

Functional Requirements for Subject Authority Records (FRSAR): A Conceptual Model of Aboutness

Maja Žumer¹, Athena Salaba², and Marcia Lei Zeng²

¹ University of Ljubljana, Slovenia
University of Ljubljana, Faculty of Arts Askerceva 2, 1000 Ljubljana Slovenia

{Maja.Zumer}@nuk.uni-lj.si

² Kent State University, Kent, Ohio, USA
314 Library, PO Box 5190, Kent, Ohio 44224, USA

{asalaba,mzeng}@kent.edu

Abstract. Provides a brief overview of the activities of the IFLA Functional Requirements for Subject Authority Records (FRSAR) working group. Introduces the group's terms of reference and the work completed so far, including definitions of user tasks and subject entities. Discusses the development of the entity-relationship conceptual model of subject entities in the bibliographic universe.

Keywords: FRSAR, IFLA, subject access, knowledge organization systems (KOS), conceptual model, subject authority data, subject authority records.

1 Background

IFLA's Working Group on the Functional Requirements for Bibliographic Records (FRBR)¹ developed an entity-relationship conceptual model of the bibliographic universe [1]. The working group began with identifying essential user tasks and then defining the relevant entities, their attributes, and relationships. The basic elements of the FRBR model are the result of a logical analysis undertaken by members of the Study Group of the data typically reflected in bibliographic records. The entities are divided into three groups:

- Group 1 consists of four entities that are the product of intellectual or artistic endeavors: *work*, *expression*, *manifestation*, and *item*.
- Group 2 consists of entities that are actors, those who are responsible for the intellectual or artistic content, the physical production and dissemination, or the custodianship, of Group 1 entities: *person*, *corporate body*
- Group 3 consists of entities that serve as subjects of intellectual or artistic endeavor.

¹ IFLA WG on Functional Requirements for Bibliographic Records (FRBR), <http://www.ifla.org/VII/s13/wgfrbr/finalreport.htm>

We can illustrate Group 1 entities using the example of *The Da Vinci Code* by Dan Brown. *Items* are individual copies of the book (for instance, my personal copy vs. a copy owned by the Slovenian National Library with call number 601547, etc.). A particular publication is called a *manifestation* by FRBR (e.g., the edition published by Bantam in 2003). However, books are more than just physical objects and the intellectual content is really our main focus. For instance, one may be interested in the original text, in the Slovenian translation, or in an abridged version. These are all *expressions* of the same work. Therefore, the Bantam 2003 edition is a *manifestation* that contains Brown's original text (an *expression*) of *The Da Vinci Code* (the *work*).

In most structured retrieval systems, information regarding the bibliographic universe is not recorded exclusively in bibliographic records. Authority records are used to record information about all controlled access points that are currently included in bibliographic records or have the potential to be assigned as access points in bibliographic records. Controlled access points include names of entities identified by FRBR such as members of Group 2 (*persons, corporate bodies*), titles of Group 1 entities (*works, expressions, manifestations and items*), and terms for Group 3 entities. In the FRBR model the entities of all three groups are defined, but the main focus is on the first group.

The second IFLA Working Group, the Functional Requirements and Numbering of Authority Records (FRANAR)² is charged with the task of continuing the work of FRBR by developing a conceptual model for authority records. In the 2007 draft of the *Functional Requirements for Authority Data* (FRAD) model, the group defines authority records as aggregates of information regarding entities that are assigned as controlled access points in bibliographic records and focuses on Group 1 and Group 2 entities (IFLA, 2007).

A third IFLA Working Group, co-chaired by the authors, was formed in April 2005 and charged with the task of developing a conceptual model for the Functional Requirements for Subject Authority Records (FRSAR)³. All controlled access points related to all three entity groups as defined by the FRBR conceptual model have the potential to be the topic of a work. In other words, Group 1, 2 and 3 entities can have an "is-the-subject-of" relationship with *work*. FRSAR's terms of reference are:

- Build a conceptual model of Group 3 entities within the FRBR framework as they relate to the aboutness of works,
- Provide a clearly defined, structured frame of reference for relating the data that are recorded in subject authority records to the needs of the users of those records, and
- Assist in an assessment of the potential for international sharing and use of subject authority data both within the library sector and beyond.

² IFLA WG on the Functional Requirements and Numbering of Authority Records (FRANAR), <http://www.ifla.org/VII/d4/wg-franar.htm>

³ IFLA WG on the Functional Requirements for Subject Authority Records (FRSAR), <http://www.ifla.org/VII/s29/wgfrsar.htm>

2 Users and Context of Use of Subject Authority Records/Data

During the process of developing an entity-relationship conceptual model of subject authority records, the FRSAR Working Group initially analyzed who the users of subject authority data are, identified contexts of the use of the data, and described some of the use scenarios. Possible subject authority record data user groups include a) information professionals who create metadata, b) reference and public services librarians and other information professionals who are searching for information as intermediaries, c) controlled vocabulary creators, such as catalogers, thesaurus and ontology creators, and d) end-users using information retrieval systems to fulfill their information needs.

The FRSAR Working Group felt strongly that, in order to define user tasks, an actual user study was necessary, and two studies were therefore conducted. The first was a pilot study at the 2006 Semantic Technologies Conference (San Jose, California, USA). Most study participants were either creators of semantic tools, including controlled vocabularies, taxonomies and ontologies, or developers and managers of semantic technology systems. The second study was an international survey sent to information professionals throughout the world during the months of May-September 2007. Participants included authority record creators, vocabulary creators and managers, catalogers, metadata librarians, and reference librarians among others. Participants were asked to describe their work and their use of subject authority data in different contexts, including cataloging/metadata creation, subject authority work, and searching or helping others search bibliographic information. The results of these studies enriched our understanding of subject authority data use and informed and further confirmed the FRSAR user tasks.

Based on the results from our user studies, five subject authority data user tasks, representing uses by all the above user groups, are defined as follows:

Find: To find a subject entity or set of entities corresponding to stated criteria.

Identify: To identify a subject entity based on certain attributes/characteristics.

Select: To select a subject entity.

Obtain: To obtain additional information about the subject entity and/or to obtain bibliographic records or resources about this subject entity.

Explore: To explore relationships between subject entities, correlations to other subject vocabularies and structure of a subject domain.

3 The Conceptual Model

An examination of other models covering subject data and a comparison of the current Group 3 entities served as a starting point. Group 3 entities defined by FRBR include *concept* (an abstract notion or idea), *object* (a material thing), *event* (an action or occurrence), and *place* (a location). This part of the FRBR model has been criticized and several issues regarding Group 3 entities have been raised, particularly the unsymmetrical treatment of space and time and the fact that processes are not modeled.

The working group investigated the approaches of other models, specifically the proposed model by Buizza and Guerrini [2] and <indec> [3] and analyzed possible solutions from conservative (only making minor amendments) to radical (proposing a new model). A small study was performed, in which four students and faculty members at the Kent State University School of Library and Information Science classified existing subject terms used by the NSDL (National Science Digital Library) contributors. These include about 3000 terms assigned based on a variety of subject vocabularies and free keywords. They classified terms into six categories: concrete stuff, abstract stuff, event, time, place, and other. The results show that there is a blurred distinction between concrete and abstract concepts and difficulties in the classification of named instances, which resulted in many terms being put into the ‘other’ category. This indicates that it would be difficult for any user (end user, librarian or vocabulary developer) to conduct such a task when using subject authority data. These categories also do not seem helpful or necessary to the end users.

As a result, the FRSAR Entity sub-group proposed a more abstract conceptual model:

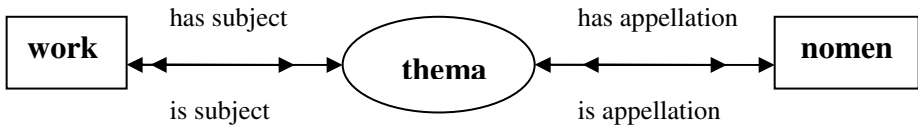


Fig. 1. Conceptual model

Thema: Anything that can be the subject of a *work*.

Nomen: Any alphanumeric, sound, visual, or any other symbol, sign, or combination of symbols by which a *thema* is known, referred to, or addressed.

This model first confirms what FRBR has already defined: WORK has subject THEMA, and proposes a new part: THEMA has appellation NOMEN. The use of Latin terms is to avoid mapping to an English term (such as subject or concept) that has been understood and translated with a different understanding. The terms for the entities and relationships are subject to change.

Thema therefore includes existing Group 1 and Group 2 entities, and, in addition, all other subjects of works. In a particular application *thema* would normally have implementation-specific types. In the current discussions the most important distinction seems to be the one between named particulars and classes.

In general, *nomen* can be domain-, community- and language-specific. Meanwhile, two important specific types of *nomen* are recognized: *identifier* (name assigned to an entity, which is usually persistent and unique within a domain) and *constructed name* (name constructed in the authority control/vocabulary maintenance process, which usually serves as an access point), for which the term ‘controlled access point’ is used in FRAD. Attributes of *nomen* serve to carry information about a particular instance and typically include but are not limited to: type, origin/source/system/vocabulary, medium, language, script, transliteration/transcription, time and place of validity, target community, and status.

In addition to the many-to-many relationships between *work* and *thema* and *thema* and *nomen*, as illustrated in Figure 1, there are *thema-to-thema* and *nomen-to-nomen* relationships. Particular *thema-to-thema* relationships are implementation-dependent but the generally applicable ones include Hierarchical (Partitive, Generic, and Instantiation) and Associative (=other) relationships. Some *nomen-to-nomen* relationships are: Partitive (parts/components of a *nomen* may be *nomen*) and Equivalence (two *nomen* are equivalent, if they have an 'is appellation' relationship with the same *thema*). The equivalence *nomen-to-nomen* relationship can further be specified. For example, replaces/is replaced by, has variant form/is variant form, has derivation/is derived from, etc.

A good example at the implementation-level is chemical substances. For example, each drug might have its chemical name(s), drug name, trade name, generic name (U.S. Adopted name), system-specific identifiers such as CAS Registry Number, classification code (alpha or numerical), as well as other unique expressions such as a flat structure diagram, structure diagram (include stereo bonds), molecular formulas, etc. The relationships between these names, identifiers, terms, and other expressions are *nomen-to-nomen* because they are how this same drug is known, referred to, and addressed by specific systems. Meanwhile, in addition to the relationship with other substances, a drug itself contains various compounds and elements, which form *thema-to-thema* relationships.

The Working Group is currently analyzing attributes and relationships in view of defined user tasks and testing with samples collected from different domains.

4 Interoperability with Other Communities

The final term of reference for the FRSAR Working Group is to assist in an assessment of the potential for international sharing and use of subject authority data both within the library sector and beyond. The challenges in true global sharing and use of subject authority data come from many technological aspects (such as heterogeneous structures), various languages and scripts, diverse construction rules and best practice guides, and dynamically developed and advanced encoding schemas, especially when other communities (museum, archive, science, education) are involved. It is important to separate what we usually call concepts (or topics or subjects) from what they are known by, referred to, or addressed. The potential value of this *work-thema-nomen* model for subject authority data is obvious.

Among the efforts to achieve global sharing and use of subject authority data, some have, in fact, focused on *nomen* (for example, a translated metadata vocabulary, a symmetrical multilingual thesaurus, a multi-access index to a vocabulary, etc.) However, many efforts actually have had to focus on the conceptual level, for example, when mapping between two thesauri. These kinds of efforts usually encounter many more challenges because they deal with the intension and extension of these concepts as well as the relationships among them.

This *thema-nomen* conceptual model also matches well the encoding schemas such as SKOS (Simple Knowledge Organisation System), OWL (Web Ontology Language), and more general, RDF encoding which uses URIs as the basis of their mechanism for identifying the subjects, predicates, and objects in statements. SKOS

defines the classes and properties sufficient to represent the common features found in a standard thesaurus. It is based on a concept-centric view of the vocabulary, where primitive objects are not terms, but abstract concepts represented by terms. Each SKOS concept is defined as an RDF resource. Each concept can have RDF properties attached, including: one or more preferred terms (at most one in each natural language), alternative terms or synonyms, and definitions and notes, with specification of their language [4].

When the DCMI Abstract Model [5] became a DCMI Recommendation in 2007, its one-to-one principle (i.e., each DC metadata description describes one, and only one, resource) was recognized or followed by other metadata standards. Under the one-to-one principle, a record can contain *description sets* that may contain *descriptions* composed by *statements* which use property-value pairs. Consequently, information can be processed, exchanged, referred to, and linked at the *statement* level. This information model is independent of any particular encoding syntax, thus facilitating the development of better mappings and cross-syntax translations (DCMI, 2007). The conceptual model proposed by the FRSAR group corresponds to this abstract model in that it allows any *thema* to be independent of any *nomen*, including any syntax a *nomen* may use. Accordingly, this conceptual model will facilitate the sharing and reuse of subject authority data among subject vocabularies and interoperability of resource metadata.

References

1. Functional Requirements for Bibliographic Records: final report. KG Saur, München (1998)
2. Buizza, P., Guerrini, M.: A Conceptual Model for the New “Soggettario”: Subject Indexing in the Light of FRBR. *Cataloging & Classification Quarterly* 34(4), 31–45 (2002)
3. Rust, G., Bide, M.: The <indec> metadata framework: Principles, model and data dictionary, Indecs Framework Ltd. (2000), <http://www.indec.org/pdf/framework.pdf>
4. SKOS Core Guide.: W3C Working Draft 2 (November 2000), <http://www.w3.org/2004/02/skos/>
5. DCMI: DCMI Abstract Model (2007), <http://dublincore.org/documents/abstract-model/>