

Conceptual models: museums & libraries: towards an object-oriented formulation of FRBR aligned on the CIDOC CRM ontology

*Maja Žumer (University of Ljubljana) & Patrick Le Bœuf (National Library of France)
ELAG 2006 “New tools and new library practices”, Bucharest, 26 April 2006*

FRBR (“Functional Requirements for Bibliographic Records”) is a name that should sound familiar to many attendants of the ELAG Conferences, as ELAG devoted regular workshops to it between 1998 and 2003. This entity-relationship model was developed on behalf of the International Federation of Library Associations and Institutions (IFLA) from 1991 to 1997, and published in 1998 by Saur. The IFLA Cataloguing Section appointed in 2002 a Review Group to maintain it and ensure its evolution, on the same pattern as the IFLA ISBD Review Group ensures the evolution of ISBDs. The FRBR model covers “textual, music, cartographic, audio-visual, graphic, three-dimensional materials” (*FRBR Final Report*, p. 7).

The key concepts of FRBR are represented by 10 entities that are distributed over 3 groups.

Group 1 of entities represents the objects primarily described in library catalogues, and comprises:

- Work: a creation of the mind, a set of concepts;
- Expression: one of the possible “texts” (in the broadest, not limited to linguistic outputs, sense of the term) that express a given creation of the mind, a set of signs;
- Manifestation: one of the possible embodiments of an Expression;
- and Item: an individual exemplar of a Manifestation.

Work is *realised* through Expression, Expression is *embodied* in Manifestation, Manifestation is *exemplified* by Item.

Group 2 of entities represents the agents who performed activities related to instances of any of the entities defined in Group 1, and comprises:

- Person: a human being;
- and Corporate Body: a group of human beings.

Group 3 of entities represents what works are about (i.e., the entities that have a “subject relationship” to the Work entity), and comprises:

- all of the entities already declared in Group 1 and Group 2;
- plus: Concept, Object, Event, and Place.

CIDOC CRM (Comité International pour la DOcumentation – Conceptual Reference Model) is not totally unknown to ELAG attendants either, as it was introduced as a topic for a workshop at ELAG 2000 (see <<http://nuk.uni-lj.si/elag2000/2000WS11.HTML>>). It was originally developed by the ICOM (International Council of Museums) CIDOC, from 1996 on, and is currently maintained by an ad hoc group named CRM-SIG (Special Interest Group). It will soon be published as ISO standard 21127. It covers any kind of data (either “descriptive” data or “authority” data) created by museums in the fields of fine arts, archaeology, natural history, etc.

The central notion in CIDOC CRM is the notion of Event: something that happens in space and time and brings about some change in the world.

An event (i.e., an instance of the Event class) can involve:

- instances of the Actor class (persons, groups...), who can play a decisive role in provoking the event or just witness it or undergo it, and who are referred to through instances of Actor Appellation;
- bits of the physical world and/or creations of the mind (i.e., instances of the class named Physical Thing and/or the class named Conceptual Object; e.g., canvass and paint on the one hand, and the image formed by the paint on the canvass, on the other hand), which are referred to through instances of Appellation (names, titles, codes, whatever).

In addition, an event:

- occurs in time, and has therefore a duration, i.e., an instance of the class named Time-Span, which is referred to through instances of Time Appellation (e.g., instances of Date);
- and occurs in space, and can as such be related to an instance of Place, which is referred to through an instance of Place Appellation.

Finally, we use the notion of Type (a further class declared in CIDOC CRM) to categorise all those things and produce a well-organised view of this complex, chaotic real world.

The idea that both the library and museum communities might benefit from harmonising the two models was first expressed in 2000, on the occasion of ELAG 2000 in Paris, with Nicholas Crofts and Dan Matei drafting on the spot a preliminary object-oriented representation of the FRBR model entities roughly mapped to CIDOC CRM classes. This idea grew up in the following years and eventually led to the formation in 2003 of the International Working Group on FRBR/CIDOC CRM Harmonisation, that brings together representatives from both communities. This group, chaired by Martin Doerr (ICS FORTH, Greece) (assisted by Patrick Le Bœuf), is affiliated at the same time to the IFLA FRBR Review Group and the CIDOC CRM Special Interest Group (CRM-SIG).

The main goal is to reach a common view of cultural heritage information with respect to modelling, standards, recommendations and practices. Libraries and museums are “memory institutions” – both strive to preserve cultural heritage objects, and information about such objects, and they often share the same users. Besides, the boundary between them is often blurred: libraries hold a number of “museum objects” and museums hold a number of “library objects;” the cultural heritage objects preserved in both types of institutions were created in the same cultural *context* or *period*, sometimes by the same *agents*, and they provide evidence of comparable *cultural features*. It seems therefore appropriate to build a common conceptualisation of the information gathered by the two types of organisations about cultural heritage.

Expressing the FRBR model in a different formalism than the one in which it was originally developed is also a good opportunity to correct some semantic inconsistencies or inaccuracies in the formulation of FRBR, that may be regarded as negligible as far as FRBR_{ER} is only used in a library catalogue context, but that prove to be quite crucial from the moment one strives to design an overall model for the integration of cultural heritage related information.

Mediation tools and Semantic Web activities require an integrated, shared ontology for the information accumulated by both libraries and museums for all the collections that they hold, seen as a continuum from highly “standardised” products such as books, CDs, DVDs, etc., to “raw” materials such as plants or stones, through “in-between” objects such as draft manuscripts or engraving plates. Besides, such typical “library objects” as books can be about museum objects, and museum objects can represent events or characters found in books (e.g., “Ophelia’s death”): such interrelationships should be either integrated in common information storage, or at least virtually integrated through mediation devices that allow a query to be

simultaneously launched on distinct information depositories, which requires common semantic tools such as FRBR_{OO} plugged into CIDOC CRM.

The CIDOC CRM model is influenced by the process of FRBR's re-formulation as well. Modelling bibliographic information highlights some issues that may have been overlooked during the development of CIDOC CRM, and the way such issues were addressed in FRBR_{OO} resulted in some cases in making changes in the CIDOC CRM model.

The Working Group also benefits from the diversity of its members. For instance, the two authors of this paper have slightly differing views about the purposes of this endeavour. For Maja, it should be an opportunity to revise in depth the very principles on which cataloguing rules are based, and to get rid of the many routines we are too lazy to shake away. It should also be an opportunity to shift at last from the card-catalogue paradigm to a real automated catalogue paradigm. For Patrick, it is at least equally important to design a tool that will enable Semantic Web applications to extract the exact meaning of the bibliographic information contained in catalogue records such as they stand, no matter how imperfect and old-fashioned they are, and to integrate that meaning with the meaning conveyed by museum information, through the alignment of FRBR on CIDOC CRM. Those two views are far from being conflicting anyway, and both purposes can be attained.

The Working Group had 6 meetings so far. Detailed reports are not yet publicly available, but a first draft definition of FRBR as an object-oriented model aligned on CIDOC CRM will be published very soon. Of course, this first draft version is likely to be modified very rapidly, but it will manifest the Group's will to share the result of its work with the interested community.

These meetings consist of "translating" FRBR entities and attributes into the object-oriented formalism used for CIDOC CRM. This translation process borrows as much as possible from CIDOC CRM, but occasionally it also gives back to the model developed by the museum community.

We follow three main principles:

- Take a user (or use)-centered approach;
- Whenever there are too many attributes declared in FRBR_{ER} for a given entity, this is an indication that the entity is too complex to be regarded as a primitive concept of the underlying ontology, which shows that such entities should be split into more "atomic" notions;
- Whenever a given attribute actually refers to an event (e.g., "date" makes no sense by itself: it is necessarily the date at which some event occurred), we strive to make the event explicit.

As mentioned before, the work done by the FRBR/CRM Harmonisation group will have consequences for our understanding of FRBR and the future development of the model. We would like to illustrate our work with some examples.

The Work entity, such as defined in FRBR, is "a distinct intellectual or artistic creation." "We recognize the *work* through individual realizations or *expressions* of the work, but the *work* itself exists only in the commonality of content between and among the various *expressions* of the *work*." This leads to several possible interpretations of the meaning of *work*:

- concepts shared by a number of individual expressions
- overall concept (intellectual content) of a publication
- intellectual content of a particular expression (=set of signs), regardless of its materialization.

FRBR_{OO} retains the general notion of "Work" as a superclass for the various possible ways of interpreting the FRBR definitions. F46 Individual Work corresponds to the concepts associated with one complete set of signs; F43 Publication Work comprises the publisher's

intellectual contribution to a given publication; and F21 Complex Work is closer to the main interpretation in FRBR. Additionally, in order to deal with aggregates, F48 Container work is defined. Due to the fact that the approach is more “event-based,” there is also an emphasis on the creation process.

In FRBR₀₀, there is a distinction between expressions that convey the complete idea of the work they realise, and expressions that convey only a part of it: that is, between instances of F20 Self-Contained Expression and instances of F23 Expression Fragment.

The Manifestation entity is defined in FRBR in such a way that it could be interpreted as something physical and conceptual at the same time: it is defined at the same time as “the *physical embodiment* of an expression of a work,” and as an entity that “*represents* all the physical objects that bear the same characteristics,” i.e., as both a physical artefact and a (mental) representation of physical artefacts (a set). The Manifestation covers either a manuscript (in which case Manifestation overlaps with Item) or a publication (in which case Manifestation is both a Type and an Information Object).

FRBR₀₀ strives to solve such logical inconsistencies, and had to “split” the Manifestation entity into two distinct classes, corresponding to the two possible ways of interpreting the ambiguous definition provided for Manifestation in FRBR, namely F3 Manifestation Product Type and F4 Manifestation Singleton. Whereas F3 Manifestation Product Type is declared as a subclass of the CIDOC CRM class E55 Type, and therefore as a subclass, too, of the CIDOC CRM class E28 Conceptual Object (a merely abstract notion), F4 Manifestation Singleton is declared as a subclass of the CIDOC CRM class E24 Physical Man-Made Thing, and therefore as a subclass, too, of the CIDOC CRM class E18 Physical Thing.

The current cataloguing rules (and practice) are based on Paris principles and ISBD. They were originally developed for card catalogues and several researchers ascribe the difficulties users have with OPACs to the fact that our current catalogues are still only an electronic copy of a card catalogue. A FRBR-based catalogue could, on the other hand, use the potential of the new technology and become a different tool.

Therefore we may question the decision to base the definition of attributes on the existing practice. In addition, the FRBR/CRM analysis showed that some attributes are assigned to wrong entities.

Future tasks will involve the examination of all other FRBR entities (Person, Group, Concept, Place, Event, and Object), of all FRAR entities that are not mentioned in FRBR, and of all relationships described in both FRBR and FRAR. In addition FRSAR, when developed, will have to be taken into account, too. The resulting picture will be formalised and stabilised, and will result in a full-length description of FRBR₀₀, which will be submitted for approval to both the CIDOC CRM SIG and the IFLA FRBR Review Group (and the IFLA Cataloguing Section). It is expected that FRBR₀₀ will be regarded as a new, “official” release of the IFLA FRBR model.

FRBR₀₀ will be used for implementation purposes, most notably in the context of integrated information system design and Semantic Web activities.