FRSAD

Functional Requirements for Subject Authority Data

Marcia Zeng
Based on the work of the IFLA FRSAR WG

SAC Meeting, ALA, 2010-06-28

Today's talk

1. Background
2. Current state
3. User tasks and the modeling approaches
4. The model --entities, attributes, relationships
5. Relationships with FRBR and FRAD
6. Implications for interoperability

Acknowledgement

- This presentation is based on the work of the FRSAR (**Functional Requirements for Subject Authority Records**) Working Group, established by the IFLA Division IV Bibliographic Control and especially the Section of Classification and Indexing.

» IFLA, OCLC, Kent State University, and Univ. of Ljubljana have provided funding, facilities, and tremendous support.

Report available at:

- **FRSAR: Functional Requirements for Subject Authority Data (FRSAD)**
  - http://nkos.slis.kent.edu/FRSAR/
- Working Group: Leda Bultrini, Lois Mai Chan, Jonathan Furner, Edward O'Neill, Gerhard Riesthuis, Athena Salaba, Diane Vizine-Goetz, Ekaterina Zaytseva, Marcia Lei Zeng, and Maja Zumer.
- Advisory Group: Victoria Francu, Hemalata Iyer, Dorothy McGarry, David Miller, Päivi Pekkarinen, and Barbara Tillett.

"subject authority data"

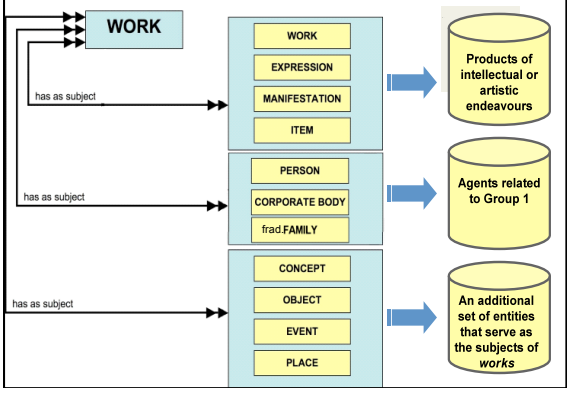
Subject authority systems
are referred to as

- "controlled vocabularies"
- "structured vocabularies"
- "concept schemes"
- "encoding schemes"
- "knowledge organization systems"
- … …

Given the purpose of this report, the discussions about subject authority data apply to all systems and structures referred to by these terms.

The study follows FRBR's approach in that it makes no *priori* assumption about the physical structure or storage of authority data.

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



FRSAR Working Group

FRSAR = Functional Requirements for Subject Authority Records

- **Terms of Reference**
  1. to build a conceptual model of Group 3 entities within the FRBR framework as they relate to the ***aboutness*** of works,
  2. 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
  3. to assist in an assessment of the potential for international sharing and use of subject authority data both within the library sector and beyond.

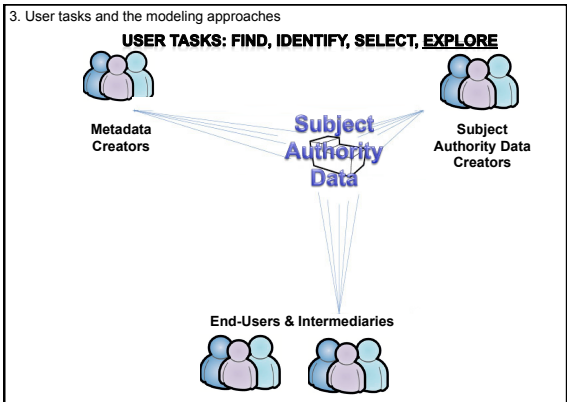
2. Current state

The "FRBR family"

- **FRBR**: the original framework
  - All entities, focusing on Group 1 entities
- **FRAD**: Functional Requirements for Authority Data
  - Focusing on Group 2 entities
  - Published in June 2009
- **FRSAD: Functional Requirements for Subject Authority Data**
  - Focusing on Group3 entities
  - FRSAR WG established in 2005
  - Draft Report on World wide review 2009 June-July
  - Final Report submitted 2010 May

Page 7

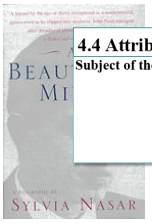
3. User tasks and the modeling approaches



Different ways of modeling – [1]

Metadata Creators

Title: A beautiful mind : a biography of John Forbes Nash, Jr., winner of the Nobel Prize in economics, 1994 /
Author: Sylvia Nasar
Publisher: New York, NY : Simon & Schuster, ©1998.
Subjects: Nash, John F., -- 1928-
Mathematicians -- United States -- Biography.
Schizophrenics -- United States -- Biography.
Call Number: 510.92 Nas-N



4.4 Attributes of a Work -- *FRAD, p.44*

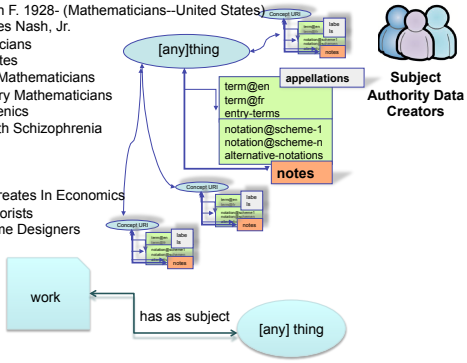
Subject of the work*	The subject aspects of the work and its content. Includes information about the subject of the work. Includes classification numbers.

* represent additions to those identified in *FRBR*

work — Author = … / Publisher = … / Subject = …

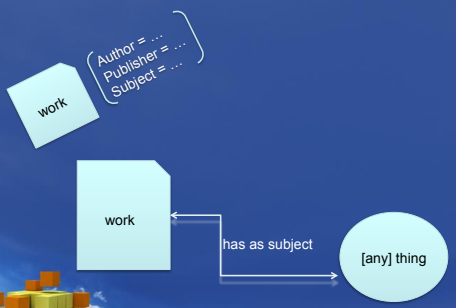
Different ways of modeling – [2]

Nash, John F., -- 1928-
Nash, John F. 1928- (Mathematicians--United States)
John Forbes Nash, Jr.
Mathematicians
United States
American Mathematicians
20th Century Mathematicians
Schizophrenics
People With Schizophrenia
Mentally ill
510.92
QA29.N25
Nobel Laureates In Economics
Game Theorists
Board Game Designers
… …



Subject Authority Data Creators

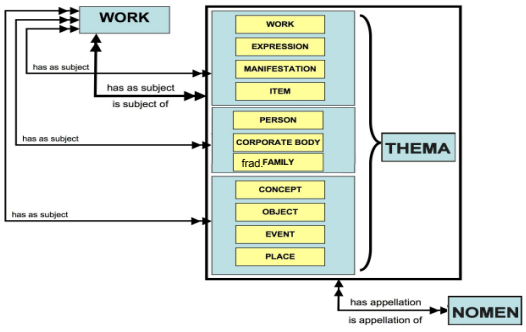
Different ways of modeling



Page 11

4. The model --entities, attributes, relationships

FRSAD's entities within the FRBR framework



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

FRSAD Conceptual Model

work — has as subject / is subject of — thema — has appellation / is appellation of — nomen

Part 1

***WORK* has as subject *THEMA* / *THEMA* is subject of *WORK*.**

work — has as subject / is subject of — thema

Thema = "any entity used as a subject of a *work*".

Different ways of modeling (cont.)

Term / code of [any] thing

SN = …
alternative label = …
term @ fr = …
caption @ zh = …
time of validity = …

[any] thing — has appellation — any sign or sequence of signs

FRSAD Conceptual Model (cont.)

work — has as subject / is subject of — thema — has appellation / is appellation of — nomen

- a new relationship
- *NOMEN* = any sign or sequence of signs (alphanumeric characters, symbols, sound, etc.) that a *thema* is known by, referred to or addressed as.
- In a given controlled vocabulary and within a domain, a *nomen* should be an appellation of only one *thema*,

Part 2

***THEMA* has appellation *NOMEN* / *NOMEN* is appellation of *THEMA*.**

thema — has appellation / is appellation of — nomen

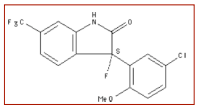
thema — has appellation / is appellation of — nomen

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

Example: one thema, many nomens

different types of nomens

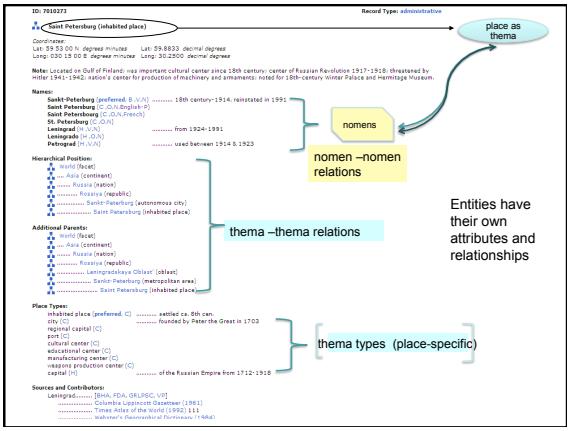
```
Accession Number    (AN): 2005:3738  USAN
Publication Year    (PY): 2002
Generic Name        (CN): Flindokalner
OTHER NAMES:
Chemical Name       (CN): 2H-Indol-2-one, 3-(5-chloro-2-methoxyphenyl)-3-fluoro-
                          1,3-dihydro-6-(trifluoromethyl)-, (3S)-
Chemical Name       (CN): (3S)-3-(5-Chloro-2-methoxyphenyl)-3-fluoro-6-
                          (trifluoromethyl)-1,3-dihydro-2H-indol-2-one
Trade Name          (CN): MaxiPost (Bristol-Myers Squibb)
Code Designation    (CN): BMS-204352
CAS Registry No.    (RN): 187523-35-9
Molecular Formula   (MF): C16 H10 Cl F4 N O2
Lin. Str. Formula   (LSF): C16 H10 Cl F4 N O2
Molecular Weight    (MW): 359.71

Absolute stereochemistry.  Rotation (+).
```



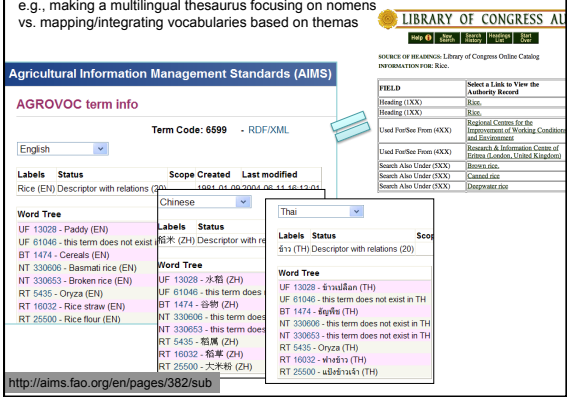
Source: STN Database Summary Sheet: USAN (The USP Dictionary of U.S. Adopted Names and International Drug Names)

The importance of the *THEMA-NOMEN* model to the subject authority data

- Separating what are usually called ***concepts*** (or *topics*, *subjects*, *classes [of concepts]*) from what they are known by, referred to, or addressed as
- A general abstract model, not limited to any particular domain or implementation
- Potential for interoperability within the library field and beyond



Interoperability efforts may focus on different entities.
e.g., making a multilingual thesaurus focusing on nomens vs. mapping/integrating vocabularies based on themas



Knowledge Organization Systems use different kinds of elements to represent *themas*

represented by:

- thesauri: terms (preferred & non-preferred)
- classification schemes: notations
- subject heading systems: terms of pre-coordinated strings
- taxonomies: category labels (w or w/t notations)
- ontologies: terms or identifiers
- picklists: terms
- … …: … …

Attributes

- Some general attributes of *thema* and *nomen* are proposed
  - *thema*: type of thema, scope note
  - *nomen*: various attributes
- In an implementation **additional** attributes may be recorded

thema types

- In an implementation *themas* can be organized based on category, kind, or type.
- This report does not suggest specific types, because they may differ depending on the implementation.

- There seems to be no universal categorization of *themas*.

e.g., models:
- Original FRBR entities
- Original FRBR entities + time
- Ranganathan's PMEST

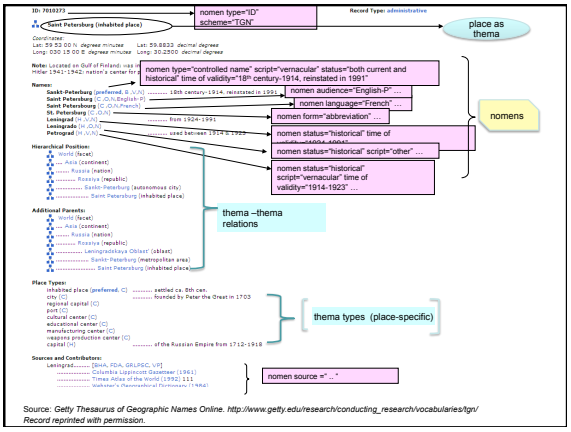
e.g., implementations:
- AAT's seven facets
- FAST's seven subject facets
- UMLS' physical & conceptual entities + events

Any attempt to declare one universal categorisation of *themas* would necessarily limit the usability of a general model.

Nomen attributes and relationships with other entities (include but not limited to)

- Type of nomen (identifier, controlled name, …)*
- Scheme (LCSH, DDC, UDC, ULAN, ISO 8601…)
- Reference Source of nomen (Encyclopedia Britannica…)
- Representation of nomen (alphanumeric, sound, visual,…)
- Language of nomen (English, Japanese, Slovenian,…)
- Script of nomen (Cyrillic, Thai, Chinese-simplified,…)
- Script conversion (Pinyin, ISO 3601, Romanisation of Japanese…)
- Form of nomen (full name, abbreviation, formula…)
- Time of validity of nomen (until xxxx, after xxxx, from… to …)
- Audience (English-speaking users, scientists, children …)
- Status of nomen (provisional, accepted, official,...)

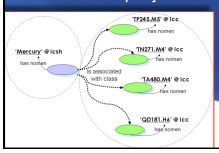
*note: examples of attribute values in parenthesis



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

Thema-to-*thema* relationships

- Hierarchical
  - The generic relationship
  - The whole-part relationship
  - The instance relationship
  - Polyhierarchical Relationship
  - Other hierarchical relationships
- Associative
  - [most commonly considered categories are listed in the report]



Other *thema*-to-*thema* relationships are domain- or implementation-dependent

Nomen-to-*nomen* relationships

- Partitive
  - A *nomen* may have components (parts).
- Equivalence
  - 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").
- The equivalence relationships of *nomens* can be specified further, e.g.,
  - Replaces/Is replaced by
    - [e.g., "integrated plant control" is replaced by "centralized control"]
  - Has variant form/Is variant form
  - Has acronym/is acronym for
    - [e.g., "VS" is acronym for "virtual storage"]
  - Has abbreviation/Is abbreviation of
  - Has transliterated form/Is transliteration of

- based on ISO/CD 25964-1 and NISO Z39.19-2005.

Relationship of FRSAD with FRBR

- The FRSAR Working Group follows FRBR in the methodology, specification, and presentation of entities and relationships.
- The "has as subject" (many-to-many) relationship is kept in its entirety in FRSAD.
- The WG also starts with a user tasks analysis and follows with the establishment of appropriate entities and relationships.

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

Relationship of FRSAD with FRAD

- Independent parallel development; no hierarchical relationship between the two models;
- FRAD was published when FRSAD was in the world-wide review;
- User tasks: "Contextualise" and "Justify" in FRAD vs. "Explore" in FRSAD;
- *Name* in FRAD vs. *Nomen* in FRSAD;
  - *Name*, *Identifier* and *Controlled access point* as separate entities in FRAD vs. values of the attribute "Type of *Nomen"* in FRSAD;
- *Rules* and *Agency* as new entities in FRAD and not explicitly modelled in FRSAD.

The future: harmonization of the FRBR family

- A new working group under the umbrella of the FRBR RG will have to develop a new model, taking FRAD and FRSAD into account

- The FRSAD model is developed with the goal to assist in an assessment of the potential for international sharing and use of subject authority data both within the library sector and beyond.
- The FRSAD model will:
  - enable the consideration of the functions of subject authority data at a higher level that is independent of any implementation, system, or specific context, and
  - allow us to focus on the semantics, structures, and interoperability of subject authority data.
- In the Linked Library Data initiatives, SKOS (Simple Knowledge Organization System) will be widely used to publish subject authority data. FRSAD model maps SKOS well in terms of SKOS classes, attributes, and semantic relationships defined in W3C standard.
  - » frsad:thema = skos:Concept
  - » frsad:nomen = skosxl:Label
  - » SKOS-XL (SKOS eXtension for Labels)