Functional Requirements for Subject Authority Data (FRSAD)

A Conceptual Model

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

1st Draft 2008-11-27
2nd Draft 2009-06-10
TABLE OF CONTENTS

MEMBERS OF THE IFLA WORKING GROUP ON THE FUNCTIONAL REQUIREMENTS FOR SUBJECT AUTHORITY RECORDS (FRSAR) ......................................................... 3

1. INTRODUCTION ............................................................................................................. 5

2. PURPOSE AND SCOPE ............................................................................................... 6
  2.1 TERMS OF REFERENCE ......................................................................................... 6
  2.2 ABOUTNESS ........................................................................................................... 7
  2.3 ABOUTNESS AND OPNESS ................................................................................. 10
  2.4 THE FRSAR WORKING GROUP .......................................................................... 10

3. USER TASKS .................................................................................................................... 11
  3.1 USERS AND USE ................................................................................................. 11
  3.2 USER STUDIES ...................................................................................................... 12
  3.3 USER TASKS ......................................................................................................... 16

4. METHODOLOGY .......................................................................................................... 19
  4.1 SUBJECT RELATIONSHIP AND GROUP 3 ENTITIES INTRODUCED IN FRBR .......... 19
  4.2 POSSIBLE APPROACHES TO THE MODEL OF ABOUTNESS ......................... 20

5. ENTITIES ...................................................................................................................................... 25
  5.1 GENERAL FRAMEWORK ................................................................................... 25
  5.2 CHOICE OF TERMS FOR FRSAD ENTITIES .................................................. 26
  5.3 THEMATICA ........................................................................................................... 26
  5.4 NOMEN ................................................................................................................ 28
  5.5 THE IMPORTANCE OF THEMATICA-NOMEN MODEL ..................................... 29

6. ATTRIBUTES .................................................................................................................. 31
  6.1 ATTRIBUTES OF THEMATICA ......................................................................... 31
  6.2 ATTRIBUTES OF NOMEN .................................................................................. 33

7. RELATIONSHIPS .......................................................................................................... 37
  7.1 WORK-TO-THEMA RELATIONSHIPS ............................................................... 37
  7.2 THEMA-TO-NOMEN RELATIONSHIPS ............................................................. 37
  7.3 THEMA-TO-THEMA RELATIONSHIPS ............................................................... 38
  7.4 NOMEN-TO-NOMEN RELATIONSHIPS ............................................................... 42

8. MAPPING OF USER TASKS, ATTRIBUTES, AND RELATIONSHIPS ..................... 44
  8.1 SUBTASKS ............................................................................................................ 44
  8.2 MAPPING OF ATTRIBUTES, RELATIONSHIPS AND USER TASKS .................. 46

9. SUBJECT AUTHORITY SYSTEMS IMPLEMENTATION EXAMPLES ..................... 47
  9.1 EXISTING MODELS OF THEMATICA TYPES ............................................... 47
  9.2 THEMATICA-THEMA RELATIONSHIPS PRESENTED IN SUBJECT VOCABULARY/AUTHORITY DATA 51
  9.3 SAME THEMATICA REPRESENTED BY NOMENS FROM DIFFERENT SCHEMES .... 63
  9.4 EXAMPLES OF DISPLAY RECORDS FROM SUBJECT VOCABULARIES/authority FILES .............. 63
  9.5 MAPPING FRSAD MODEL TO OTHER ABSTRACT MODELS ............................ 67
  9.6 CONCLUSION ........................................................................................................... 69
MEMBERS OF THE IFLA WORKING GROUP ON THE FUNCTIONAL REQUIREMENTS FOR SUBJECT AUTHORITY RECORDS (FRSAR)

Working Group (WG)
- Leda Bultrini, Italy
- Lois Mai Chan, USA
- Jonathan Furner, USA
- Edward O’Neill, USA (Liaison to OCLC FRBR and to ALA SAC)
- Gerhard Riesthuis, The Netherlands
- Athena Salaba, USA, Co-Chair, secretary
- Diane Vizine-Goetz, USA
- Ekaterina Zaytseva, Russia
- Marcia Lei Zeng, USA, Chair
- Maja Zumer, Slovenia, Co-Chair

Advisory Group (AG)
- Victoria Francu, Romania
- Hemalata Iyer, USA
- Dorothy McGarry, USA
- David Miller, USA (Liaison to ALA SAC (Subject Analysis Committee))
- Päivi Pekkarinen, Finland
- Barbara Tillett, USA

User Tasks Subgroup members:
- Athena Salaba (leader); Leda Bultrini, Lois Chan, Gerhard Riesthuis, David Miller, Dorothy McGarry, Marcia Zeng.

Subject Entities Subgroup members:
This document was prepared by:
1. INTRODUCTION

Subject access to information has been the predominant approach of users to satisfy their information needs. Research demonstrates 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 subject term—in the elements used as access points in information retrieval. For example, “World War, 1939-1945” has been established as an authorized subject heading in the Library of Congress Subject Headings (LCSH). When using LCSH, in cataloging 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 are referred to by the authorized heading. This ensures that all publications about World War II can be retrieved by and displayed under the same subject heading, either in an individual institution’s own catalog or database or in a union catalog that contains bibliographic records from a number of individual libraries or databases.

In almost all large bibliographic databases, authority control is achieved manually or semi-automatically by means of an authority file. The file contains records of headings or access points—names, titles, or subjects—that have been authorized for use in bibliographic records. In addition to ensuring consistency in subject representation, a subject authority record also records and maintains semantic relationships among subject terms and/or their labels. Records in a subject authority file are connected through semantic relationships, which may be expressed statically in subject authority records or generated dynamically according to the specific needs (e.g., presenting the broader and narrower terms) of printed or online display of thesauri, subject headings lists, classification schemes, and other knowledge organization systems.
2. PURPOSE AND SCOPE

2.1 Terms of Reference

In most structured retrieval systems, information relating to the bibliographic universe is not recorded exclusively in bibliographic records. Authority records are used to record information about controlled access points that have been or may be assigned to bibliographic records. IFLA working groups have addressed the functional requirements for bibliographic records in *Functional Requirements for Bibliographic Records* (FRBR), and the functional requirements for authority records in *Functional Requirements for Authority Data* (FRAD). The FRBR working group focused on the development of a conceptual model showing the entity-relationships of the bibliographic universe. The purpose of the FRBR model was to identify the functional requirements of information in bibliographic records to facilitate the defined 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 endeavors: *work, expression, manifestation, and item*;

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

- **Group 3 entities** are the subjects of works, intellectual or artistic endeavor.

Controlled access points include names of entities identified by FRBR such as members of Group 2 (persons, corporate bodies), titles of Group 1 entities (works, expressions, manifestations and items), and terms for Group 3 entities. In the FRBR model the entities of all three groups are defined, but the main focus is on the first group. The FRBR final report presented the model, identified entities and their attributes, and defined relationships among and across entities\(^2\).

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 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 records.

---


\(^2\) Ibid.
citations or records in a library catalogue or bibliographic file\(^3\). The primary purpose of the FRAD conceptual model is "to provide an analytical 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 it may be packaged (e.g., in authority records)." \(^4\)

While the FRANAR Working Group has included some aspects of subject authorities in the authorities model, it has not undertaken the full analysis that the FRBR Study Group envisioned for the entities that are the centre of focus for subject authorities, thesauri, and classification schemes, and of the relationships between those entities. \(^5\) As a result, the Working Group on 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 role of FRSAR was defined in the following terms of reference:

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

It should be noted that FRSAD would apply only to works with subjects, because some works may have no subject.

Some controlled vocabularies provide terminology to express other aspects of works such as form, genre, and target audience of resources. These form/genre aspects describe what the item is rather than what it is about. The form/genre aspects are beyond the charge of FRSAR and are not covered in this model.

### 2.2 Aboutness

Any attempt to model the fundamental classes of bibliographic entities necessarily faces 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


\(^4\) Ibid., p. 1.

\(^5\) Ibid., p. iii.
in the literature of library and information science (LIS), while others have been made by philosophers of logic and language.

The LIS author who has engaged most productively with philosophical analyses of aboutness is Patrick Wilson, in Chapter V, “Subjects and the sense of position,” of his Two kinds of power. Wilson's general conclusion is that “[t]he notion of the subject of a writing is indeterminate.” Wilson contrasts works (which he calls here “writings”) with physical objects that are “determinate in every respect,” and that “must have some definite shape and size and so on, at any moment,” whether or not we are able to discover what the values of those variables are. For Wilson, works are not like physical objects, and the subjects of works are not like the shapes or sizes of physical objects. Not only is it the case that different methods of determining the subjects of a work produce different results: there is no way in principle of deciding which of two equally specific or equally exhaustive subject statements is the “correct” statement, because (for Wilson) works do not “have” subjects in the same way that physical objects have shapes. Wilson claims that “being on a given subject” contains a “quasi-technical term which is nowhere explained” in libraries or in the LIS literature. The claim is that, “to the librarian, ‘being on a given subject’ means [nothing more nor less than] ‘being the sort of writing which our methods of assigning single locations assign to the position with such and such a name’.”

Wilson's position exemplifies a family of views of the nature of aboutness that may be placed at one of the two ends of a spectrum of such kinds of view. At this pole—which might conveniently be called “nominalist” or (following Hjørland) “idealist”—we may locate views that comprise some or all of the following component claims:

- aboutness cannot sensibly be conceived as a property of works, but rather as a relation between sets of works, subjects, agents, and dates;

- what we call subjects are merely linguistic expressions that serve as labels or names for sets of works or for positions within a sequence or hierarchical structure;

- it makes no sense to speak of works “having” subjects, or of “the” subject of a

---

12 Wilson, op. cit.
13 Ibid., p. 89.
14 Ibid., p. 90.
15 Ibid., p. 92.
16 Ibid., p. 92.
17 Hjørland, op. cit.
work (unless the intention is to designate the expression that happens to be attributed to the work by a particular agent on a particular date); and

- it is never possible to determine the truth of the sentence “work \( w \) is about subject \( z \)” on any objective basis.

Conversely, at the opposite pole—which (again following Hjørland\(^{18}\)) we might call “realist”—lie views made up of some or all of the following elements:

- aboutness is a property class whose instances are predicates of classes of works;
- what we call subjects are the things designated by the linguistic expressions that comprise subject statements: such things may be concrete (existing in space-time) or abstract, and they may be particulars (not instantiable) or universals;
- we may speak sensibly of works “having” subjects, and of “the” subject(s) of any given work; and
- it is possible to determine the truth of the proposition that work \( w \) is about subject \( z \) objectively, by specifying a regular procedure by which work \( w \) may be analyzed in order to discover its subject.

The fact that the nature of aboutness continues to be the subject of such debate in the field is a result not primarily of the inability of proponents of views at the two poles to persuade their opponents of the merits of those views, but rather of the largely unacknowledged influence of the realist view on the activity of designers and users of knowledge organization systems. It is difficult to find well-reasoned defenses of the realist view in the literature, yet 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, continue to act as if the realist view is the correct one.

Ultimately, the FRSAR WG does not take a philosophical position on the nature of aboutness, but rather 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 catalogers and indexers.

\(^{18}\) Hjørland, op. cit.
2.3 Aboutness and Ofness

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) of such works. 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 “about” the passage of time. Standard metadata schemata for cultural objects correspondingly allow for distinctions to be made between (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. The FRSAR WG recognizes that any statement of the "ofness" of a work is just as much a subject statement as any statement of the work’s "aboutness", and is just as likely (if not more so) to be the target of a catalog user’s search. It is in this sense that "ofness" is included in this conceptual model.

2.4 The FRSAR Working Group

The FRSAR Working Group established two sub-groups: User Tasks and Subject Entities. The first focused on user studies and the definition of user tasks. The second 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 (with the relationship WORK HAS AS SUBJECT);

(b) possible sub-entities in the Group 3 cluster; and

(c) additional entities related to the Group 3 cluster.

---


3. USER TASKS

3.1 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 metadata, b) reference and public services librarians and other information professionals who search for information as intermediaries, c) controlled vocabulary creators, such as catalogers and thesaurus and ontology creators; and d) end-users who use information retrieval systems to fulfill their information needs. Figure 3.1 illustrates the three-point perspective of subject authority data users.

![Figure 3.1 Users of subject authority data](image)

In her study of professional searchers, Fidel\(^\text{21}\) identified strategies for selecting terms from the indexing language using semantic relationships such as broader, narrower and synonymous terms. Of all the cases of term selection, 75% percent included indexing language consultation by the searchers. Researchers have examined end-user queries when searching online catalogs and the use of authority data displayed to users as sources for term selection.\(^\text{22}\) A number of more

---


recent studies have focused on the effects of thesaurus-enhanced information retrieval systems and user interactions with and understanding of indexing language information (including behaviors in term selection) and found that indexing languages benefit users in finding terms for query expansion and in improving search performance.  

3.2 User Studies

The FRSAR Working Group felt strongly that, in order to define user tasks, it was necessary to conduct one or more user studies. Two studies were conducted. The first was a pilot study at the 2006 Semantic Technologies Conference (San Jose, California, USA). Most participants in this study were either creators of semantic tools (including controlled vocabularies, taxonomies, and ontologies) or developers and managers of semantic technology systems. Participants were asked to describe their work and the ways they use semantic tool data. Table 3.1 provides a summary of participant responses.

<table>
<thead>
<tr>
<th>In what ways do you use the semantic tools?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>(ranked according the number of answers)</strong></td>
</tr>
<tr>
<td>1   to find/identify relevant information resources using specified terms</td>
</tr>
<tr>
<td>2   to access metadata information</td>
</tr>
<tr>
<td>3   to find/identify appropriate terms when searching for information</td>
</tr>
<tr>
<td>3   to navigate or modify search queries through broader and narrower terms</td>
</tr>
<tr>
<td>4   to assign subject terms and/or class labels in metadata records</td>
</tr>
<tr>
<td>5   to explore a topic through browsing synonymous and related terms</td>
</tr>
<tr>
<td>6   other:</td>
</tr>
<tr>
<td>-To discover relationships.</td>
</tr>
<tr>
<td>-It is called systems engineering.</td>
</tr>
<tr>
<td>-Growing and operating business.</td>
</tr>
<tr>
<td>-Software development.</td>
</tr>
<tr>
<td>-Data integration/sys. eng.</td>
</tr>
<tr>
<td>-Creating tools to process and use ontologies for data integration.</td>
</tr>
<tr>
<td>-Research</td>
</tr>
</tbody>
</table>

For the second study, a survey was sent to information professionals throughout the world during the months of May-September 2006. Participants included authority record creators, vocabulary creators and managers, catalogers, metadata librarians, and reference librarians among others. Participants were asked to describe their work and their use of subject authority data in different contexts, including cataloging/metadata creation, subject authority work, and searching or helping others search bibliographic information. The results of these studies enriched the working group’s understanding of how subject authority data is used and informed and provided further confirmation of the FRSAD user tasks.

Of the 798 participants, the largest group represented professionals working in the creation of metadata and cataloging records (79%), followed by those involved with quality control of metadata and cataloging projects (60%), supervising these types of projects (46%), and creating and maintaining subject authority records (43%). About 30% of all participants assist others in finding information (reference work). For more details see Figure 3.2 below.
When these information professionals were asked how they use subject authority data in cataloging and metadata creation, about 95% of respondents indicated that they use the data to select and verify terms for cataloging and indexing. Other uses include (see Figure 3.3) exploring (69%), verifying, and understanding relationships among subject terms (72%).
Among the participants who use subject authority data in subject authority work, a majority use the data to normalize and standardize terms (87%). They also modify authority data (62%), maintain data for future use (58%), and establish and update term relationships (53%) as represented in Figure 3.4.

Finally, Figure 3.5 shows that of participants who use subject authority data in searching bibliographic information for their own needs or when helping others (end-users) to find bibliographic information, the majority use the data available to select and verify appropriate terms for use in search queries (87%). Other uses include finding relevant documents on a specified topic (77%), modifying search queries by utilizing semantic relationships (65%), navigating bibliographic information by using semantic relationships (64%), and exploring and understanding subject areas and the relationships of domain subject terminology (62%).
3.3 User Tasks

Based on the results from the two user studies, four tasks for subject authority data have been defined as follows:

**Find:** find an entity (the subject itself or its representation) or set of entities that correspond(s) to user’s stated criteria (i.e., to locate a single entity or a set of entities using an attribute, set of attributes, or relationships between entities)

**Identify:** identify an entity (the subject itself or its representation) based on specific attributes or characteristics (i.e., to confirm that the entity found corresponds to the entity sought, or to distinguish between two or more entities with similar characteristics)

**Select:** select an entity (the subject itself or its representation) appropriate to a user’s needs (i.e., to choose an entity that meets the user’s requirements or to reject an entity as being inappropriate to the user’s needs)
Explore: explore relationships between entities (subjects or their representations) (i.e., to explore relationships in order to understand the structure of a subject domain and its terminology)

Examples of the above user tasks are illustrated in chapter 8, section 8.1.

Figure 3.6 shows a comparison of user tasks as defined in FRBR, FRAD and FRSAD.

<table>
<thead>
<tr>
<th>FRBR</th>
<th>FRAD</th>
<th>FRSAD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Find</strong> entities of Group 1 that have entities from Group 1, 2, 3 as their subject</td>
<td><strong>Find</strong> one entity or entities from Group 1</td>
<td><strong>Find</strong> an entity or entities from Group 1</td>
</tr>
<tr>
<td><strong>Identify</strong></td>
<td></td>
<td><strong>Identify</strong></td>
</tr>
<tr>
<td><strong>Select</strong></td>
<td></td>
<td><strong>Select</strong></td>
</tr>
<tr>
<td><strong>Obtain</strong></td>
<td></td>
<td><strong>Obtain</strong></td>
</tr>
<tr>
<td><strong>Contextualize</strong></td>
<td></td>
<td><strong>Contextualize</strong></td>
</tr>
<tr>
<td><strong>Justify</strong></td>
<td></td>
<td><strong>Justify</strong></td>
</tr>
</tbody>
</table>

Figure 3.6 Comparison of user tasks as defined in FRBR, FRAD and FRSAD

The *explore* user task is a new task introduced in FRSAD; whereas the *find*, *identify*, and *select* user tasks have been previously introduced in FRBR and/or FRAD conceptual models. As demonstrated by the subject authority data use survey, a large number of participants (69%) use subject authority data to explore relationships among terms during cataloging 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 and metadata records. This reflects 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.

Chapter 8 offers a more detailed analysis of subtasks involved in the use of subject authority data.
4. METHODOLOGY

4.1 Subject Relationship and Group 3 Entities Introduced in FRBR

The subject relationship introduced in FRBR can be illustrated as:

Figure 4.1 Extension of FRBR Figure 3.3

The diagram in FRBR depicts the “subject” relationships between works and entities in the first, second, and third groups. The entities in the third group 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).
### 4.2 Possible Approaches to the Model of Aboutness

The FRSAR working group had, 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 itself 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.”\(^{25}\) In the years following the publication of the FRBR model, some of the discussions focused on Group 3 entities, particularly the fact that time is not included.\(^{26}\) 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”.\(^{27}\) Delsey followed up with a presentation of a version of his paper at IFLA satellite meeting in Järvenpää, Finland, before the General Conference in Oslo in August 2005, and his ideas have provoked much discussion among the members of the FRSAR WG.

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’”;\(^{28}\)

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 strings”\(^ {29}\) (all italic emphases added).

---


\(^{28}\) The IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR) is charged with the design of a conceptual model of Functional Requirements of Authority Data (FRAD; formerly Functional Requirements of Authority Records, or FRAR).

\(^{29}\) Delsey, *op. cit.*, p. 50.
Delsey identifies two “key questions” that arise in the context of his first “broad objective,” which has to do with 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.” 30 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” way? 31

In preparation the working group performed a pilot study, in which four students and faculty members at Kent State University School of Library and Information Science classified existing subject terms used by the NSDL (National Science Digital Library) contributors. These include about 3000 terms assigned based on a variety of subject vocabularies and free keywords. They classified terms into six categories: concrete stuff, abstract stuff, event, time, place, and others. The same method was also applied by one of the WG 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. In addition, there were difficulties in the classification of 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 study the working group discussed several possible approaches to the development of a theoretical framework of aboutness.

**Scenario 1**

Keep FRBR Group 3 entities (concept, object, event, and place) and only analyse attributes and relationships. The advantage is that we keep an existing framework; the disadvantage is that there seems to be a general agreement that group 3 entities need to be revisited, as demonstrated in the above pilot study. An additional argument for rejecting this scenario is that the original categorisation of Group 3 entities into four classes goes too far down the road prescribing a particular way of structuring the subject languages that are used to provide access to works. Any subject language that lacks a faceted structure able to clearly distinguish between concepts, objects, events, and places can be modelled only with difficulty. The FRSAD model, as a conceptual model, ought not to

---

30 Delsey, *op. cit.*, p. 50.
impose any constraint on the forms that subject languages take in particular implementations.

**Scenario 2**
Add *time* to the FRBR list. This solves one part of the problem, but the resulting model still does not account for processes, activities, or situations.

**Scenario 3**
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, has been justified theoretically and covers well all areas of aboutness. 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 4**
Take the `<indecs>`\(^{32}\) as the basis of the new framework. The main focus of the `<indecs>` model is intellectual property and rights management, but it also overlaps significantly with FRBR. The basic `<indecs>` entities are defined as:

- **Percept**: an entity which is perceived directly with at least one of the five senses.
  - **Being**: an entity which has characteristics of animate life; anything which lives and dies
  - **Thing**: an entity without the characteristics of animate life
- **Concept**: an entity which cannot be perceived directly through the mode of one of the five senses; and 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 <indecs> model and the FRBR model are (a) the subtyping of **Percept** in the <indecs> model into **Being** and **Thing**, and in the FRBR model into **Item**, **Person**, and **Object**, (b) the absence of a FRBR entity that directly corresponds to the <indecs> entity **Situation**, and (c) the absence of an <indecs> entity that directly corresponds to the FRBR entity **Place**.

As Delsey 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?

The dual prospect of (a) multiplying the number of Group 3 entity classes beyond the original four, and (b) renaming certain entity classes with labels that are yet more obscure (cf. **Percept**), raises doubts about the usability and utility of an <indecs>-esque revision of the FRBR model. The results of the user study noted above are also relevant here.

**Scenario 5**

Make a pragmatic list of entities. One example of such a list has been created by Buizza and Guerrini.\(^{33}\) They define 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). The **concept** is further categorised as:

- Object (material thing)
- Abstraction
- Living organism
- Person
- Corporate body
- Work
- Matter/material
- Property/quality
- Action
- Process
- Event
- Place
- Time

The problem with such lists is that the entities are not mutually exclusive, have a lot of overlap and rely on individual common everyday definitions of the entities. This is a serious disadvantage for a theoretical model.

**Scenario 6**

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

This last scenario (6) was the decision taken by the WG, 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 languages that are used to provide access to works. A good model should allow for any domain-specific structure 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, which enables the treatment of attributes and relationships on a higher level.
5. ENTITIES

5.1 General Framework

**Figure 5.1** FRSAD’s relation to FRBR

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

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

**Figure 5.2** 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 the list, presented here as terms in English, there are other possible nomens in other languages, controlled vocabularies such as subject heading lists, classification systems, etc.

While some works may have no thema as subject (such as musical works), the cases of thema without a nomen are beyond the scope of this model.

5.2 Choice of Terms for FRSAD Entities

The working group proposes the use of Latin terms, thema (plural themata or themas) and nomen (plural nomina or nomens), because they have no pre-existing meaning in our context and do not require translation. From here on, in this document the plural forms themas and nomens will be used. Other possible (English) terms include "subject", "topic", and "concept"; but even the discussions within the group proved that there are very different views on granularity (some see "subject" and "topic" as synonyms; for others, "topic" is a component of "subject"). In addition, "name" is often understood as "proper name".

5.3 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 THEM / THEM is subject of WORK. (Figure 5.3)

![Figure 5.3 Work-Thema relationship](image)

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'.”34 We may therefore see thema from both the end-user and librarian point of view. From the user point of view thema comprises the aboutness of the

---

(possibly unknown) resources, which will satisfy the information need. From the indexer’s point of view one or more *themas* capture the aboutness of a particular resource.

Within the FRBR framework, *thema* includes existing Group 1 and Group 2 entities and, in addition, all other subjects of *works* (Figure 5.4). While an entity on its own, it can be viewed as a super-entity or super-class of all FRBR entities, enabling us to model relationships and attributes on a more general and abstract level.

![Diagram showing the relationship between work, thems, and nomen](image)

**Figure 5.4** Within the FRBR framework, *Thema* includes existing Group 1 and Group 2 entities and, in addition, all other subjects of *works*.

In any particular application *thema* would normally have implementation-specific types, but, based on the pilot study, as explained in Section 4, there seems to be no generally applicable categorization of *thema*. See chapter 9 for examples of implementation scenarios.

*Thema* can vary substantially in complexity or simplicity. For some *works* "is subject of" is a one-to-one relationship, meaning that the totality of the aboutness is encompassed in a single *thema*. For other *works* the relationship is one-to-many, meaning that the aboutness of the *work* is captured in more *themas*. It is virtually impossible to define what the "atomic" level of a *thema* is, because any *thema* can be fragmented further. The argument can be reversed: simple *themas* may be combined or aggregated, resulting in more complex *themas* or *themas*. To some extent the granularity of a *thema* is also dependent on 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 rules for creating or
establishing *nomen* is not directly reflected in the complexity of the *thema*, nor is it completely independent. Some types of controlled vocabularies (such as subject heading systems) enable the establishment of complex *thema*, while others (such as thesauri) usually encourage more atomic *thema*.

### 5.4 NOMEN

This model also proposes a new relationship: *THEMA* has appellation *NOMEN*/*NOMEN* is appellation of *THEMA* (Figure 5.5).

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

---

**Figure 5.5 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 *nomen* and there may be a *nomen* referring to more than one *thema*. In a given controlled vocabulary and within a domain, though, a *nomen* should be an appellation of only one *thema*, as illustrated in Figure 5.6.

---

**Figure 5.6 Thema-Nomen relationship within a controlled vocabulary**

The double arrow in Figure 5.6 means that the controlled vocabulary may include variants in addition to the preferred form of a *nomen*.
5.5 The Importance of THEMANNOMEN Model

As early as 1923, Ogden & Richards\(^{35}\) published their famous triangle of meaning which illustrated the relationship between language, thought content, and referent. The graph (Figure 5.7) implies that the referent of an expression (a word or another sign or symbol) is relative to different language users. The theoretical foundation of it can be traced back to Aristotle who distinguished objects, the words that refer to them, and the corresponding experiences in the psyche, as well as Frege who distinguished between two types of meaning: thought content and referent in his essay “Über Sinn und Bedeutung”.\(^{36, 37}\) “In Plato's Cratylus, Socrates argues that it is not enough to try to understand what a thing is, based on its name, because the name-givers may have been living in ancient times, and the name reflects only what the name-givers thought was the nature of reality then; however, they may have been wrong. Thus, it has been historically recognized that multiple terms may refer to the same object or idea, a single term may refer ambiguously to more than one object or idea, and terms may be confusing because they are out of date”.

![Figure 5.7 Ogden's semiotic triangle. Ogden and Richards, 1923, p.11](http://en.wikisource.org/wiki/On_Sense_and_Reference (accesses May 22, 2009)).

The model was also adopted by researchers in library and information science as the basis for building knowledge organization systems.\(^{39, 40}\)

---


\(^{38}\) Ogden and Richards. op. cit.
"Metalanguage consists of signs that signify something about other signs, but what they signify depends on what relationships those signs have to each other, to the entities they represent, and to the agents who use those signs to communicate with other agents".\(^{41}\)

The importance of the \textit{THEMA-NOMEN} model for the subject authority data is to separate \textit{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 \textit{nomen} (for example, a translated metadata vocabulary, a symmetrical multilingual thesaurus, or a multi-access index to a vocabulary). However, most efforts have focused on the conceptual level, e.g., mappings between two thesauri or between a classification scheme and a thesaurus. Such efforts usually encounter much greater challenges because they are concerned with the subjects as well as the relationships among the subjects.


\(^{40}\) Campbell et al., \textit{op. cit.}

6. ATTRIBUTES

6.1 Attributes of **THEMA**

In the FRSAD model, *thema* is defined in a very abstract and general way; its attributes are implementation-dependent. “Type” is therefore the only general attribute, but its values are, again, implementation-dependent.

6.1.1 Type

In most implementations we can organize *thema* according to their kind or type. Two examples from existing implementations are presented below. (Detailed explanations of these schemes and systems can be found in Chapter 9).

a) UMLS (Unified Medical Language System) 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

b) Art and Architecture Thesaurus (AAT) facets

---


These examples show very different approaches to define types of \textit{thema}. In the UMLS, \textit{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 the \textit{Art and Architecture Thesaurus (AAT)}, all \textit{themas} are categorized in seven types: "Associated Concepts", "Physical Attributes", "Styles and Periods", "Agents", "Activities", "Materials", and "Objects".

Obviously since \textit{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. So will be the attributes for "Styles and Periods", "Agents", and "Materials" in AAT.

Since all FRBR entities may be included as members of the class \textit{thema}, they can also be seen as another approach to typing theme. The FRBR entities all have been defined with unique attributes, e.g., the attributes for "Person" vs. "Place" vs. "Object".

In summary, in any implementation there will normally be additional attributes of \textit{thema} than \textit{type}. Those attributes will be dependent on the type of \textit{thema} and the subject or application domain.

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

For example:

\begin{center}
\begin{tabular}{ll}
\textbf{Class} & \textbf{Instance} \\
Palaces & Buckingham Palace \\
Ships & The \textit{Lusitania} \\
Battles & The Battle of Hastings \\
\end{tabular}
\end{center}
6.2 Attributes of \textit{NOMEN}

The attributes of \textit{nomen} listed (in no particular order) below represent the most common generally applicable attributes. Not all are applicable in all cases, and the list is not comprehensive. While all listed attributes are applicable to instances of \textit{nomens}, some may also be used for an entire scheme. The examples of attribute values are illustrative only and should not be seen as prescriptive.

\textbf{Type}

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

- Identifier – the name, assigned to an entity, which is persistent and unique within a domain
- Controlled name – the name constructed in the authority control/vocabulary maintenance process, which usually serves as an access point (note: called \textit{Controlled access point} in FRAD)

If needed, the values of the type attribute may be further refined; for example, further refinement to include different kinds/formats of identifiers.

\textbf{Scheme}

The scheme in which the \textit{nomen} is established. Includes value encoding schemes (subject heading lists, thesauri, classification systems, name authority lists, etc.) and syntax encoding schemes (standards for encoding dates, etc.). In a particular implementation the actual values of this attribute may come from a list and may be coded.

Examples of attribute values:

- LCSH
- UDC
- ULAN
- ISO 8601

\textbf{Reference Source}

The source in which the \textit{nomen} is found. Used as justification for including the \textit{nomen} in the system.

Examples of attribute values:

- Encyclopedia Britannica
- Webster's Third New International Dictionary (1961)
Representation

The way in which the *nomen* is expressed such as picture, letters, symbols, sound. In a particular implementation the actual values of this attribute may come from a list and may be coded.

Examples of values:
- alphanumeric
- sound
- graphic

Language

The language of a *nomen*. In a particular implementation the actual values of this attribute may come from a list and may be coded.

Examples of values:
- Greek
- Chinese
- Slovenian

Script

The script used for the *nomen*. In a particular implementation the actual values of this attribute may come from a list and may be coded.

Examples of values:
- Cyrillic
- Thai
- Chinese-simplified
- Chinese-unsimplified

Transcription/transliteration

The rule, system, or standard used to render the *nomen* in a different representation. Usually used for conversion to a different script. In a particular implementation the actual values of this attribute may come from a list and may be coded.

Examples of values:
- Pinyin
**Encoding**

Any character set, used for a *nomen*, will be encoded according to a particular standard. In a particular implementation the actual values of this attribute may come from a list and may be coded.

Examples of values:
- ASCII
- Latin-1
- UTF-8
- GB 18030 (Chinese simplified)
- Big5 (Chinese unsimplified)

**Form**

Any additional information, which helps to interpret the *nomen*. In a particular implementation the actual values of this attribute may come from a list and may be coded.

Examples of attribute values:
- Full name
- Abbreviation
- Formula

**Time of validity**

The time period, in which a particular instance of *nomen* is/was used or is/was valid within a scheme. Should not be confused with the temporal aspect of a *thema*, which is a *thema* attribute. In a particular implementation the actual values of this attribute may come from a list and may be coded.

Examples of values:
- until May 11, 1949
- after 1945
- 1945 - 1967

**Audience**

The community or user group for which this *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 or rule.

Examples of values:
- English-speaking users
- Scientists
- Children

**Status**

The status of a particular *nomen* in a controlled vocabulary. Should not be confused with management of schemes (including or excluding a *thema*).

Examples:
- Proposed
- Accepted
- Obsolete

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

Current subject authority records typically include other elements such as notes and/or administrative data. Also, current authority records merge data describing both *thema* and *nomen* (or even a set of *nomen*) into one record. For example scope note/definition is often included and it applies to *thema*. Since this is a conceptual model, we are not dealing with these aspects of data modeling.
7. RELATIONSHIPS

The FRSAD model establishes two sets of relationships:

(1) Relationships between different types of entities: WORK-TEEMA and THEMA-NOMEN. These are the primary relationships and are illustrated in Chapter 5 when the entities are presented.

(2) Relationships between entities of the same type: THEMA-THEMA and NOMEN-NOMEN.

7.1 WORK-to-THEMA Relationships

Work-to-Theme relationships are discussed in Section 5.1. To review, in the FRSAD model, thema includes existing Group 1 and Group 2 entities and, in addition, all other subjects of works. Their relationships can be illustrated as:

*Work has subject Thema / Thema is subject of Work*

(From Figure 5.3 Work-Thema relationship)

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

7.2 THEMA-to-NOMEN Relationships

The THEMA-to-NOMEN relationships can be abstracted with the following statements:

*Thema has appellation Nomen / Nomen is appellation of Thema*

As stated in Section 5.4, in general (i.e., in natural language or when mapping different vocabularies) the relationship "has appellation/is appellation of" is a many-to-many
7.3 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," the thema-to-thema relationship types are discussed in the context of subject authority systems.

7.3.1 Hierarchical Relationships

Hierarchical relationships show 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. "The primary function of the hierarchical relationship is to convey the same concept, but at different levels of specificity." Hierarchical structures are found in classification schemes, subject heading systems, and thesauri. The hierarchical relationship includes the following types.

7.3.1.1 Generic Relationship

Genus-species hierarchies are used in logic and definition. The generic relationship is the logical relationship of inclusion. "Of limited domain and range, it is strictly defined in terms of the properties of reflexivity, anti-symmetry, and transitivity." 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 relationship between

---

44 Delsey, op. cit.
parrots and pets does not exist in logic. 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).

7.3.1.2 Whole-Part Relationship

Whole-part forms another major type in hierarchical structures. The hierarchical whole-part relationship covers situations in which one concept is inherently included in another, regardless of context, so that the terms can be organized into hierarchies, with the "whole" treated as a broader term. The relationship is usually specified as "PartOf". For example, in a personal computer there is a motherboard or system board with slots for expansion cards and holding parts such as Central Processing Unit (CPU) and Random Access Memory (RAM).

In addition to physical component part relationships, "whole and part" can be applied to several types of situations such as, geographical regions, organizational structures, human anatomies, and conceptual topic-subtopic relationships. 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.

7.3.1.3 Instance Relationship

The instance relationship identifies the link between a general category of things or events, expressed by a common noun, and an individual instance of that category, often a proper name. This type of relationship is expressed as "InstanceOf". For example, Mydoom and ILOVEYOU are two instances of computer worms (Worms (computer)), expressed by proper names.

7.3.1.4 Perspective Hierarchical Relationship

Perspective hierarchies, which do not have the logical properties of genus-species hierarchies, are seen more 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). 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 (Arthropoda), it can belong to as many perspective hierarchies as there are aspects of insects to be studied. In a classification, an insect can be looked at, or studied, from the point of view of agricultural pests, disease carriers, food, art representation, and control technology. Other reasons

---

47 Svenonius, op. cit.
49 Svenonius, op. cit.
to employ perspective hierarchies are that concepts and terms like "beauty" are poly-
semantic, vague, or ambiguous. Hence there might be no agreement of what genus (class)
such concepts belong to.

7.3.1.5 Polyhierarchical Relationship

Some concepts can belong to more than one superordinate concept at the same time and
are therefore considered to possess polyhierarchical relationships. These relationships
can be generic, e.g., "organ" belongs under both the "wind instrument" hierarchy and the
"keyboard instrument" hierarchy; whole-part, e.g., "biochemistry" is part of "biology"
and is also part of "chemistry"; or more than one types, e.g., "skull" belongs under the
"bones" (kind-of) and also under the "head" (part-of) hierarchies.  

Hierarchical structures reveal relationships between and among concepts and classes of
concepts. Used in the bibliographical universe, hierarchical relationships provide
disambiguation functions to satisfy the find and identify user tasks. Yet they are the most
effective in furthering the linking and navigation objectives and satisfying the select, and
especially the explore user tasks. They are of particular help to users with undefined or
very broad information needs and they will also allow users to improve their searching.

7.3.2 Associative Relationships

Associative relationships cover relations that are beyond hierarchical yet semantically or
conceptually associated and co-occurring. Associative relationships between thema are
made explicit in some of the subject authority systems.

In general, associative relation links are established among the 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 the
categories listed in Figure 7.1.  

<table>
<thead>
<tr>
<th>Associative Relationships</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cause / Effect</td>
<td>Accident / Injury</td>
</tr>
<tr>
<td>Process / Agent</td>
<td>Velocity measurement / Speedometer</td>
</tr>
<tr>
<td>Action / Product</td>
<td>Writing / Publication</td>
</tr>
<tr>
<td>Action / Patient</td>
<td>Teaching / Student</td>
</tr>
<tr>
<td>Concept or Thing / Properties</td>
<td>Steel alloy / Corrosion resistance</td>
</tr>
<tr>
<td>Thing or Action / Counter-agent</td>
<td>Pest / Pesticide</td>
</tr>
</tbody>
</table>

ISO (2009). ISO/CD 25964-1, Information and documentation — Thesauri and interoperability with
Group.

Information Resources Press.

NISO. op.cit.

Dearborn.
In particular implementations, a decision would be made as to whether and which associative relationships to include and at what level of specificity.

7.3.3 Other Approaches to Semantic Relationships

In 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 ALA conference, exemplified over a hundred associative relationships and 26 hierarchical relationships. Over 40 in the associative group and 20 in the hierarchical group have been identified by other sources. The Unified Medical Language System (UMLS) classified semantic relationship types in 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 other cases, such relationship types for just geographical regions are identified as:

- Inherently spatial
- Containment
- Overlap
- Proximity
- Directional

Explicitly stated

---


These examples illustrate implementation-dependent relationship typing.

7.4 NOMEN-to-NOMEN Relationships

7.4.1 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 the same natural language can be found in five general situations:

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 subject vocabularies.

In addition, equivalence relationships exist between *nomens* in different languages (e.g., in multilingual vocabularies) 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 these types 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

---

7.4.2 Partitive Relationship

Partitive relationships also exist between nomens. A nomen may have components (parts). These components may or may not be nomen on their own. The composition of such nomen may be governed by rules, such as citation order in faceted classification schemes, or the order of subdivisions in a subject heading.

Other nomen-to-nomen relationships may also be established.
8. MAPPING OF USER TASKS, ATTRIBUTES, AND RELATIONSHIPS

8.1 Subtasks

When using subject authority data, a user may need to find, identify, and select a subject entity or entities. In addition, a user may choose to explore a subject domain, its terminology, the relationships that exist among these terms and the correlation of subject terminology of one schema or vocabulary to the respective representation (term, notation, etc.) of this topic in another schema or vocabulary.

The following is a list of subtasks that result from placing the primary subject authority data user tasks in the contexts of different user groups (end users, cataloguers, vocabulary creators…) and both the direct and indirect use of these data as discussed in section 3.1.

User subtasks in direct and indirect use of subject authority data include:

**FIND: using the data to find entities that correspond to the user's stated search criteria**

1. Using attributes and relationships of *thema* 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* that belongs to the medical field and is of the type *substance*.  
   - A user is looking for a *thema* of the *type* medical condition(s) for which “chloromadinone acetate” (*thema*) is used as medication (relationship).
   - A cataloguer is looking for instances (set of *thema*) of a particular artistic style (type) using *thema-to-thema* hierarchical relationships.

2. Using attributes and relationships 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 cataloger is looking for the preferred *nomen* in Library of Congress Subject Headings, a particular subject authority system, *nomen* for the *thema* “lilac flower” (as commonly referred to in English) to assign it as a subject access point in a bibliographic record.

3. Using attributes and relationships of a *nomen* to find work(s) that have as subject the *thema* this *nomen* is representing. For example:
   - An end-user is using the *nomen* from a particular subject authority system to find materials about the *thema* “dragonflies”.

44
IDENTIFY: using the data retrieved to identify an entity

4. Using subject authority data to identify a thema, i.e., confirm that the thema found is the one sought by the user or to distinguish between two similar thema. For example:
   ○ A user is using subject authority data to identify if the thema “clothing” or the thema “Costume” is more appropriate for a specific information need.

5. Using subject authority data to identify a nomen, i.e., confirm that the nomen found is the one sought by the user or to distinguish between two similar nomen. For example:
   ○ A user is using subject authority data to identify if the nomen “craftsman style” and the nomen “craftsman movement” are representing the same thema.

SELECT: using the data to select an entity that is appropriate to the user's needs

6. Using subject authority data to select a thema from the set of thema found. For example:
   ○ Select thema at the appropriate level of specificity from a hierarchy of related themas. For example, a cataloguer is using subject authority data to select the thema “Amazon River” as a more appropriate subject access point in a bibliographic record rather than the broader thema “rivers”.

7. Using subject authority data to select a nomen from the set of nomen found. For example:
   ○ Select the preferred nomen for a thema within a subject authority system to use in searching or in assigning access points. For example, a user is using subject authority data to select “ale glass” among the AAT nomen “ale glasses,” “glass, beer,” and “malt-beverage glass”.

EXPLORE: using the data in order to explore the entities and the relationships among them

8. Using subject authority data to explore the relationships between two or more thema within the same subject authority system. For example, a user is using subject authority data to explore associative relationships of the thema “digital libraries”.

9. Using subject authority data to explore the relationships between two or more nomen within the same subject authority system. For example, a user is using subject authority data to explore the relationship of the nomen “vase carpets” and the nomen “vase rugs”.

10. Using subject authority data to explore the correlation of themas among two or more subject authority systems. For example, a user is using subject authority data to explore the correlation of the thema “domestic cats” among LCSH and the Sears subject headings lists.

11. Using subject authority data to explore the correlation of nomens among two or more subject authority systems. For example, a user is using subject authority data to explore whether the MeSH nomen “cataract” correlates to the NLM Classification notation (nomen) “WW 260”.

45
12. 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 structured within the ASIS&T Thesaurus.

### 8.2 Mapping of Attributes, Relationships and User Tasks

Tables 8.1 and 8.2 map the attributes and relationships defined in chapters 6 and 7 to the defined set of user tasks. The mapping is intended to clarify which attributes and relationship are required to support each particular user task. Only attributes and relationship specified in the FRSAD model are mapped. The decision as to which attributes and relationships to include or indicate as mandatory is application or implementation specific.

#### THEMA

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Find</th>
<th>Identify</th>
<th>Select</th>
<th>Explore</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relationship</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hierarchical relationship</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Associative relationship</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

= strong importance  □ = moderate importance

**Table 8.1 Mapping for Thema**

#### NOMEN

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Find</th>
<th>Identify</th>
<th>Select</th>
<th>Explore</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scheme</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reference Source</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Representation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Language</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Script</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transcription/transliteration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Encoding</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Form</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time of validity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Audience</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relationship</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equivalence relationship</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Partitive relationship</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

= strong importance  □ = moderate importance

**Table 8.2 Mapping for Nomen**
9. SUBJECT AUTHORITY SYSTEMS IMPLEMENTATION EXAMPLES

This chapter provides examples found in implementations of subject authority systems.

9.1 Existing Models of THEMA Types

In Chapter 6 Attributes, type is defined as the only general attribute of thema because all other attributes are implementation-dependent. In any particular application, thema would normally have particular implementation-specific types. Based on our preliminary study, there seems to be no generally applicable categorization of thema. This is also supported by the following examples. These systems and schemes range from general coverage (Faceted Application of Subject Terminology) to more special 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 9.1.1 Faceted Application of Subject Terminology (FAST) subject facets:

FAST (Faceted Application of Subject Terminology) is an adaptation of the Library of Congress Subject Headings (LCSH) with a simplified syntax. LCSH headings form the basis for FAST authority records, which are accessible through the OCLC FAST Test Databases Web site at http://fast.oclc.org/. FAST employs a faceted approach by defining headings according to their functions and categorized all headings into eight facets. Seven of them are subject facets and one is for form (genre) that is out of the scope of the FRSAD model. The subject facets include:

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

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

---

Example 9.1.2. *Unified Medical Language System*® (UMLS) semantic types

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 to facilitate 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 meaning of terms, the same concept 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**, that exist between Semantic Types. The over 130 semantic types and more than 50 semantic relationships defined by the UMLS can be found in the UMLS 2004 AB Documentation. 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 from 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.

---

62 ibid.
Example 9.1.3. The Foundational Model of Anatomy Ontology (FMA) semantic types

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 schemes those sources use for arranging their anatomical terms. 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 by placing an emphasis on the highest level of granularity of its concepts and the large number and specificity of the structural relationships that exist between the references of these concepts. According to project documentation, the FMA consists of 75,000 anatomical classes, 130,000 unique terms, over 205,000 frames, and 174 unique slots showing different types of relations, attributes, and attributed relationships. FMA is a typical example of disciplined modeling showing how semantic types for a subject vocabulary 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 9.1.4. *Art and Architecture Thesaurus (AAT) facets*\(^{65}\)

*Art and Architecture Thesaurus* (AAT) is a controlled vocabulary for fine art, architecture, decorative arts, archival materials, and material culture for the purposes of indexing, cataloging, searching, as well as research tools. The facets 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

---
\(^{65}\) *Art and Architecture Thesaurus Online, Hierarchy Display, op. cit.*
locations in the AAT hierarchies rather than a defined portion that is specific only for Renaissance painting. 66

The examples above indicate that in actual implementations there are always attempts to define some fundamental facets or atoms to accommodate all types of themas. However, in reality, thema would normally have implementation-specific types.

9.2 THEMÆ-THEMÆ Relationships presented in Subject Vocabulary/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 metadata,
- reference and public services librarians and other information professionals who are searching for information as intermediaries,
- subject authority systems creators and maintainers, and
- end-users using information retrieval systems to fulfill 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.

9.2.1 Thema-Thema relationships presented by individual vocabularies

The focus of this section is on the semantic relations presented in vocabularies. The following examples demonstrate how thema-to-thema relationships are presented for the same thema, "Mercury" (as a liquid metal and/or as an element), in various subject vocabularies. The same object can be viewed from the perspective of different properties, i.e., a polyhierarchical characteristic (same concept or object belonging in different hierarchies). Webster’s definition of mercury is: “a heavy silvery toxic univalent and bivalent metallic element; the only metal that is liquid at ordinary temperatures.” 67 68

Example 9.2.1.1. LC Subject Authority

Thema: mercury (as a liquid metal)
[Note: in the following entry:
  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;
Sa = Topical term or geographic name entry element
Sw = Control subfield;  Broader Terms are not currently specified.

Several semantic relations 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 9.2.2.)
Example 9.2.1.2. Art and Architecture Thesaurus:

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

**Identification Information:**
- **ID:** 300011026
- **Record Type:** concept

**Terms:**
- mercury (preferred, C,D,U,L,C,English-P)
- Hg (C,U,F,I,U,A,English)
- argentol (C,U,F,I,U,English)

**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)>

**Figure 9.3** An online display record of the AAT concept “Mercury”

Example 9.2.1.2 shows a screen captured from the Art and Architecture Thesaurus (AAT) online version. Hierarchical relationships of concepts “mercury” and both “nonferrous metal” (under the node “<metal by composition or origin>”) and “elements” (under the node “<materials by chemical form>”) are presented in the hierarchies.
Figure 9.4 Illustration of the semantic relations between the *themes*
presented in Figure 9.3
Example 9.2.1.3. Medical Subject Headings (MeSH): Standard Display

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

<table>
<thead>
<tr>
<th>MeSH Heading</th>
<th>Mercury</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tree Number</td>
<td>D01.268.556.504</td>
</tr>
<tr>
<td>Tree Number</td>
<td>D01.268.956.437</td>
</tr>
<tr>
<td>Tree Number</td>
<td>D01.552.544.504</td>
</tr>
</tbody>
</table>

See Also: 
- Mercury Isotopes
- Mercury Radioisotopes
- 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

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]

Figure 9.5 Extracted portion from a MeSH record indicating semantic relations

Example 9.2.1.3 shows data derived from a Standard Display of a MeSH record found through the MeSH Browser. The information indicates that the *thema* represented by a *nomen*, "Mercury", has associative relationships ("see also") with *themas* which are represented by *nomens* "Mercury Isotopes", "Mercury Radioisotopes", and "Organomercury Compounds", as illustrated in Figure 9.6 below.
There are also two immediate hierarchical relationships (see Figure 9.7): (1) between thesaurus represented by the nomens "Mercury" and "Metals, Heavy". The latter can be traced up to two upper classes; (2) between thesaurus represented by nomens "Mercury" and "Transition Elements". (See illustration in Figure 9.7; notational form of nomens are not included). Allowable qualifiers enable the concept to be further limited to specific aspects (e.g., "administration & dosage" (AD), "isolation & purification" (IP), and "toxicity" (TO)) and form specific subject headings to represent different thesaurus (e.g., "Mercury – toxicity", or "Mercury – isolation & purification").
Example 9.2.1.4. *Dewey Decimal Classification*

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

<table>
<thead>
<tr>
<th>Class Number:</th>
<th>669.71</th>
</tr>
</thead>
<tbody>
<tr>
<td>Segmented Number:</td>
<td>669/71</td>
</tr>
<tr>
<td>Caption:</td>
<td>Mercury</td>
</tr>
</tbody>
</table>

**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 9.8a. Screen captured from OCLC Connexion WebDewey for classes related to "Mercury" as a metal

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

<table>
<thead>
<tr>
<th>Class Number:</th>
<th>546.663</th>
</tr>
</thead>
<tbody>
<tr>
<td>Segmented Number:</td>
<td>546/.663</td>
</tr>
<tr>
<td>Caption:</td>
<td>*Mercury</td>
</tr>
</tbody>
</table>

**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 9.8b. Screen captured from OCLC Connexion WebDewey for classes related to "Mercury" as an element

It should be noticed that although the relationships are similar with what are 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) that representing the themes, not the captions, as one can find from the above figure where both captions are "Mercury" although they are affiliated with two different class in DDC. The two pairs of hierarchical relationships are illustrated in the following figure: Figure 9.9a is for thema "mercury as a metal" and Figure 9.9b is for thema "mercury as an element".

Figure 9.9a Illustration of the hierarchical relationships (through the classificatory structure) between the DDC classes shown in Figure 9.8a

Figure 9.9b Illustration of the hierarchical relationships (through the classificatory structure) between the DDC classes shown in Figure 9.8b
9.2.2 Inter-system *Thema* crosswalking through the *Nomens* of the *Thema*

**Example 9.2.2.1 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]

<table>
<thead>
<tr>
<th>Nomen</th>
<th>CC</th>
<th>DI</th>
<th>PT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mercury</td>
<td>A9630D</td>
<td>January 1971</td>
<td>planets</td>
</tr>
<tr>
<td>planets</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>transits</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Nomen</th>
<th>Planets and satellites</th>
<th>CC</th>
<th>DI</th>
</tr>
</thead>
<tbody>
<tr>
<td>A9630</td>
<td>(exc. the Moon) for Earth, see A91... for celestial mechanics, see A9510...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A9630D</td>
<td>Mercury</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Figure 9.10* Extracted entries from *INSPEC Thesaurus* (top) and *INSPEC Classification* (bottom) Showing inter-system *thema's* crosswalking between the two authority systems

Example 9.2.2.1 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 9.11.
Figure 9.11 Illustration of the inter-system thesaurus crosswalking between INSPEC Thesaurus and INSPEC Classification shown in Figure 9.10
Example 9.2.2.2. 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:
  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;
  Broader Terms are not currently available.]

Figure 9.12. An authority record screen captured from OCLC Connexion Authorities

In example 9.2.2.2, 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, which have the nomens "QD181.H6" (in Chemistry), "TA480.M4" (in Engineering materials), "TN271.M4" (in Prospecting), and "TP245.M5" (in Chemical technology). (See also Figure 9.13.)
Figure 9.13. Illustration of the inter-system themas' crosswalking between LCSH and LCC showing in Figure 9.12.
9.3 Same THEMA Represented by Nomens from Different Schemes

The following example 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 the examples demonstrated, schemes may have different levels of specificity, structure, and syntax for the *nomens* they have established.

9.4 Examples of Display Records from Subject Vocabularies/Authority Files

As shown in section 9.2 above, 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., to information professionals, subject authority systems creators and maintainers, and end-users). Following are captured screens of records displayed to end-users. They contain mixed information regarding *thema*, *nomen*, relationships between a *thema* and *nomens*, as well as among different *themas*. In addition, they demonstrate that *thema* types are implementation dependent one and vary in different domains.
Example 9.4.1. 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.

Source: STN Database Summary Sheet: USAN (The USP Dictionary of U.S. Adopted Names and International Drug Names)
Example 9.4.2. 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 *nomen*, 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 9.4.3. 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 9.2.1.3 and Figure 9.6 and 9.7. 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.
In summary, this Expanded Concept View presents various types of semantic relationships among *themas*:

1. 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". (Figure 9.6)

2. Two immediate hierarchical relationships: (1) between *themas* represented by the *nomens* "Mercury" and "Metals, Heavy". The latter can be traced up to two upper classes "; (2) between *themas* represented by *nomens* "Mercury" and "Transition Elements". (The same is true for these *themas* and their *nomens* with notational forms).

3. Allowable qualifiers enable the concept to be further limited to specific perspectives (e.g., administration & dosage (AD), isolation & purification (IP), and toxicity (TO)) and forming specific subject headings to represent different *themas* (e.g., Mercury -- toxicity).

4. 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 those in natural languages name and in specific identification number. Various attributes of *nomens* are also presented.

### 9.5 Mapping FRSAD Model to Other Abstract 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 vocabulary, as well as ontologies. SKOS defines classes and properties sufficiently for representing the common features found in a standard thesaurus and other KOS structures. 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, which include: one or more preferred terms (at most one in each natural language); alternative terms or synonyms; and definitions and notes with
These can be matched to what have been defined in the FRSAD model, in terms of *thema*, *nomen* and their attributes. SKOS has also specific properties to represent all the semantic relationships, as described in Chapter 7.

For the issues of the complexity and granularity of *themas* and comprehensive semantic relationships between and among *themas* that FRSAD attempted to cover, OWL has even better matches. OWL ontologies provide classes, properties, individuals, and data values and are stored as Semantic Web documents. OWL 1 was mainly focused on constructs for expressing information about classes and individuals. OWL 2, the newest W3C working draft, offers new constructs for expressing additional restrictions on properties, new characteristics of properties, incompatibility of properties, properties chains and key properties. OWL 2 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 important, 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.

Moreover, 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) has been recognized or followed by other metadata standards, e.g., the VRA Core 4.0 released by the Visual Resources Association in 2007. According to the DCMI model, a record can contain *description sets*, which may contain *descriptions* composed of *statements*, which use *property-value* pairs. This results in information that can be processed, exchanged, referred to, and linked to at the statement level. When a record contains descriptions of the resource, the individual descriptions also can be linked to the authority data that manages the values associated with those properties (e.g., the subject authority data, the property name authority data, or the geographic authority data). Such an information model is independent of any particular encoding syntax and facilitates the development of better mappings and cross-syntax translations.

---


74 DCMI Abstract Model. *op.cit.*
conceptual model proposed by the FRSAR Working Group and described in this report corresponds to this abstract model by allowing any *thema* to be independent of any *nomen*, including any syntax that a *nomen* may use. Thus this conceptual model will facilitate the sharing and reuse of subject authority data amongst not only the subject vocabularies themselves, but also metadata resources.

9.6 Conclusion

The FRSAR Working Group has presented in this report a conceptual model of Group 3 entities within the FRBR framework as they relate to the aboutness of *works* and defined a structured frame of reference for relating the data that are recorded in subject authority records to the needs of the users of those records. The FRSAD model is developed with the goal to assist in an assessment of the potential for international sharing and use of subject authority data both within the library sector and beyond. It enables the consideration of the functional requirements of subject authority data at a level that is independent of any implementation, system, or specific context. Putting the subject authority data in the context of the Semantic Web developments, especially in the perspective of the Web of Data, subject authority data that are modeled based on FRSAD and encoded in SKOS and OWL will be able to become part of the linked open data and contribute to the development of the Semantic Web.