

LexEVS 6.0 Mapping Discussion

Contents of this Page

- [1. Object Model Considerations](#)
 - [1.1 Map Set](#)
 - [1.2 Group of Mappings \(a mapping subset, or search results, or a page of listings\)](#)
 - [1.3 Mapping](#)
 - [1.4 Mapping attributes](#)
- [2. Searching Scenarios](#)
 - [2.1 By name](#)
 - [2.2 By code](#)
 - [2.3 Restrictions](#)
 - [2.4 Default Preferred Names](#)
- [3. Browsing/Discovery Scenarios](#)
 - [3.1 Grouping, Categorization](#)
 - [3.2 Selections](#)
 - [3.3 Misc](#)
- [4. Presentation Scenarios](#)
 - [4.1 Views](#)
 - [4.2 Other Considerations](#)
- [5. Loader Considerations](#)
- [6. Obtaining or Generating Data](#)
 - [6.1 Obtaining](#)
 - [6.2 Generating Mappings](#)
- [7. Maintenance and Legacy Scenarios](#)

1. Object Model Considerations

1.1 Map Set

- Connects two versioned coding schemes
- Is versioned itself.
- Should be named in a descriptive fashion (e.g. MDR12 to SNOMEDCT_2010_07_31 Mappings).
- Significant amount of metadata exists at this level
- Characterize map sets (complexity, completeness, content domain, "officialness", ..)
- Consider metadata to track whether mappings are curated or not.
- Consider metadata to track whether mappings are generated (and not reviewed) or not.
- Map set metadata should indicate whether map rank is important, useful, or even present.

1.2 Group of Mappings (a mapping subset, or search results, or a page of listings)

This is more sophisticated with complex rule-based map sets where a collection of mapping entries may actually represent only a single mapping.

1.3 Mapping

- "from" and "to" codes (or expressions of codes)
- Represented as an association with qualifiers for other semantically important information
- If target is an expression, represented as an "association to data"
- Have "default preferred name" in the event that a "from"/"to" code cannot be resolved in a loaded coding scheme
- "map rank" may be a standard part of the model. If so, values should be normalized, so that 1 always means "best" and increasing numbers represent lowering quality (exactly how much lower and why is map set dependent). A "map rank" threshold can be chosen by an application, so that it pays attention only to the "highest quality" mappings as defined by that application. Because the values are map set and application specific - algorithms/decisions used are use-case specific.

1.4 Mapping attributes

Any information about a mapping than what directly fits the "association" object model will be rendered as attributes. Hopefully these attributes will have standard names across different types of map set loader. In other words – we can define the semantics of mapping association qualifiers so that particular ones are used to always represent the same aspect of some kind of mapping semantics.

2. Searching Scenarios

2.1 By name

- Find mappings where the "from" code has a name matching a search string
- Find mappings where the "to" code has a name matching a search string
- Find mappings where the "from" code has a name matching a search string AND the target concept has a name matching a different string

2.2 By code

- like by name but for matching codes

2.3 Restrictions

- Restrict by vocabulary ("from" vocabulary, "to" vocabulary, or map set vocabulary)
- Restrict by "current" – only retrieve mappings whose endpoints can be resolved in current versions of terminologies
- Restrict by cardinality – e.g. only retrieve 1-1 mappings (that meet other criteria) - (1-1, n-1, 1-n, n-n)
- Restrict by association type – e.g. only retrieve "synonymous" mappings

2.4 Default Preferred Names

- In some cases older versions of map sets will be loaded that contain "to" or "from" codes that cannot be resolved among the versions of loaded vocabularies
- In this case, a "default preferred name" will be available so that browsers can still effectively visualize the data.
- This default name could be indexed so that searches retrieve these results, even though the codes no longer exist

3. Browsing/Discovery Scenarios

3.1 Grouping, Categorization

- How authoritative is it?
- Is it actively maintained?
- Does it connect "current version" terminologies loaded into LexEVS?
- How important or heavily used is it?
- What "kind" of mapping set is it (i.e. what is it for) Group by "from" terminology or "to" terminology

3.2 Selections

- Pre-select certain map sets for searching based on well-defined criteria

3.3 Misc

- When searching across multiple map sets, consider options like grouping all mappings with the same "from" terminology together to make cognitive task easier.
- Support ability to identify map sets that are "to" or "from" a particular terminology
- E.g. get all map sets where the "from" coding scheme is NCI2010_07

4. Presentation Scenarios

4.1 Views

- Map Set view (including all known metadata)
- Map Group view (where needed) – e.g. one "subset" in a complex rule-based mapping.
- Mappings view (sortable table, paging capabilities).
 - Standard column headers come from LexEVS view.
 - Need to consider "extended" columns that may not apply to all map sets (e.g. MAP_RANK, or other MRMAP fields like MAPPRIORITY, etc)
- Individual mapping view – link "from" and "to" codes to the concept pages for those things – may not need this
- "Mappings" tab on concept view – render all mappings "to" or "from" that concept
 - need to determine which map sets to search in
 - retrieve this info only when user actually clicks on this tab

4.2 Other Considerations

- Make good decisions about when clicking should open a new window vs. reload in the same page.
- Support a "view all mappings" function for simple browsing of small sets (without requiring search).
- Handling expression-based mappings (may need to parse – can do based on grammar or "style").
- Consider when a new page should open and when content should be reloaded in the current tab/browser window.

5. Loader Considerations

- Need to specify map set metadata (switches or prefs file)
- Need to indicate end point coding schemes for "from" and "to" codes.
- If we know they cannot be resolved, placeholder names for endpoints should be provided and loaded.
- WARNING: be careful using Excel to export to CSV because of conversion things like "078.12" to "78.12"
- Consider loading a characterization of the "type" or "category" of a map set for organization in a list and pre-selection (see presentation)

- Consider how many map sets to load into a single coding scheme, and what criteria would be used to make that decision.
- Handle cases where we know the "from"/"to" codes cannot be resolved due to different versions and load default preferred names.

6. Obtaining or Generating Data

6.1 Obtaining

- MRMAP data from UMLS (e.g. SNOMEDCT-ICD9CM)
- CTCAE 3->4 mappings
- Go/BiomedGT Mappings
- Identify major mapping efforts, start gathering data – formats can drive further loader considerations

6.2 Generating Mappings

- Data sets created as part of NCI-META export (e.g. PDQ-NCIt)
- Map sets generated as a special project (e.g. NCIt Neoplasms to SNOMEDCT)
- On-demand generation of mappings from an integrated terminology resource (like UMLS or NCI-META).
 - Criteria-based candidate selection
 - Criteria-based ranking and filtering
 - See report (posted to gforge)
 - API performance considerations.

7. Maintenance and Legacy Scenarios

It may be important to know what map sets have mappings whose "from" or "to" codes will not resolve.

Managing versions:

- "from" coding scheme,
- "to" coding scheme
- Map set coding scheme itself

Learn what we need to from various mapping maintenance environments about authoring, data models, visualization, etc.

- IHTSDO stand-alone tool (and eventual workbench)
- CogZ (protégé tool)
- Oboedit
- Ad-hoc data manipulation