APPENDIX D. EXAMPLES FROM SUBJECT AUTHORITY SYSTEMS

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

D.1 Existing Models of THEMA Types

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

Example D.1.1 *Faceted Application of Subject Terminology* (FAST) subject facets[56]:

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

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

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

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

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

The *Unified Medical Language System*® (UMLS), developed, maintained, and distributed by the National Library of Medicine of the United States, provides a unified system for correlating a large number of biomedical terminologies and facilitates the development of computer systems that behave as if they "understand" the meaning of the language of biomedicine and health. In order to facilitate the establishment of correspondences in the meanings of terms, the same concepts occurring in different constituent vocabularies are assigned to high level semantic types encompassed within the *UMLS Semantic Network*. It consists of: (a) a set of broad subject categories, or **Semantic Types***,* that provide a consistent categorization of all concepts represented in the *UMLS Metathesaurus*®, and (b) a set of useful and important relationships, or **Semantic Relations**, which exist between Semantic Types. More than 130 semantic types and 50 semantic relationships defined by the UMLS can be found in the UMLS 2004 AB Documentation[59]. The following are the high level semantic types:

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

Events
- Activity
- Phenomenon or Process

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

---

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

Example D.1.3. *The Foundational Model of Anatomy Ontology* semantic types[60]

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

Anatomical Entity
- Non-physical anatomical entity
- Physical anatomical entity

Attribute Entity
- Cell morphology
- Cell shape type
- Cell surface feature
- Concept name
- Miscellaneous term
- Organ part phenotype
- Physical attribute relationship
- Physical state
- Structural relationship value

Dimensional Entity
- Line
- Point
- Surface
- Volume

As a domain ontology, the FMA represents deep knowledge of the structure of the human body. It emphasis is on the highest level of granularity of the concepts. Meanwhile it also presents a great number of specific structural relationships between the references of these concepts. According to project documentation[61], the FMA consists of approximately 75,000 anatomical classes, 130,000 unique terms, 205,000 frames, and 170 unique slots showing different types of relations, attributes, and attributed relationships. FMA is a typical example of modeling that shows how **semantic types** for a concept scheme can be defined. It not only encompasses the diverse entities that make up the human body but is also capable of modeling a great deal of knowledge relating these entities.

Example D.1.4. *Art and Architecture Thesaurus* (AAT) facets[62]

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

[60] *The Foundational Model of Anatomy ontology* (FMA). 2006--. School of Medicine, University of Washington. Available at: http://sig.biostr.washington.edu/projects/fm/index.html (accessed 2010-01-26).
[61] About FMA. [2006]. School of Medicine, University of Washington. Available at: http://sig.biostr.washington.edu/projects/fm/AboutFM.html (accessed 2010-01-26).
[62] *Art and Architecture Thesaurus Online*. *Hierarchy Display. op. cit.*

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

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

The conceptual framework of facets is not subject-specific. One example is the subject "Renaissance painting". Terms to describe Renaissance paintings will be found in many locations in the AAT hierarchies rather than a defined portion that is specific only for Renaissance painting[63].

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

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

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

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

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

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

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

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

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

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

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

Example D.2.1.1. ***LC Subject Authority***

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

[Note: in the following entry, MARC21 coding is used:
010 = Library of Congress control number
040 = Cataloging source
053 = LC Classification Number
$c = Explanatory term (specifying topic)
150 = Heading--Topical term
450 = *See* from Tracing--Topical term (unauthorized form/variant of term)
550 = *See Also* From Tracing--Topical Term;
$a = Topical term or geographic name entry element
$w = Control subfield; g - Broader term.]

LC Control Number: sh 85083794
HEADING: Mercury
000 00558cz a2200217n 450
001 4734282
005 19900221112154.6
008 860211i| anannbabn |a ana
035 __ |**a** (DLC)sh 85083794
906 __ |**t** 8528 |**u** fk03 |**v** 0
010 __ |**a** sh 85083794
040 __ |**a** DLC |**c** DLC |**d** DLC
053 _0 |**a** QD181.H6 |**c** Chemistry
053 _0 |**a** TA480.M4 |**c** Engineering materials
053 _0 |**a** TN271.M4 |**c** Prospecting
053 _0 |**a** TP245.M5 |**c** Chemical technology
150 __ |**a** Mercury
450 __ |**a** Hydrargyrum
450 __ |**a** Quicksilver
550 __ |**w** g |**a** Liquid metals
953 __ |**a** xx00 |**b** fg07

[Note: in this captured screenshot, subfield signs are displayed as a vertical bar.]
Figure D.1: A record from the LC Subject Authority File

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

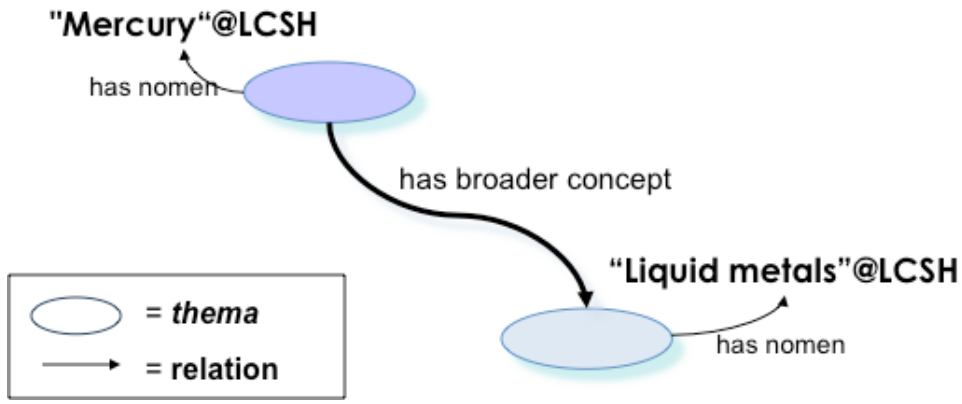


Figure D.2 Illustration of the semantic relations between two *themas* represented in Figure D.1

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

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

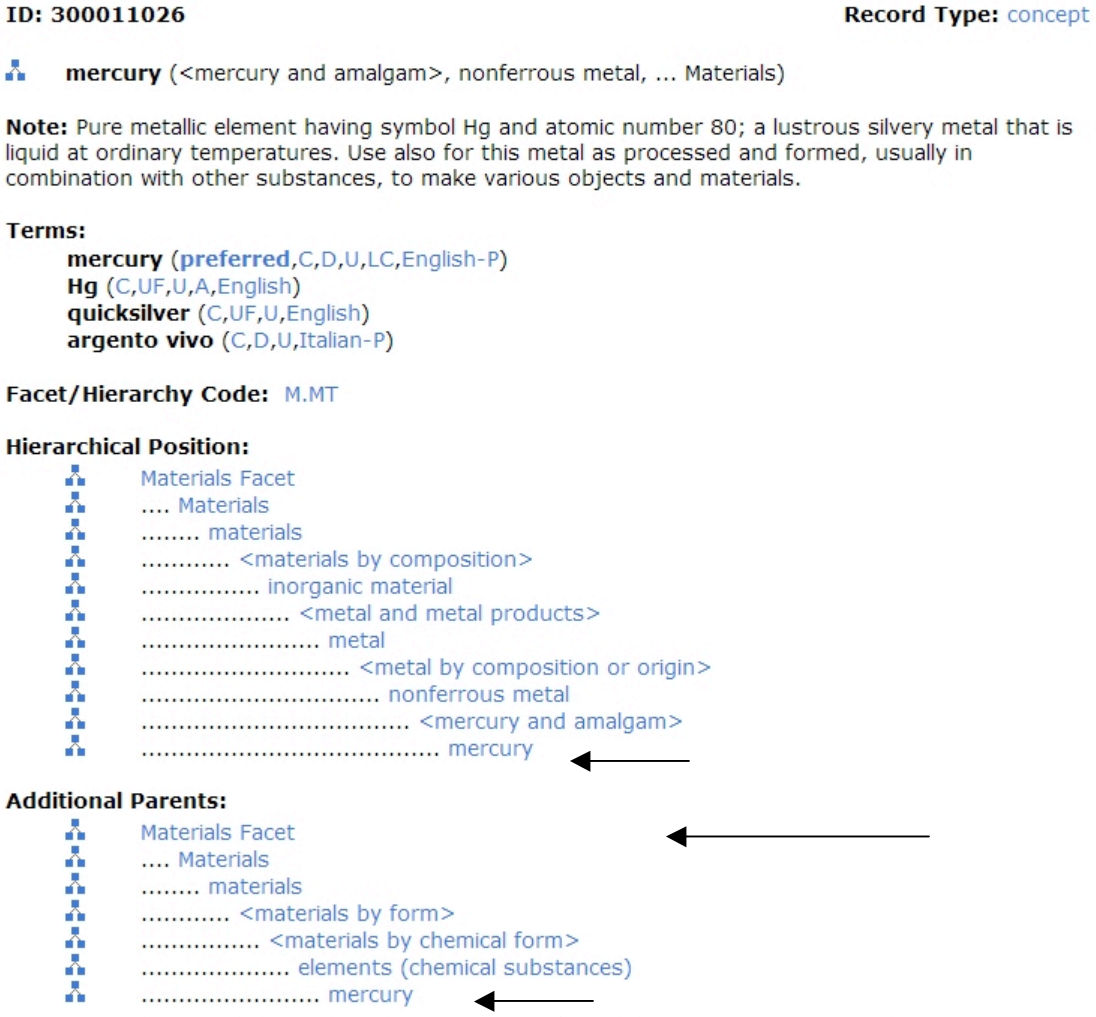


Figure D.3 An online display record of the AAT concept "Mercury"

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

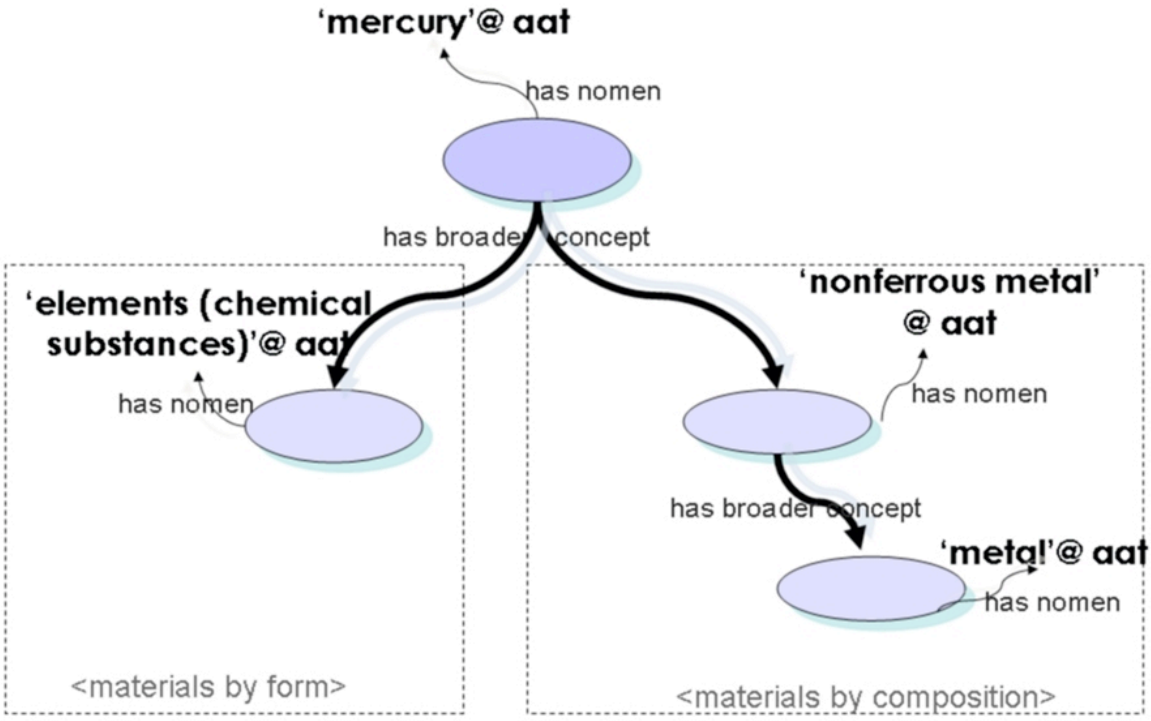


Figure D.4 Illustration of the semantic relations between the *themas* presented in Figure D.3

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

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

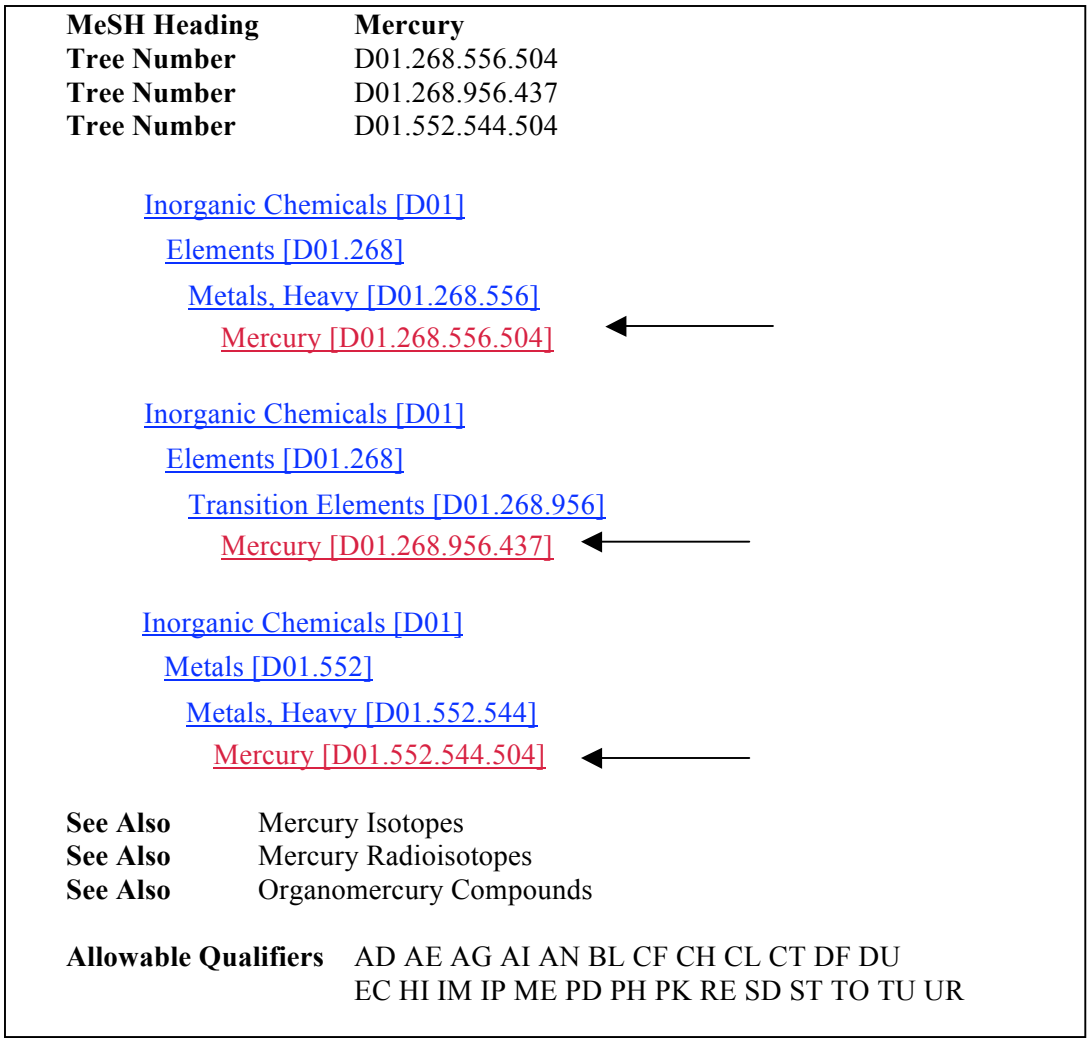


Figure D.5 Extracted portion from a MeSH record indicating semantic relations

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

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

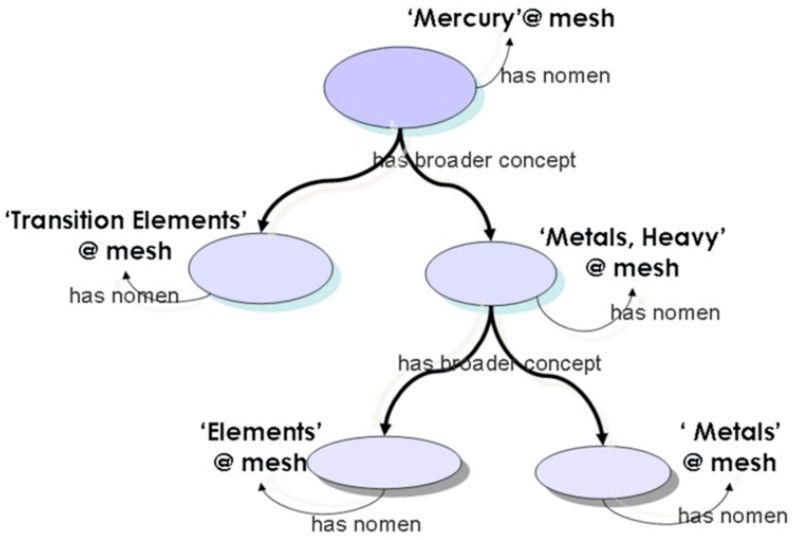


Figure D.6 Illustration of the hierarchical relationships (through the tree structure) from the extracted MeSH Heading record shown in Figure D.5

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

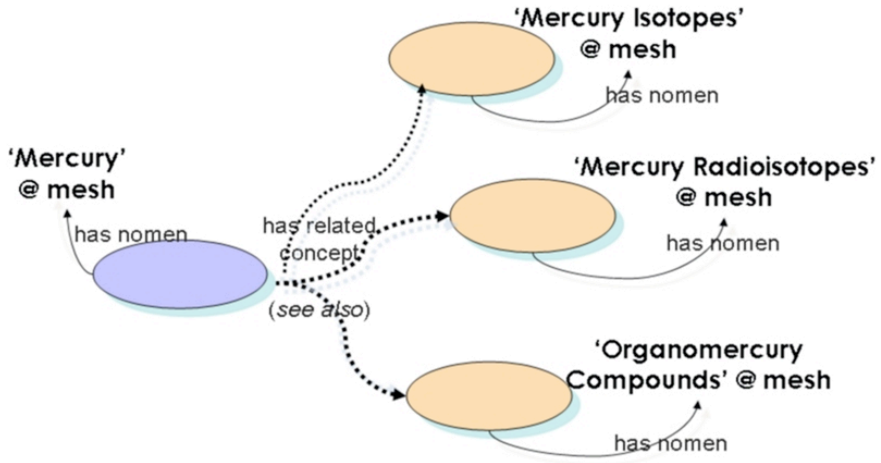


Figure D.7 Illustration of the associative relationships ("see also") from the extracted MeSH record shown in Figure D.5

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

Example D.2.1.4 . ***Dewey Decimal Classification***

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

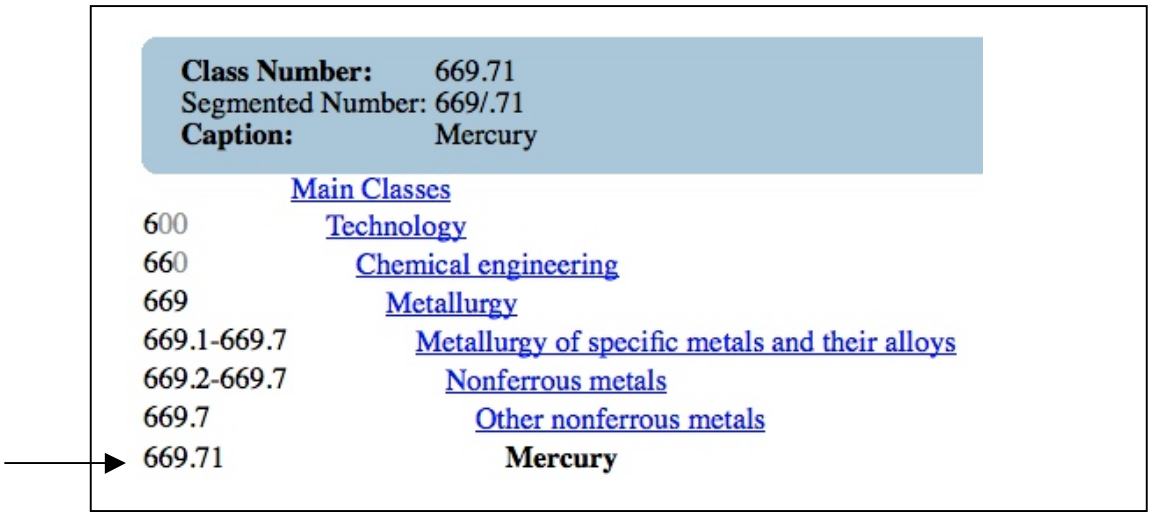


Figure D.8a. Screen captured from OCLC Connexion WebDewey for classes related to "mercury (as a metal)"

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

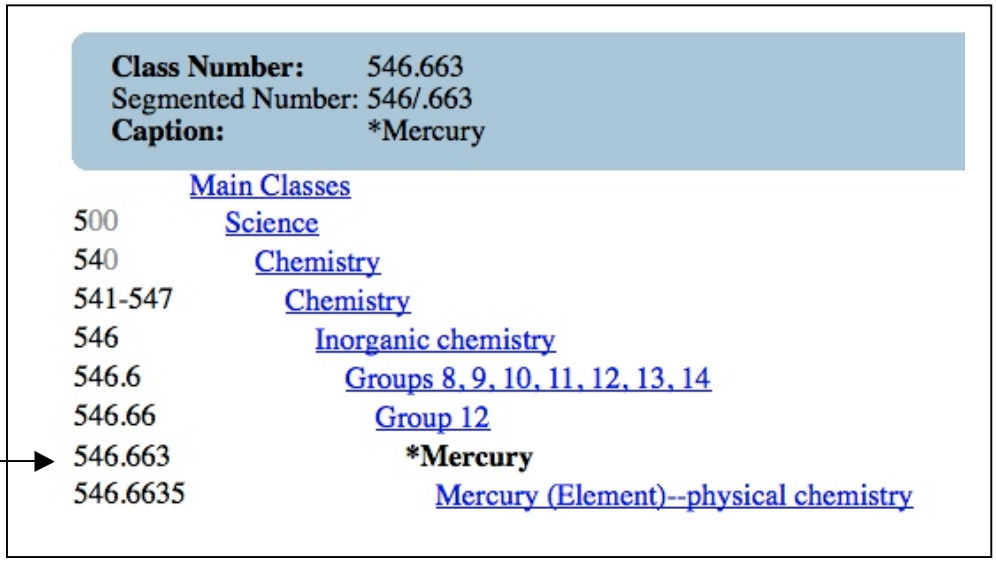


Figure D.8b. Screen captured from OCLC Connexion WebDewey for classes related to "mercury (as an element)"

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

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

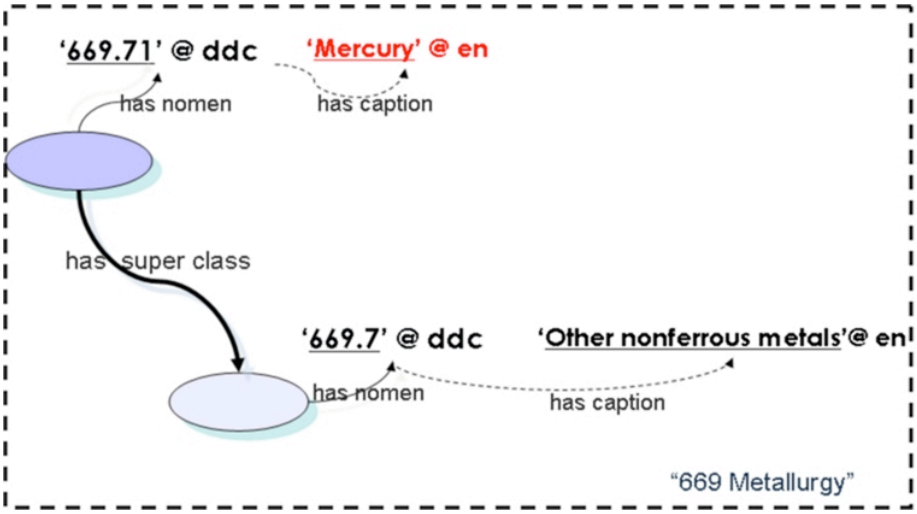


Figure D.9a Illustration of the hierarchical relationships (through the classificatory structure) between the DDC classes shown in Figure D.8a

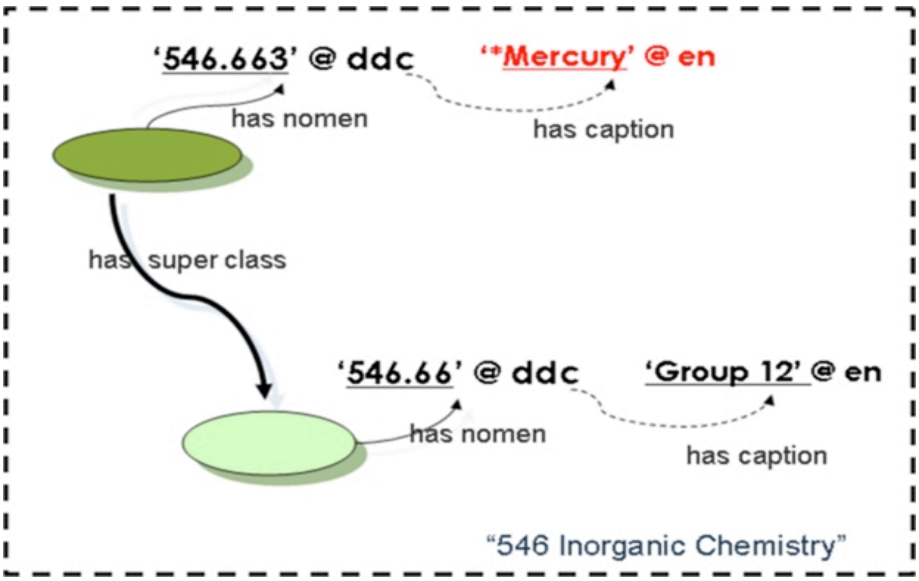


Figure D.9b Illustration of the hierarchical relationships (through the classificatory structure) between the DDC classes shown in Figure D.8b

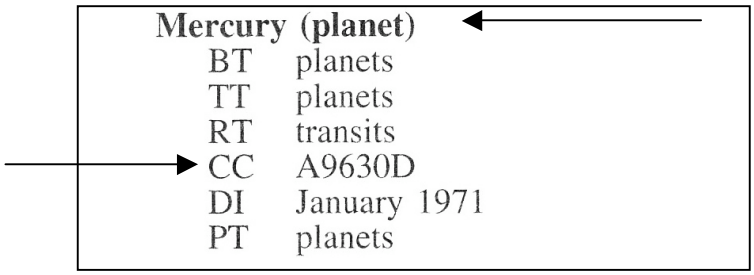
D.2.2 Inter-system *THEMA* crosswalking through *NOMENs*

Example D.2.2.1 *INSPEC Thesaurus and INSPEC Classification*

Thema: **mercury (planet)**

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

From *INSPEC Thesaurus* (2004, pg. h76)**:**
[Note: CC= Classification Code]



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

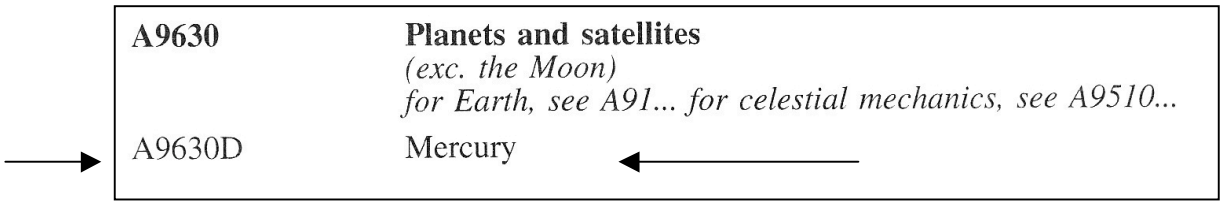


Figure D.10 Extracted entries from *INSPEC Thesaurus* (top) and *INSPEC Classification* (bottom) showing inter-system *thema* crosswalking

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

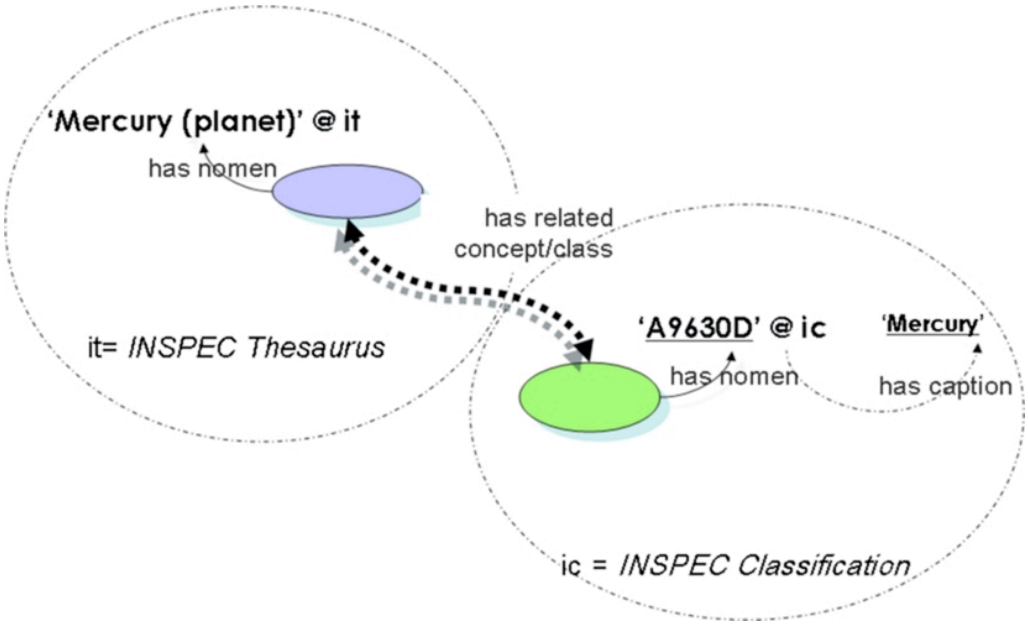


Figure D.11 Illustration of the inter-system *themas*' crosswalking between *INSPEC Thesaurus* and *INSPEC Classification* shown in Figure D.10

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

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

Example taken from Library of Congress Subject Authority File:

[Note: in the following entry, MARC21 coding is used:
010 = Library of Congress control number
040 = Cataloging source
053 = LC Classification Number
$c = Explanatory term (specifying topic)
150 = Heading--Topical term
450 = *See* from Tracing--Topical term (unauthorized form/variant of term)
550 = *See Also* From Tracing--Topical Term;
$a = Topical term or geographic name entry element
$w = Control subfield; g = Broader term.]

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

LC Control Number: sh 85083794
HEADING: Mercury
000 00558cz a2200217n 450
001 4734282
005 19900221112154.6
008 860211i| anannbabn |a ana
035 __ |**a** (DLC)sh 85083794
906 __ |**t** 8528 |**u** fk03 |**v** 0
010 __ |**a** sh 85083794
040 __ |**a** DLC |**c** DLC |**d** DLC
053 _0 |**a** QD181.H6 |**c** Chemistry
053 _0 |**a** TA480.M4 |**c** Engineering materials
053 _0 |**a** TN271.M4 |**c** Prospecting
053 _0 |**a** TP245.M5 |**c** Chemical technology
150 __ |**a** Mercury
450 __ |**a** Hydrargyrum
450 __ |**a** Quicksilver
550 __ |**w** g |**a** Liquid metals
953 __ |**a** xx00 |**b** fg07

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

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

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

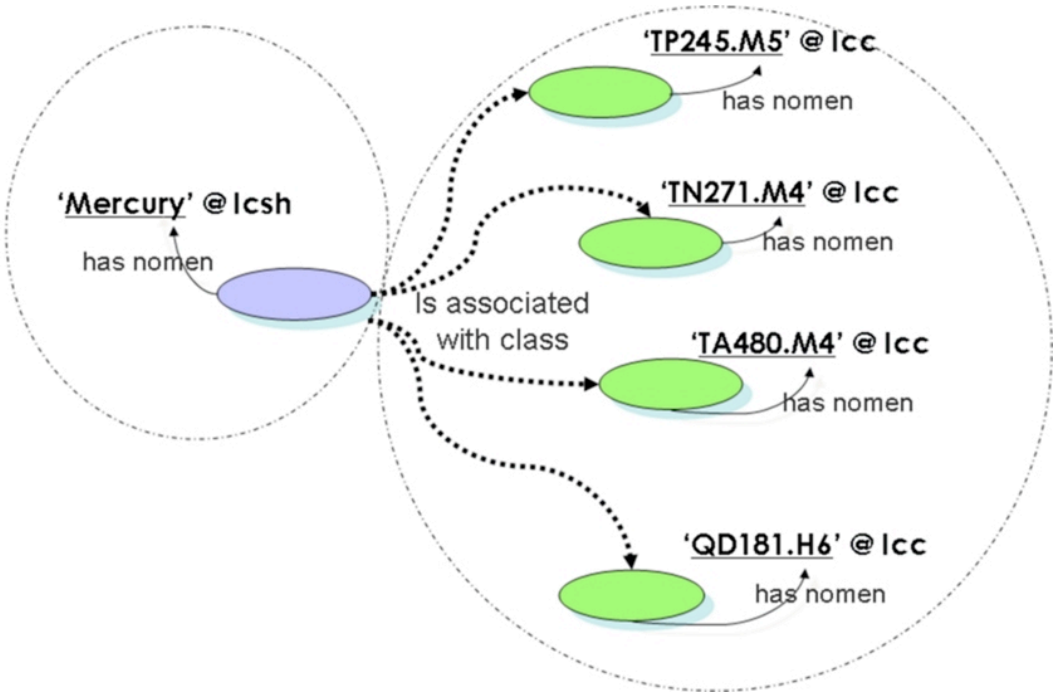


Figure D.13. Illustration of the inter-system *themas*' crosswalking between LCSH and LCC showing in Figure D.12

D.3 Same THEMA Represented by NOMENs from Different Schemes

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

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

DDC: "331.881102770943"
Constructed/combined from:
331.8811 – labor unions in industries and occupations other than extractive, manufacturing, construction
-027.7 – academic libraries
-0943 – Germany

LCSH: "Library employees--Labor unions—Germany"
"Universities and colleges--Employees--Labor unions—Germany"
"Collective bargaining--Academic librarians--Germany"
"Libraries and labor unions--Germany"

FAST:
"Library employees--Labor unions"
"Universities and colleges--Employees--Labor unions"
"Collective bargaining--Academic librarians"
"Libraries and labor unions"
"Germany"

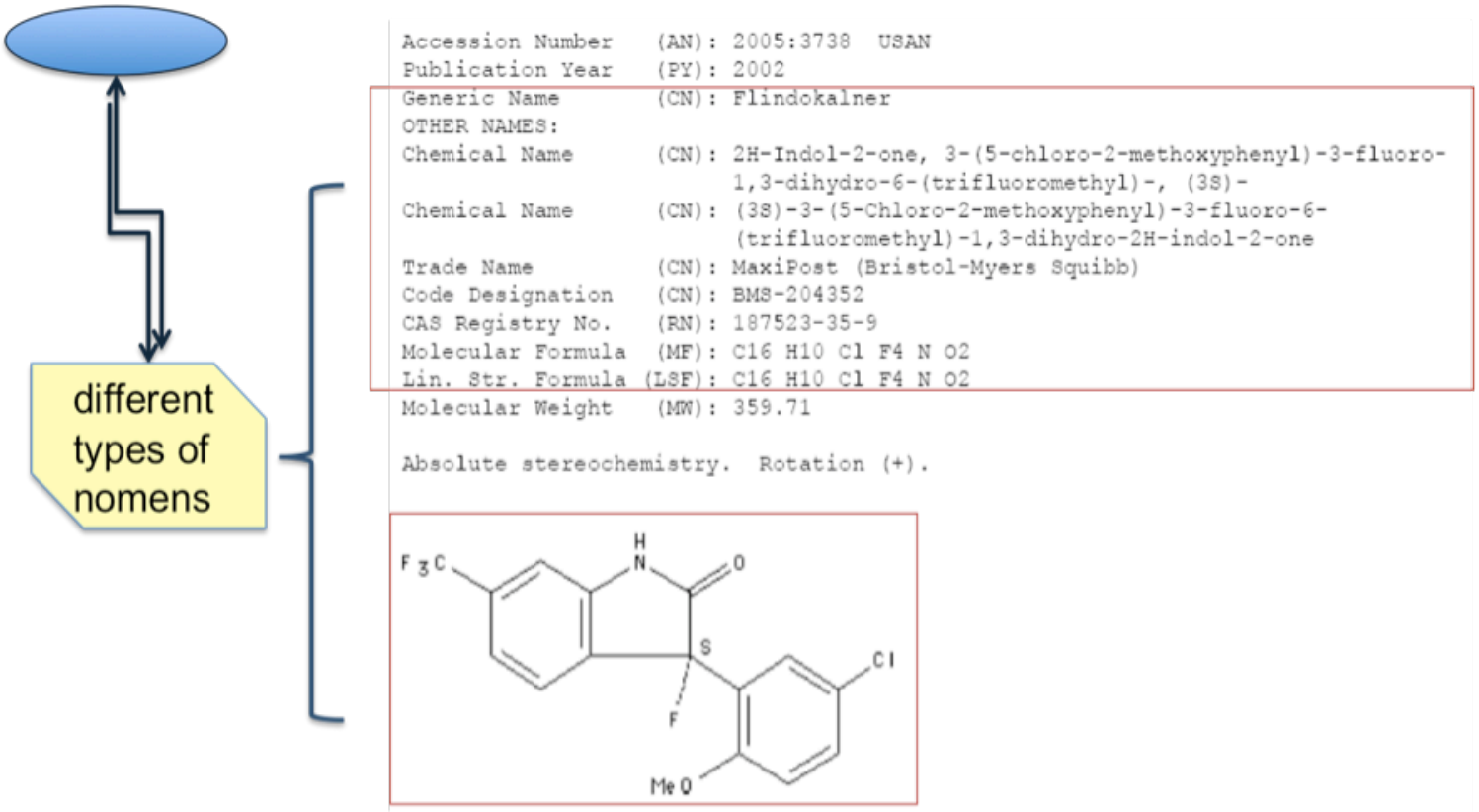
As this example demonstrates, schemes may allow the representation of *themas* at different levels of specificity through the structure and syntax of the *nomen*s they have established.

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

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

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

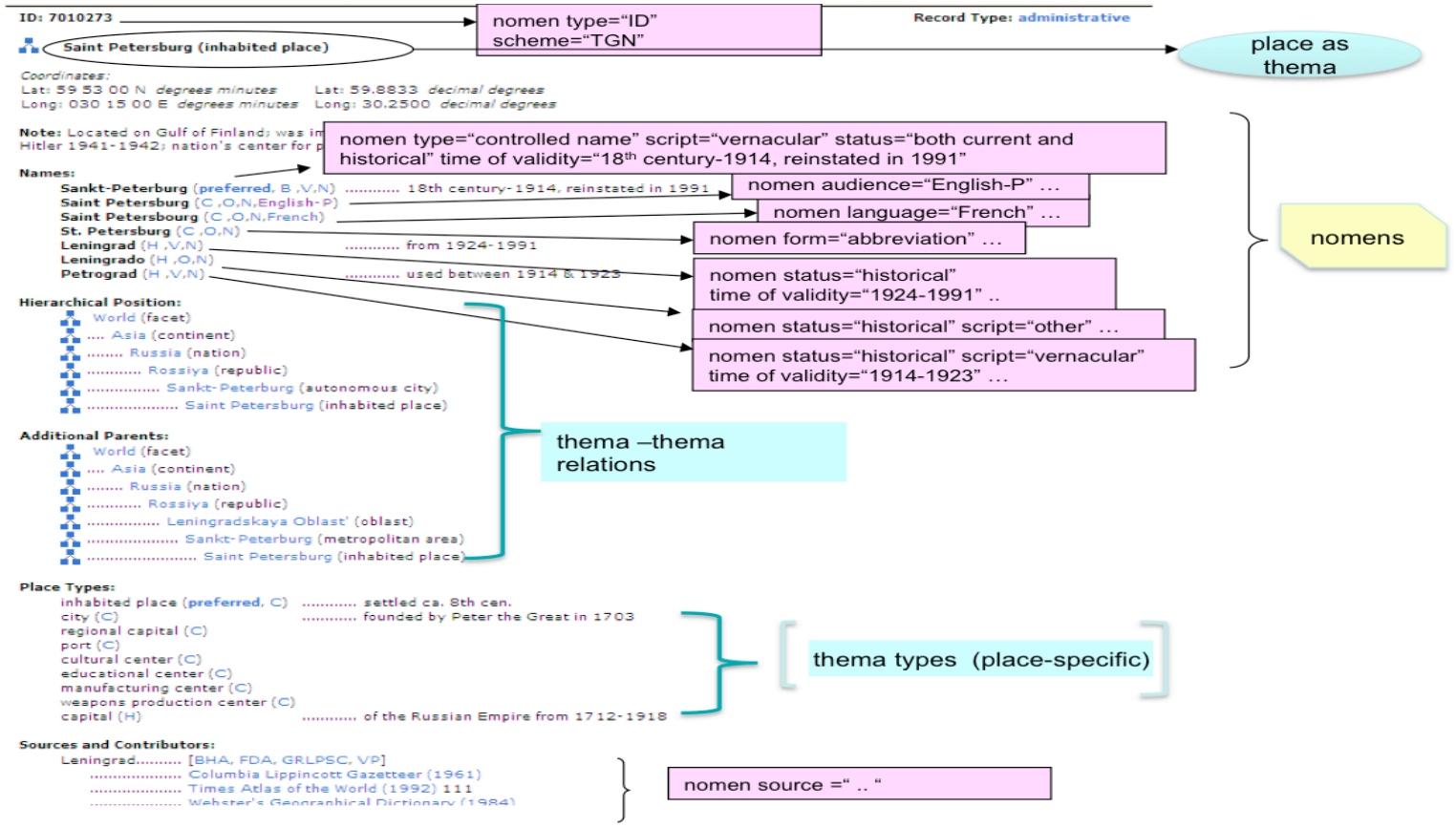
The figure below demonstrates how a *thema* could have various *nomen*s in the context of specific systems. The forms of the *nomen*s for this chemical compound are not only in various names represented in natural language, but also those represented in artificial languages such as codes, formulas and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



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

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

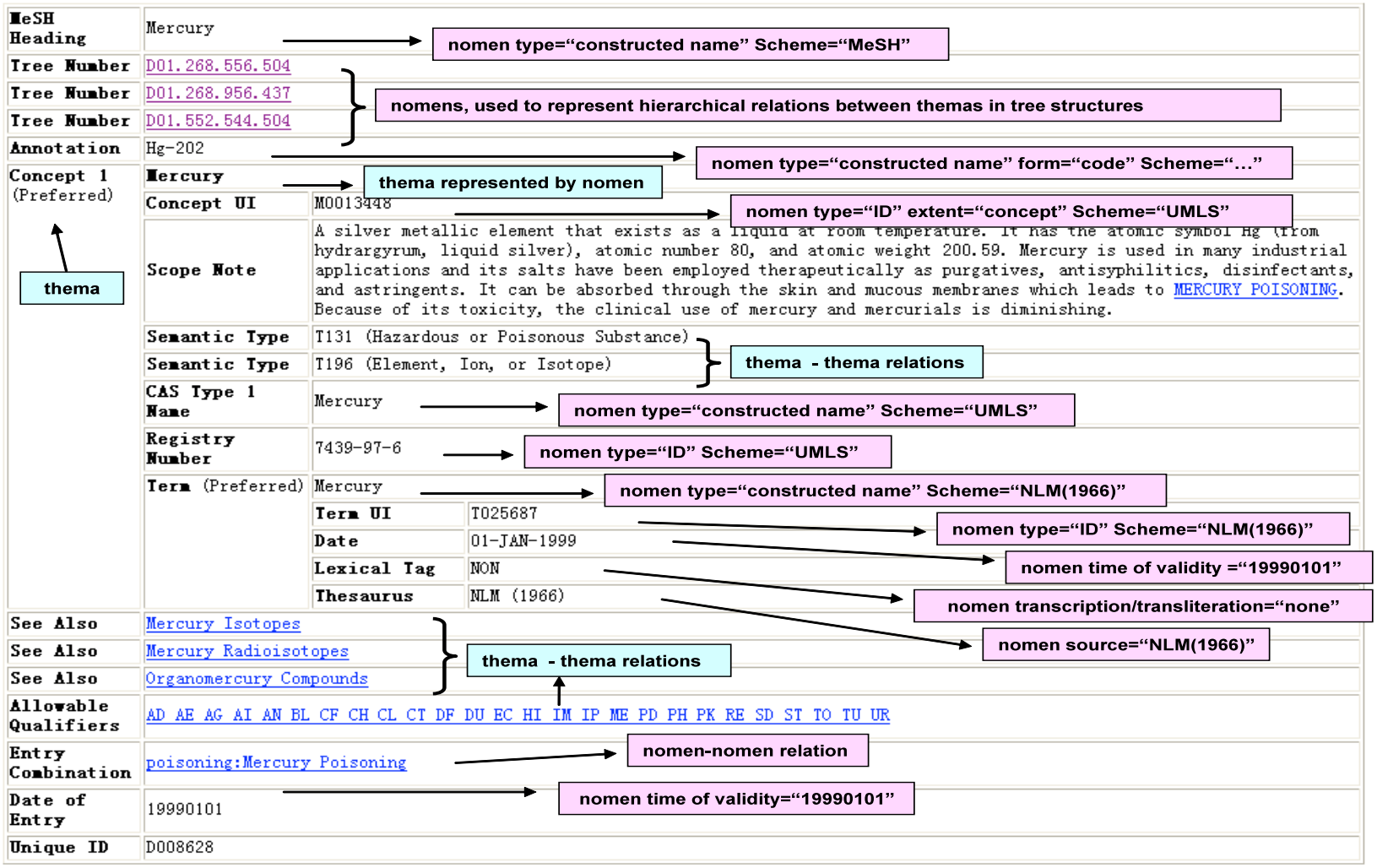
This example presents: (1) the hierarchical relationships of a *thema* (in this case a place) with other *thema*s, i.e., the "whole-part" relationships; (2) various *nomen*s, to be chosen as preferred terms in different contexts, with attributes regarding the form, time of validity, status, audience,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



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

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

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



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

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

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

b) Associative relationships between “Mercury” (as a liquid metal and as an element) and other *thema*s represented by *nomen*s “Mercury Isotopes”, “Mercury Radioisotopes”, and “Organomercury Compounds”.

c) Allowable qualifiers enable the concept to be further limited to specific perspectives (e.g., “administration & dosage (AD)”, “isolation & purification (IP)”, and “toxicity (TO)”). These facilitate the forming of specific subject headings (e.g., “Mercury – TO”, or “Mercury – IP”) to represent different *themas*.

d) The semantic types of this *thema*: “T131 (Hazardous or Poisonous Substance)” and “T196 (Element, Ion, or Isotope)” as defined by UMLS.

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