This paper is a pre-print version presented at the ISKO UK 2009 conference, 22-23 June, prior to peer review and editing. For published proceedings see special issue of Aslib Proceedings journal.

Mapping FRSAD Model and Other Abstract Models

Marcia Zeng, Kent State University, USA
Maja Zumer, University of Ljubljana, Slovenia

IFLA FRBR Group 3 entities "represent an additional set of entities that serve as the subjects of works" (IFLA, 1999: 16). A third IFLA Working Group of the FRBR family, FRSAR (Functional Requirements for Subject Authority Records), was formed in April 2005 and charged with the task of developing functional requirements and a conceptual model for subject authority records. One of the terms of reference is to build a conceptual model of Group 3 entities within the FRBR framework as they relate to the "aboutness" of works. In this framework all three entity groups as defined by the FRBR conceptual model have the potential to be the subject of a work. In other words, Group 1, 2 and 3 entities all can have an "is-subject-of" relationship with a work. The FRSAR Working Group proposed an abstract conceptual model and presented it at the IFLA Conference in August 2007. The model was further discussed and developed by the Working Group in 2008. The draft report prepared by the FRSAR Working Group has indicated that the focus of the model is on the authority ***data*** instead of authority ***records***, hence the abbreviation used in the report is FRSAD, i.e., Functional Requirements for Subject Authority Data.

1. The FRSAD Conceptual Model

First of all, the FRSAD model confirms what FRBR has already defined: **WORK has subject THEMA**. Here **THEMA** is the term used to refer to *anything that can be subject of a work*. Thema includes any FRBR entities. The relationship is shown below:

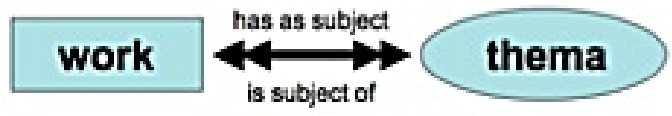


Figure 1. FRSAD Conceptual Model -- Part 1: WORK has subject THEMA

- Presented in the FRBR model, thema includes any FRBR entities, including existing Group 1 and Group 2 entities, and in addition, all other subjects of works.
- Presented with entity-relationship model, WORK-THEMA is a many-to-many: relationship: any work can have one or more thema, and any thema may be the subject of one or more works.

FRSAD WG has proposed another entity related to THEMA, i.e., **NOMEN**. Consequently this model also proposes a new relationship: **THEMA has appellation NOMEN**.

Figure 2. FRSAD Conceptual Model -- Part 2: Thema has appellation Nomen

NOMEN is a term used to refer to *any sign or sequence of signs (alphanumeric characters, symbols, sound, etc.) by which a thema is known, referred to or addressed.* (FRSAD Draft Report, Version 0.1, pp. 26.)

In general, (i.e. in natural languages or when mapping different vocabularies), the *has appelation/is appellation of* relationship is a many-to-many relationship. A thema has one or more nomen and there may be more than one thema referred to by the same nomen. However, in controlled vocabularies and within a domain, a nomen normally is an appellation of only ONE thema, as illustrated in Figure 3. (The attributes identified for nomen are not included in this discussion.)

Figure 3: Thema-nomen relationship in controlled vocabularies

The attributes of the entities are defined by FRSAD but will not be discussed in this paper.

2. Thema and Nomen

It is the nature of themas that they can vary substantially in complexity (or simplicity). It is virtually impossible to define what the 'atomic' level of a thema is, because any thema can be fragmented further. On the other side we intuitively associate some thema with single concepts and others as combinations of concepts. In the latter case the thema can also be seen as a relationship between concepts.

Often the complexity of a thema is associated with the complexity of the nomen by which it is represented. Since the proposed model introduces a clear split between the thema ('the thing') and the nomen ('the label' used to refer to it), the complexity of the semantic rules for creating or establishing nomen is not directly reflected in the complexity of the thema. Therefore in addition to the many-to-many relationships between work and thema and between thema and nomen, there are thema-to-thema and nomen-to-nomen relationships.

The importance of the THEMA-NOMEN model for the subject authority data is to separate what are usually called *concepts* (or *topics* and *subjects*) from what they are known by, referred to, or addressed. Among the efforts to achieve a global sharing and use of subject authority data, some efforts have focused on nomen (for example, a translated metadata vocabulary, a symmetrical multilingual thesaurus, or a multi-access index to a vocabulary). However, most efforts have focused on the conceptual level, e.g. mappings between two thesauri.

Such efforts usually encounter much greater challenges because they are concerned with the concepts as well as the relationships among the concepts.

3. Mapping with Other Abstract Models

This approach to separate thema from what it is known by, referred to, or addressed (i.e., nomen) can find its root in the classical model and literature reviewed below.

Ogden & Richard's (1923) famous triangle of meaning illustrated the relationship between language, thought content, and referent. The graph implies that the referent of an expression (a word or another sign or symbol) is relative to different language users. The theoretical foundation of it can be traced back to Aristotle who distinguished objects, the words that refer to them, and the corresponding experiences in the *psyche*, as well as Frege who distinguished between two types of meaning: thought content and referent in his essay “On Sinn und Bedeutung” (Campbell et al., 1998). "In Plato's *Cratylus*, Socrates argues that it is not enough to try to understand what a thing is, based on its name, because the name-givers may have been living in ancient times, and the name reflects only what the name-givers thought was the nature of reality then; however, they may have been wrong. Thus, it has been historically recognized that multiple terms may refer to the same object or idea, a single term may refer ambiguously to more than one object or idea, and terms may be confusing because they are out of date" (Campbell, et al., 1998).

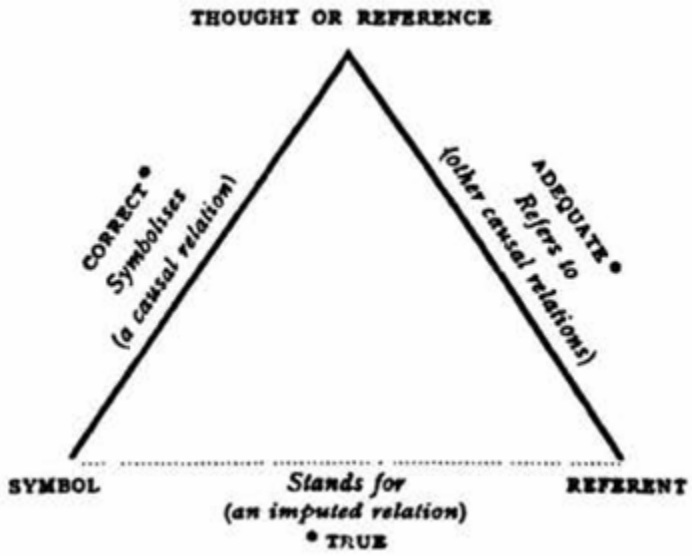
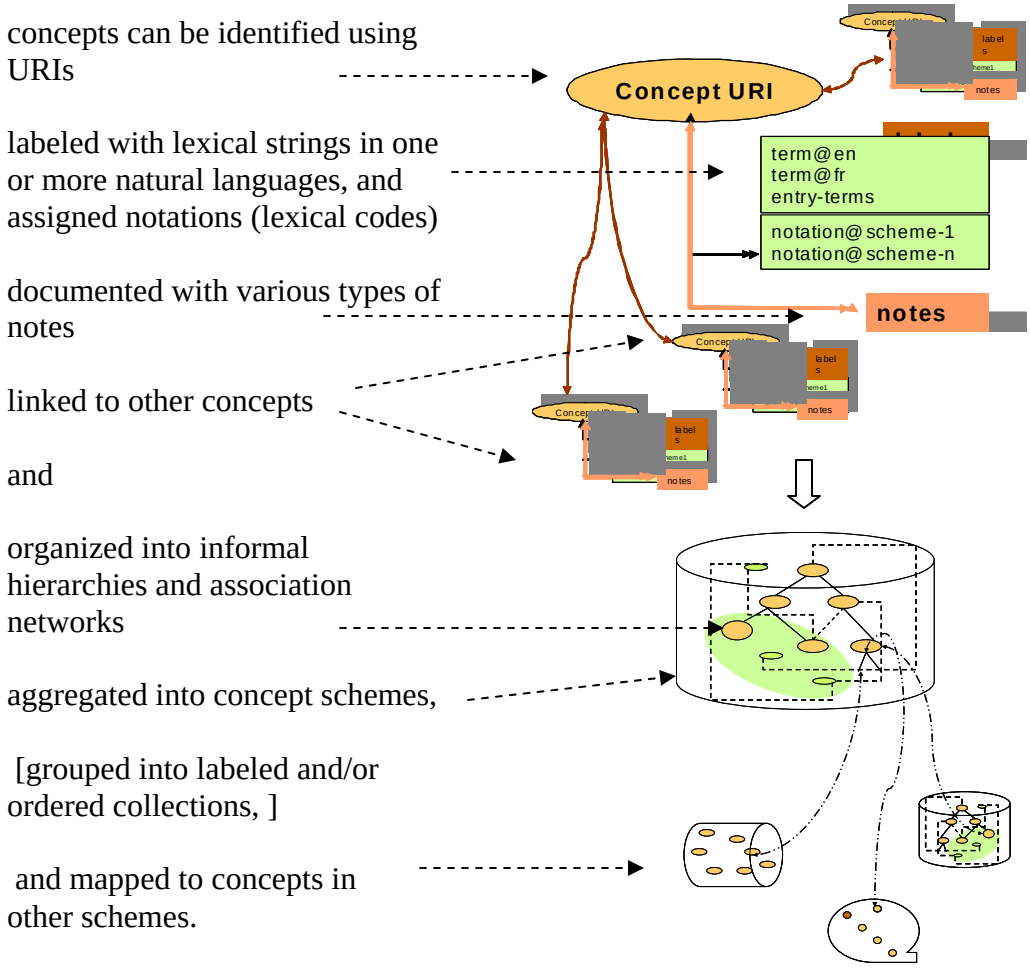


Figure 12. Ogden' semiotic triangle. Ogden and Richards, 1923, p.11

The triangle was also adopted by researchers in library and information science as the basis for building knowledge organization systems (Dahlberg, 1992, Campbell et al, 1998). "Metalanguage consists of signs that signify something about other signs, but what they signify depends on what relationships those signs have to each other, to the entities they represent, and to the agents who use those signs to communicate with other agents" (Sowa, 2000, Section 2).

This thema-nomen conceptual model also matches well with the data models behind the encoding schemas such as SKOS and OWL. SKOS defines classes and properties sufficiently for representing the common features found in a standard thesaurus. It is based on a *concept-centric* view of vocabulary, where primitive objects are not *terms*; rather, they are *abstract concepts* represented by terms. Each SKOS concept is defined as an RDF resource and each concept can have RDF properties attached, which include: 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 (SKOS Core Guide, 2008). In a simplified figure, the synopsis can be illustrated as: Using SKOS --



The consistency of the FRSAD model and SKOS is obvious: for both the basic unit in the subject authority data is thema (or concept) and its appellations (or labels). Relationships between and among the concepts are the basic semantic relationships. Using multilingual thesauri as an example, the work can be done at the nomen level, i.e., only for the translations and/or additions of terms representing the concepts. The work also can be done at thema level, i.e., dealing with the semantic relationships (hierarchical and associate relationships) regardless of the labels of them.

For the issues of the complexity and granularity of themas and comprehensive semantic relationships between and among themas that FRSAD attempted to cover, OWL Web Ontology Language has great matches.

According to DCMI Abstract Model (DCMI Abstract Model, 2007), a metadata record can contain more than one *description set*, which may contain *descriptions* composed by *statements*, which use property-value pairs. This results in information, which 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 (DCMI Abstract Model, 2007). The conceptual model proposed by the FRSAR group corresponds to this 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.

With the development of the Linked Data, abstract models like FRSAR and that behind SKOS will facilitate the implementation of the Semantic Web. Using RDF and applying URIs (Uniform Resource Identifier) as the basic mechanism for identifying subjects, predicates, and objects in metadata statements, subject authority data that are modeled as FRSAD and encoded in SKOS and OWL will be able to become part of the linked data for the next generation Web.

References:

Campbell, Keith E., Diane E. Oliver, Kent A. Spackman, and Edward H. Shortliffe. (1998). Representing Thoughts, Words, and Things in the UMLS. Journal of the American Medical Informatics Association, 5(5): 421–431.

Dahlberg, Ingetraut. (1992) Knowledge Organization and Terminology: Philosophical and Linguistic Bases. International Classification, Vol. 19, 1992: 65-71.

DCMI Abstract Model. Dublin Core Metadata Initiative, 2007; http://dublincore.org/documents/abstract-model/ (accessed January 2009).

IFLA. (1998). Functional Requirements for Bibliographic Records: Final Report. München: KG Saur.

IFLA FRSAR Working Group. (2008). Functional Requirements for Subject Authority Data, Draft Report, Version 0.1 (internal document)

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

Ogden, C. K. and I. A. Richards, I. A. (1923). The Meaning of Meaning: A Study of the Influence of Language Upon Thought and of the Science of Symbolism. London: Routledge & Kegan Paul. TUSC general coll BF355 .O4; with supplementary essays by B. Malinowski and F.G. Crookshank ; edited by John Constable 2001

Sowa, John F. (2000). Ontology, Metadata, and Semiotics. In: Conceptual structures : logical, linguistic, and computational issues : 8th International Conference on Conceptual Structures, ICCS 2000, Darmstadt, Germany, August 2000 : proceedings / Bernhard Ganter, Guy W. Mineau (eds.) pp. 55-81. http://users.bestweb.net/~sowa/peirce/ontometa.htm (accessed January 2009).