



International Federation of Library Associations and Institutions

Functional Requirements for Subject Authority Data (FRSAD)

A Conceptual Model

IFLA Working Group on the
Functional Requirements for Subject Authority Records (FRSAR)

Editors: Marcia Lei Zeng, Maja Žumer, Athena Salaba

Approved by the Standing Committee of the IFLA Section on Classification and Indexing

June 2010

CONTENTS

MEMBERS OF THE IFLA WORKING GROUP ON THE FUNCTIONAL REQUIREMENTS FOR SUBJECT AUTHORITY RECORDS (FRSAR)	4
1. BACKGROUND.....	6
2. PURPOSE AND SCOPE.....	8
2.1 PURPOSE	8
2.2 SCOPE	9
2.3 ABOUTNESS AND OFNESS.....	10
2.4 METHOD	12
2.5 COMPONENTS OF THE STUDY	12
3. ENTITIES.....	14
3.1 DIAGRAMMING CONVENTIONS.....	14
3.2 GENERAL FRAMEWORK.....	14
3.3 CHOICE OF TERMS FOR FRSAD ENTITIES	16
3.4 THEMA.....	16
3.5 NOMEN	18
4. ATTRIBUTES	19
4.1 ATTRIBUTES OF A THEMA.....	19
4.2 ATTRIBUTES OF A NOMEN	21
5. RELATIONSHIPS.....	25
5.1 WORK-TO-THEMA RELATIONSHIP	25
5.2 THEMA-TO-NOMEN RELATIONSHIP	25
5.3 THEMA-TO-THEMA RELATIONSHIPS.....	26
5.4 NOMEN-TO-NOMEN RELATIONSHIPS	31
6. USER TASKS	33
6.1 USERS AND USE	33
6.2 USER TASKS	34
6.3 ASSESSING VALUES RELATIVE TO USER TASKS	35
6.4 MAPPING OF ATTRIBUTES, RELATIONSHIPS, AND USER TASKS.....	37
7. CONCLUSION.....	38
APPENDIX A. MODELING ABOUTNESS.....	39
A.1 SUBJECT RELATIONSHIP AND GROUP 3 ENTITIES INTRODUCED IN FRBR	39
A.2 POSSIBLE APPROACHES TO THE MODEL OF ABOUTNESS	40
APPENDIX B. RELATIONSHIP OF FRSAD WITH FRBR AND FRAD	45
B.1 RELATIONSHIP OF FRSAD WITH FRBR.....	45
B.2 RELATIONSHIP OF FRSAD WITH FRAD	46
APPENDIX C. FRSAD MODEL AND OTHER MODELS.....	48
C.1 THE IMPORTANCE OF THE THEMA-NOMEN MODEL	48
C.2 MAPPING THE FRSAD MODEL TO OTHER MODELS.....	49
C.3 CONCLUSION	50
APPENDIX D. EXAMPLES FROM SUBJECT AUTHORITY SYSTEMS.....	51
D.1 EXISTING MODELS OF THEMA TYPES.....	51
D.2 THEMA-THEMA RELATIONSHIPS PRESENTED IN SUBJECT AUTHORITY DATA.....	55

D.3	SAME THEMA REPRESENTED BY NOMENS FROM DIFFERENT SCHEMES.....	68
D.4	EXAMPLES OF DISPLAY RECORDS FROM CONTROLLED VOCABULARIES OR SUBJECT AUTHORITY FILES	68
REFERENCES	73

MEMBERS OF THE IFLA WORKING GROUP ON THE FUNCTIONAL REQUIREMENTS FOR SUBJECT AUTHORITY RECORDS (FRSAR)

Working Group

Leda Bultrini
ARPA Lazio (Regional Environment Protection Agency), Italy

Lois Mai Chan
University of Kentucky, USA

Jonathan Furner
University of California Los Angeles, USA

Edward O'Neill
OCLC, USA

Gerhard Riesthuis
University of Amsterdam, The Netherlands

Athena Salaba (Co-Chair and Secretary)
Kent State University, USA

Diane Vizine-Goetz
OCLC, USA

Ekaterina Zaytseva
Russian National Public Library for Science and Technology, Russia

Marcia Lei Zeng (Chair)
Kent State University, USA

Maja Žumer (Co-Chair)
University of Ljubljana, Slovenia

Advisory Group

Victoria Francu (2005-06)
Central University, Bucharest, Romania

Hemalata Iyer (2008)
Albany State University of New York, USA

Dorothy McGarry
University of California, Los Angeles, USA

David Miller
Curry College, USA

Päivi Pekkarinen
National Library of Health Sciences, Finland

Barbara Tillett
Library of Congress, USA

This document was prepared by:

Marcia Lei Zeng, Maja Žumer, Athena Salaba, Jonathan Furner, Lois Mai Chan, Ed O'Neill, Diane Vizine-Goetz

1. BACKGROUND

The IFLA Study Group on the Functional Requirements for Bibliographic Records (FRBR) developed a conceptual model showing the entities and relationships of the bibliographic universe in 1997. The purpose of the FRBR model is to identify the functional requirements of information in bibliographic records to facilitate the specified user tasks.¹ The basic entities of the FRBR model are the result of a logical analysis of the data typically represented in bibliographic records. The entities are divided into three groups:

Group 1 entities are defined as the products of intellectual or artistic endeavours that are named or described in bibliographic records: *work, expression, manifestation, and item*.

Group 2 entities are those responsible for the intellectual or artistic content, the physical production and dissemination, or the custodianship of the Group 1 entities: *person, corporate body, and family*².

Group 3 entities represent an additional set of entities that serve as the subjects of *works: concept, object, event, and place*.

The FRBR final report presents the entity-relationship model, identifies entities and their attributes, and defines relationships among entities. Although in the FRBR model the entities of all three groups are defined, the main focus is on the first group. The developers of FRBR envisioned that its extensions would cover the additional data that are normally recorded in authority records.

The Working Group on Functional Requirements and Numbering of Authority Records (FRANAR) was established in April 1999. It was charged to continue the work of FRBR by developing a conceptual model for entities described in authority records. Authority data in the context of their work is defined as “the aggregate of information about a *person, family, corporate body, or work* whose name is used as the basis for a controlled access point for bibliographic citations or records in a library catalogue or bibliographic file”³. The primary purpose of the Functional Requirements for Authority Data (FRAD) conceptual model is “to provide a framework for the analysis of functional requirements for the kind of authority data that is required to support authority control and for the international sharing of authority data. The model focuses on data, regardless of how they may be packaged (e.g., in authority records)”⁴. While the FRANAR Working Group has

¹ *Functional Requirements for Bibliographic Records: Final Report*. (1998). IFLA Study Group on the Functional Requirements for Bibliographic Records. München: KG Saur

² "Family" entity was added in *Functional Requirements for Authority Data - A Conceptual Model*. (2009).

³ *Functional Requirements for Authority Data - A Conceptual Model*. (2009). IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR), ed. by Glenn E. Patton. München: K.G. Saur. p.15.

⁴ *Ibid.*, p. 13.

included some aspects of subject data in their model, they have not undertaken the full analysis of the entities and relationships relevant to subject authorities⁵.

As a result, the IFLA Working Group on the Functional Requirements for Subject Authority Records (FRSAR) was formed in 2005 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 FRSAR Working Group (2005 to present) and the FRANAR Working Group (1999 to 2009) both worked in parallel to develop models within the FRBR framework. By the time FRANAR released its final report in June 2009, FRSAR also had released its first draft report of the *Functional Requirements for Subject Authority Data* (FRSAD) for world-wide review. Since the two reports were developed independently, the relationship between the FRSAD and FRAD models is explained in Appendix B of this report.

⁵ *Functional Requirements for Authority Data - A Conceptual Model*. (2009). IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR), ed. by Glenn E. Patton. München: K.G. Saur. p. 8.

2. PURPOSE AND SCOPE

2.1 Purpose

Subject access to information has been a significant approach of users to satisfy their information needs. Research results have demonstrated that the integration of controlled vocabulary information with an information retrieval system helps users perform more effective subject searches. This integration becomes possible when subject authority data (information about subjects from authority files) are linked to bibliographic files and are made available to users.

The purpose of authority control is to ensure consistency in representing a value—a name of a person, a place name, or a term or code representing a subject—in the elements used as access points in information retrieval. For example, “World War, 1939-1945” has been established as an authorised subject heading in the *Library of Congress Subject Headings* (LCSH). When using LCSH, in cataloguing or indexing, all publications about World War II are assigned the established heading regardless of whether a publication refers to the war as the “European War, 1939-1945”, “Second World War”, “World War 2”, “World War II”, “WWII”, “World War Two”, or “2nd World War”. The synonymous expressions lead to the authorised heading. This ensures that all publications about World War II can be retrieved by and displayed under the same subject heading, either in local catalogue or database or in a union catalogue.

In almost all large bibliographic databases, authority control is achieved manually or semi-automatically by means of an authority file. The file contains data about access points – names, titles, or subject terms – that have been authorised for use in bibliographic records. In addition to ensuring consistency in subject representation, a subject authority system may also record established semantic relationships among subject concepts and/or their labels. Data in a subject authority system are connected through semantic relationships, which may be expressed in subject authority records or generated according to specific needs (e.g., presenting the broader and narrower concepts) in printed or online displays of thesauri, subject headings lists, classification schemes, and other subject authority systems. Such systems have been referred to as "controlled vocabularies", "structured vocabularies", "concept schemes", "encoding schemes", and "knowledge organization systems" interchangeably depending on their function and structure, as well as according to the communities that use them. Given the purpose of this report, the discussions about subject authority data apply to all systems and structures referred to by these terms. The study follows FRBR's approach in that it makes no *priori* assumption about the physical structure or storage of authority data.

2.2 Scope

The primary purpose of this study is to produce a framework that will provide a clearly stated and commonly shared understanding of what the subject authority data/record/file aims to provide information about, and the expectation of what such data should achieve in terms of answering user needs. The role of the FRBR Working Group 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 that data;
- To assist in an assessment of the potential for international sharing and use of subject authority data both within the library sector and beyond.

To fulfil these terms of reference, the FRBR Working Group established two sub-groups: User Tasks Sub-Group and Subject Entities Sub-Group.

The User Tasks Sub-Group focused on user studies and the definition of user tasks. For the purposes of this study, the users of subject authority data include information professionals who create and maintain subject authority data, information professionals who create and maintain metadata, intermediaries and end users who search for information to fulfil information needs. The functional requirements for subject authority data are defined in relation to the following general tasks that are performed by these users:

Find one or more subjects and/or their appellations, that correspond(s) to the user's stated criteria, using attributes and relationships;

Identify a subject and/or its appellation based on its attributes or relationships (i.e., to distinguish between two or more subjects or appellations with similar characteristics and to confirm that the appropriate subject or appellation has been found);

Select a subject and/or its appellation appropriate to the user's needs (i.e., to choose or reject based on the user's requirements and needs);

Explore relationships between subjects and/or their appellations (e.g., to explore relationships in order to understand the structure of a subject domain and its terminology).

The Subject Entities Sub-Group focused on the Group 3 entities including the study of current FRBR Group 3 entities and alternatives in order to define:

- a) entities that can serve as subjects of a *work* (the “has as subject” relationship);
- b) possible sub-entities in the Group 3 cluster; and
- c) additional entities related to the Group 3 cluster.

The FR SAR Working Group is aware that some controlled vocabularies provide terminology to express other aspects of *works* in addition to subject (such as form, genre, and target audience of resources). While very important and the focus of many user queries, these aspects describe **isness** or what class the *work* belongs to based on form or genre (e.g., novel, play, poem, essay, biography, symphony, concerto, sonata, map, drawing, painting, photograph, etc.) rather than what the *work* is about. Some of these aspects are explicitly covered by the FRBR model, for example, “form of work,” “intended audience,” etc. as attributes of *work*. While the Group acknowledges that there are cases where a vocabulary provides terminology, or has been used, also for **isness**, the focus of the FR SAD model is on **aboutness** (the FRBR-defined relationship *work* “has as subject ...”). On the other hand, any case of a *work* about a form or genre (e.g. about romance novels, about dictionaries) clearly falls within the aboutness category.

2.3 Aboutness and Ofness

When modelling the fundamental classes of bibliographic entities we necessarily face the challenge of carrying out the most appropriate analysis of aboutness—i.e., the relation between a work and its subject matter. Aboutness is a concept that is central to the field of knowledge organization, and many authors have made significant contributions to our understanding of the nature of work–subject relations. Some of these contributions appear in the literature of library and information science (LIS),^{6, 7, 8} while others have been made by philosophers of logic and language^{9, 10, 11}. A review of these literatures will show that there is not as much consensus on the nature of aboutness as one might hope to find: in fact, there are a wide variety of views.

At the risk of oversimplifying what is undoubtedly a complex situation, we might consider that it is possible to place views on aboutness on a spectrum whose poles represent the two extremes of “nominalism” (or, following Hjørland,¹² “idealism”) and “realism.” For the thoroughgoing nominalist, it does not make sense to talk of works “having” or “being about” subjects—aboutness should be conceived not as a property of

⁶ Wilson, P. (1968). *Two kinds of power: An essay on bibliographic control*. Berkeley, CA: University of California Press.

⁷ Hutchins, W. J. (1977). On the problem of “aboutness” in document analysis. *Journal of Informatics*, 1 (1): 17-35.

⁸ Hjørland, B. (1992). The concept of “subject” in information science. *Journal of Documentation*, 48 (2): 172-200.

⁹ Ryle, G. (1933). “About.” *Analysis*, 1 (1): 10-12.

¹⁰ Putnam, H. (1958). Formalization of the concept “about.” *Philosophy of Science*, 25 (2): 125-130.

¹¹ Goodman, N. (1961). About. *Mind*, 70 (277): 1-24.

¹² Hjørland, *op. cit.*

works but rather as a relation, constructed by a particular person at a particular time, between a particular set of works and a particular linguistic expression (i.e., a name or label). The realist, on the other hand, is content to proceed on the assumption that subjects are real things that exist separately from the linguistic expressions that we use to name them, and that it is possible to determine “the” subject(s) of any given work. Of course, there are other points of view on aboutness that may be located either at intermediate points between these two poles, or on different spectrums. But it is probably fair to say that most people who are actively engaged in the tasks of designing bibliographic classification schemes, indexing documents in accordance with such schemes, and using those schemes as tools for finding the kinds of documents we want, act in accordance with assumptions that are consistent with some version of the realist viewpoint.

Ultimately, the FRSAR Working Group does not take a philosophical position on the nature of aboutness, rather, it looks at the problem from the user’s point of view. When confronted with an information need that can potentially be met by finding and using a document about a certain subject, the user both expects to be able to formulate a search statement specifying the subject, and expects that the tools and services at hand are capable of comparing such search statements with the subject statements generated by cataloguers and indexers.

Those LIS authors who have focused on the subjects of visual resources, such as artworks and photographs, have often been concerned with how to distinguish between the “aboutness” and the “ofness” (both specific and generic depiction or representation) of such works.¹³ In this sense, “aboutness” has a narrower meaning than that used above. A painting of a sunset over San Francisco, for instance, might be analyzed as being (generically) “of” sunsets and (specifically) “of” San Francisco, but also “about” the passage of time. Standard metadata schemas for cultural objects correspondingly allow for distinctions to be made among (a) **description** of the **kinds** of things depicted in *works*, (b) **identification** of the **particular** people, objects, events, and places depicted, and (c) **interpretation** of the **meanings** of *works*¹⁴. The FRSAR Working Group recognizes that any statement on the “ofness” of a work can be a **subject statement** and is likely to be the target of a catalogue user’s search. The FRSAD model is therefore applicable to situations where ofness statements (depiction) are included in subject access.

¹³ Shatford, S. (1986). Analyzing the subject of a picture: A theoretical approach. *Cataloging & Classification Quarterly*, 6 (3): 39-62.

¹⁴ *Categories for the Description of Works of Art* (CDWA). (2000). Eds. Baca, M. and Harpring, P. The J. Paul Getty Trust and College Art Association. Los Angeles, CA: Getty Research Institute. Available at: http://www.getty.edu/research/conducting_research/standards/cdwa/index.html (accessed 2010-01-20).

2.4 Method

The method used to build this conceptual model is the entity analysis technique used in FRBR. The development of such a conceptual model consists of the following steps¹⁵:

- 1) The analysis starts with the **user tasks** as well as the key objects that are of interest to users in a specific domain.
- 2) The attention is not on individual data but on the “things” the data describe. Each of the objects of interest or **entities** defined for the model serves as the focal point for a cluster of data.
- 3) At a higher level, an entity diagram depicts the **relationships** that normally hold between one type of entity and another type of entity.
- 4) Important characteristics or **attributes** of each entity are then identified.
- 5) Each attribute and relationship is mapped to the user tasks. Relative values of importance are assigned to each attribute and relationship with specific reference to the task performed and the entity that is the object of the user’s interest.

These steps were followed in the development of this model. The background of the initial conceptual analysis of user tasks and entities performed by the Group is explained in detail in Appendix A.

2.5 Components of the Study

The remainder of the report is divided into two major segments: the main body of the report follows the FRBR report structure and presents the entity-relationship model; the second segment contains four appendices that explain the methodology and implementation considerations.

The remaining body of the report consists of four chapters:

- Chapter 3 of the study identifies and defines the entities used in the model.
- Chapter 4 analyses the attributes associated with each of the entities defined in the model.
- Chapter 5 analyses the relationships used in the model, including the relationships operating both at the general level and between specific instances of entities.
- Chapter 6 presents the user tasks and then maps the attributes associated with each entity to the four generic user tasks the subject authority data is intended to support, showing the relevance of each attribute or relationship to each of the user tasks.

The appendices contain additional materials:

¹⁵ *Functional Requirements for Bibliographic Records: Final Report*. (1998). IFLA Study Group on the Functional Requirements for Bibliographic Records. München: KG Saur. p.9. Section 2.3 Methodology.

- Appendix A discusses the initial analysis used as basis in building the conceptual model for aboutness. It analyses the subject relationship and Group 3 entities introduced in FRBR as well as possible approaches to the model of aboutness.
- Appendix B explains the relationship of the FRSAD model with the FRBR and the FRAD models, given the fact that both FRAD and FRSAD are based on FRBR but were developed in parallel.
- Appendix C furthers the discussion of the importance of the FRSAD model and maps it with related developments including the ISO standard for thesauri, the W3C's Simple Knowledge Organization System (SKOS) and OWL Web Ontology Language, and the Dublin Core Abstract Model.
- Appendix D contains examples from existing subject authority systems through the perspective of the FRSAD model.

3. ENTITIES

3.1 Diagramming Conventions

FRSAD follows the conventions used in both FRBR and FRAD:



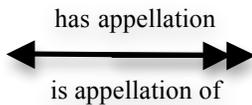
A rectangle represents an entity.



A single-headed arrow on a line represents a relationship in which any given instance of the entity at the beginning of the line may be associated with only one instance of the entity to which the arrow is pointing.



A double-headed arrow on a line represents a relationship in which any given instance of the entity at the beginning of the line may be associated with one or more instances of the entity to which the arrow is pointing.



A relationship above the line indicates left-to-right direction; a relationship below the line indicates a right-to-left direction.

3.2 General Framework

The FRSAR Working Group proposes a generalisation of FRBR, as seen in Figure 3.1. This diagram is based on the original FRBR Figure 3.3 that depicts the “subject” relationships between works and entities in Group 1, Group 2, and Group 3. A new entity, *family*, added by FRAD in Group 2, is also reflected in this general framework. The entities in the third group represent an additional set of entities that serve as the subjects of *works*. The group includes, according to the FRBR model, *concept* (an abstract notion or idea), *object* (a material thing), *event* (an action or occurrence), and *place* (a location).

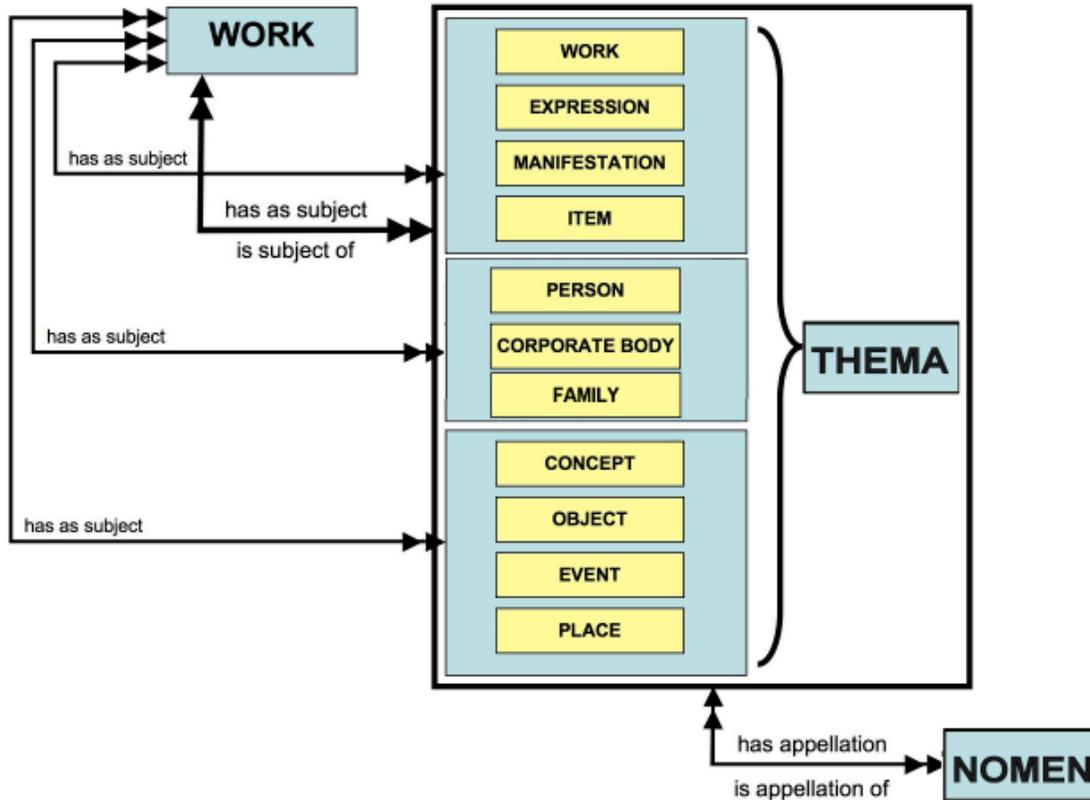


Figure 3.1 FRSAD's relation to FRBR (with the addition of FRAD entity *family*)

The FRSAR Working Group introduced the following two entities:

Thema: any entity used as a subject of a *work*

Nomen: any sign or sequence of signs (alphanumeric characters, symbols, sound, etc.) that a *thema* is known by, referred to, or addressed as.

The Functional Requirements for Subject Authority Data (FRSAD) model is presented as:



Figure 3.2 FRSAD conceptual model

Both “has as subject/is subject of” and “has appellation/is appellation 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*. 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”, etc. There are many other *works* about any of these

themas. For any of the *themas* in this list (presented here as terms in English) there are other possible *nomens* in other languages and in different controlled vocabularies.

Some *works* are perceived as having no *thema* as subject (such as certain musical works or abstract artwork), and no subject access is provided to them. These cases are not covered by FRSAD. The cases of a *thema* without a *nomen* are also beyond the scope of this model.

3.3 Choice of Terms for FRSAD Entities

The Working Group chose Latin terms, *thema* (plural *themata* or *themas*) and *nomen* (plural *nomina* or *nomens*), because they have no pre-existing meaning in our context, are culturally neutral and do not require translation. For *thema*, other possible (English) terms include “subject”, “topic”, and “concept”; however, even discussions within the Working Group proved that there are very different views on granularity (some see “subject” and “topic” as synonyms, while others see “topic” as a component of “subject”). The Working Group needed to distinguish *thema* from the previously defined FRBR entity *concept* because *thema* is a superclass of all FRBR entities (to be explained in the next section). For *nomen*, it is the case that the term “name” is often considered synonymous with proper name. In addition, the Working Group needed to distinguish *nomen* from the FRAD entity *name* because *nomen* includes FRAD entities *name*, *identifier*, and *controlled access point*.

3.4 THEMA

Thema is defined as “any entity used as a subject of a *work*”. Therefore this model confirms one of the basic relationships defined in FRBR: **WORK has as subject THEMA / THEMA is subject of WORK**.

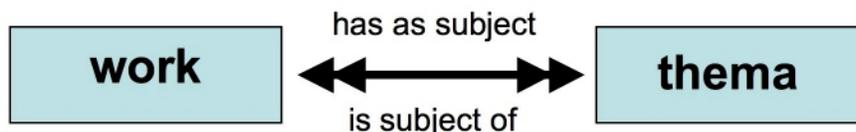


Figure 3.3 *Work-Thema* relationship

According to Delsey, the first broad objective of FRSAD is to ensure that the scope of the entities defined is sufficient to cover everything that a user of a library catalogue might view as a “subject”¹⁶. We may therefore see *thema* from different points of view. From the point of view of end-users and intermediaries, *thema* comprises the aboutness of the

¹⁶ Delsey, T. (2005). Modeling subject access: Extending the FRBR and FRANAR conceptual models. *Cataloging & Classification Quarterly* 39 (3/4): 49–61.

(possibly unknown) resources that will satisfy the information need. From the point of view of information professionals who create metadata, one or more *themas* capture the aboutness of a particular resource.

Within the FRBR framework, *thema*, which can be viewed as an entity on its own and also as a super-entity or superclass, includes existing Group 1 and Group 2 entities, and additionally, all others that serve as the subjects of *works* (i.e., Group 3). In other words, *thema* is a superclass of all FRBR entities (Figure 3.4). Defining *thema* as a super entity enables modelling of relationships and attributes on a more general and abstract level.

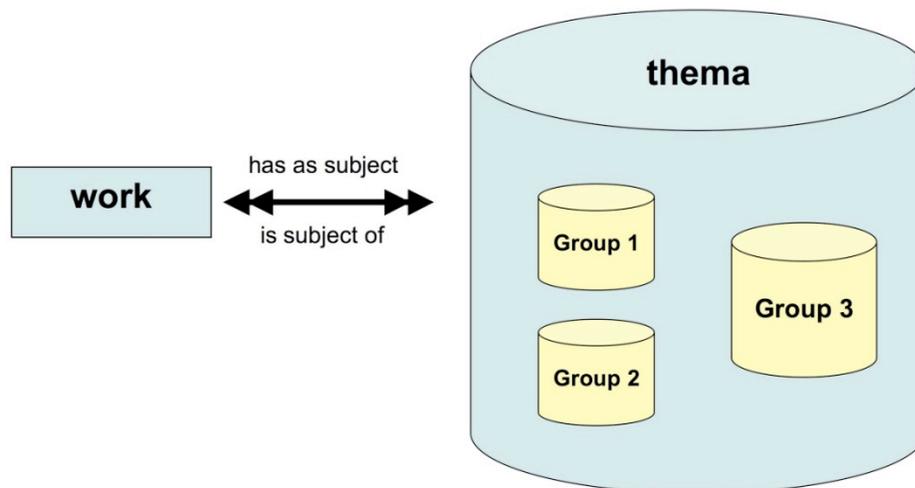


Figure 3.4 Within the FRBR framework, *thema* includes existing Group 1 and Group 2 entities and, in addition, all others that serve as the subjects of *works* (i.e., Group 3)

While the original FRBR Group 3 entities (*object, concept, event, place*) may be used in a given implementation, the Working Group does not propose them as a universally applicable set of Group 3 entities. In a particular application *themas* would normally have implementation-specific types; but, based on the pilot study, as explained in Appendix A, there seems to be no generally applicable categorization of *thema*. Appendix D provides examples of existing implementations through the perspective of the FRSAD model.

Themas can vary substantially in complexity or simplicity. Depending on the circumstances (the subject authority system, user needs, the nature of the *work*, etc.) the aboutness of a *work* can be expressed as a one-to-one relationship between the *work* and the *thema*; this means that the totality of the aboutness is encompassed in a single *thema*. In other circumstances the relationship is one-to-many, meaning that the aboutness of the *work* is captured in two or more *themas*. It is virtually impossible to define what the universal “atomic” level of a *thema* might be, because any *thema* can be fragmented further. The argument can be reversed: simple *themas* may be combined or aggregated, resulting in more complex *thema(s)*. In each particular implementation the atomic level is specified and rules guide the creation of *nomens* for complex *themas*.

To some extent the granularity of a *thema* also depends upon the controlled vocabulary used for its appellation(s). 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 and syntactic rules for creating or establishing a *nomen* is not directly reflected in the complexity of the *thema*, nor is it completely independent. Some types of controlled vocabularies (such as subject headings systems) enable the establishment of complex *themas* (e.g., by creating pre-coordinated strings), while others (such as thesauri) are mainly conceived for the use of more atomic *themas*.

3.5 NOMEN

The FRSAD model proposes a new relationship: **THEMA has appellation NOMEN/NOMEN is appellation of THEMA** (Figure 3.5).

Nomen is defined as “any sign or sequence of signs (alphanumeric characters, symbols, sound, etc.) that a *thema* is known by, referred to, or addressed as. Examples include “love,” “∞,” or “595.733.” A *Nomen* can be human-readable or machine-readable. *Nomen* is a superclass of the FRAD entities *name*, *identifier*, and *controlled access point*.

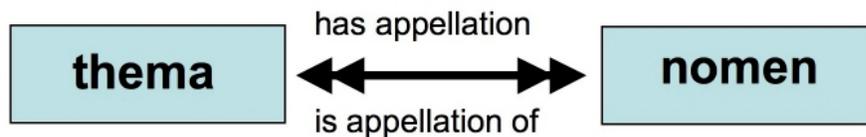


Figure 3.5 *Thema-Nomen* relationship

In general (i.e., in natural language or when mapping different vocabularies) the “has-appellation/is appellation of” relationship is a many-to-many relationship. A *thema* has one or more *nomens* and there may be a *nomen* referring to more than one *thema*. In a given controlled vocabulary, however, a *nomen* should be an appellation of only one *thema*, as illustrated in Figure 3.6. See Appendix D for examples from subject authority systems.

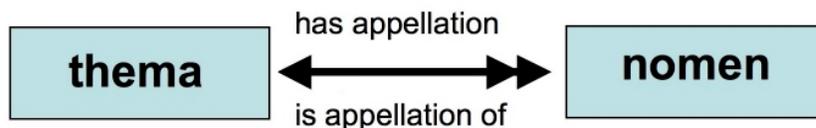


Figure 3.6 *Thema-Nomen* relationship within a controlled vocabulary

4. ATTRIBUTES

4.1 Attributes of a *THEMA*

In the FRSAD model, the entity *thema* is defined in a very abstract and general way. Attributes of a *thema* are implementation-dependent and will vary. “Type” and “scope note” can be considered general attributes, but particular values of “type” are, again, implementation-dependent. In any implementation there will normally be additional attributes of a *thema* other than “type” and “scope note”. Those attributes will be dependent on both the type of *themas* and the application domain.

4.1.1 Type of *thema*

The category to which a *thema* belongs in the context of a particular knowledge organisation system.

In an implementation *themas* can be organised based on category, kind, or type. This report does not suggest specific types, because they may differ depending on the implementation (see Appendix A).

For example, in some implementations the original FRBR entities *work*, *expression*, *manifestation*, *item*, *person*, *family*, *corporate body*, *concept*, *object*, *event*, and *place* can be used as types, perhaps even adding *time* as has been suggested (see Appendix A.2). In general, any entity defined in FRBR and/or FRAD may become a type of *thema*. And their attributes (also defined in FRBR and FRAD) will equally apply.

In other implementations, a different set of types may be defined. Two examples from existing implementations, UMLS and AAT, are presented below. Detailed explanations of these subject authority systems can be found in Appendix D.

1) *Unified Medical Language System (UMLS) semantic types*¹⁷

Entities

Physical Object

Organism

Anatomical Structure

¹⁷ National Library of Medicine. (2003-). *Unified Medical Language System. Current Semantic Types*. UMLS 2004AB Documentation. Last updated: 21 March 2008. Available at: http://www.nlm.nih.gov/research/umls/META3_current_semantic_types.html (accessed 2010-01-20).

Manufactured Object
Substance
Conceptual Entity
Idea or Concept
Finding
Organism Attribute
Intellectual Product
Language
Occupation or Discipline
Organization
Group Attribute
Group

Events

Activity
Phenomenon or Process

2) *Art and Architecture Thesaurus* (AAT) facets¹⁸

[Top of the AAT hierarchies]
Associated Concepts
Physical Attributes
Styles and Periods
Agents
Activities
Materials
Objects

These examples show very different approaches to defining types of *themas*. In the UMLS, *themas* are first differentiated as “Entity” or “Event.” The types of UMLS “Entity” are “Physical Object” or “Conceptual Entity.” The types of “Events” are grouped into “Activity” and “Phenomenon or Process.” In AAT, all *themas* are categorized into seven types: “Associated Concepts,” “Physical Attributes,” “Styles and Periods,” “Agents,” “Activities,” “Materials,” and “Objects.”

Clearly since *themas* are very different, they will also necessarily have different attributes. In the first example “Substance” (a physical object) will have very different attributes from “Organization” (a conceptual entity) in the UMLS. The same is true for the attributes of “Styles and Periods,” “Agents,” and “Materials” in the AAT.

¹⁸ Art and Architecture Thesaurus Online Hierarchy Display. In: *Art and Architecture Thesaurus*. (2000-). Los Angeles: J. Paul Getty Trust, Vocabulary Program. Available at: <http://www.getty.edu/vow/AATHierarchy?find=&logic=AND¬e=&english=N&subjectid=300000000> (accessed 2010-01-20).

Another possible distinction can be made at the *thema* level between **Classes** and **Instances**. These two types of *thema* are fundamental and many subject authority systems recognize them. The **Class/Instance** distinction is essentially equivalent to the universal/particular distinction, typically made on the basis of instantiability (and hence is sometimes characterized in philosophical literature as a kind/instance distinction).

For example:

<u>Class</u>	<u>Instance</u>
Palaces	Buckingham Palace
Ships	The Lusitania
Battles	The Battle of Hastings

4.1.2 Scope Note

A text describing and/or defining the *thema* or specifying its scope within the particular subject authority system.

4.2 Attributes of a *NOMEN*

The attributes of a *nomen* listed below represent the most common (general) attributes. Not all are applicable in every case, and the list is not comprehensive. While all listed attributes are applicable to individual instances of *nomens*, some may also be used for an entire subject authority system and declared on that level. The examples of attribute values are illustrative only and should not be seen as prescriptive. In any particular implementation the actual values of an attribute are selected from a controlled list and/or are coded.

In addition to the proposed general attributes, presented here, there may be additional, implementation-specific attributes.

Current subject authority records typically include other elements such as administrative data. Also, current authority system may allow merging data -- which describe both *thema(s)* and *nomen(s)* -- into one record. As this is a conceptual model, such aspects of implementation are not discussed.

4.2.1 Type of nomen

Category to which the *nomen* belongs.

In addition to other implementation-specific types, there are two important values of this attribute:

- identifier – the sign or sequence of signs assigned to an entity that is persistent and unique within a domain
- controlled name – the name constructed during the authority control or vocabulary maintenance process that usually serves as an access point (note: labelled as *controlled access point* in FRAD).

If needed, the values of the type attribute may be further refined; for example, additional refinement may include different kinds/formats of identifiers (e.g., URI, ISBN).

4.2.2 Scheme

The scheme in which the *nomen* is established, including value encoding schemes (subject heading lists, thesauri, classification systems, name authority lists, etc.) and syntax encoding schemes (standards for encoding dates, etc.).

Examples of attribute values:

- LCSH
- DDC
- UDC
- ULAN
- ISO 8601

4.2.3 Reference Source of nomen

The source in which the *nomen* is found. It may also be modelled as a relationship with the appropriate Group 1 entity.

Examples of attribute values:

- Encyclopaedia Britannica
- Webster's Third New International Dictionary (1961)
- Columbia Gazetteer

4.2.4 Representation of nomen

The data type in which the *nomen* is expressed.

Examples of values:

- alphanumeric
- sound
- graphic

4.2.5 Language of nomen

The language in which the *nomen* is expressed.

Examples of values:

- English
- Greek
- Chinese
- Slovenian

4.2.6 Script of nomen

The script in which the *nomen* is expressed.

Examples of values:

- Cyrillic
- Thai
- Chinese (Simplified)
- Chinese (Traditional)

4.2.7 Script conversion

The rule, system, or standard used to render the *nomen* in a different representation.

Examples of values:

- Pinyin
- ISO 3602, 1998, Romanisation of Japanese (kana script)

4.2.8 Form of nomen

Any additional information that helps to interpret the *nomen*.

Examples of attribute values:

- Full name
- Abbreviation
- Formula

4.2.9 Time of validity of nomen

The time period, in which the *nomen* is/was used or is/was valid within a subject vocabulary system.

This should not be confused with the temporal aspect of a *thema*.

Examples of values:

- until May 11, 1949
- after 1945
- 1945 - 1967

4.2.10 Audience

The community or user group for which the *nomen* is the preferred form.

In the global environment it is usually impossible to declare one *nomen* of a *thema* to be the preferred form. The notion of “preferred” form can, in general, be tied only to a particular community, defined by name, rule, or convention.

Examples of values:

- English-speaking users
- Scientists
- Children

4.2.11 Status of nomen

The status of the *nomen* in a subject authority system.

This should not be confused with the management of a subject authority system (e.g., including or excluding a *thema*).

Examples:

- Proposed
- Accepted
- Obsolete

5. RELATIONSHIPS

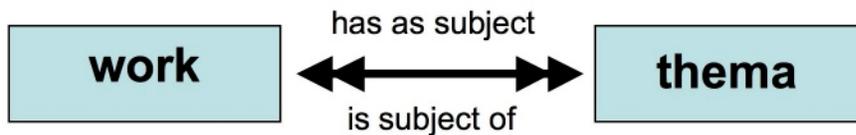
The FRSAD model establishes two sets of relationships:

- 1) Relationships between different types of entities: *WORK*-to-*THEMA* and *THEMA*-to-*NOMEN*. These are the primary relationships and are also illustrated in Chapter 3 where the entities are presented.
- 2) Relationships between entities of the same type: *THEMA*-to-*THEMA* and *NOMEN*-to-*NOMEN*. These are presented in detail in this chapter.

5.1 *WORK*-to-*THEMA* Relationship

The *Work*-to-*Thema* relationship is discussed in Section 3.2. In the FRSAD model, *thema* includes existing Group 1 and Group 2 entities and all entities that serve as the subjects of *works*. Their relationships can be illustrated as:

Work has as subject Thema / Thema is subject of Work



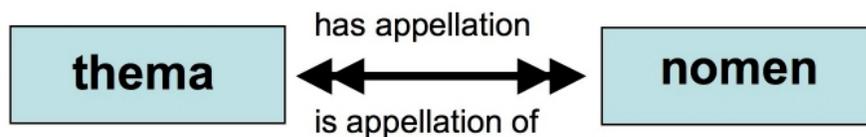
(From Figure 3.3 *Work-Thema* relationship)

Thema refers to anything that can be the subject of a *work*. Presented within the entity-relationship model, the *WORK*-to-*THEMA* relationship is many-to-many: any *work* can have one or more *themas*, and any *thema* may be the subject of one or more *works*.

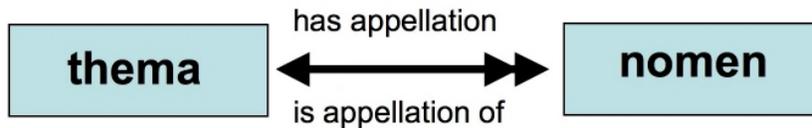
5.2 *THEMA*-to-*NOMEN* Relationship

The *THEMA*-to-*NOMEN* relationship is specified with the following statements:

Thema has appellation Nomen / Nomen is appellation of Thema



(From Figure 3.5 *Thema-Nomen* relationship)



(From Figure 3.6 *Thema-Nomen* relationship within a controlled vocabulary)

As stated in Section 3.4, in general (i.e., in natural language or when mapping different vocabularies) the relationship “has appellation/is appellation of” is a many-to-many relationship. Any *thema* may have more than one *nomen* (see Figure 3.5); and any *nomen* may be the appellation of more than one *thema*. In subject authority-related processes such as integrating or mapping controlled vocabularies, this condition may result in ambiguity regarding the meaning, scope, and definition of *themas* represented by the same *nomen*. Therefore, in a given controlled vocabulary, a *nomen* normally is an appellation of only one *thema* but a *thema* can have more than one *nomen* (see Figure 3.6). Based on this general principle, a more complex *nomen*, for example, with added qualifiers, must be constructed when necessary in order to eliminate ambiguity.

5.3 *THEMA-to-THEMA* Relationships

Only relationships directly applicable for subject access are analyzed here. The FRBR and FRAD models cover additional entity-to-entity relationships such as relationships between *works*.

In order to ensure that (1) the attributes relevant to the construction and use of subject authority data are adequately covered, and (2) the model provides a clear and pragmatic representation of the relationships that are “reflected through subject access points in bibliographic records as well as those reflected in the syndetic structure of thesauri, subject headings lists, and classification schemes and in the syntactic structure of indexing strings”¹⁹, the *thema-to-thema* relationship types are discussed in the context of subject authority systems.

5.3.1 Hierarchical Relationships

Hierarchical structures show relationships between and among concepts and classes of concepts. Hierarchical relationships reveal degrees or levels of superordination and subordination, where the superordinate term represents a class or a whole, and subordinate terms refer to its members or parts. *Hierarchical structures* are found in classification schemes, subject heading systems, thesauri, and other knowledge organization systems. Used in the bibliographic universe, hierarchical relationships provide disambiguation functions to assist with the *identify* user task. Yet they are the

¹⁹ Delsey, T. (2005). Modeling subject access: Extending the FRBR and FRANAR conceptual models. *Cataloging & Classification Quarterly* 39 (3/4): 49–61.

most effective in furthering linking and navigation objectives, and satisfying the *select*, and especially, the *explore* user tasks. They are of particular aid to users with undefined or very broad information needs and they also allow users to improve their searching.

Typically, a hierarchical relationship may be one of three types: *the generic relationship*, *the hierarchical whole-part relationship*, and *the instance relationship*²⁰. Some concepts can belong to more than one superordinate concept simultaneously. They are considered to have *polyhierarchical relationships*. Other perspective hierarchical relationships also exist (as explained in Section 5.3.1.5).

5.3.1.1 The Generic Relationship

The *generic relationship* is the logical relationship of inclusion. The primary function of the hierarchical relationship is to convey the same concept, but at different levels of specificity²¹. “Of limited domain and range, it is strictly defined in terms of the properties of reflexivity, anti-symmetry, and transitivity”²². It is sometimes represented as the “all-some” relationship. For example, all parrots are birds, and some birds are parrots. But not all parrots are pets therefore the genus-species relationship between parrots and pets does not exist in logic²³. In the computer science literature and formal ontology construction, the characteristic of “inheritance” of genus-species relationships is also widely presumed. This “hierarchical force” assumes that what is true of a given class (e.g., furniture) is true of all member-classes it subsumes (chairs, tables, and so on.)

5.3.1.2 The Whole-Part Relationship

The *whole-part relationship* covers situations where one concept is inherently included in another, regardless of context, so that concepts can be organized into hierarchies (with the “whole” treated as a broader term). For example, blood vessels are part of the cardiovascular system in anatomy.

In addition to physical component part relationships, “whole and part” can be applied to several common types of situations such as geographical regions, hierarchical organizational structures, disciplines or fields of discourse. Because such relationships, being synthetic rather than analytic, are not necessarily or logically true in subject authority systems they may be differentiated as special hierarchical relationships (rather than genus-species and perspective hierarchies) or as associative relationships.

²⁰ ISO (2009). *ISO/CD 25964-1, Information and documentation — Thesauri and interoperability with other vocabularies — Part 1: Thesauri for information retrieval*. ISO/TC 46 / SC 9 ISO 25964 Working Group.

²¹ Clarke, S.G. (2001). Thesaural relationships. In: *Relationships in Knowledge Organization*. Eds. Bean, C.A. and Green, R. Dordrecht: Kluwer. p. 42.

²² Svenonius, E. (2000). *The Intellectual Foundation of Information Organization*. Cambridge, Mass.: MIT Press, p.151.

²³ Svenonius, *op. cit.*

5.3.1.3 The Instance Relationship

The *instance relationship* identifies the link between a general class of things or events, expressed by a common noun, and an individual instance of that category, expressed by a proper noun. For example, “Mydoom” and “ILOVEYOU” are two instances of “computer worms” that are expressed by proper nouns.

5.3.1.4 Polyhierarchical Relationship

Some concepts can belong to more than one superordinate concept and are therefore considered to possess *polyhierarchical relationships*. These relationships can be (a) generic, e.g., music instrument “organ” belongs under both the “wind instrument” hierarchy and the “keyboard instrument” hierarchy; (b) whole-part, e.g., “biochemistry”, is part of “biology” and is also part of “chemistry”; or (c) more than one type, e.g., “skull”, belongs under the “bones” (kind-of), and also under the “head” (part-of), hierarchies²⁴.

5.3.1.5 Other Hierarchical Relationship

Other hierarchical relationships, which do not have the logical properties of the above hierarchies, are seen often in subject authority systems. This may be partially due to the requirements of literary warrant (the natural language used to describe content objects), user warrant (the language of users), and sometimes, organizational warrant (the needs and priorities of the organization)²⁵. Their value is that they provide points of view about a concept and the aspect under which it is considered. For instance, although an insect can belong to only one genus-species hierarchy (e.g., Arthropoda), it can belong to as many perspective hierarchies as there are aspects of insects to be studied. In a classification scheme, an insect can be looked at, or studied, from the point of view of agricultural pests, disease carriers, food, and control technology²⁶. Other reasons to employ such hierarchies are that concepts and terms like “happiness” are poly-semantic, vague, or ambiguous. Hence there might be no agreement as to what genus (class) such concepts belong to.

5.3.2 Associative Relationships

Associative relationships cover affiliations between pairs of *themas* that are not related hierarchically yet are semantically or conceptually connected and co-occurring. Associative relationships between *themas* are made explicit in some of the subject authority systems.

²⁴ Examples are from ISO (2009). *ISO/CD 25964-1, Information and documentation — Thesauri and interoperability with other vocabularies — Part 1: Thesauri for information retrieval*. ISO/TC 46 / SC 9 ISO 25964 Working Group.

²⁵ NISO. (2005). ANSI/NISO Z39.19-2005. *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*. Bethesda, Maryland: NISO Press.

²⁶ Svenonius, *op. cit.*

In general, associative relation links are established among *themas* belonging to different hierarchies, or among overlapping *themas* within the same array on a particular level of the hierarchy. Most commonly considered associative relationships fall into these categories^{27, 28, 29}:

Associative Relationships	Examples
Cause / Effect	accident / injury
Process / Agent	velocity measurement / speedometer
Action / Product of the action	weaving/cloth
Action / Patient or Target	teaching / student
Concept or Thing / Properties	steel alloy / corrosion resistance
Thing or Action / Counter-agent	pest / pesticide
Thing/ Its parts (if it does not qualify for the hierarchical whole-part relationship)	car / engine
Raw material / Product	grapes / wine
Action / Property	communication/communication skills
Field of study/ Objects or phenomena studied	forestry/forests

In each particular implementations, a decision would be made about whether to include associative relationships and if so, which ones to include and at what level of specificity.

5.3.3 Other Approaches to Semantic Relationships

In the literature and in practice, other approaches to differentiate semantic relation types have been used. A taxonomy of subject relationships, compiled in 1996 and shared at an American Libraries Association (ALA) conference, listed over a hundred associative relationships and 26 hierarchical relationships³⁰. Over 40 in the associative group and 20 in the hierarchical group have been verified by other sources³¹.

²⁷ Lancaster, F.W. (1986). *Vocabulary Control for Information Retrieval*. 2nd ed. Arlington, Virginia: Information Resources Press.

²⁸ NISO. *op.cit.*

²⁹ Aitchison, J., A. Gilchrist, and D. Bawden (2000). *Thesaurus Construction and Use: A Practical Manual*. 4th ed. London: Fitzroy Dearborn.

³⁰ Michel, D. (1996). Taxonomy of Subject Relationships. Appendix B (Part 2), Final Report to the ALCTS/CCS Subject Analysis Committee, submitted by Subcommittee on Subject Relationships/Reference Structures. ALA Association for Library Collections and Technical Services (ALCTS) Cataloging and Classification Section (CCS) Subject Analysis Committee (SAC). Available at: <http://www.ala.org/ala/mgrps/divs/alcts/mgrps/ccs/cmtes/subjectanalysis/subjectrelations/msrscu2.pdf> Also available is the hierarchical display of this Taxonomy, at: <http://www.ala.org/ala/mgrps/divs/alcts/mgrps/ccs/cmtes/sac/inact/subjectrelations/appendixbpartii.cfm> (accessed 2010-03-31).

³¹ Subcommittee on Subject Relationships/Reference Structures. (1997). Final Report to the ALCTS/CCS Subject Analysis Committee. ALA Association for Library Collections and Technical Services (ALCTS) Cataloging and Classification Section (CCS) Subject Analysis Committee (SAC). Available at: <http://www.ala.org/ala/mgrps/divs/alcts/mgrps/ccs/cmtes/sac/inact/subjectrelations/finalreport.cfm>

The *Unified Medical Language System (UMLS)*³² classified semantic relationship types into two main groups and a number of sub-groups:

- isa
- associated_with
 - physically_related_to
 - spatially_related_to
 - functionally_related_to
 - temporally_related_to
 - conceptually_related_to

Spatial relationship types in UMLS include location_of, adjacent_to, surrounds, and traverses.

Whereas in another case, such relationship types for geographical regions only are identified as³³:

- Inherently spatial
 - Containment
 - Overlap
 - Proximity
 - Directional
- Explicitly stated
 - PartOf
 - AdministrativePartOf
 - AdministrativePartitionMemberOf
 - AdministrativeSeatOf
 - ConventionallyQualifiedBy
 - SubfeatureOf
 - GeophysicalPartitionMemberOf
- PhysicallyConnectedTo
 - FlowsInto

These examples illustrate implementation-dependent relationship typing.

³² National Library of Medicine. (2004) Unified Medical Language System. Current relations in the semantic network. In: NLM. *Unified Medical Language System-Semantic Network Documentation, Section 3. Semantic Networks*. Available at: http://www.nlm.nih.gov/research/umls/META3_current_relations.html (accessed 2010-01-20).

³³ Hill, L. (1999). Content standards for digital gazetteers. Presentation at the *JCDL2002 NKOS Workshop "Digital gazetteers--Integration into distributed digital library services"*, July 18, 2002, Portland, Oregon. Available at: <http://nkos.slis.kent.edu/DL02workshop.htm> (accessed 2010-01-20).

5.4 *NOMEN-to-NOMEN* Relationships

Only the equivalence and whole-part relationships are discussed below. Other *nomen-to-nomen* relationships may also be established.

5.4.1 Equivalence Relationship

Equivalence of *nomen* is a very important notion in subject access. Two *nomens* are equivalent if they are appellations of the same *thema*. The equivalence relationships in a monolingual controlled vocabulary can be found in five general situations:^{34,35}

- a) the *nomens* are synonyms
- b) the *nomens* are near or quasi-synonyms
- c) the *nomens* have lexical variants
- d) a *nomen* is regarded as unnecessarily specific and it is represented by another *nomen* with broader scope
- e) a *nomen* is regarded as unnecessarily specific and it is represented by a combination of two or more terms (known as “compound equivalence”).

It is obvious that equivalence relationships do not assume exact equivalence. Inexact and partial equivalence are often found in controlled vocabularies. In reality, the *nomens* in the above d) and e) situations represent different *themas*. But since in some controlled vocabularies these *nomens* are connected as preferred and alternative terms, it may be interpreted that these *nomens* are assumed to represent the same *thema* in a controlled vocabulary.

In addition, equivalence relationships exist between *nomens* in different languages and across schemes. For example, “iron” (a term in English), “železo” (a term in Slovenian), and “Fe” (a chemical symbol) are all *nomens* for the same metal and are therefore considered equivalent.

The equivalence relationships of *nomens* can be specified further. For example:

- Replaces/Is replaced by
[e.g., “integrated plant control” is replaced by “centralized control”]
- Has variant form/Is variant form
 - Has acronym/is acronym for
[e.g., “VS” is acronym for “virtual storage”]
 - Has abbreviation/Is abbreviation of
 - Has transliterated form/Is transliteration of

5.4.2 The Whole-Part Relationship

The *whole-part relationship* also exists between *nomens*. A *nomen* may have components (parts). These components may or may not be a *nomen* on their own. The composition of

³⁴ ISO. (2009). ISO/CD 25964-1. *op. cit.*

³⁵ NISO. (2005). Z39.19-2005. *op.cit.*

such *nomen* may be governed by rules, for example, the citation order in faceted classification schemes or the order of subdivisions in a subject heading.

6. USER TASKS

6.1 Users and Use

During the early stages of developing the entity-relationship conceptual model of subject authority records, the FRSAR Working Group considered it essential to analyze the users of subject authority data, to identify the contexts in which the data is used, and to characterize different usage scenarios.

Potential user groups include:

- a) information professionals who create and maintain subject authority data, including cataloguers and controlled vocabulary creators;
- b) information professionals who create and maintain metadata;
- c) reference services librarians and other information professionals who search for information as intermediaries; and
- d) end-users who search for information to fulfil their information needs.

Intermediaries (group c) act on behalf of end-users (group d). They interact with the bibliographic data in a similar way as end-users (although at a higher level of expertise). Therefore, for the purpose of this use analysis the intermediaries and end-users are considered belonging to the same end-user group.

Figure 6.1 illustrates the three-point perspective of subject authority data users.

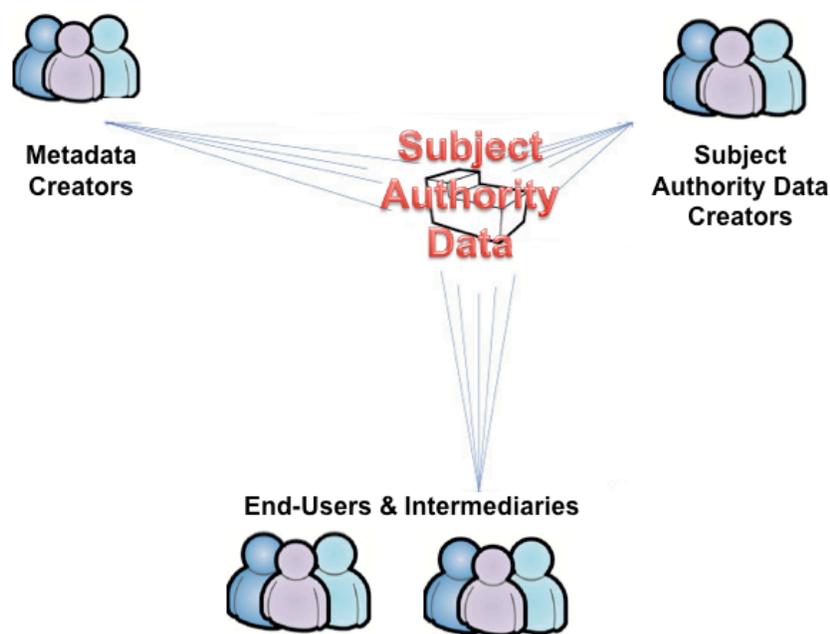


Figure 6.1 Users of subject authority data

6.2 User Tasks

When using subject authority data, a user may need to *find*, *identify*, and *select* a subject entity or entities. A user may also choose to *explore* a subject domain and its terminology as well as the relationships that exist among the *themas*. In addition, the user may explore the correlation of the *nomen(s)* of a *thema* in one subject authority system to the respective *nomen(s)* of the same *thema* in another subject authority system.

Based on the results from the two user studies conducted by the User Tasks Sub-Group, four tasks for subject authority data have been defined:

- **Find** one or more subjects and/or their appellations, that correspond(s) to the user's stated criteria, using attributes and relationships;
- **Identify** a subject and/or its appellation based on their attributes or relationships (i.e., to distinguish between two or more subjects or appellations with similar characteristics and to confirm that the appropriate subject or appellation has been found);
- **Select** a subject and/or its appellation appropriate to the user's needs (i.e., to choose or reject based on the user's requirements and needs);
- **Explore** relationships between subjects and/or their appellations (e.g., to explore relationships in order to understand the structure of a subject domain and its terminology).

The *explore* user task is a new task introduced in FRSAD; whereas the *find*, *identify*, and *select* user tasks have been previously introduced in the FRBR and/or FRAD conceptual models. The subject authority data use survey conducted by FRSAR's User Tasks Sub-Group indicates that a large number of participants (69%) use subject authority data to explore relationships among terms during cataloguing and metadata creation. In addition, 62% of participants use subject authority data to explore relationships while searching for bibliographic resources, and 64% use these data to navigate and browse bibliographic descriptions. These numbers reflect a major use of subject authority data for a task that was not present in the FRAD and FRBR models; therefore, the group deemed it important to add the *explore* user task.

Figure 6.2 shows a comparison of user tasks as defined in FRBR, FRAD, and FRSAD.

User Tasks

FRBR	FRAD	FRSAD
Find	Find	Find
Identify	Identify	Identify
Select		Select
Obtain		Explore
	Contextualize	
	Justify	

Figure 6.2 Comparison of user tasks as defined in FRBR, FRAD, and FRSAD

6.3 Assessing Values Relative to User Tasks

While in some cases the users' information needs are limited to authority data only, in most cases the users will utilise subject authority data to *find*, *identify*, *select*, and/or *obtain* Group 1 entities as specified by the FRBR user tasks.

The following is a list of tasks that result from placing the primary subject authority data user tasks (*find*, *identify*, *select*, and *explore*) in the context of different user groups as it relates to interacting only with subject authority data. The activities of using subject authority data to access bibliographic data are covered by FRBR. In the following examples, text within double quotation marks represents a nomen; a thema is referred by English text within curly brackets. Single quotations marks are used for thema types.

FIND: using the data to find one or more subjects and/or their appellations, that correspond(s) to the user's stated criteria, using attributes and relationships

- 1) Using subject authority data to **find** a *thema* or a set of *themas* based on the user's search criteria. For example:
 - A user is looking for a *thema* of the type 'substance' within the medical field.
 - A user is looking for a *thema* that is a 'medical condition(s)' for which {chloromadinone acetate} is used for treatment.
 - A user is looking for *themas* that are particular artistic styles (for example, {modernism}) using *thema-to-thema* hierarchical relationships.
- 2) Using subject authority data to **find** a *nomen* or a set of *nomens* for a *thema*. For example:
 - A user is looking for the *Dewey Decimal Classification* number (*nomen*) for the *thema* {dragonflies} (as it is referred to in English).

- A user is looking for the preferred *nomen* in the *Library of Congress Subject Headings* for the *thema* {lilac flower} (as commonly referred to in English).

IDENTIFY: using the data to identify a subject and/or its appellation based on their attributes or relationships

- 3) Using subject authority data to **identify** a *thema*, i.e., to confirm that the *thema* found is the one sought by the user, or to distinguish between two similar *themas*.
For example:
 - A user is employing subject authority data to identify whether the *thema* {clothing} or the *thema* {costume} is more appropriate for a specific information need.
- 4) Using subject authority data to **identify** a *nomen*, i.e., to confirm that the *nomen* found is the one sought by the user, or to distinguish between two similar *nomens*.
For example:
 - A user employs subject authority data to verify whether the *nomen* “craftsman style” is the appropriate *nomen* in a particular system.

SELECT: using the data to select a subject and/or its appellation appropriate to the user’s needs

- 5) Using subject authority data to **select** a *thema* from the set of *themas* found.
For example:
 - Select a *thema* at the appropriate level of specificity from a hierarchy of related *themas*: A user is using subject authority data to select the *thema* {volley ball} as a more appropriate subject access point in a bibliographic record rather than the broader *thema* {ball games}.
- 6) Using subject authority data to **select** a *nomen* from the set of *nomens* found.
For example:
 - Select the preferred *nomen* for a *thema* within a subject authority system to use in searching or in assigning access points: A user is using subject authority data to select “ale glasses” among the *nomens* “ale glasses,” “glass, beer,” and “malt-beverage glass” found in the *Art and Architecture Thesaurus*.

EXPLORE: using the data in order to explore relationships between subjects and/or their appellations

- 7) Using subject authority data to **explore** the relationships between two or more *themas* within the same subject authority system.
For example:
 - A user is using subject authority data to explore associative relationships of the *thema* {digital libraries} and other *themas*.
- 8) Using subject authority data to **explore** the relationships between two or more *nomens* within the same subject authority system.
For example:
 - A user is using subject authority data to explore the relationship of the *nomen* “ladybugs” and the *nomen* “ladybirds” in LCSH.

- 9) Using subject authority data to **explore** the correlation of *themas* between two or more subject authority systems.
For example:
- A user is using subject authority data to explore the correlation of the *thema* {domestic cats} between the LCSH and the *Sears List of Subject Headings*.
- 10) Using subject authority data to **explore** the correlation of *nomens* between two or more subject authority systems.
For example:
- A user is using subject authority data to explore whether the *Medical Subject Headings* (MeSH) *nomen* “cataract” correlates to the *National Library of Medicine (NLM) Classification’s nomen* “WW 260.”
- 11) Using subject authority data to **explore** the structure of a subject domain within a subject authority system.
For example:
- A user is using subject authority data to explore how the domain {computer science} is represented within the *American Society for Information Science and Technology (ASIS&T) Thesaurus*.

6.4 Mapping of Attributes, Relationships, and User Tasks

Tables 6.1 and 6.2 map the attributes and relationships defined in chapters 4 and 5 to the defined set of user tasks. The mapping is intended to clarify which attributes and relationships are required to support each particular user task. Only attributes and relationships specified in the FRSAD model are mapped. The decision as to which attributes and relationships to include or indicate as mandatory and, to some extent, the determination of importance, are application- or implementation-specific. The degree of importance included in Table 6.1 followed the FRBR report and is based on an analysis of common library and other information agencies’ practice.

THEMA

Attributes	Find	Identify	Select	Explore
Type of thema	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Scope note		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Thema-to-Thema Relationships				
Hierarchical relationship	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	■
Associative relationship	<input type="checkbox"/>		<input type="checkbox"/>	■

■ = strong importance □ = moderate importance Blank = not important

Table 6.1 Mapping for *Thema*

NOMEN

Attributes	Find	Identify	Select	Explore
Type of nomen	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Scheme	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Reference Source of nomen		<input type="checkbox"/>		
Representation of nomen	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Language of nomen	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Script of nomen	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
Script conversion	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
Form of nomen	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Time of validity of nomen	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Audience	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
Status of nomen	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
Nomen-to-Nomen Relationships				
Equivalence relationship	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Whole-part relationship	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

= strong importance
 = moderate importance
Blank = not important

Table 6.2 Mapping for *Nomen*

7. CONCLUSION

In this report, the FRSAR Working Group has presented a conceptual model within the FRBR framework as it relates to the aboutness of *works*. The report defines a structured frame of reference for relating the data that are recorded in subject authority records to the needs of the users of these data. The FRSAD model is developed with the goal of assisting in an assessment of the potential for international sharing and use of subject authority data both within the library sector and beyond. It enhances consideration for the functional requirements for subject authority data at a level that is independent of any implementation, system, or specific context.

APPENDIX A. MODELING ABOUTNESS

A.1 Subject Relationship and Group 3 Entities Introduced in FRBR

The subject relationship introduced in the FRBR model is illustrated in Figure A.1:

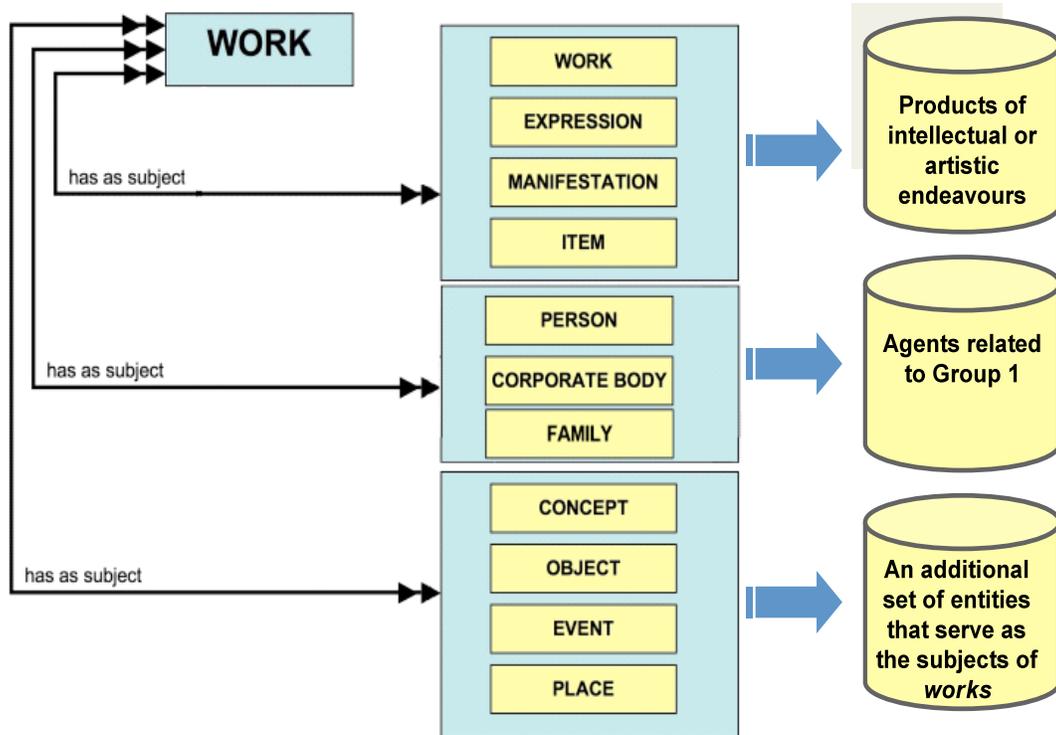


Figure A.1 Extension of FRBR Figure 3.3 "Group 3 entities and 'subject' relationships"

The diagram in FRBR Figure 3.3 depicts the “subject” relationships between *works* and entities in Group 1, Group 2, and Group 3. These three groups are represented as the components on the right side of the above figure. The left and centre components in the figure are based on the FRBR Figure, with the *Family* entity added in Group 2 according to the FRAD model.

The entities in Group 3 represent an additional set of entities that serve as the subjects of *works*. The FRBR report specified Group 3 entities under Figure 3.3 as³⁶:

3.1.3 Group 3 Entities: Concept, Object, Event, Place

³⁶ *Functional Requirements for Bibliographic Records: Final Report*. (1998). IFLA Study Group on the Functional Requirements for Bibliographic Records. München: KG Saur, p. 17.

The entities in the third group (outlined in bold in Figure 3.3) represent an additional set of entities that serve as the subjects of *works*. The group includes *concept* (an abstract notion or idea), *object* (a material thing), *event* (an action or occurrence), and *place* (a location).

The diagram depicts the “subject” relationships between entities in the third group and the *work* entity in the first group. The diagram indicates that a *work* may have as its subject one or more than one *concept*, *object*, *event*, and/or *place*. Conversely, a *concept*, *object*, *event*, and/or *place* may be the subject of one or more than one *work*.

The diagram also depicts the “subject” relationships between *work* and the entities in the first and second groups. The diagram indicates that a work may have as its subject one or more than one *work*, *expression*, *manifestation*, *item*, *person*, and/or *corporate body*.

A.2 Possible Approaches to the Model of Aboutness

The FRSAR Working Group has, as the central part of its terms of reference, the goal of building a conceptual model of Group 3 entities within the FRBR framework as they relate to the **aboutness** of *works*.

It is mentioned in the FRBR study that “further analysis is needed of the entities that are the centre of focus for subject authorities, thesauri, and classification schemes, and of the relationships between those entities”³⁷. In the years following the publication of the FRBR model, some researchers focused on Group 3 entities, particularly on the fact that *time* is not included.³⁸ Consequently, time and space are not treated symmetrically. Some discussions brought attention to the lack of coverage of activities and processes.

Tom Delsey, in his paper published in *Cataloging & Classification Quarterly* in 2005, highlighted the aspects of the FRBR model that “will need to be re-examined as part of a more intensive analysis of subject access”³⁹. Delsey followed up with a presentation of a paper at the IFLA satellite meeting in Järvenpää, Finland, before the IFLA General Conference in Oslo in August 2005. His presentation has provoked much discussion among the members of the FRSAR Working Group.

Delsey identified three “broad objectives” to be met by re-examination of the ways in which the FRBR model analyzes data relevant to subject access:

- 1) “to ensure that the scope of the *entities* defined in the [FRBR and FRAD] models is sufficient to cover everything that a user of a library catalogue might view as a ‘subject’”;

³⁷*Ibid*, p. 7.

³⁸ Heaney, M. (1997). Time is of the essence. Available at: <http://www.bodley.ox.ac.uk/users/mh/time978a.htm> (accessed 2010-01-20).

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

- 2) “to ensure that the *attributes* that come into play in the construction and use of subject access points and subject authority records are adequately covered”; and
- 3) “to ensure that the models provide a clear and pragmatic representation of the *relationships* that are reflected through subject access points in bibliographic records as well as those reflected in the syndetic structure of thesauri, subject heading lists, and classification schemes and in the syntactic structure of indexing strings⁴⁰ (emphases added)”.

Delsey identified two “key questions” related to entities: “The first [of the key questions] is whether the entities are defined in sufficiently broad terms to cover fully what we might characterize as the “subject” universe. The second is whether the categorizations represented by the entities defined in the models are appropriate and meaningful for the purposes of clarifying the bibliographic conventions through which that “subject” universe is reflected”⁴¹. In other words:

1. Are the entity classes *collectively exhaustive*? Does the model cover the whole universe of subject-related entity classes?
2. Are the entity classes *individually appropriate*? Does the model carve up the universe of subject-related entity classes in the “right” way⁴²?

As a first step, the FRSAR Entities Sub-Group performed a pilot study, 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 included 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 others. The same method was also applied by one of the Working Group members to another set of subject terms from controlled vocabularies used in two library science textbooks. The results show that there is a blurred distinction between concrete and abstract concepts; for example, the distinction between a particular chair as a physical object and the concept of chairs. In addition, there were difficulties in classifying named instances (proper names), which resulted in many terms being put into the “others” category. The results of this test indicate 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 do not seem helpful or necessary to the end users either.

Following the pilot study, the Working Group discussed several possible previously identified approaches to the development of a theoretical framework of aboutness.

⁴⁰ Delsey, *op. cit.*, p. 50.

⁴¹ Delsey, *op. cit.*, p. 50.

⁴² See also: Furner, J. (2006). The ontology of subjects of works. Paper presented at ASIS&T 2006: Annual Meeting of the American Society for Information Science and Technology, Austin, TX, November 3–8, 2006.

Scenario 1

Keep FRBR Group 3 entities (*concept, object, event, and place*) and only analyze attributes and relationships. The advantage is that the Working Group uses an existing framework. However, as demonstrated in the pilot study of the FRSAR Entities Sub-Group, the Group 3 entities need to be revisited. Adding *time* to the FRBR list solves one part of the problem, but the resulting model still does not cover processes, activities, or situations.

An additional argument for rejecting this scenario is that the original categorisation of Group 3 entities into four classes goes too far towards prescribing a particular way of structuring the subject languages that are used to provide access to *works*. Any subject authority system that lacks a faceted structure to distinguish clearly between *concepts, objects, events, and places* can be modelled only with difficulty. Rather than taking a stand on exactly which aspects to identify for the entire information community, the Working Group felt it was important to provide a higher level, more theoretical approach and not to impose any constraint on the forms that subject authority systems take in particular implementations. This modelling does not limit any community from implementing the original FRBR Group 3 entities; on the contrary, it allows for more flexibility.

Scenario 2

Take Ranganathan's facets as the basis of the new framework. The facets would become entities:

- Personality
- Matter
- Energy
- Space
- Time

The advantage is that this approach is well known in the library community, has been justified theoretically, and covers all areas of aboutness quite well. The issues are whether we would still have problems defining some of the entities, and whether librarians and end users would have trouble understanding and applying them.

Scenario 3

Take the <indec>⁴³ model as the basis of the new framework. The main focus of the <indec> model is intellectual property and rights management, but it also overlaps significantly with FRBR. The basic <indec> entities are defined as:

- **Percept:** an entity that is perceived directly with at least one of the five senses.

⁴³ Rust, G. and Bide, M. (2000). The <indec> metadata framework: Principles, Model and Data Dictionary. Version2. Indecs Framework Ltd. Available at: http://www.doi.org/topics/indec/indec_framework_2000.pdf (accessed 2010-01-20).

- **Being**: an entity that has characteristics of animate life; anything which lives and dies
- **Thing**: an entity without the characteristics of animate life
- **Concept**: an entity that cannot be perceived directly through the mode of one of the five senses; an abstract entity, a notion or idea; an abstract noun; an unobservable proposition, which exists independently of time and space
- **Relation**: the interaction of percepts and/or concepts; a connection between two or more entities
 - **Event**: a dynamic relation involving two or more entities; something that happens; a relation through which an attribute of an entity is changed, added or removed
 - **Situation**: a static relation involving two or more entities; something that continues to be the case; a relation in which the attributes of entities remain unchanged

Being and **Thing** together correspond to a supertype of the FRBR entity *object*; **Concept** roughly corresponds to the FRBR entity *concept*; and **Event** corresponds to the FRBR entity *event*. Thus, the three main differences between the <indec> model and the FRBR model are (a) the subtyping of **Percept** in the <indec> model into **Being** and **Thing**, and in the FRBR model into *item*, *person*, and *object*, (b) the absence of an FRBR entity that directly corresponds to the <indec> entity **Situation**, and (c) the absence of an <indec> entity that directly corresponds to the FRBR entity *place*.

As Delsey⁴⁴ notes, these differences raise corresponding questions about the possibility of making changes to the set of Group 3 entities defined in the original FRBR model: (a) Should the original entity **Object** be subtyped into two entities—e.g., **Inanimate object**, and **Animate object**? (b) Should **Situation** be added as an entity? (c) Should the FRBR entity *place* be removed? Note that in the FRBR report *places* are treated as entities only to the extent that they are the subject of a *work*.

Scenario 4

Make a pragmatic list of entities. Buizza and Guerrini created one example of such a list⁴⁵ for the Italian project Nuovo soggettario. Two logical entities, the *subject* (the topic, the basic theme of the work, the summarisation of its main contents) and *concept* (a unit of thought, each of the single elements which make up the subject), were defined. The list shows, as an example, what can be a concept in a specific implementation and draws on categories, roles and relationships from the report of the project:

- Object (material thing)
- Abstraction
- Living organism
- Person
- Corporate body

⁴⁴ Delsey, op. cit., p. 51 - 52

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

- Work
- Matter/material
- Property/quality
- Action
- Process
- Event
- Place
- Time

The problem with such lists is that the entities are not mutually exclusive, have overlaps, and rely on individual common everyday definitions of the entities. The authors' original purpose was to show the compatibility of those categories with the model. It is also a warning that making a pragmatic list of entities would be a disadvantage for a theoretical model.

Scenario 5

Do not make any recommendation on categorisation of subjects. This approach is a more abstract view and does not pose restrictions on any implementations. It also allows a more abstract, general view.

This last scenario (5) was the decision taken by the Working Group, based on comparative analysis of all scenarios and the pilot user study. None of scenarios 1-5 are ideal for all situations, while each may be a good solution for particular implementations. Any further categorization of Group 3 entities would prescribe a particular way of structuring the subject authority systems that are used to provide access to works. A good model should allow for any multiple domain-specific structures and should be flexible enough to accommodate different implementations. This can be achieved only by a more abstract theoretical model, completely independent of any implementation that enables the treatment of attributes and relationships on a more general level.

APPENDIX B. RELATIONSHIP OF FRSAD WITH FRBR AND FRAD

Two models, FRAD (by FRANAR Working Group) and FRSAD (by FRSAR Working Group), complement and further develop some aspects of the original model of FRBR. The three models together have been labelled the ‘FRBR family’, suggesting that they are all considered parts of a larger general model. There exist some differences among them, though; the respective FRBR, FRANAR and FRSAR working groups have made different modelling decisions during their independent model development. Eventually FRBR, FRAD and FRSAD will have to be harmonised and a resultant new consolidated model will be developed. In order to facilitate this process the FRSAR Working Group takes this opportunity to list the most important differences below.

B.1 Relationship of FRSAD with FRBR

The FRSAR Working Group follows FRBR in the methodology, specification, and presentation of entities and relationships. The “has as subject” (many-to-many) relationship, established between the *work* and the entity(ies) representing the aboutness of the *work*, is kept in its entirety in FRSAD. As in FRBR, the FRSAD model also starts with a user tasks analysis and follows with the establishment of appropriate entities and relationships. The four areas where some differences were introduced in FRSAD are:

- The addition of the “Explore” task;
- *Thema* is introduced as a superclass of all entities that can be subjects of a *work*. Attributes and relationships of *thema* are presented;
- No entities are explicitly predefined in Group 3;
- *Nomen* is introduced (including attributes and relationships) and is defined as a separate entity instead of an attribute.

The inclusion of the “explore” task is based on the findings of the user study conducted by the FRSAR Working Group. Users of subject authority data also use these data to explore a domain, to get acquainted with the terminology, and to identify semantic relationships. The FRSAR Working Group is confident that the same is true for bibliographic information in general, and recommends that the ‘explore’ task be added to the general model.

In Fig. 3.3 of the FRBR report, the depicted “subject” relationship has three boxes representing all three groups of entities respectively, on the right side of the ‘has as subject’ relationship. FRSAD has developed this further by creating a superclass (*thema*), thus enabling the modelling of the "has as subject" relationship on a more general level. *Thema* includes Group 1, Group 2 and all other entities that can be the subjects of a *work*. Therefore, the subject relationship can easily be modelled as "*work* has as subject *thema*."

FRBR defines four entities in Group 3: *concept*, *object*, *event*, and *place*. The FRSAR Working Group, based on the pilot user study, literature review, and independent analysis, decided to avoid any predefined subclasses. There seems to be no universal categorisation of *themas* and any attempt to declare one would necessarily limit the usability of a general model. Each particular implementation will need to define the categories or types of *themas*. The original FRBR Group 3 entities are, therefore, only one possible scenario. (Please refer to Section 4.1.1.)

FRSAD introduces a differentiation between a thing itself and its appellation. The appellation (name, label, etc.) is often modelled as an attribute of the entity it refers to (also in FRBR). While simpler, this approach makes it impossible to introduce the attributes (e.g. language) and relationships (e.g. the relationship between a former and current name) of the appellation itself, because in an E-R model one may not have attributes of an attribute. *Nomen* is therefore introduced in FRSAD as an entity, rather than an attribute, to enable appropriate modelling.

B.2 Relationship of FRSAD with FRAD

The FRANAR Working Group was established in 1999 with the mandate of developing FRBR further in the area of authority files. Later the decision was made by FRANAR to focus on Group 2 entities and *work* only. As a consequence, the FRSAR Working Group was established to cover the ‘has as subject’ relationship and the appropriate entities. The FRAD and FRSAD models were therefore developed independently. The working groups, although both following FRBR and its modelling approach, have made several different decisions. The most significant ones are:

- User tasks: “Contextualise” and “Justify” in FRAD vs. “Explore” in FRSAD;
- *Name* in FRAD vs. *Nomen* in FRSAD;
- *Name*, *Identifier* and *Controlled access point* as separate entities in FRAD vs. values of the attribute “Type of *Nomen*” in FRSAD;
- *Rules* and *Agency* as new entities in FRAD and not explicitly modelled in FRSAD.

B.2.1 User tasks

The Working Group believes that “explore” is a generalisation of “contextualise” and expresses better the user task of browsing, getting acquainted, becoming familiar with, and discovering.

“Justify,” on the other hand, is a task of information professionals and not end-users. It is an important task on its own, but falls within metadata creation and not metadata use. Since FRSAD follows the FRBR approach that has not extended its model to cover such tasks, it is not included in the FRSAD model.

B.2.2 Name and Nomen

Although similar at first glance, the two entities are different: FRAD *name* is defined as “a character or group of words and/or characters by which an entity is known in the real

world”. The FRSAD *nomen* is a more general entity, comprising any (textual or other) appellation both in the real world and in artificial systems. In relation to FRAD, *nomen* is a superclass of FRAD *name*, *identifier* and *controlled access point*.

B.2.3 Separate appellation entities in FRAD vs. *nomen* entity and “type” attribute in FRSAD

Nomen is the general appellation entity in FRSAD with specific types. This allows the introduction of any type that will be necessary for an implementation. In addition, some possible general values of the attribute “type” are already suggested, such as “identifier”. FRAD “name,” “identifier” and “controlled access point” are therefore possible types of *nomens*. This approach allows flexibility; even particular kinds of identifiers (URI, ISBN, etc.) can be defined as values of “type” of a *nomen*.

B.2.4 Rules and Agency

Rules and *agency* are not specifically modelled in FRSAD. The position of the Working Group is that the focus of the model is not on the cataloguing process and it is not necessary to include that level of detail. If needed, *rules* (which are applied in all phases of cataloguing, not only in creation of controlled access points) should be considered instances of *work*. *Agencies*, which apply the rules, should be considered instances of *corporate body*. If modelled, they are in a relationship with the attribute assignment event.

APPENDIX C. FRASAD MODEL AND OTHER MODELS

C.1 The Importance of the THEMA-NOMEN Model

As early as 1923, Ogden and Richards⁴⁶ published a famous triangle of meaning that illustrated the relationship between language, thought content, and referent. The graph (Figure D.1) 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, and the words that refer to them, and the corresponding experiences in the *psyche*. Equally, Frege distinguished between two types of meaning: thought content and referent, in his essay *Über Sinn und Bedeutung*. It is not enough to try to understand what a thing is, based on its name, because it may have been named in ancient times, and the name reflects only what the name-givers thought was the nature of reality then. Therefore multiple terms may refer to the same object or idea, a single term may refer ambiguously to more than one object or idea, and outdated terms may be confusing⁴⁷.

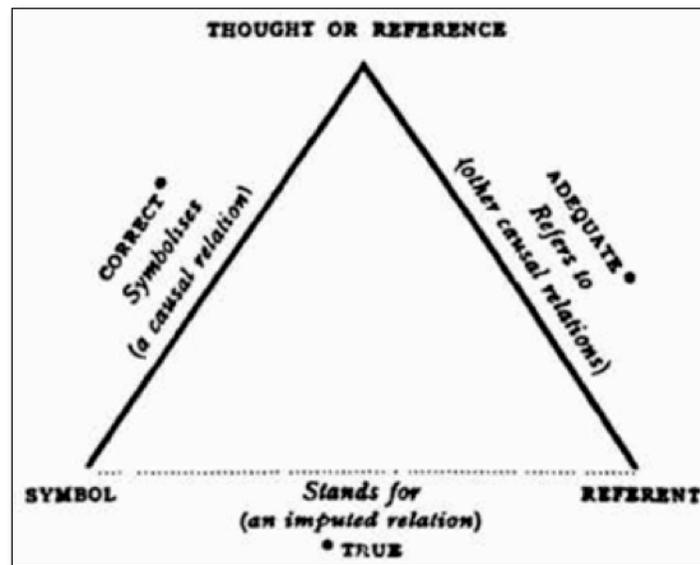


Figure D.1 Ogden's Semiotic Triangle. (Ogden and Richards, 1923,⁴⁸ p.11)

Ogden's model was also adopted by researchers in library and information science as the basis for building subject authority systems^{49,50}.

⁴⁶ Ogden, C. K., and 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.

⁴⁷ Frege, G. (1892). Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, NF 100. 1892, S. 25–50. Available at: <http://www.gavagai.de/HHP31.htm> (accessed 2009-05-22).

⁴⁸ Ogden and Richards. *op. cit.*

The importance of the *thema-nomen* model for subject authority data is to separate *subjects* 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 *nomen*, e.g., 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 or between a classification scheme and a thesaurus. Such efforts usually encounter much greater challenges because they are concerned with the subject mappings in terms of their meaning as well as the relationships among the subjects.

C.2 Mapping the FRSAD Model to Other Models

This *thema-nomen* conceptual model matches well with encoding schemas such as SKOS Simple Knowledge Organization System (SKOS) and OWL Web Ontology Language (OWL), which provide models for expressing the basic structure and content of knowledge organization systems (KOS) such as thesauri, classification schemes, subject heading lists, taxonomies and other similar types of controlled vocabularies, as well as ontologies. SKOS defines classes and properties sufficiently for representing the common features found in a standard thesaurus and other KOS structures. The SKOS model is based on a concept-centric view of vocabulary, where primitive objects are not labels; rather, they are concepts represented by labels. 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. These 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 language⁵¹. Each of these can be matched to what have been defined in the FRSAD model in terms of *thema*, *nomen*, and their attributes. SKOS also has specific properties to represent all of the semantic relationships that are described in Chapter 5.

Regarding issues of complexity and granularity of the *themas* and comprehensive semantic relationships between and among *themas* that FRSAD attempts to cover, OWL has even better matches. OWL ontologies provide classes, properties, individuals, and data values and are stored as Semantic Web documents⁵². OWL 1 mainly focused on constructs for expressing information about classes and individuals. OWL 2, the newest

⁴⁹ Dahlberg, I. (1992). Knowledge organization and terminology: philosophical and linguistic bases. *International Classification*. 19(2):65-71.

⁵⁰ Campbell et al., *op. cit.*

⁵¹ *SKOS Simple Knowledge Organization System Reference* (2009). Eds. Miles, A. and Bechhofer, S. . W3C Candidate Recommendation 17 March 2009. Available at: <http://www.w3.org/TR/skos-reference/> (accessed 2010-01-20).

⁵² *OWL 2 Web Ontology Language Structural Specification and Functional-Style Syntax*. (2009). Eds. Motik, B, Patel-Schneider, P.F. and Parsia, B. W3C Working Draft 21 April 2009. Available at: <http://www.w3.org/TR/owl2-syntax/> (accessed 2010-01-20).

W3C working draft, offers new constructs for expressing additional restrictions on properties, new characteristics of properties, incompatibility of properties, properties chains, and key properties⁵³. 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.

When the DCMI Abstract Model 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. According to the DCMI model, a record can contain *description sets*, which may contain *descriptions* composed of *statements*, which use *property-value* pairs.⁵⁴ 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 translations⁵⁵. The conceptual model proposed by the FRSAR Working Group corresponds to this abstract model by allowing any *thema* to be independent of any *nomen*, including any syntax that a *nomen* may use. Therefore this conceptual model will facilitate the sharing and reuse of subject authority data amongst not only the subject authority systems themselves, but also metadata resources.

C.3 Conclusion

Putting the subject authority data within the context of the Semantic Web developments, especially from the perspective of the Web of Data, subject authority data that are modelled based on FRSAD and encoded in SKOS and OWL will be able to become part of linked open data and contribute to the further development of the Semantic Web.

⁵³ *OWL 2 Web Ontology Language New Features and Rationale*. (2009). Eds. Golbreich, C. and Wallace, Evan K. W3C Working Draft 21 April 2009. Available at: <http://www.w3.org/TR/owl2-new-features/> (accessed 2010-01-20).

⁵⁴ *DCMI Abstract Model*. (2007). Eds. Powell, A., Nilsson, M., Naeve, A. Johnston, P. and Baker, T. Available at: <http://dublincore.org/documents/abstract-model/> (accessed 2010-01-20).

⁵⁵ *Ibid.*

APPENDIX D. EXAMPLES FROM SUBJECT AUTHORITY SYSTEMS

This appendix provides examples found in implementations of existing subject authority systems through the perspective of the FRSAD model, presented in four parts: 1) existing models of *thema* types; 2) *thema-thema* relationships presented in subject authority data (both in individual vocabularies and cross-schemes); 3) same *thema* represented by *nomens* from different schemes; and 4) examples of display records from controlled vocabularies or subject authority systems.

D.1 Existing Models of THEMA Types

In Chapter 4 Attributes, “type” is defined as a general attribute of *thema* because other attributes are usually implementation-dependent. In any particular application, *themas* would normally have particular implementation-specific types. Based on our preliminary study, there seems to be no generally applicable categorization of *themas*. This is also supported by the following examples, ranging from general (*Faceted Application of Subject Terminology*) to more specialized subject domains such as biomedical and health sciences (*Unified Medical Language System* and *The Foundational Model of Anatomy Ontology*) and art and architecture (*Art and Architecture Thesaurus*).

Example D.1.1 *Faceted Application of Subject Terminology* (FAST) subject facets⁵⁶:

Faceted Application of Subject Terminology (FAST) is an adaptation of the *Library of Congress Subject Headings* (LCSH) with a simplified syntax. LCSH headings form the basis for FAST authority file. FAST employs a faceted approach by defining headings according to their functions and categorizes all headings into eight facets. Seven of them are subject facets and one is form (genre) facet. The subject facets include:

- Topical
- Personal Names (as Subjects)
- Corporate Names (as Subjects)
- Geographics
- Periods
- Titles
- Events

Headings in the FAST database include both single-concept and multiple-concept headings. Each FAST heading or heading-string belongs to a single facet.

⁵⁶ FAST: *Faceted Application of Subject Terminology*. [2001-]. OCLC Online Computer Library Center. Available at: <http://www.oclc.org/research/activities/fast/default.htm> (accessed 2010-01-26).

Example D.1.2. *Unified Medical Language System*[®] (UMLS) semantic types^{57, 58}

The *Unified Medical Language System*[®] (UMLS), developed, maintained, and distributed by the National Library of Medicine of the United States, provides a unified system for correlating a large number of biomedical terminologies and facilitates the development of computer systems that behave as if they “understand” the meaning of the language of biomedicine and health. In order to facilitate the establishment of correspondences in the meanings of terms, the same concepts occurring in different constituent vocabularies are assigned to high level semantic types encompassed within the *UMLS Semantic Network*. It consists of: (a) a set of broad subject categories, or **Semantic Types**, that provide a consistent categorization of all concepts represented in the *UMLS Metathesaurus*[®], and (b) a set of useful and important relationships, or **Semantic Relations**, which exist between Semantic Types. More than 130 semantic types and 50 semantic relationships defined by the UMLS can be found in the UMLS 2004 AB Documentation⁵⁹. The following are the high level semantic types:

Entities

- Physical Object
 - Organism
 - Anatomical Structure
 - Manufactured Object
 - Substance
- Conceptual Entity
 - Idea or Concept
 - Finding
 - Organism Attribute
 - Intellectual Product
 - Language
 - Occupation or Discipline
 - Organization
 - Group Attribute
 - Group

Events

- Activity
- Phenomenon or Process

The scope of the *UMLS Semantic Network* is broad, allowing for the semantic categorization of a wide range of terminology in multiple domains. The top level types are **Entities** (including “Physical Object” and “Conceptual Entity”) and **Events** (including “Activity” and “Phenomenon or Process”). Looking at its major groupings of semantic types (such as organisms, anatomical structures, biologic function, chemicals, events, physical objects, and concepts or ideas) it is obvious that they are designed to be especially applicable in the domain of biomedical and health areas.

⁵⁷ National Library of Medicine. (2003-). *Unified Medical Language System. Current Semantic Types. UMLS 2004AB Documentation.* Last updated: 21 March 2008. Available at: http://www.nlm.nih.gov/research/umls/META3_current_semantic_types.html (accessed May 22, 2009).

⁵⁸ UMLS Factsheet. Available at: <http://www.nlm.nih.gov/pubs/factsheets/umls.html> (accessed May 22, 2009).

⁵⁹ *ibid.*

Example D.1.3. *The Foundational Model of Anatomy Ontology* semantic types⁶⁰

The Foundational Model of Anatomy (FMA) initially developed as an enhancement of the anatomical content of UMLS, is a domain ontology of the concepts and relationships that pertain to the structural organization of the human body. It is found that while there is considerable correspondence in the meaning of anatomical terms in the UMLS sources, there is very little similarity in the arrangement of anatomical terms among the source schemas. It is important that the underlying semantic structure of these abstractions must also be aligned. The top-level semantic types are **Anatomical Entity**, **Attribute Entity**, and **Dimensional Entity**:

- Anatomical Entity**
 - Non-physical anatomical entity
 - Physical anatomical entity
- Attribute Entity**
 - Cell morphology
 - Cell shape type
 - Cell surface feature
 - Concept name
 - Miscellaneous term
 - Organ part phenotype
 - Physical attribute relationship
 - Physical state
 - Structural relationship value
- Dimensional Entity**
 - Line
 - Point
 - Surface
 - Volume

As a domain ontology, the FMA represents deep knowledge of the structure of the human body. Its emphasis is on the highest level of granularity of the concepts. Meanwhile it also presents a great number of specific structural relationships between the references of these concepts. According to project documentation⁶¹, the FMA consists of approximately 75,000 anatomical classes, 130,000 unique terms, 205,000 frames, and 170 unique slots showing different types of relations, attributes, and attributed relationships. FMA is a typical example of modeling that shows how **semantic types** for a concept scheme can be defined. It not only encompasses the diverse entities that make up the human body but is also capable of modeling a great deal of knowledge relating these entities.

Example D.1.4. *Art and Architecture Thesaurus* (AAT) facets⁶²

Art and Architecture Thesaurus (AAT) is a controlled vocabulary for fine art, architecture, decorative arts, archival materials, and material culture for the purposes of

⁶⁰ *The Foundational Model of Anatomy ontology* (FMA). 2006---. School of Medicine, University of Washington. Available at: <http://sig.biostr.washington.edu/projects/fm/index.html> (accessed 2010-01-26).

⁶¹ About FMA. [2006]. School of Medicine, University of Washington. Available at: <http://sig.biostr.washington.edu/projects/fm/AboutFM.html> (accessed 2010-01-26).

⁶² *Art and Architecture Thesaurus Online. Hierarchy Display. op. cit.*

indexing, cataloging, and searching, as well as research tools. It was developed for literature about art and architecture and for records describing works of art and architecture. The facets in AAT are conceptually organized in a scheme that proceeds from abstract concepts to concrete, physical artifacts. These facets are: “Associated Concepts”, “Physical Attributes”, “Styles and Periods”, “Agents”, “Activities”, “Materials”, and “Objects”. Homogeneous groupings of terminology, or hierarchies, are arranged within the seven facets of the AAT:

- Top of the AAT hierarchies
 - **Associated Concepts Facet**
 - Associated Concepts
 - **Physical Attributes Facet**
 - Attributes and Properties
 - Conditions and Effects
 - Design Elements
 - Color
 - **Styles and Periods Facet**
 - Styles and Periods
 - **Agents Facet**
 - People
 - Organizations
 - Living Organisms
 - **Activities Facet**
 - Disciplines
 - Functions
 - Events
 - Physical and Mental Activities
 - Processes and Techniques
 - **Materials Facet**
 - Materials
 - **Objects Facet**
 - Object Groupings and Systems
 - Object Genres (Hierarchy Name)
 - Components (Hierarchy Name)
 - Built Environment (Hierarchy Name)
 - Furnishings and Equipment
 - Visual and Verbal Communication

The conceptual framework of facets is not subject-specific. One example is the subject “Renaissance painting”. Terms to describe Renaissance paintings will be found in many locations in the AAT hierarchies rather than a defined portion that is specific only for Renaissance painting⁶³.

In summary, all examples in this section indicate that in actual implementations there are always attempts to define some fundamental facets or atoms to accommodate all types of

⁶³About the AAT. Los Angeles: J. Paul Getty Trust, Vocabulary Program. Revised 12 November 2008. Available at: http://www.getty.edu/research/conducting_research/vocabularies/aat/about.html (accessed May 22, 2009).

themas. However, the resulting *themas* “types” differ from implementation to implementation.

D.2 THEMA-THEMA Relationships presented in Subject Authority Data

Authority **records** can be stored and displayed differently within a system, and they may also have various combinations of components when displayed to:

- information professionals who create and maintain subject authority data, including cataloguers and controlled vocabulary creators;
- information professionals who create and maintain metadata;
- reference services librarians and other information professionals who search for information as intermediaries; and
- end-users who search for information to fulfil their information needs.

Therefore, it is the authority **data**, not the **records**, which will be the focus in the examples presented in the following sections.

D.2.1 *Thema-Thema* relationships presented by individual vocabularies

The emphasis of this section is on the semantic relations presented in vocabularies. The following examples demonstrate how *thema-to-thema* relationships are presented in different vocabularies for the same *thema*, “mercury” (as a liquid metal and/or as an element). The same object can be viewed from different perspectives and therefore it may belong under different hierarchies (polyhierarchical relationship). Webster’s definition of mercury is: “a heavy silvery toxic univalent and bivalent metallic element; the only metal that is liquid at ordinary temperatures”⁶⁴.

[Note: In the figures in this section, an oval shaped node is used to represent a *thema*.]

⁶⁴ *Webster's Online Dictionary*. Definition: Mercury. Available at: <http://www.websters-online-dictionary.org/definition/mercury> (accessed July 2008).

Example D.2.1.1. LC Subject Authority

Thema: **mercury (as a liquid metal)**

[Note: in the following entry, MARC21 coding is used:

010 = Library of Congress control number

040 = Cataloging source

053 = LC Classification Number

\$c = Explanatory term (specifying topic)

150 = Heading--Topical term

450 = *See* from Tracing--Topical term (unauthorized form/variant of term)

550 = *See Also* From Tracing--Topical Term;

\$a = Topical term or geographic name entry element

\$w = Control subfield; g - Broader term.]

LC Control Number: sh 85083794

HEADING: Mercury

000 00558cz a2200217n 450

001 4734282

005 19900221112154.6

008 860211i| anannbabn |a ana

035 __ |a (DLC)sh 85083794

906 __ |t 8528 |u fk03 |v 0

010 __ |a sh 85083794

040 __ |a DLC |c DLC |d DLC

053 _0 |a QD181.H6 |c Chemistry

053 _0 |a TA480.M4 |c Engineering materials

053 _0 |a TN271.M4 |c Prospecting

053 _0 |a TP245.M5 |c Chemical technology

150 __ |a Mercury

450 __ |a Hydrargyrum

450 __ |a Quicksilver

550 __ |w g |a Liquid metals

953 __ |a xx00 |b fg07



[Note: in this captured screenshot, subfield signs are displayed as a vertical bar.]

Figure D.1: A record from the LC Subject Authority File

Several semantic relationships are indicated in this record. There is a semantic relationship between this *thema*, which has a *nomen* “Mercury”, and another *thema*, which has a *nomen* “Liquid metals” (see illustration below). This can be recognized by the field tag 550, which means “see also”. (Inter-system relationships will be explained later in section D.2.2.)

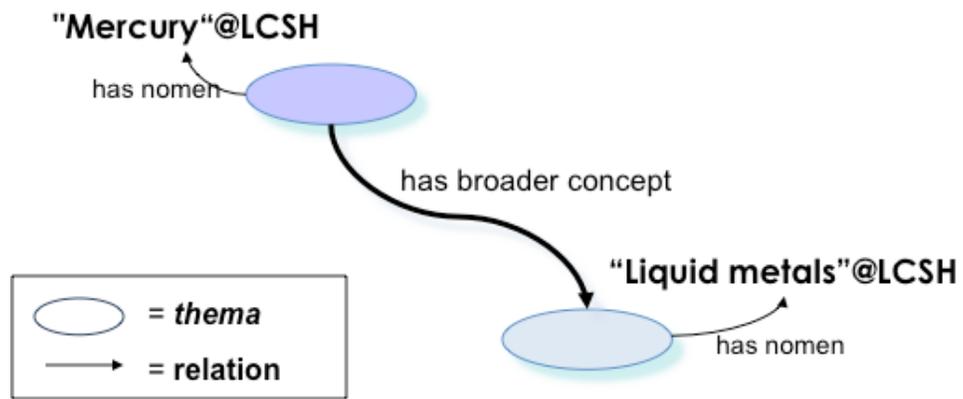


Figure D.2 Illustration of the semantic relations between two *themas* represented in Figure D.1

Example D.2.1.2. *Art and Architecture Thesaurus*:

Thema: mercury (as a liquid metal and as an element)

ID: 300011026

Record Type: concept

 **mercury** (<mercury and amalgam>, nonferrous metal, ... Materials)

Note: Pure metallic element having symbol Hg and atomic number 80; a lustrous silvery metal that is liquid at ordinary temperatures. Use also for this metal as processed and formed, usually in combination with other substances, to make various objects and materials.

Terms:

mercury (preferred,C,D,U,LC,English-P)

Hg (C,UF,U,A,English)

quicksilver (C,UF,U,English)

argento vivo (C,D,U,Italian-P)

Facet/Hierarchy Code: M.MT

Hierarchical Position:

 Materials Facet
... Materials
..... materials
..... <materials by composition>
..... inorganic material
..... <metal and metal products>
..... metal
..... <metal by composition or origin>
..... nonferrous metal
..... <mercury and amalgam>
..... mercury ←

Additional Parents:

 Materials Facet
... Materials
..... materials
..... <materials by form>
..... <materials by chemical form>
..... elements (chemical substances)
..... mercury ←

←

Figure D.3 An online display record of the AAT concept “Mercury”

Figure D.3 shows a screen captured from the *Art and Architecture Thesaurus* (AAT) online version. Hierarchical relationships of the *themas* represented by *nomens* “mercury”, “elements (chemical substances)”, and “nonferrous metal” are presented in the hierarchies. Such semantic relationships are illustrated in the following figure (Figure D.4).

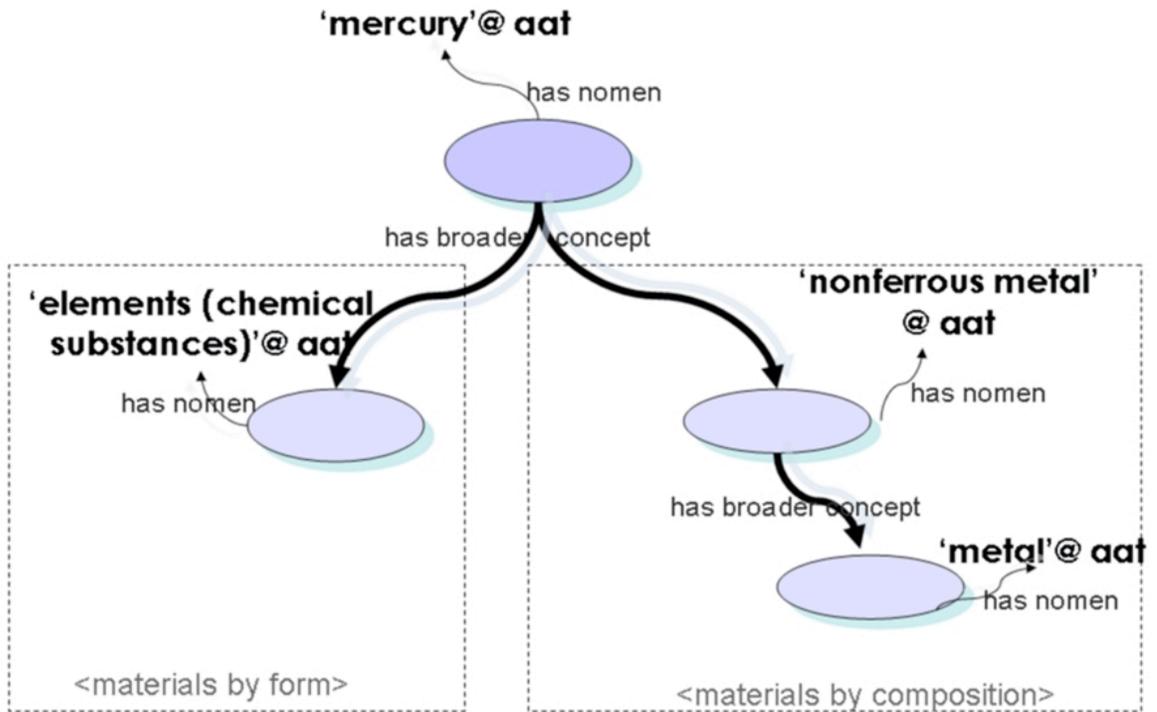


Figure D.4 Illustration of the semantic relations between the *themas* presented in Figure D.3

Example D.2.1.3. Medical Subject Headings (MeSH): Standard Display

Thema: **mercury (as a liquid metal and as an element):**

MeSH Heading	Mercury
Tree Number	D01.268.556.504
Tree Number	D01.268.956.437
Tree Number	D01.552.544.504
Inorganic Chemicals [D01]	
Elements [D01.268]	
Metals, Heavy [D01.268.556]	
	Mercury [D01.268.556.504] ←
Inorganic Chemicals [D01]	
Elements [D01.268]	
Transition Elements [D01.268.956]	
	Mercury [D01.268.956.437] ←
Inorganic Chemicals [D01]	
Metals [D01.552]	
Metals, Heavy [D01.552.544]	
	Mercury [D01.552.544.504] ←
See Also	Mercury Isotopes
See Also	Mercury Radioisotopes
See Also	Organomercury Compounds
Allowable Qualifiers	AD AE AG AI AN BL CF CH CL CT DF DU EC HI IM IP ME PD PH PK RE SD ST TO TU UR

Figure D.5 Extracted portion from a MeSH record indicating semantic relations

Figure D.5 shows data derived from a Standard Display of a MeSH record found through the MeSH Browser. It can be viewed from three segments:

a) The hierarchical relationships can be traced following the “Tree Numbers”. Analysis reveals two immediate hierarchical relationships (see Figure D.6; notational form of *nomens* are not included): (1) between *themas* represented by *nomens* “Mercury” and “Transition Elements”; (2) between *themas* represented by the *nomens* “Mercury” and “Metals, Heavy”. The latter can be traced up to two upper classes”.

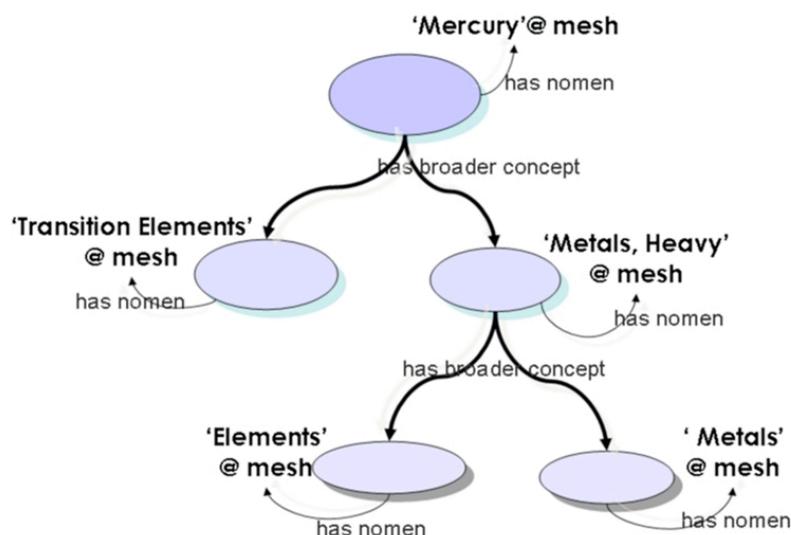


Figure D.6 Illustration of the hierarchical relationships (through the tree structure) from the extracted MeSH Heading record shown in Figure D.5

b) The information indicates that the *thema* represented by a *nomen*, “Mercury”, has associative relationships (“see also”) with *themas* represented by *nomens* “Mercury Isotopes”, “Mercury Radioisotopes”, and “Organomercury Compounds”, as illustrated in Figure D.7:

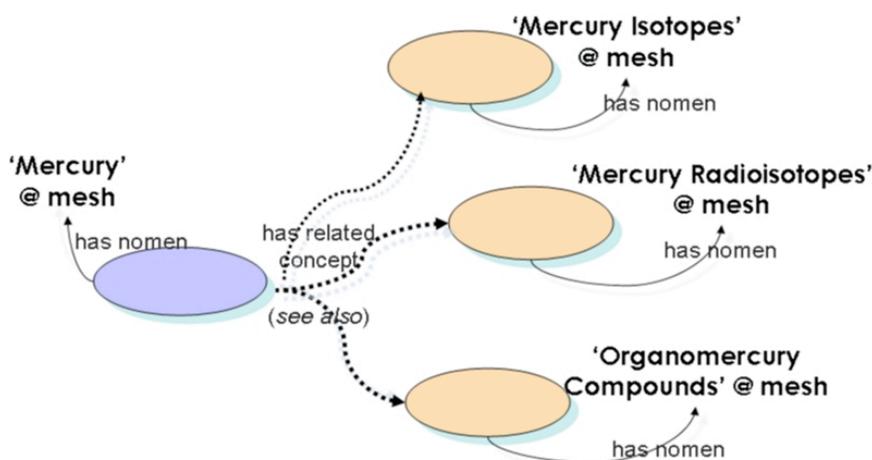


Figure D.7 Illustration of the associative relationships (“see also”) from the extracted MeSH record shown in Figure D.5

c) The MeSH record also provides allowable qualifiers to enable the forming of more complex concepts. In this example, the concept can be further limited to specific aspects: “administration & dosage (AD)”, “isolation & purification (IP)”, “toxicity (TO)”, etc. These facilitate the forming of specific subject headings (e.g., “Mercury – TO”, or “Mercury – IP”) to represent different *themas*.

Example D.2.1.4 . Dewey Decimal Classification

Thema: **mercury (as a metal)**

Class Number: 669.71
Segmented Number: 669/.71
Caption: Mercury

Main Classes

- 600 Technology
- 660 Chemical engineering
- 669 Metallurgy
- 669.1-669.7 Metallurgy of specific metals and their alloys
- 669.2-669.7 Nonferrous metals
- 669.7 Other nonferrous metals
- 669.71 **Mercury**

Figure D.8a. Screen captured from OCLC Connexion WebDewey for classes related to “mercury (as a metal)”

Thema: **mercury (as an element)**

Class Number: 546.663
Segmented Number: 546/.663
Caption: *Mercury

Main Classes

- 500 Science
- 540 Chemistry
- 541-547 Chemistry
- 546 Inorganic chemistry
- 546.6 Groups 8, 9, 10, 11, 12, 13, 14
- 546.66 Group 12
- 546.663 ***Mercury**
- 546.6635 Mercury (Element)--physical chemistry

Figure D.8b. Screen captured from OCLC Connexion WebDewey for classes related to “mercury (as an element)”

It should be noted that although the relationships are similar to what is presented in other thesauri (shown before), in a classification scheme such relationships are presented through the notational codes associated with *themas*, which reflect the conceptual hierarchy of a scheme. Hence it is the **notations** (669.71 and 546.663), not the **captions**, that represent the *themas*, as one can find from the above figures where both captions are “Mercury” although they are affiliated with two different classes in DDC. The two pairs

of hierarchical relationships are illustrated in the following figures: Figure D.9a is for *thema* “mercury as a metal” and Figure D.9b is for *thema* “mercury as an element”.

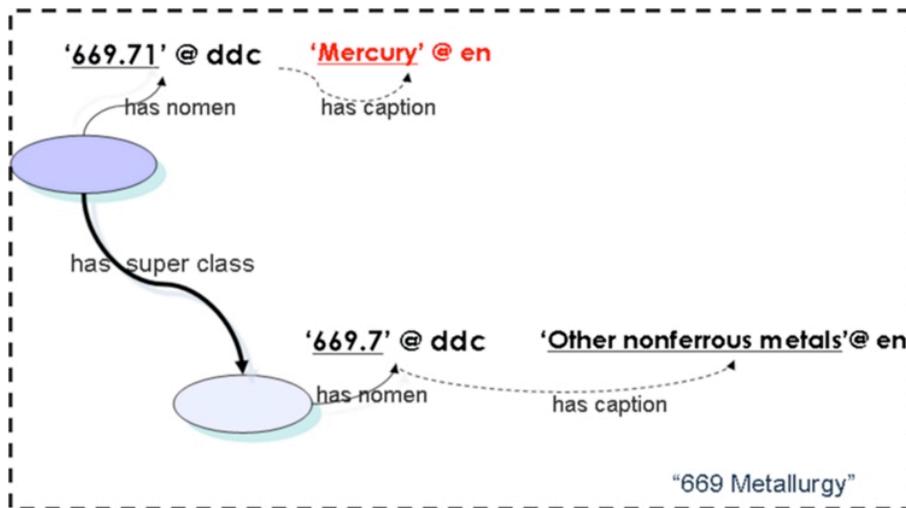


Figure D.9a Illustration of the hierarchical relationships (through the classificatory structure) between the DDC classes shown in Figure D.8a

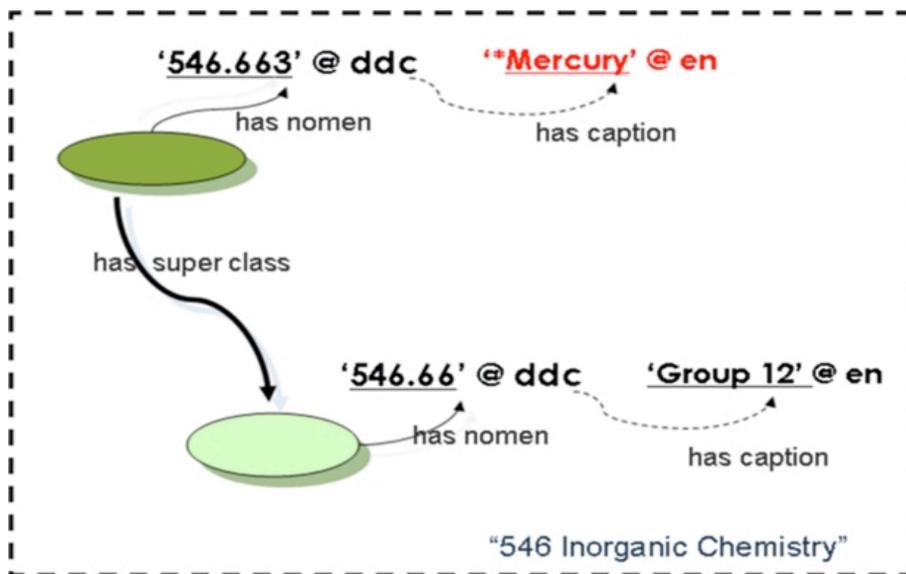


Figure D.9b Illustration of the hierarchical relationships (through the classificatory structure) between the DDC classes shown in Figure D.8b

D.2.2 Inter-system *THEMA* crosswalking through *NOMENS*

Example D.2.2.1 *INSPEC Thesaurus* and *INSPEC Classification*

Thema: mercury (planet)

Note: Although the term “Mercury” has multiple meanings and is a good example of homographs, the focus of this section is **not** on homograph control.

From *INSPEC Thesaurus* (2004, pg. h76):

[Note: CC= Classification Code]

	Mercury (planet)	←
	BT planets	
	TT planets	
	RT transits	
→	CC A9630D	
	DI January 1971	
	PT planets	

From *INSPEC Classification* (2004 pg. 84):

	A9630	Planets and satellites	
		<i>(exc. the Moon)</i>	
		<i>for Earth, see A91... for celestial mechanics, see A9510...</i>	
→	A9630D	Mercury	←

Figure D.10 Extracted entries from *INSPEC Thesaurus* (top) and *INSPEC Classification* (bottom) showing inter-system *thema* crosswalking

Example D.2.2.1 demonstrates that a *thema*, “planet Mercury”, can be crosswalked through the *nomens* in two different authority systems, where “Mercury (planet)” is a *nomen* (in a form of a thesaurus term) from the *INSPEC Thesaurus* and “A9630D” is a *nomen* (in a form of a notation in a classification) from the *INSPEC Classification*. This is illustrated in Figure D.11.

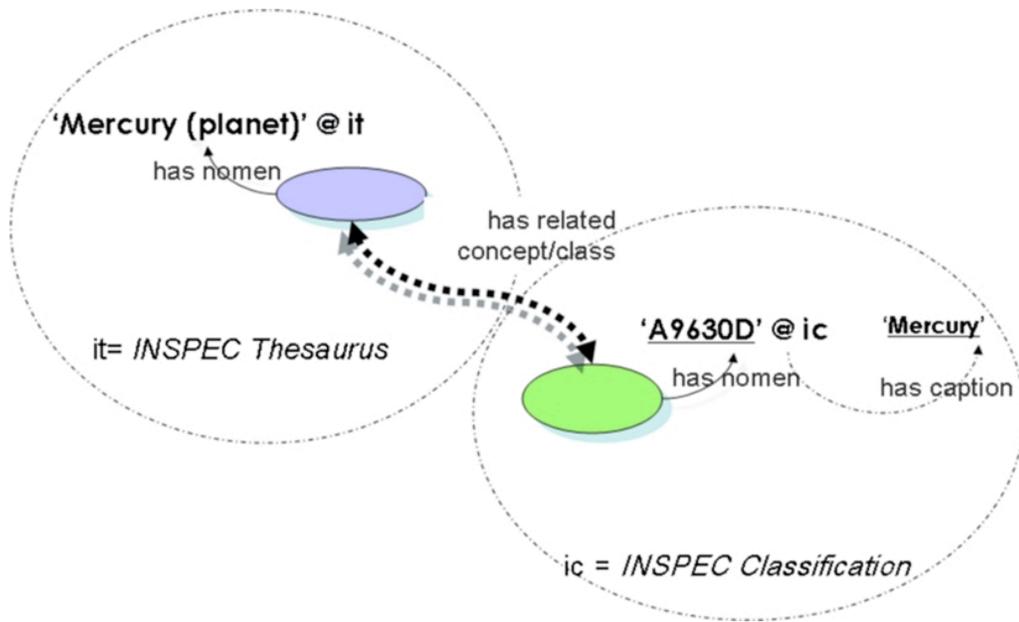


Figure D.11 Illustration of the inter-system *themas*' crosswalking between *INSPEC Thesaurus* and *INSPEC Classification* shown in Figure D.10

Example D.2.2.2. LCSH and *Library of Congress Classification* (LCC)

Thema: **“Mercury” (as a metal and an element)**

Example taken from Library of Congress Subject Authority File:

[Note: in the following entry, MARC21 coding is used:

010 = Library of Congress control number

040 = Cataloging source

053 = LC Classification Number

\$c = Explanatory term (specifying topic)

150 = Heading--Topical term

450 = *See* from Tracing--Topical term (unauthorized form/variant of term)

550 = *See Also* From Tracing--Topical Term;

\$a = Topical term or geographic name entry element

\$w = Control subfield; g = Broader term.]

This same record is also used in a previous section (D.2.1) when semantic relationships between *themas* from the same scheme are presented. In the following example, the relationships of *themas* from different schemes are further explored.

LC Control Number: sh 85083794

HEADING: Mercury

000 00558cz a2200217n 450

001 4734282

← 005 19900221112154.6

008 860211i| ananbabn |a ana

035 __ |a (DLC)sh 85083794

906 __ |t 8528 |u fk03 |v 0

010 __ |a sh 85083794

040 __ |a DLC |c DLC |d DLC

053 _0 |a QD181.H6 |c Chemistry

053 _0 |a TA480.M4 |c Engineering materials

053 _0 |a TN271.M4 |c Prospecting

053 _0 |a TP245.M5 |c Chemical technology

150 __ |a Mercury

450 __ |a Hydrargyrum

450 __ |a Quicksilver

550 __ |w g |a Liquid metals

953 __ |a xx00 |b fg07

[Note: in this captured screenshot, subfield signs are displayed as a vertical bar.]

Figure D.12. A record from the LC Subject Authority File

In this example, the *thema* “mercury” (as a metal and an element), represented by the *nomen* “Mercury” in LCSH, is crosswalked to the *Library of Congress Classification* (LCC) where the *thema* is placed in different classes that have the *nomens* “QD181.H6” (in Chemistry), “TA480.M4” (in Engineering materials), “TN271.M4” (in Prospecting), and “TP245.M5” (in Chemical technology). Figure D.13 illustrates such relationships.

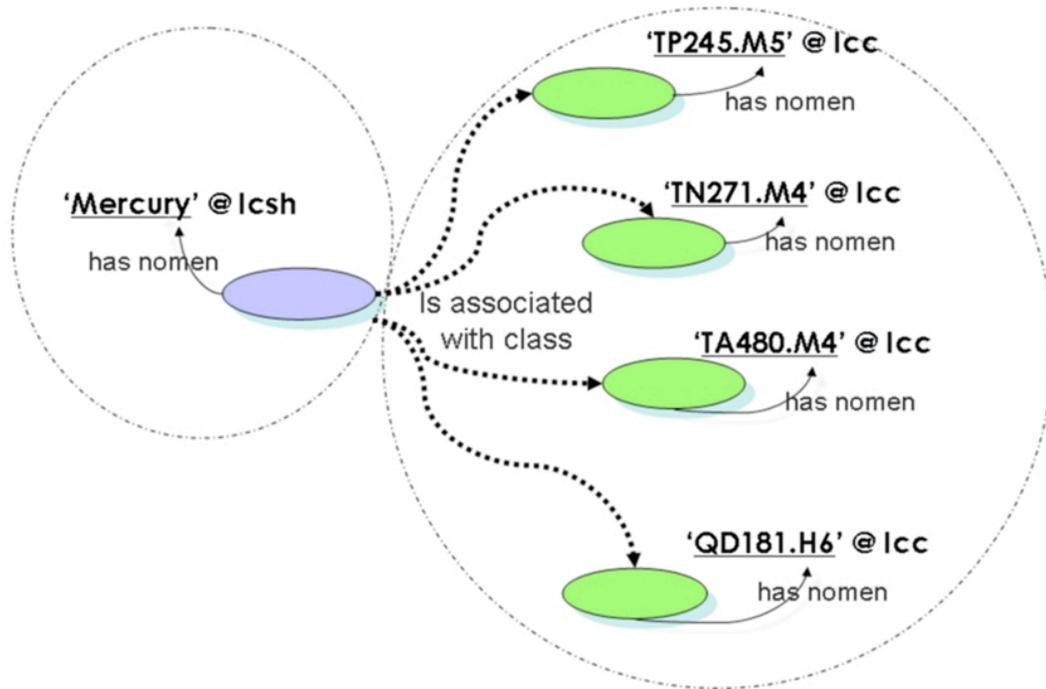


Figure D.13. Illustration of the inter-system *themas'* crosswalking between LCSH and LCC showing in Figure D.12

D.3 Same THEMA Represented by NOMENs from Different Schemes

The following case demonstrates that, to some extent, the granularity of a *thema* is also dependent on its appellations in a particular scheme.

For example, a resource is about “academic library labor unions in Germany”. The *thema* will be represented by the *nomens* established in different schemes such as:

DDC: “331.881102770943”

Constructed/combined from:

331.8811 – labor unions in industries and occupations other than
extractive, manufacturing, construction
-027.7 – academic libraries
-0943 – Germany

LCSH: “Library employees--Labor unions—Germany”

“Universities and colleges--Employees--Labor unions—Germany”

“Collective bargaining--Academic librarians--Germany”

“Libraries and labor unions--Germany”

FAST:

“Library employees--Labor unions”

“Universities and colleges--Employees--Labor unions”

“Collective bargaining--Academic librarians”

“Libraries and labor unions”

“Germany”

As this example demonstrates, schemes may allow the representation of *themas* at different levels of specificity through the structure and syntax of the *nomens* they have established.

D.4 Examples of Display Records from Controlled Vocabularies or Subject Authority Files

As shown in section D.2, authority **records** can be displayed differently within a particular system; furthermore, they can also have various combinations of authority **data** when displayed to different users (e.g., subject authority data creators and maintainers, metadata creators and end-users). Following are captured screens of records displayed online. They contain mixed information regarding *thema*, *nomen*, relationships between a *thema* and its *nomens*, as well as among different *themas*. In addition, they demonstrate that *thema* types are implementation-dependent and vary in different domains.

Example D.4.1. A chemical substance and its *NOMEN* -- A display record from *The USP Dictionary of U.S. Adopted Names and International Drug Names*

The figure below demonstrates how a *thema* could have various *nomens* in the context of specific systems. The forms of the *nomens* for this chemical compound are not only in various names represented in natural language, but also those represented in artificial languages such as codes, formulas and a graph.

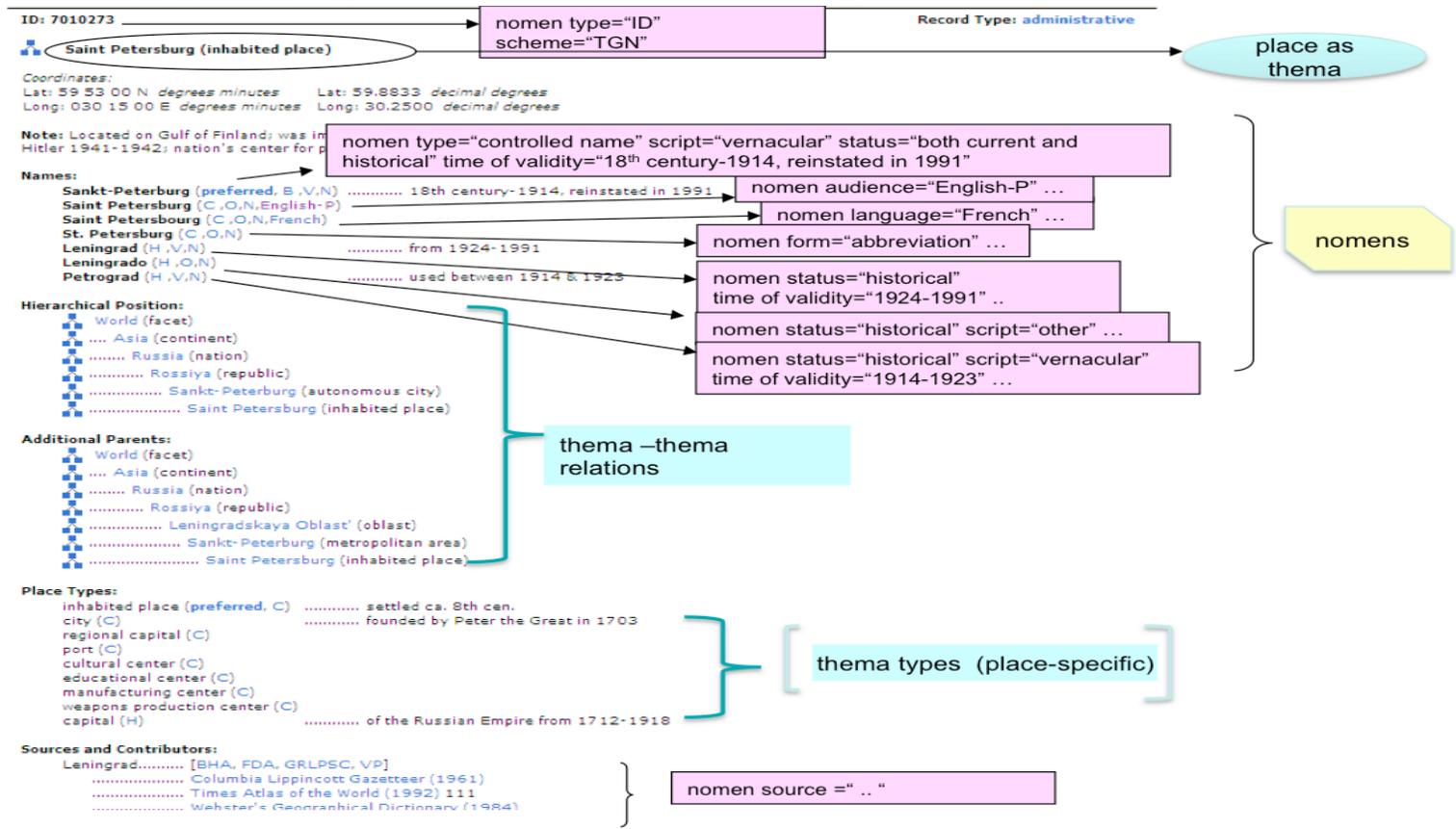
different types of nomens

Accession Number	(AN): 2005:3738	USAN
Publication Year	(PY): 2002	
Generic Name	(CN): Flindokalner	
OTHER NAMES:		
Chemical Name	(CN): 2H-Indol-2-one, 3-(5-chloro-2-methoxyphenyl)-3-fluoro-1,3-dihydro-6-(trifluoromethyl)-, (3S)-	
Chemical Name	(CN): (3S)-3-(5-Chloro-2-methoxyphenyl)-3-fluoro-6-(trifluoromethyl)-1,3-dihydro-2H-indol-2-one	
Trade Name	(CN): MaxiPost (Bristol-Myers Squibb)	
Code Designation	(CN): BMS-204352	
CAS Registry No.	(RN): 187523-35-9	
Molecular Formula	(MF): C ₁₆ H ₁₀ Cl F ₄ N O ₂	
Lin. Str. Formula	(LSF): C ₁₆ H ₁₀ Cl F ₄ N O ₂	
Molecular Weight	(MW): 359.71	
Absolute stereochemistry. Rotation (+).		

Source: STN Database Summary Sheet: USAN (The USP Dictionary of U.S. Adopted Names and International Drug Names)
<http://www.cas.org/ASSETS/773D56DEC03E4769BF0E1BC206BB371E/usan.pdf>, p.5. Record reprinted with permission.

Example D.4.2. A place as a thema – A display record from *Getty Thesaurus of Geographic Names (TGN)*

This example presents: (1) the hierarchical relationships of a *thema* (in this case a place) with other *themas*, i.e., the “whole-part” relationships; (2) various *nomens*, to be chosen as preferred terms in different contexts, with attributes regarding the form, time of validity, status, audience, and source of a particular *nomen*; and (3) *thema* types that are place-specific.



Source: *Getty Thesaurus of Geographic Names Online*. http://www.getty.edu/research/conducting_research/vocabularies/tgn/
Record reprinted with permission.

Example D.4.3. A display record (Extensive Concept View) from *Medical Subject Headings* (MeSH)

Thema-thema relationships presented in the *Medical Subject Headings* (MeSH) have been explained in a previous section with Example D.2.1.3 and Figure D.6 and D.7. The following Expanded Concept View displays an additional component for “Concept 1: Mercury.” The summary of the semantic relationships displayed in this record is presented below the figure.

MeSH Heading	Mercury	→	nomen type="constructed name" Scheme="MeSH"	
Tree Number	D01.268.556.504	}	nomens, used to represent hierarchical relations between themas in tree structures	
Tree Number	D01.268.956.437			
Tree Number	D01.552.544.504			
Annotation	Hg-202	→	nomen type="constructed name" form="code" Scheme="..."	
Concept 1 (Preferred)	Mercury	→	thema represented by nomen	
	Concept UI	M0013448	→	nomen type="ID" extent="concept" Scheme="UMLS"
thema	Scope Note	A silver metallic element that exists as a liquid at room temperature. It has the atomic symbol Hg (from hydrargyrum, liquid silver), atomic number 80, and atomic weight 200.59. Mercury is used in many industrial applications and its salts have been employed therapeutically as purgatives, antisyphilitics, disinfectants, and astringents. It can be absorbed through the skin and mucous membranes which leads to MERCURY POISONING . Because of its toxicity, the clinical use of mercury and mercurials is diminishing.		
	Semantic Type	T131 (Hazardous or Poisonous Substance)	} themata - themata relations	
Semantic Type	T196 (Element, Ion, or Isotope)			
CAS Type 1 Name	Mercury	→	nomen type="constructed name" Scheme="UMLS"	
Registry Number	7439-97-6	→	nomen type="ID" Scheme="UMLS"	
Term (Preferred)	Mercury	→	nomen type="constructed name" Scheme="NLM(1966)"	
	Term UI	T025687	→	nomen type="ID" Scheme="NLM(1966)"
	Date	01-JAN-1999	→	nomen time of validity = "19990101"
	Lexical Tag	NON	→	nomen transcription/transliteration="none"
	Thesaurus	NLM (1966)	→	nomen source="NLM(1966)"
See Also	Mercury Isotopes	}	thema - themata relations	
See Also	Mercury Radioisotopes			
See Also	Organomercury Compounds			
Allowable Qualifiers	AD AE AG AI AN BL CF CH CL CT DF DU EC HI IM IP ME PD PH PK RE SD ST TO TU UR			
Entry Combination	poisoning:Mercury Poisoning	→	nomen-nomen relation	
Date of Entry	19990101	→	nomen time of validity="19990101"	
Unique ID	D008628			

Source: Medical Subject Headings on MeSH Browser (2008 MeSH); <http://www.nlm.nih.gov/mesh/2008/MBrowser.html>

This Expanded Concept View presents various types of semantic relationships among *themas*:

- a) Two immediate hierarchical relationships: (1) between *themas* represented by *nomens* “Mercury” and “Transition Elements”. The same is true for these *themas* and their *nomens* with notational forms; (2) between *themas* represented by the *nomens* “Mercury” and “Metals, Heavy”. The latter can be traced up to two upper classes.
- b) Associative relationships between “Mercury” (as a liquid metal and as an element) and other *themas* represented by *nomens* “Mercury Isotopes”, “Mercury Radioisotopes”, and “Organomercury Compounds”.
- c) Allowable qualifiers enable the concept to be further limited to specific perspectives (e.g., “administration & dosage (AD)”, “isolation & purification (IP)”, and “toxicity (TO)”). These facilitate the forming of specific subject headings (e.g., “Mercury – TO”, or “Mercury – IP”) to represent different *themas*.
- d) The semantic types of this *thema*: “T131 (Hazardous or Poisonous Substance)” and “T196 (Element, Ion, or Isotope)” as defined by UMLS.

Thema-nomen relationships are clearly presented in the record, including the *nomens* in natural languages and as specific identification numbers. Various attributes of *nomens* are also presented.

REFERENCES

Aitchison, J., Gilchrist, A. and Bawden, D. (2000). *Thesaurus Construction and Use: A Practical Manual*. 4th ed. London: Fitzroy Dearborn.

Art and Architecture Thesaurus. (2000-). Los Angeles: J. Paul Getty Trust, Vocabulary Program. Available at:http://www.getty.edu/research/conducting_research/vocabularies/aat/ (accessed 2010-01-20).

Buizza, P. and Guerrini, M. 2002. A conceptual model for the New Soggetario: Subject indexing in the light of FRBR. *Cataloging & Classification Quarterly*, 34(4): 31–45.

Campbell, K. E., Oliver, D.E., Spackman, K.A. and Shortliffe, E.H. (1998). Representing thoughts, words, and things in the UMLS. *Journal of the American Medical Informatics Association*, 5(5): 421–431.

Categories for the Description of Works of Art (CDWA). (2000). Eds. Baca, M. and Harpring, P. The J. Paul Getty Trust and College Art Association. Los Angeles, CA: Getty Research Institute.

Available at:

http://www.getty.edu/research/conducting_research/standards/cdwa/index.html (accessed 2010-01-20)

Clarke, S.G. (2001). Thesaural relationships. In: *Relationships in Knowledge Organization*. Eds. Bean, C.A. and Green, R. Dordrecht: Kluwer.

Dahlberg, I. (1992). Knowledge organization and terminology: philosophical and linguistic bases. *International Classification*. 19(2):65-71.

DCMI Abstract Model. (2007). Eds. Powell, A., Nilsson, M., Naeve, A. Johnston, P. and Baker, T. Available at: <http://dublincore.org/documents/abstract-model/> (accessed 2010-01-20).

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

Frege, G. (1892). Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, NF 100. 1892, S. 25–50. Available at: <http://www.gavagai.de/HHP31.htm> (accessed 2009-05-22).

Functional Requirements for Authority Data - A Conceptual Model. (2009). IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR), ed. by Glenn E. Patton. München: K.G. Saur.

- Functional Requirements for Bibliographic Records: Final Report*. (1998). IFLA Study Group on the Functional Requirements for Bibliographic Records. München: KG Saur. Available at: <http://www.ifla.org/files/cataloguing/frbr/frbr.pdf> (accessed 2010-01-20).
- Furner, J. (2006). The ontology of subjects of works. Paper presented at ASIS&T 2006: Annual Meeting of the American Society for Information Science and Technology, Austin, TX, November 3–8, 2006.
- Goodman, N. (1961). About. *Mind*, 70 (277): 1-24.
- Heaney, M. (1997). Time is of the essence. Available at: <http://www.bodley.ox.ac.uk/users/mh/time978a.htm> (accessed 2010-01-20).
- Hill, L. (1999). Content standards for digital gazetteers. Presentation at the *JCDL2002 NKOS Workshop "Digital gazetteers--Integration into distributed digital library services"*, July 18, 2002, Portland, Oregon. Available at: <http://nkos.slis.kent.edu/DL02workshop.htm> (accessed 2010-01-20).
- Hjørland, B. (1992). The concept of “subject” in information science. *Journal of Documentation*, 48 (2): 172-200.
- Hutchins, W. J. (1977). On the problem of “aboutness” in document analysis. *Journal of Informatics*, 1 (1): 17-35.
- ISO (2009). *ISO/CD 25964-1, Information and documentation — Thesauri and interoperability with other vocabularies — Part 1: Thesauri for information retrieval*. ISO/TC 46 / SC 9 ISO 25964 Working Group.
- Lancaster, F.W. (1986). *Vocabulary Control for Information Retrieval*. 2nd ed. Arlington, Virginia: Information Resources Press.
- Michel, D. (1996). Taxonomy of Subject Relationships. Appendix B (Part 2), Final Report to the ALCTS/CCS Subject Analysis Committee, submitted by Subcommittee on Subject Relationships/Reference Structures. ALA Association for Library Collections and Technical Services (ALCTS) Cataloging and Classification Section (CCS) Subject Analysis Committee (SAC). Available at: <http://www.ala.org/ala/mgrps/divs/alcts/mgrps/ccs/cmtes/subjectanalysis/subjectrelations/mrsrscu2.pdf>. Also available is the hierarchical display of this Taxonomy, at: <http://www.ala.org/ala/mgrps/divs/alcts/mgrps/ccs/cmtes/sac/inact/subjectrelations/appen dixbpartii.cfm> (accessed 2010-03-31).
- National Library of Medicine. (2003-). *Unified Medical Language System*. Available at: <http://www.nlm.nih.gov/research/umls/> (accessed 2010-01-20).
- NISO. (2005). ANSI/NISO Z39.19-2005. *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*. Bethesda, Maryland: NISO Press.

Ogden, C. K., and 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.

OWL 2 Web Ontology Language Structural Specification and Functional-Style Syntax. (2009). Eds. Motik, B, Patel-Schneider, P.F. and Parsia, B. W3C Working Draft 21 April 2009. Available at: <http://www.w3.org/TR/owl2-syntax/> (accessed 2010-01-20).

OWL 2 Web Ontology Language New Features and Rationale. (2009). Eds. Golbreich, C. and Wallace, Evan K. W3C Working Draft 21 April 2009. Available at: <http://www.w3.org/TR/owl2-new-features/> (accessed 2010-01-20).

Putnam, H. (1958). Formalization of the concept “about.” *Philosophy of Science*, 25 (2): 125-130.

Rust, G. and Bide, M. (2000). The <indecs> metadata framework: Principles, Model and Data Dictionary. Version2. Indecs Framework Ltd. Available at: http://www.doi.org/topics/indecs/indecs_framework_2000.pdf (accessed 2010-01-20).

Ryle, G. (1933). “About.” *Analysis*, 1 (1): 10-12.

Shatford, S. (1986). Analyzing the subject of a picture: A theoretical approach. *Cataloging & Classification Quarterly*, 6 (3): 39-62.

SKOS Simple Knowledge Organization System Reference (2009). Eds. Miles, A. and Bechhofer, S. W3C Candidate Recommendation 17 March 2009. Available at: <http://www.w3.org/TR/skos-reference/> (accessed 2010-01-20).

Subcommittee on Subject Relationships/Reference Structures. (1997). Final Report to the ALCTS/CCS Subject Analysis Committee. ALA Association for Library Collections and Technical Services (ALCTS) Cataloging and Classification Section (CCS) Subject Analysis Committee (SAC). Available at: <http://www.ala.org/ala/mgrps/divs/alcts/mgrps/ccs/cmtes/sac/inact/subjectrelations/finalreport.cfm> (accessed 2010-01-20).

Svenonius, E. (2000). *The Intellectual Foundation of Information Organization*. Cambridge, Mass.: MIT Press.

Wilson, P. (1968). *Two kinds of power: An essay on bibliographic control*. Berkeley, CA: University of California Press.