



Introducing FRSAD and Mapping it with SKOS and other models

(Based on the work of the FRSAR Working Group)

Marcia Lei Zeng
Kent State University
Kent, Ohio, USA
E-mail: mzeng@kent.edu

Maja Žumer
University of Ljubljana
Ljubljana, Slovenia
E-mail: maja.zumer@ff.uni-lj.si

Meeting:

200. Classification and Indexing

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>

Abstract:

The Functional Requirements for Subject Authority Records (FRSAR) Working Group was formed in 2005 as the third IFLA working group of the FRBR family to address subject authority data issues and to investigate the direct and indirect uses of subject authority data by a wide range of users. This paper introduces the Functional Requirements for Subject Authority Data (FRSAD), the model developed by the FRSAR Working Group, and discusses it in the context of other related conceptual models defined in the specifications during recent years, including the British Standard BS8723-5: Structured vocabularies for information retrieval – Guide Part 5: Exchange formats and protocols for interoperability, W3C's SKOS Simple Knowledge Organization System Reference, and OWL Web Ontology Language Reference. These models enable the consideration of the functions of subject authority data and concept schemes at a higher level that is independent of any implementation, system, or specific context, while allowing us to focus on the semantics, structures, and interoperability of subject authority data.

Functional Requirements for Subject Authority Records (FRSAR) Working Group (WG) is the third IFLA working group of the FRBR family. Formed in April 2005, it is charged with the task of developing a conceptual model of FRBR Group 3 entities within the FRBR framework as they relate to the “aboutness” of *works*. This paper introduces the Functional Requirements for Subject Authority Data (FRSAD), the model developed by the FRSAR WG, and discusses it in the context of other related conceptual models developed along with the movement of the Linked Data for the Semantic Web during recent years.

1. Background

IFLA FRBR Group 3 entities are recognized as the subjects of *works* (i.e. the results of intellectual or artistic endeavour). They "represent an **additional** set of entities that serve as the **subjects** of works" (IFLA, 1999: 16 and IFLA, 2008: 17, emphasis added by authors), in addition to Group 1 and 2 entities, which can also be subjects of *works*. FRBR Group 3 entities includes concept (an abstract notion or idea), object (a material thing), event (an action or occurrence), and place (a location) (IFLA, 1999). The FRSAR Working Group was formed to address subject authority data issues and to investigate the direct and indirect uses of subject authority data by a wide range of users. The role of FRSAR WG was defined in the following terms of reference:

- to build a conceptual model of Group 3 entities within the FRBR framework as they relate to the aboutness of *works*,
- to 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
- to assist in an assessment of the potential for international sharing and use of subject authority data both within the library sector and beyond.

Two FRSAR sub-groups were formed. The User Tasks subgroup focused on user studies and the definition of user tasks. The Subject Entities subgroup mainly focused on the Group 3 entities including the study of current FRBR Group 3 entities and alternatives. The sub-groups conducted two user tasks surveys and one small experiment in 2006 and 2007. About twenty group meetings

were held in addition to numerous online discussions. A draft report was reviewed by the Advisory Group members in late 2008 and early 2009. The revised report is to undergo a world-wide review before IFLA 2009 Milan conference.

2. The FRSAD Model

From the time the FRSAR Working Group was formed, there seems to have been a general agreement in the FRBR research community that Group 3 entities should be revisited (Delsey, 2005). In the beginning, the FRSAR WG was simply thinking of enhancing the existing model based on FRBR Group 3 entities and lengthy discussions on what should be added (e.g., 'time') occurred. The WG investigated the approaches of other existing models, which include: the <indecs> model (Rust and Bide, 2000), Ranganathan's facets (Ranganathan, 1962), and the pragmatic list of entities developed by two Italian researchers, Buizza and Guerrini (2002). These models present solid references for revising the FRBR conceptual model. The WG analyzed and discussed possible solutions based on each of these models, from conservative (making minor amendments of FRBR Group 3) to radical (proposing a completely new model). However the WG found that none of the models and the ones based on them could be universal enough to reflect the needs of today's subject authority data, considering particularly different domains and subject access tools. By 2007, the FRSAR WG shifted focus to the development of a new conceptual model of Group 3 entities within the FRBR framework as they relate to the "aboutness" of *works*.

In this framework, 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, all of the Group 1, 2 and 3 entities can have an "is-subject-of" relationship with the *work*. The FRSAR Subject Entity sub-group proposed a more abstract conceptual model and presented it at the 2007 IFLA Conference. As presented in the following Figure, the model should be understood with two key points of view:

1. This model confirms one of the basic relationships defined in FRBR: **WORK has as subject THEMA / THEMA is subject of WORK.**

1.1 **THEMA** is the term used to refer to anything that can be subject of a *work*. It is defined as "any FRBR entity as used as a subject of a *work*".

1.2 **THEMA** includes any of the FRBR entities -- existing Group 1 and Group 2 entities and, in addition, all other subjects of *works*. While an entity on its own, it can be viewed as a super-entity or super-class of all FRBR entities, enabling us to model relationships and attributes on a more general and abstract level.

2. This model also proposes a new relationship: **THEMA has appellation NOMEN / NOMEN is appellation of THEMA.**

2.1 **NOMEN** is defined as any sign or sequence of signs (alphanumeric characters, symbols, sound, etc.) by which a *thema* is known, referred to or addressed. For example, "love", "∞", or "595.733".

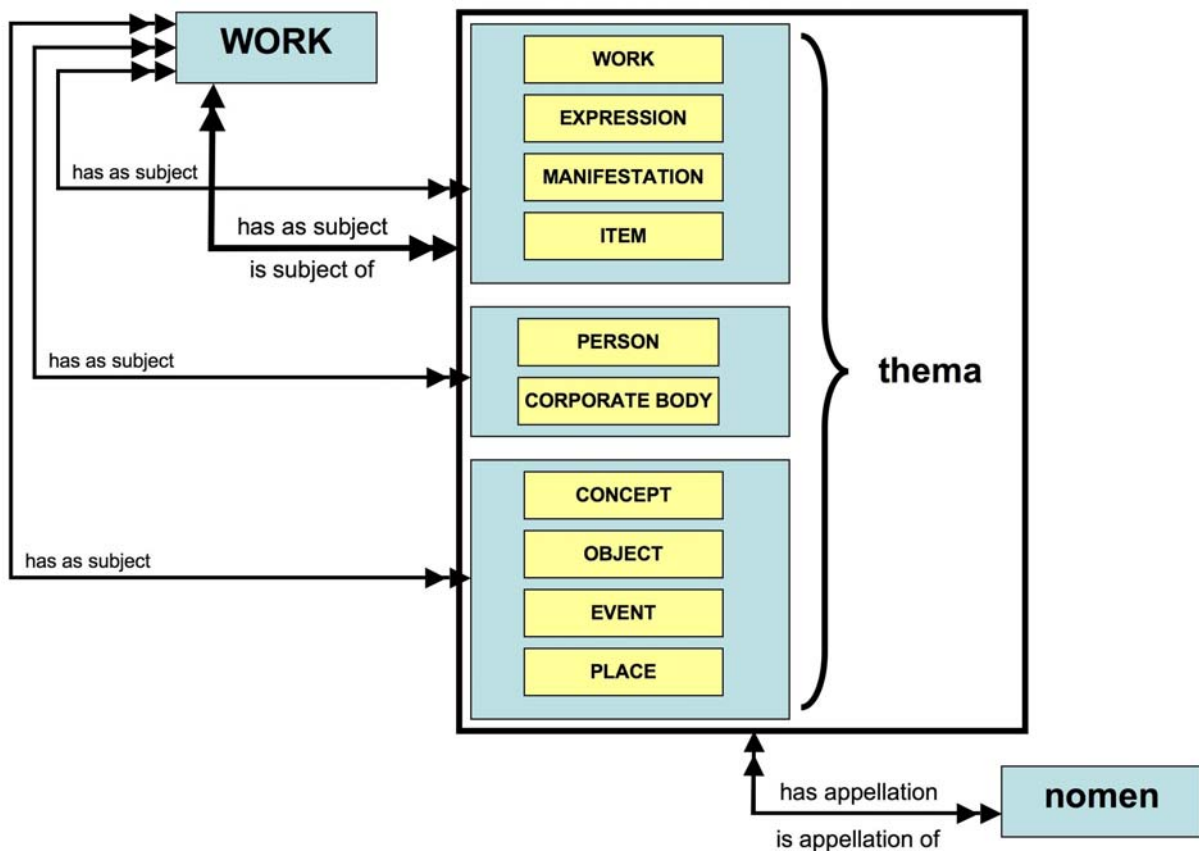


Figure 1. FRSAD's relation to FRBR

To simplify the above figure, the FRSAD model can be presented by the following illustration:



Figure 2. FRSAD Conceptual Model.

Thus the relationships between entities are to be explained further:

- The "has as subject/is subject of" relationships are many-to-many relationships. Any *work* can have more than one *thema*; and any *thema* can be the subject of more than one *work*.
- In general (i.e. in natural language or when mapping different vocabularies) the "has-appellation/is appellation of" relationship is also a many-to-many relationship. A *thema* has one or more *nomen* and there may be a *nomen* referring to more than one *thema*.
- It is important to note that, in a given controlled vocabulary and within a domain, though, a *nomen* should be an appellation of only one *thema*,

We can take “A brief history of time: from the big bang to black holes” by Stephen W. Hawking as an example. The *work* has several *themas*: cosmology, space and time, unification of physics, black holes, big bang, history of time, universe. There are many other *works* about any of these *themas*. For any of the *themas* in the list, *nomens* presented here are terms in English, but there are other possible *nomens* in other languages or formed according to other controlled vocabularies such as subject heading lists, classification systems, etc.

The issues of (a) the complexity and granularity of *themas*, (b) attributes of *thema* and *nomen*, and (c) relationships between and among *themas*, *nomens*, and *thema* to *nomen* are all discussed in the FRSAD draft report.

The importance of the *THEMA-NOMEN* model for the subject authority data is to separate *themas* (or "subjects", "concepts", "classes (of concepts)", "topics", etc.) from what they are known as, referred to, or addressed as. Among the efforts to achieve global sharing and use of

subject authority data, some efforts have focused on *nomens* (for example, a translated metadata vocabulary, a symmetrical multilingual thesaurus, or a multiple-access index to a vocabulary). However, most efforts have focused on the conceptual level, e.g., the semantic mapping between two thesauri or between a classification scheme and a thesaurus. Such efforts usually encounter much greater challenges because they are concerned with the *themes* as well as the relationships among the *themes*.

3. Mapping FRSAD with SKOS and other models

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 sharing come from many aspects: heterogeneous structures, various languages and scripts, diverse construction rules and practices, and dynamically developed encoding schemas. A preliminary comparison of FRSAD and other models will enable us to consider at a higher level that is independent of any implementation, system, or specific context, and will allow us to focus on the semantics, structures, and interoperability.

BS8723 and ISO 25964: A model for structured vocabularies (more specifically, thesauri) was defined by the British standard *BS8723-5: Structured vocabularies for information retrieval – Guide. Part 5: Exchange formats and protocols for interoperability* (DD 8723-5:2008). (The model, XML Schema, and examples are available at the BS8723 Official Development Website). This model has been slightly revised and included in the *ISO/CD 25964-1 Information and documentation — Thesauri and interoperability with other vocabularies — Part 1: Thesauri for information retrieval* (2008: 92) which went out for ballot at the beginning of 2009. It includes what is needed for modeling: (1) a whole thesaurus, (2) arrays of thesaurus concepts, and (3) records that document a thesaurus entry. In the model, each *concept* in a thesaurus is represented by one preferred term per language, and by any number of nonpreferred terms. The notation, scope note and broader/narrower/related term relationships apply to the concept as a whole, rather than to its preferred term. A unique identifier may be assigned to each concept (ISO/CD 25964-1, 2008). Overall, both this model and the FRSAD model represent these relationships: (1) *thema-and-nomen* (a record documenting a concept and its *nomen(s)*), (2) *thema-and-thema*

(hierarchical (broader, narrower, and top concepts)) and associative (related concepts), and (3) *nomen-and-nomen* (preferred and non-preferred, variant lexical forms, and in various languages).

SKOS: *SKOS Simple Knowledge Organization System Reference* (2009) defines classes and properties sufficiently for representing the common features found in a knowledge organization system such as thesaurus, taxonomy, controlled term lists, and other KOS structures. "Using SKOS, concepts can be identified using URIs, labeled with lexical strings in one or more natural languages, assigned notations (lexical codes), documented with various types of notes, linked to other concepts and organized into informal hierarchies and association networks, aggregated into concept schemes, grouped into labeled and/or ordered collections, and mapped to concepts in other schemes" (SKOS Reference, 2009: Synopsis). As an application of the RDF (Resource Description Framework), SKOS allows concepts to be composed and published on the World Wide Web, linked with data on the Web, and integrated into other concept schemes. Each SKOS concept is defined as an *RDF resource* and each concept can have *RDF properties* attached. SKOS model is based on a concept-centric view of vocabulary, where primitive objects are not labels; rather, they are concepts represented by labels. These can be matched to what have been defined in the FRSAD model, in terms of *thema*, *nomen* and their attributes. SKOS has also specific properties to represent all the semantic relationships, which matches the ones defined by FRSAD as well.

OWL: OWL Web Ontology Language is an ontology language for the Semantic Web with formally defined meaning. Ontologies are formalized vocabularies of terms (classes and properties), often covering a specific domain and shared by a community of users. They specify the definitions of terms by describing their relationships with other terms in the ontology (OWL 2 Web Ontology Language Document Overview, 2009). OWL ontologies provide classes, properties, individuals, and data values and are stored as Semantic Web documents. OWL 1 was mainly focused on constructs for expressing information about classes and individuals. OWL 2, the newest W3C working draft, offers new constructs for expressing additional restrictions on properties, new characteristics of properties, incompatibility of properties, properties chains and key properties (OWL 2 Web Ontology Language New Features and Rationale, 2009). OWL 2 provides axioms (statements that say what is true in the domain) that allow relationships to be established between class expressions, including: *SubClassOf*, *EquivalentClasses*, *DisjointClasses*,

and DisjointUnion. More importantly, in OWL 2, classes and property expressions are used to construct class expressions, (sometimes also called descriptions, and, in the description logic literature, complex concepts). It provides for enumeration of individuals and all standard Boolean connectives AND, OR, and NOT. The ObjectIntersectionOf, ObjectUnionOf, and ObjectComplementOf class expressions provide for the standard set-theoretic operations on class expressions. The ObjectOneOf class expression contains exactly the specified individuals (OWL 2 Web Ontology Language Structural Specification and Functional-Style Syntax, 2009). For the issues of the complexity and granularity of *themas* and comprehensive semantic relationships between and among *themas* that FRSAD attempted to cover, OWL has great matches.

DCMI Abstract Model: When the DCMI Abstract Model (DCMI AM) became a DCMI Recommendation in 2007, its one-to-one principle (i.e., each DC metadata description describes one, and only one, resource) has been recognized or followed by other metadata standards, e.g., the VRA Core 4.0 released by the Visual Resources Association in 2007 (VRA Core 4.0, 2007). According to the DCMI model, a *record* can contain *description sets*, which may contain *descriptions* composed of *statements*, which use *property-value* pairs (DCMI Abstract Model, 2007). This results in information that can be processed, exchanged, referred to, and linked to at the statement level. When a record contains descriptions of the resource, the individual descriptions also can be linked to the authority data that manages the values associated with those properties (e.g., the subject authority data, the property name authority data, or the geographic authority data). Such an information model is independent of any particular encoding syntax and facilitates the development of better mappings and cross-syntax translations. The FRSAD model corresponds to the DCMI Abstract Model by allowing any *thema* to be independent of any nomen, including any syntax that a *nomen* may use. Thus this conceptual model will facilitate the sharing and reuse of subject authority data amongst not only the subject vocabularies themselves, but also metadata resources.

In conclusion, the FRSAD model is developed with the goal 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 FRSAD model and other models developed along with the progress of the Semantic Web during the recent years enable the consideration of the functions of subject

authority data and concept schemes at a higher level that is independent of any implementation, system, or specific context, and will allow us to focus on the semantics, structures, and interoperability of subject authority data. Putting the subject authority data in the context of the Semantic Web developments, especially in the perspective of Linked Data, subject authority data that are modeled based on FRSAD and encoded in SKOS and OWL will be able to become part of the Linked Data and contribute to the development of the Semantic Web.

References:

BS8723 Official Development Website. (2008). <http://schemas.bs8723.org/Home.aspx> (accessed May 15, 2009).

Buizza, P. and Guerrini, M. A. (2002). Conceptual model for the new "Soggettario": Subject indexing in the light of FRBR. *Cataloging & Classification Quarterly*, 34(4): 31-45.

Delsey, T. (2005). Modeling subject access: Extending the FRBR and FRANAR conceptual models. *Cataloging & Classification Quarterly*, 39(3/4): 49-61.

DCMI Abstract Model. (2007). Powell, A., Nilsson, M., Naeve, A. Johnston, P. and Baker, T. eds. <http://dublincore.org/documents/abstract-model/> (accessed May 15, 2009).

DD 8723-5:2008 Structured vocabularies for information retrieval. Guide. Exchange formats and protocols for interoperability. (2008). The British Standards Institution.

IFLA. (1998). *Functional Requirements for Bibliographic Records: Final Report*. IFLA Study Group on the Functional Requirements for Bibliographic Records. München: KG Saur. <http://www.ifla.org/publications/functional-requirements-for-bibliographic-records> (accessed May 15, 2009).

IFLA. (2008). *Functional Requirements for Bibliographic Records: Final Report*. IFLA Study Group on the Functional Requirements for Bibliographic Records. Approved by the Standing Committee of the IFLA Section on Cataloguing September 1997, As amended and corrected through February 2008. <http://www.ifla.org/publications/functional-requirements-for-bibliographic-records> (accessed May 15, 2009).

ISO/CD 25964-1 Information and documentation — Thesauri and interoperability with other vocabularies — Part 1: Thesauri for information retrieval. (2008). ISO TC 46/SC 9/WG 8. Ballot version.

OWL 2 Web Ontology Language Structural Specification and Functional-Style Syntax. (2009). Motik, B, Patel-Schneider, P.F. and Parsia, B. eds. W3C Working Draft 21 April 2009. <http://www.w3.org/TR/owl2-syntax/> (accessed May 15, 2009).

OWL 2 Web Ontology Language New Features and Rationale. (2009). Golbreich, C., Wallace, E.K. eds. W3C Working Draft 21 April 2009. <http://www.w3.org/TR/owl2-new-features/> (accessed May 15, 2009).

OWL 2 Web Ontology Language Document Overview. (2009). W3C OWL Working Group. W3C Working Draft 21 April 2009. <http://www.w3.org/TR/owl2-overview/> (accessed May 15, 2009).

OWL Web Ontology Language Reference. (2004). Dean, M. and Schreiber, G. eds. W3C Recommendation 10 February 2004. <http://www.w3.org/TR/owl-ref/> (accessed May 15, 2009).

Ranganathan, S.R., (1962). Facet analysis: Fundamental categories. In: Ranganathan, S.R. *Elements of Library Classification*. 3rd. ed., Bombay, New York: Asia Publishing House. pp. 82-89.

Rust, G. and Bide, M. (2000). The <indecs> metadata framework : Principles, model and data dictionary. Indecs Framework Ltd. http://www.doi.org/topics/indecs/indecs_framework_2000.pdf (accessed May 15, 2009).

SKOS Simple Knowledge Organization System Reference. (2009). W3C Candidate Recommendation 17 March 2009; <http://www.w3.org/TR/2009/CR-skos-reference-20090317/> (accessed May 15, 2009).

VRA Core 4.0, (2007). Visual Resources Association Data Standards Committee. <http://www.vrweb.org/projects/vracore4/index.html> (accessed May 15, 2009).

***Acknowledgement:** This paper is based on the work of the FRSAR Working Group and suggestions from the Advisory Group, established by the IFLA Division IV. Bibliographic Control and especially the Section of Classification and Indexing. The Working Group members are: Leda Bultrini, Lois Mai Chan, Jonathan Furner, Edward O'Neill, Gerhard Riesthuis, Athena Salaba, Diane Vizine-Goetz, Ekaterina Zaytseva, Marcia Lei Zeng, and Maja Zumer. The Advisory Group members are: Victoria Francu, Hemalata Iyer, Dorothy McGarry, David Miller, Päivi Pekkarinen, and Barbara Tillett. The communication and sharing with the FRANAR Working Group chaired by Glenn Patton have also been very helpful. IFLA, OCLC, and Kent State University have provided funding, facilities, and tremendous support.