and Žumer, M. (2010b). Mental models of the bibliographic universe. Part 2: Comparison task and conclusions. *Journal of Documentation*, 66 (5), 668-680
12. Pisanski, J., Žumer, M. and Aalberg, T. (2009). Frbrisation: towards a bright new future for national bibliographies. *World Library and Information Congress: 75th IFLA General Conference and Council, 23-27 August 2009, Milan, Italy*. <http://www.ifla.org/files/hq/papers/ifla75/77-pisanski-en.pdf>
13. Pisanski, J., Žumer, M. and Aalberg, T. (2010). Identifiers: bridging language barriers. *World Library and Information Congress: 76th IFLA General Conference and Assembly, 10-15 August 2010, Gothenburg, Sweden*. <http://www.ifla.org/files/hq/papers/ifla76/93-pisanski-en.pdf>
14. Working Group on the Future of Bibliographic Control (2008). *On the Record: Report of the Library of Congress Working Group on the Future of Bibliographic Control*. Retrieved 19.1.2009 from: <http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf>.

15. Zhang, Y. and Salaba, A. (2007). Critical issues and challenges facing FRBR research and practice. *Bulletin of the American Society for Information Science and Technology*, 33 (6), 30-31.
16. Zhang, Y. and Salaba, A. (2009). What is next for FRBR? A Delphi study. *The Library Quarterly*, 79 (2), 233-255.