Exchanging ISO 25964-1 thesauri data using RDF, SKOS and SKOS-XL

Reported by Johan De Smedt

TPDL – 2012-09-27 - workshop

Presentation Overview

- Objective and rationale of the mapping
- Thesaurus of Concepts
- Concept relationships
  - custom concept attributes
- Terms and Term relations
  - custom term attributes
- Array
- Group
- Notes on concepts and on terms
  - hyperlinks in notes
- Thesaurus and data-set version history

Objective (1/2)

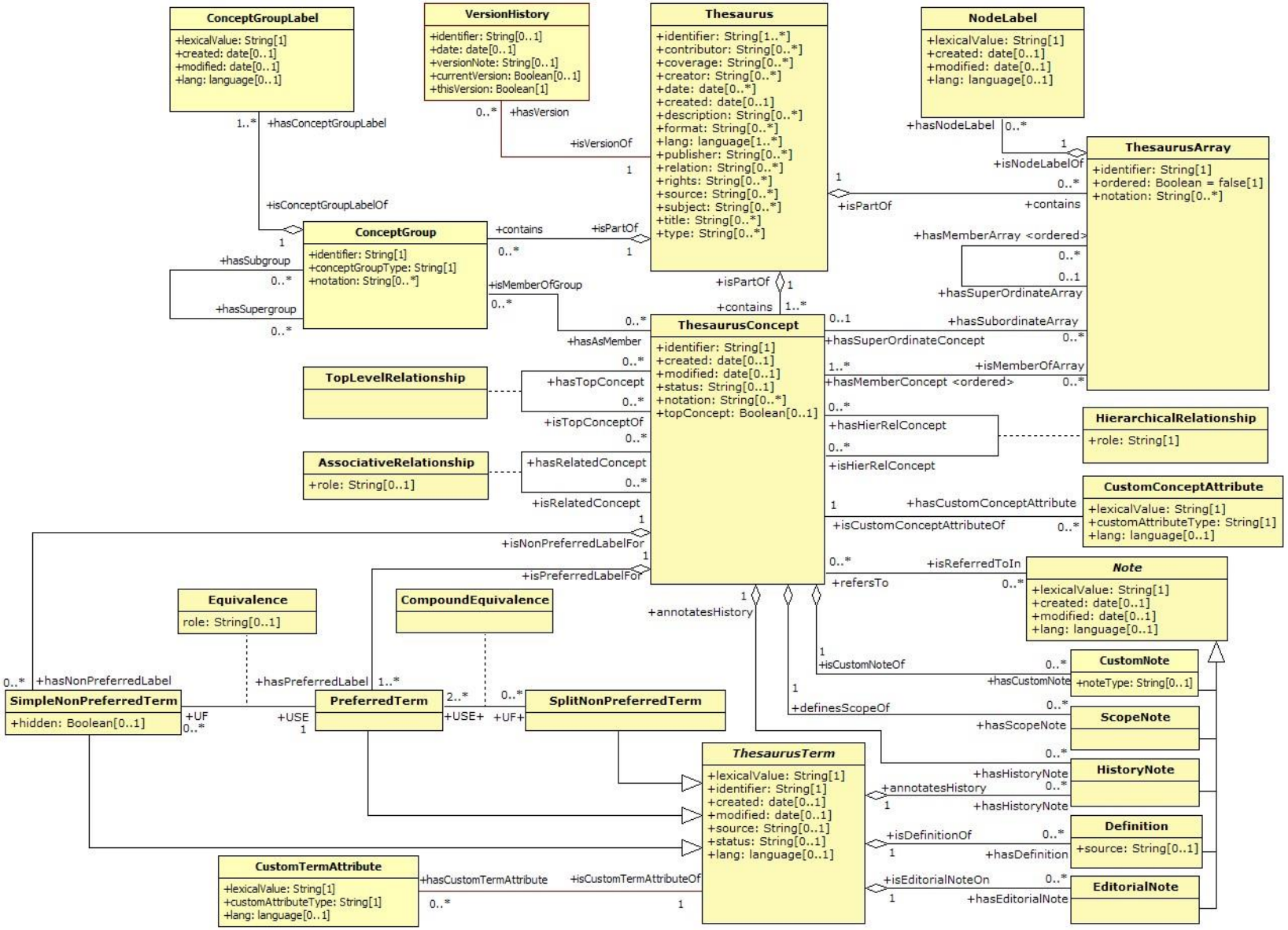
- Make correction and/or update to the Appendix "Correspondences between ISO-2788/5964 and SKOS constructs" of the SKOS Simple Knowledge Organization System Primer
  - The update is needed because ISO 25964-1:2011 has been published, replacing the earlier ISO standards ISO 2788:1986 and ISO 5964:1985
- Provide a reference framework facilitating integration and data exchange of ISO 25964 thesaurus data using RDF as a representation language
- Approach
  - **Minimal extensions** are made, using SKOS, SKOS-XL and Dublin Core where possible
- Contributors
  - ISO TC46 WG 8 working on the ISO- 25964 standard about Thesauri
    - Stella Dextre Clarke, Jutta Lindenthal, Marcia Lei Zeng, Johan De Smedt, Douglas S. Tudhope, Leonard Will
  - Antoine Isaac: Co-Editor of the SKOS Recommendations and **MADS/SKOS mapping**

Objective (2/2)

- Methodology and Levels of representation compliance
  - Always align with the specified SKOS semantics
    - SKOS semantics are a top level agreement for sharing KOS information
      - Not requiring label relations
      - Limited concept relationships
  - Always align with the SKOS-XL semantics
    - SKOS-XL semantics are a top level agreement for sharing KOS information
      - Simple label relations
  - Keep with the iso-25964 proposed extensions patterns
    - Compound and simple label relationships, thesaurus evolution, arrays, groups, notes and facets

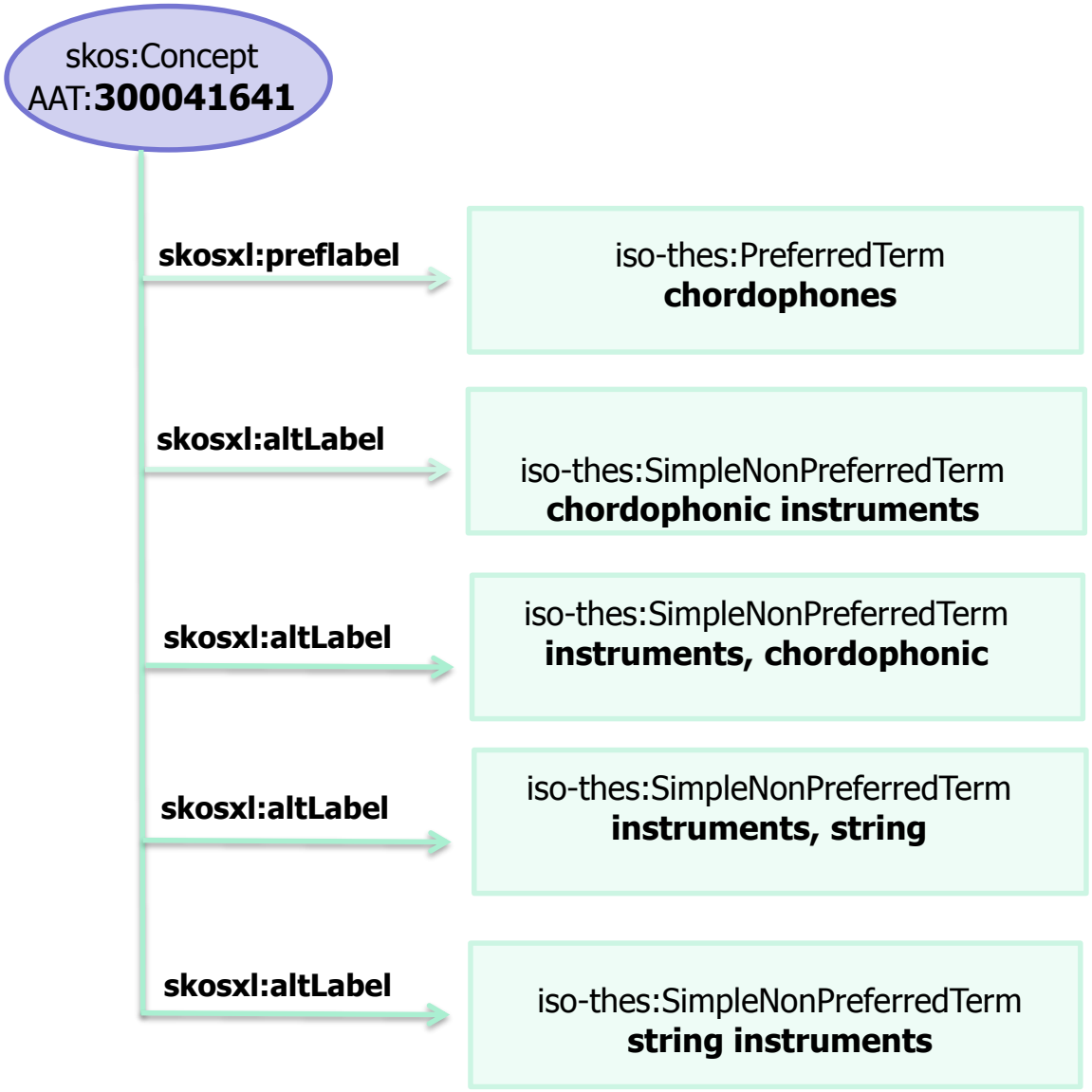
- Rationale
  - Provide general representation paradigms
    - **To date, varying implementations are found for label relations, concept relationships, arrays and groups.**
    - Example illustrations of varying mappings for same abstractions are presented in the slide stack
    - Jutta Lindenthal's presentation illustrates ways of dealing with features that are sometimes found in thesauri although not recommended in ISO 25964.
  - Allow for maximal tooling, support and exchange
    - Application specific extensions limit the use of the exchanged information to users knowing the application (profile)

ISO-25964 mapping reference



See http://www.willpowerinfo.co.uk/Will_ISKO2012_paper.doc and http://www.willpowerinfo.co.uk/LWill-ISKO2012.pdf for an introduction.

ThesaurusTerm (2/8 - equivalence example)



(Art & Architecture Thesaurus)

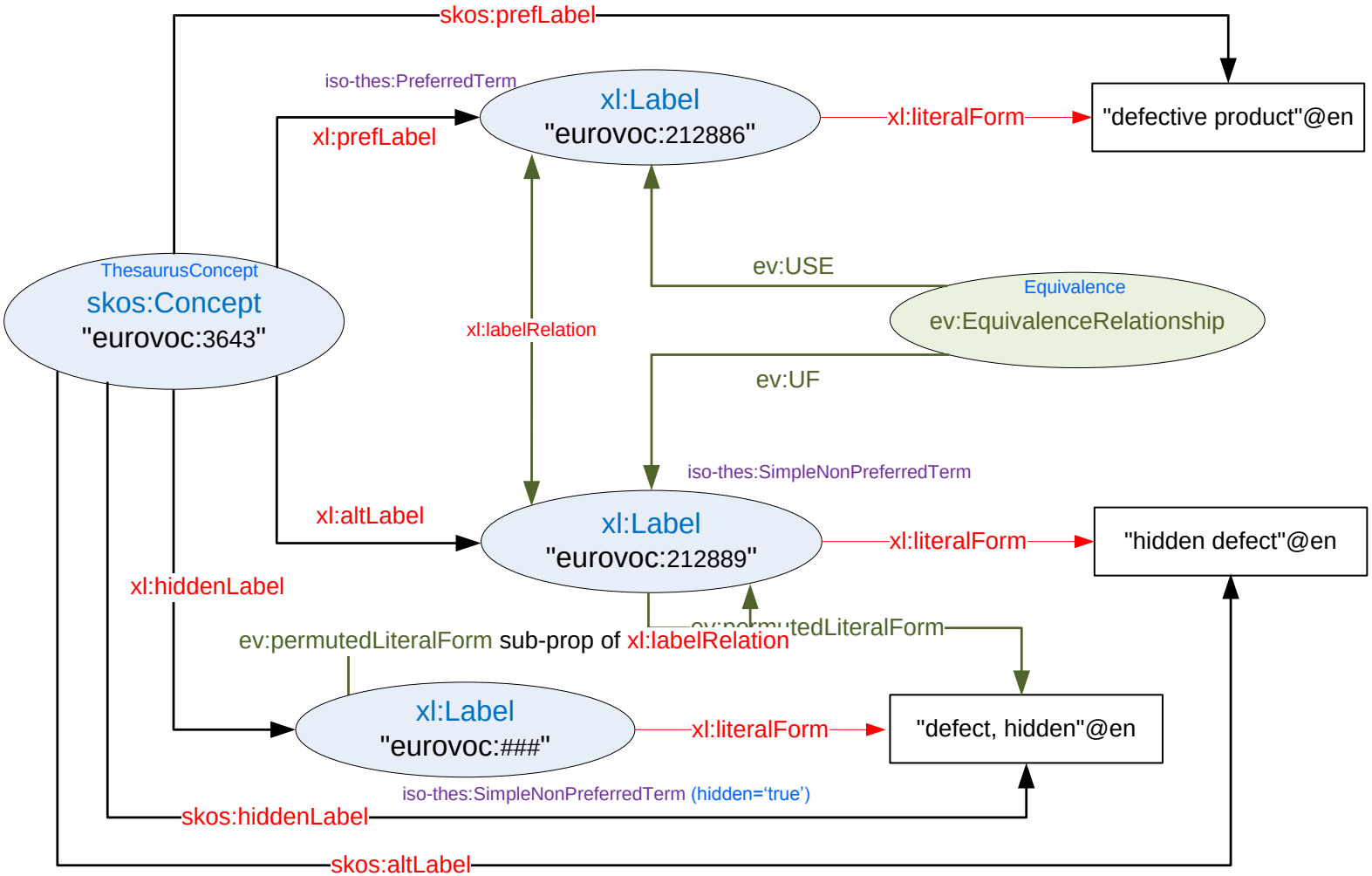
ISSUE 1: equivalence as inferred relation

ISSUE 2: semantics of sub-properties of xl:labelRelation

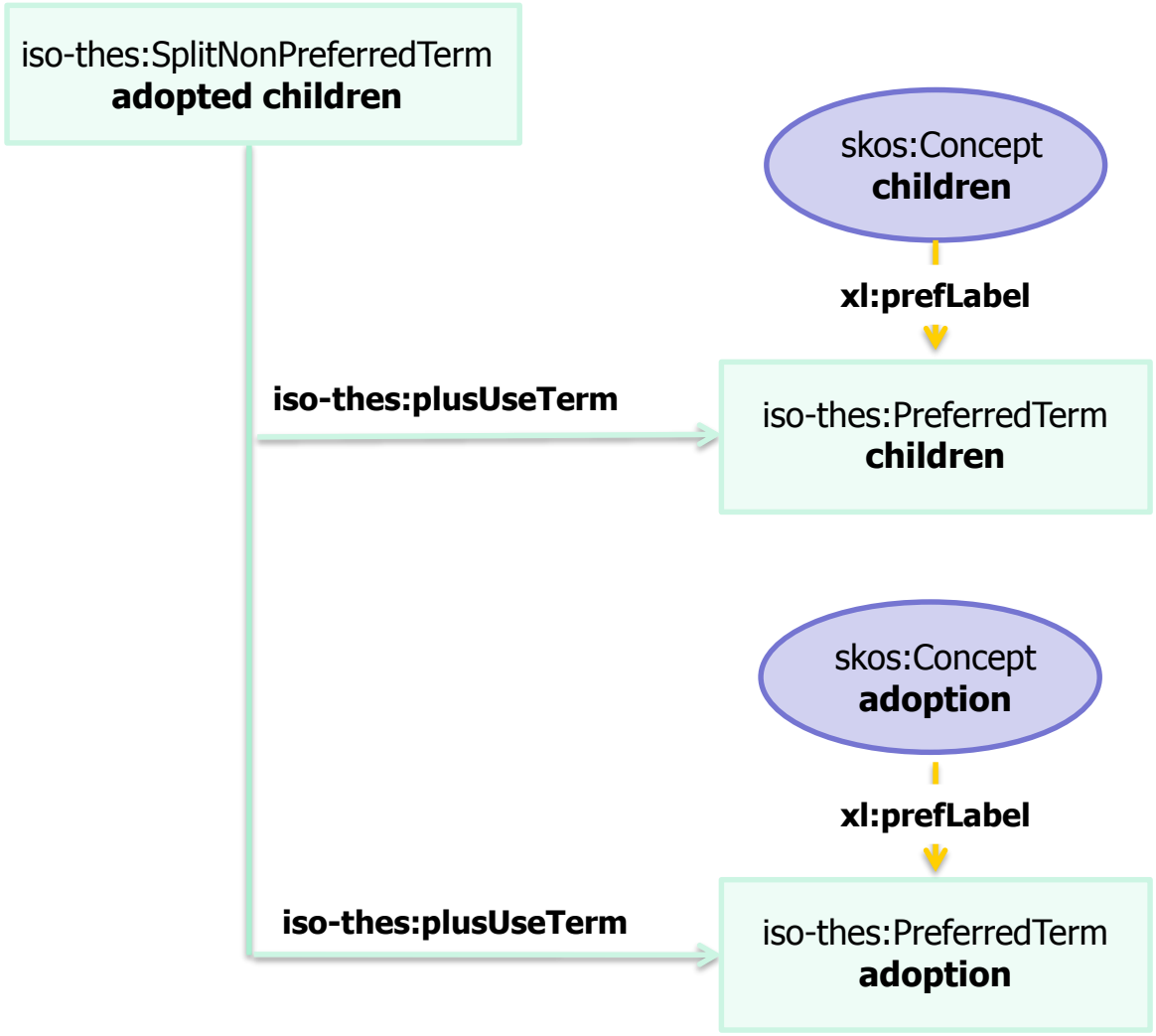
Legend:
iso 25964, skos or skos-xl classes,
skos and skos-xl properties
iso-thes extension to skos or skos-xl
Application extension (specialization)
- Eurovoc specific
- Elaborate alternative (more skos semantics)

ThesaurusTerm (4/8)

complex example EuroVoc



ThesaurusTerm (6/8 -compound equivalence example)

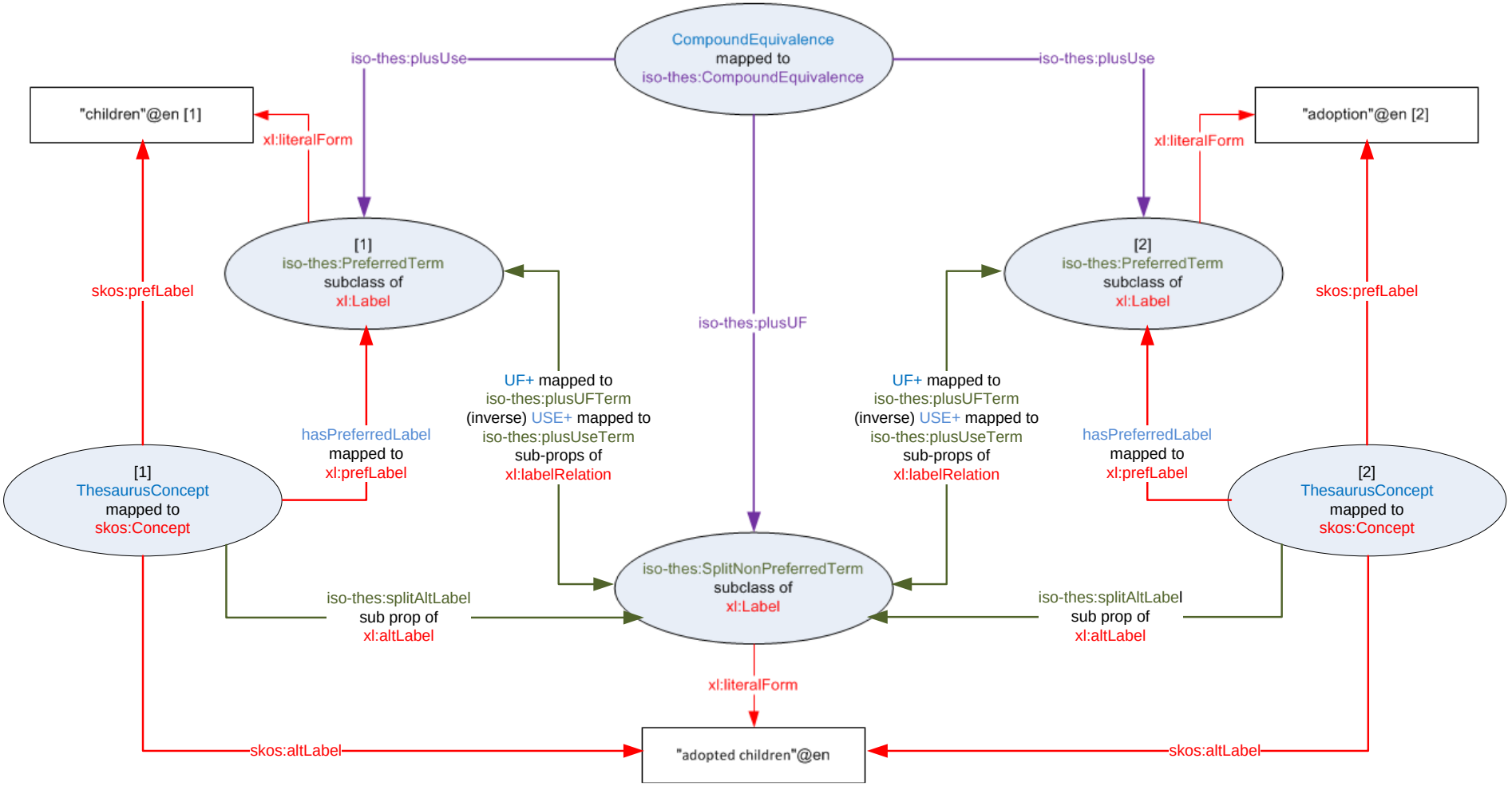


Example: Thesaurus for the Social Sciences (Thesaurus Sozialwissenschaften)
http://www.gesis.org/sowiport/suche/thesaurus.html

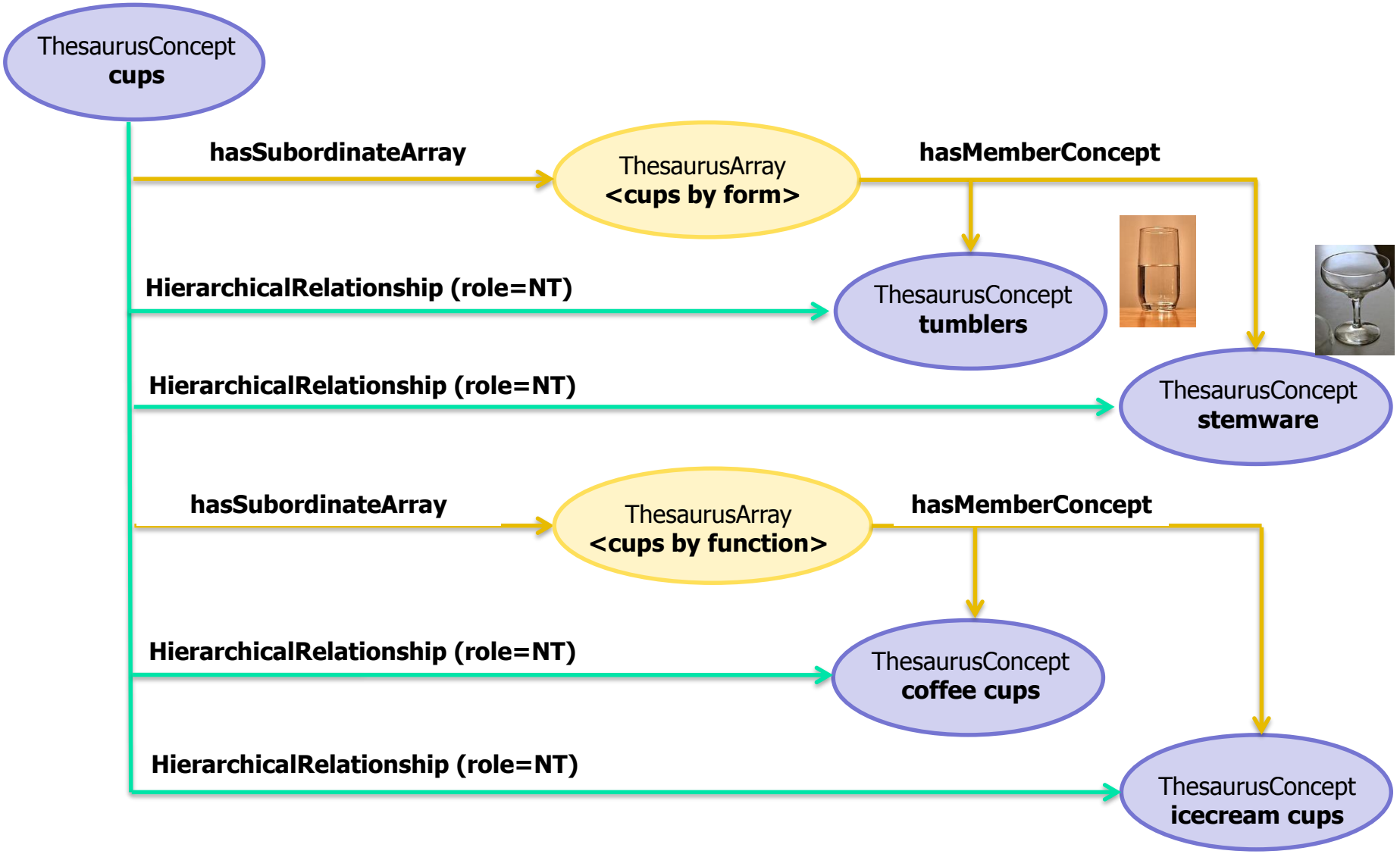
ThesaurusTerm (7/8-compound equivalence example)

"adopted children" ↠ "children" [1], "adoption" [2]

Legend:
mapped ISO-25964 aspect
skos and skos-xl specification
iso-thes extension to skos/skos-xl
- compound equivalence (new)
- compound equivalence (specialization)

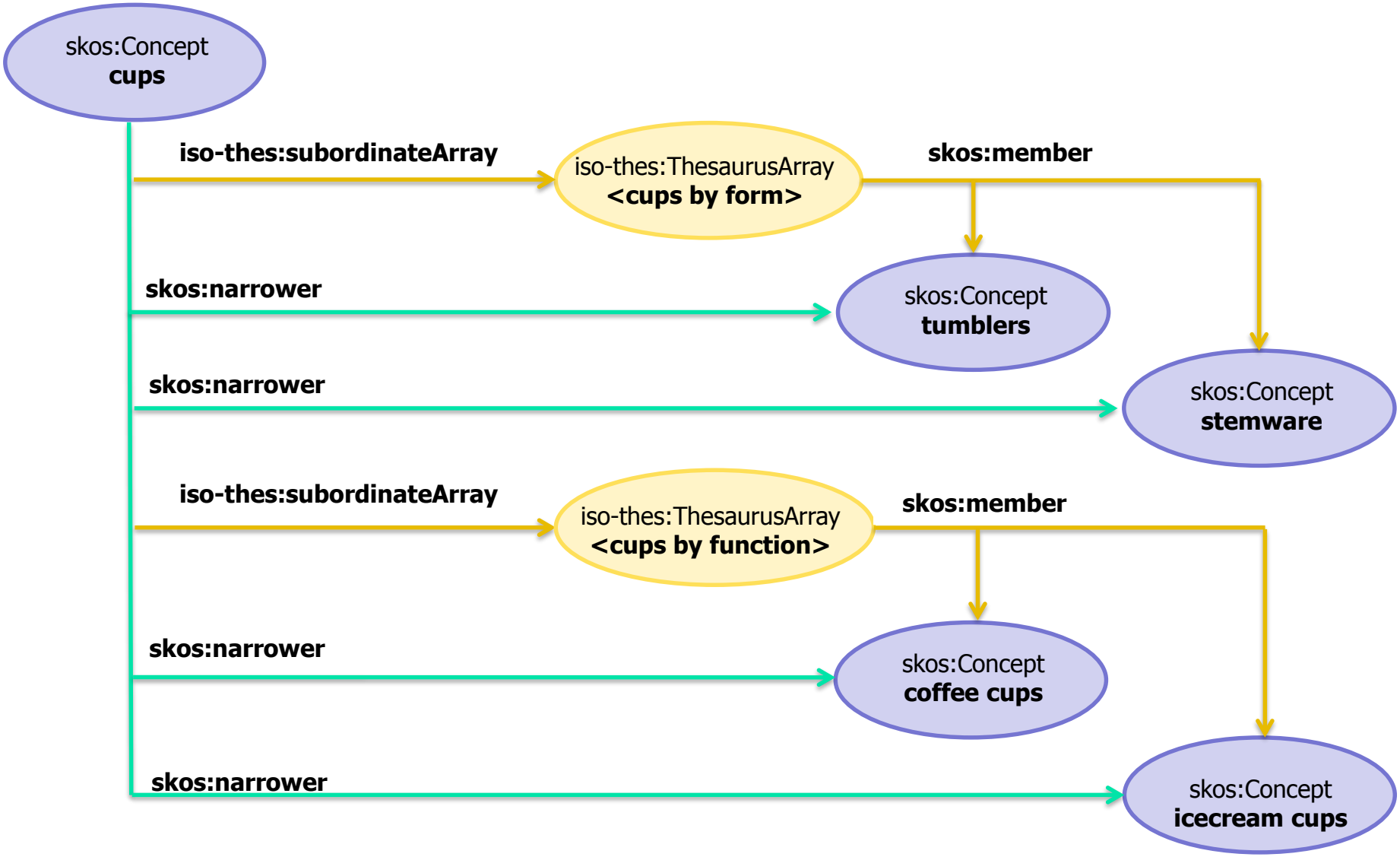


ThesaurusArray (1/4 – not ordered)

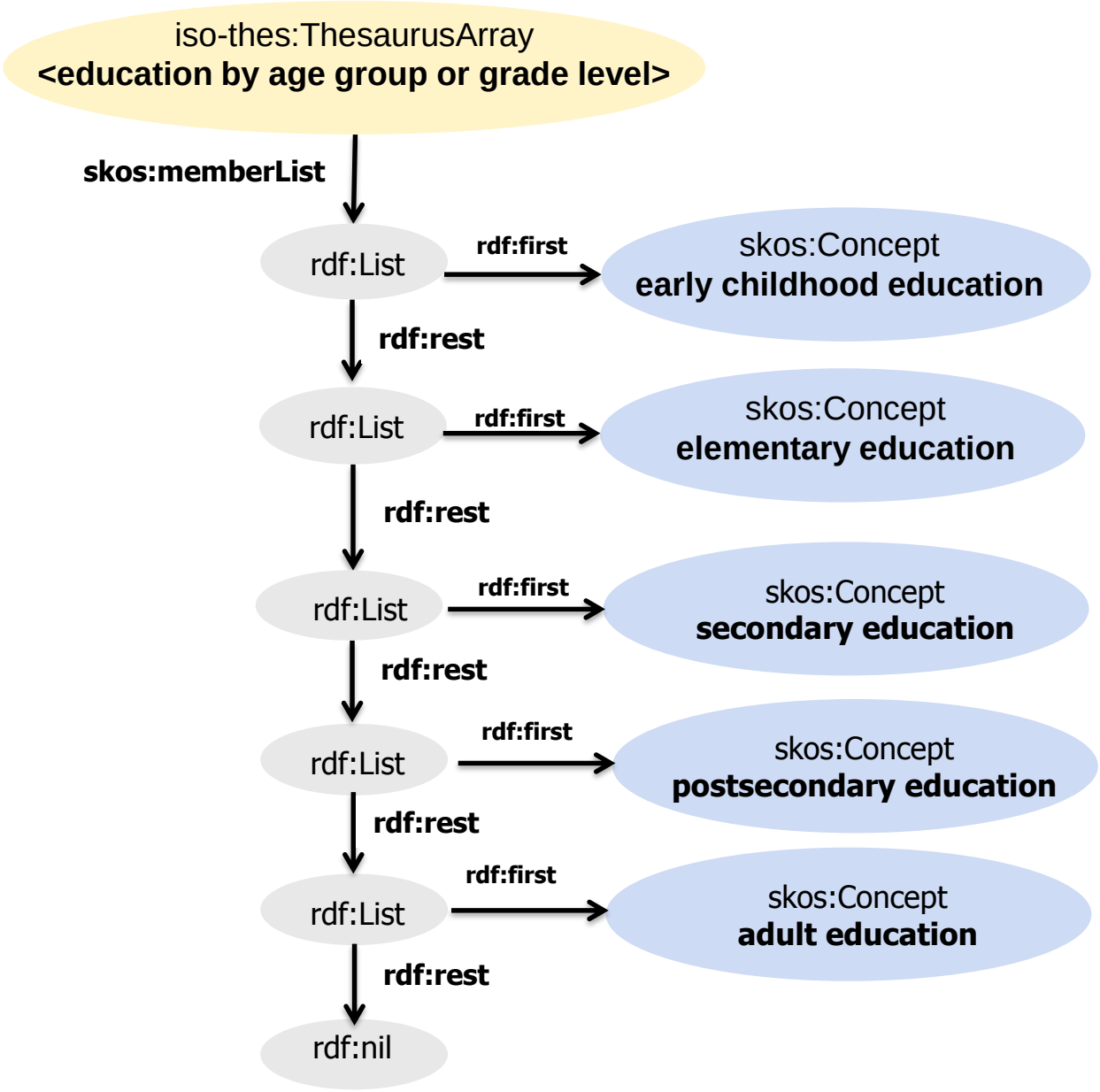


images from wikipedia

ThesaurusArray (1/4 – not ordered)

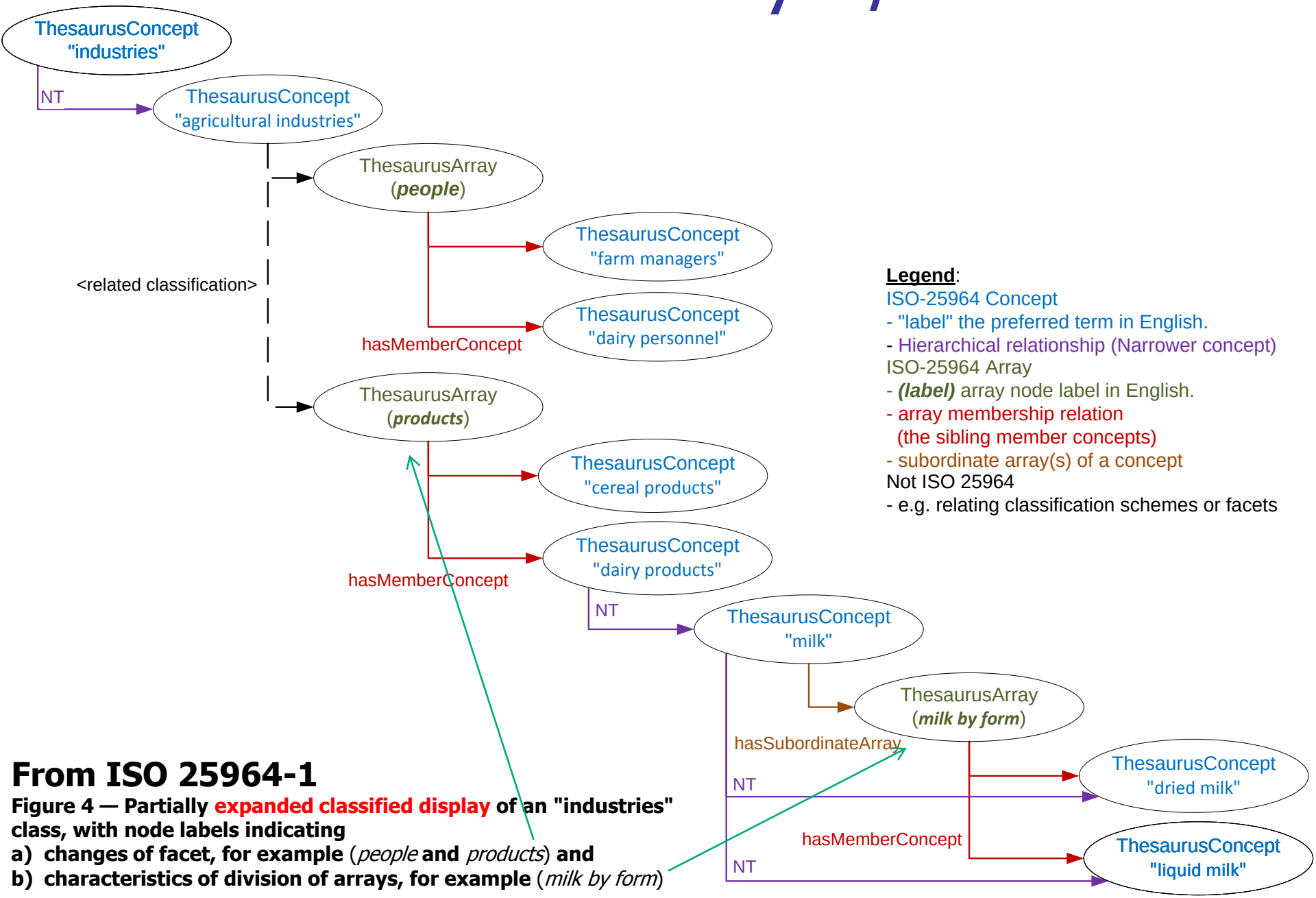


ThesaurusArray (2/4 – ordered)



(Art & Architecture Thesaurus)

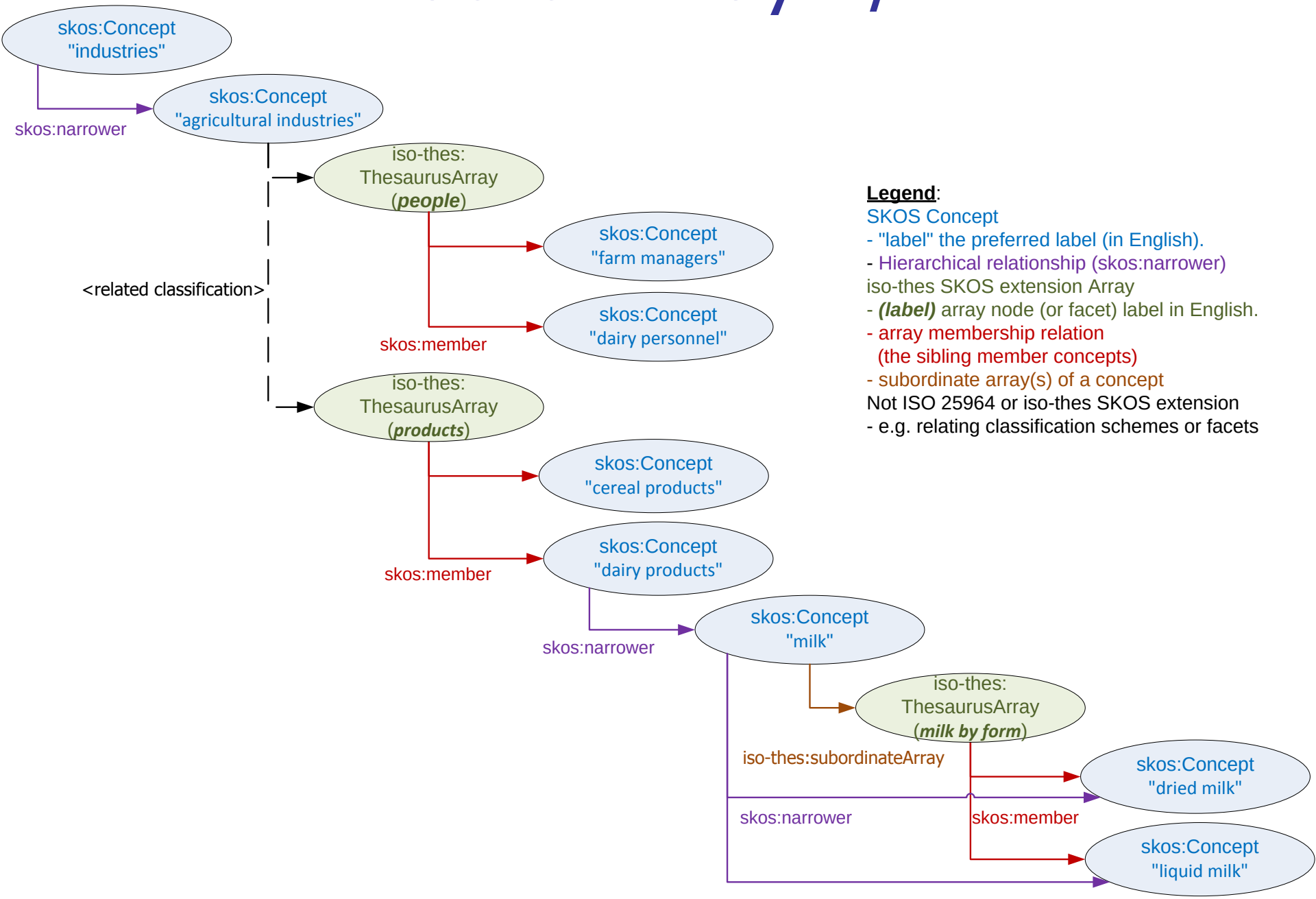
ThesaurusArray 3/4 – facet vs narrower



From ISO 25964-1

Figure 4 — Partially expanded classified display of an "industries" class, with node labels indicating
a) changes of facet, for example (*people* and *products*) and
b) characteristics of division of arrays, for example (*milk by form*)

ThesaurusArray 4/4 – facet vs narrower



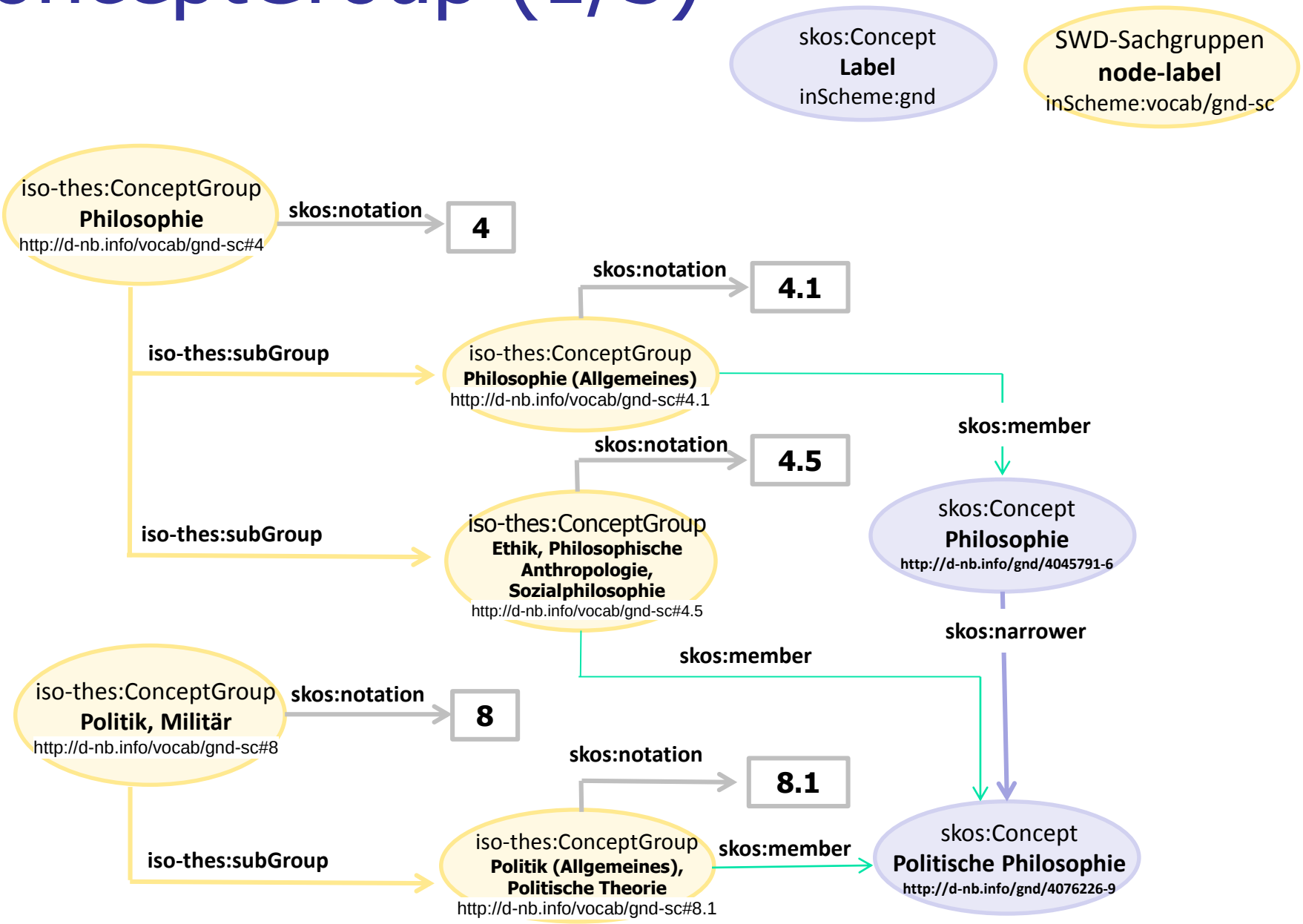
Legend:
SKOS Concept
- "label" the preferred label (in English).
- Hierarchical relationship (skos:narrower)

iso-thes SKOS extension Array
- ***(label)*** array node (or facet) label in English.
- array membership relation
  (the sibling member concepts)
- subordinate array(s) of a concept

Not ISO 25964 or iso-thes SKOS extension
- e.g. relating classification schemes or facets

ConceptGroup (1/3)

(Deutsche Nationalbibliothek)
ConceptScheme „Gemeinsame Normdatei (GND)“

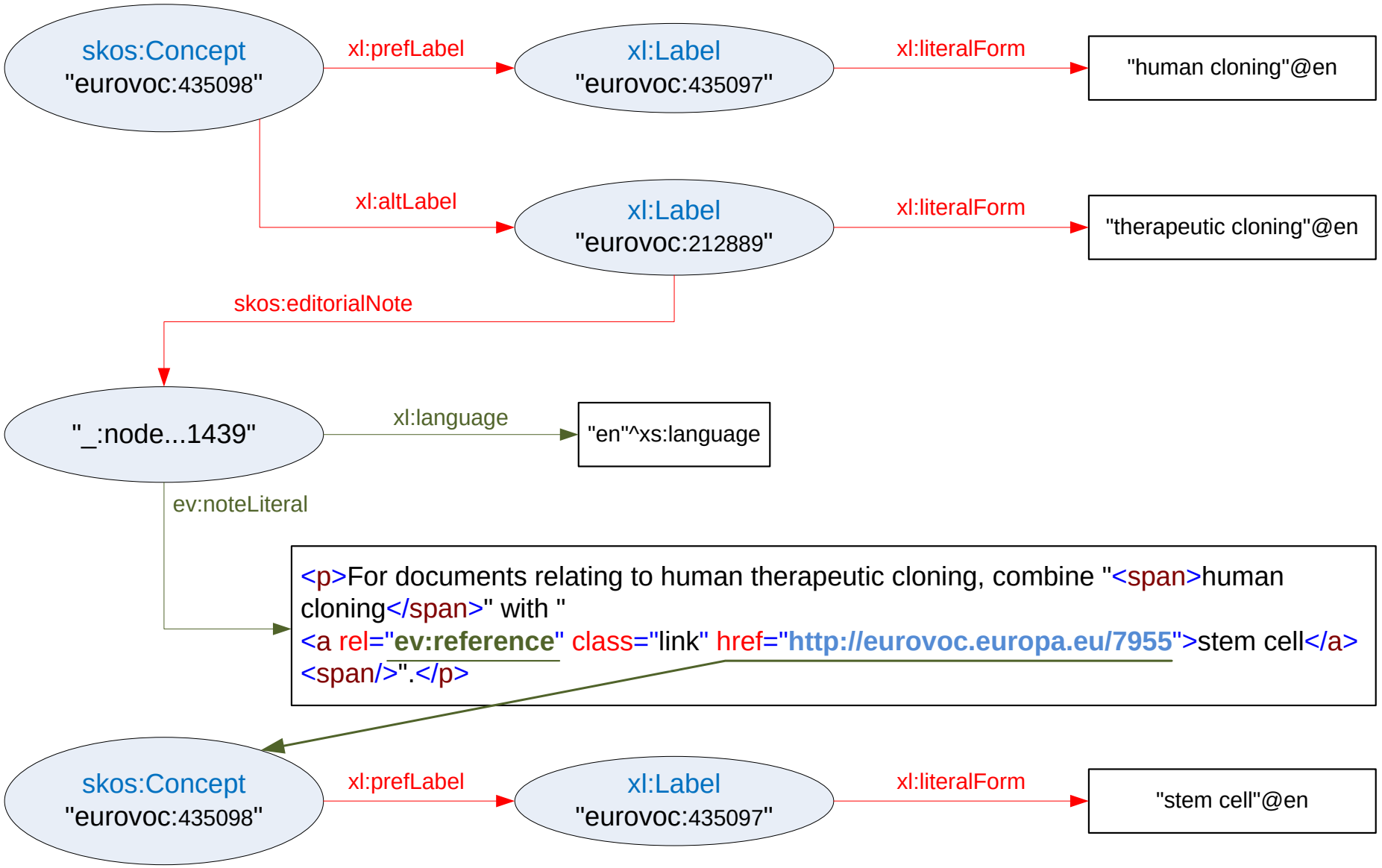


ConceptGroup (3/3)

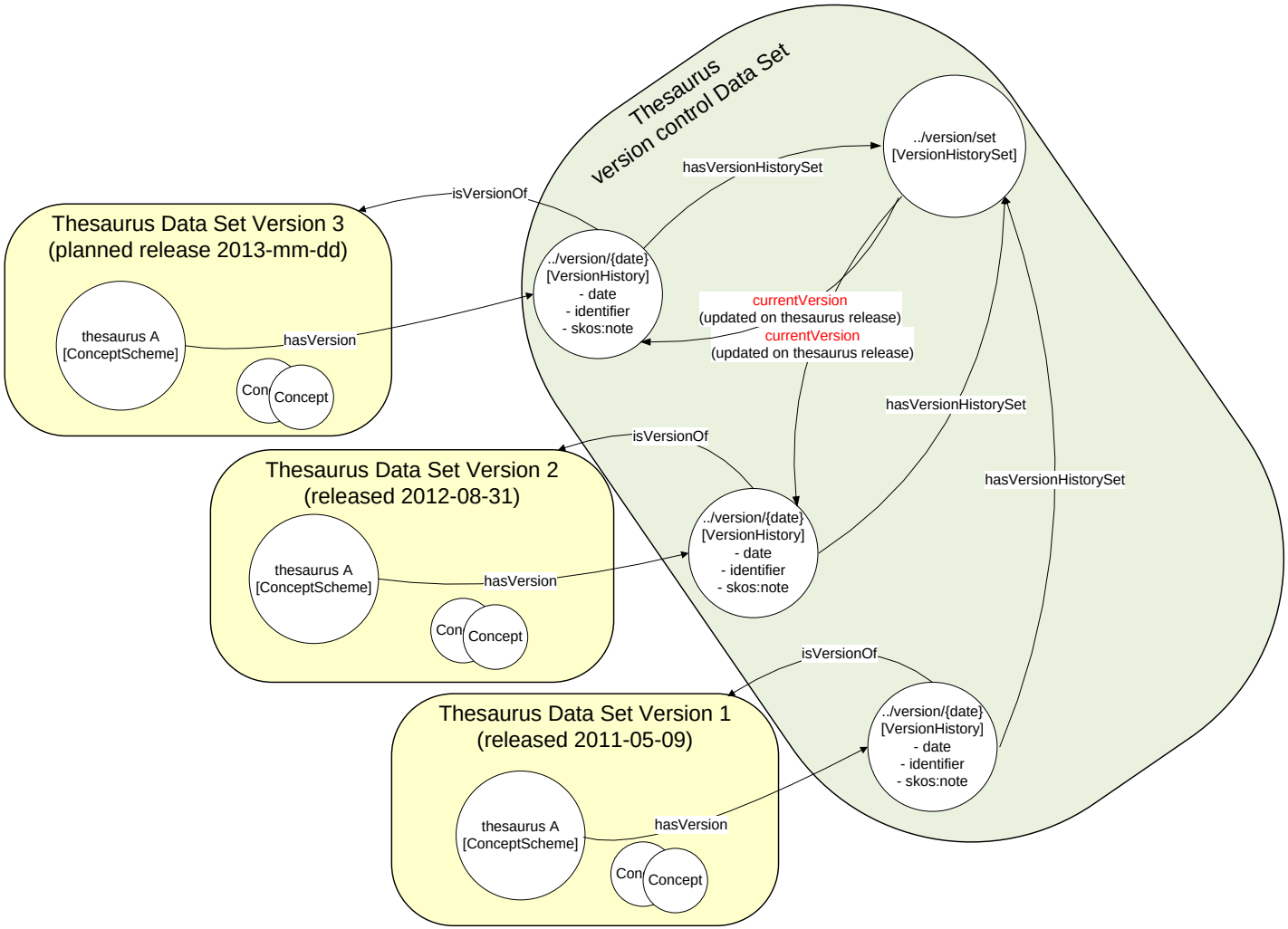
ISO-25964	SKOS/SKOS-XL/extension	Remark
ThesaurusGroup	iso-thes:ConceptGroup	a subclass of skos:Collection
- hasSupergroup	iso-thes:superGroup	domain = range = iso-thes:ConceptGroup
- hasSubGroup	iso-thes:subGroup	inverse of iso-thes:superGroup
- hasAsMember	skos:member	
- isPartOf	skos:inScheme	
- notation	skos:notation	
e.g. conceptGroupType micro-thesaurus	iso-thes:microThesaurusOf a sub-property of skos:inScheme	domain = iso-thes:ConceptGroup range = skos:ConceptScheme

- The new class is required to distinguish between ThesaurusArray (having sibling concepts) and ThesaurusGroup (grouped concepts need not be sibling concepts)

Notes (2/2 – eurovoc example)



Thesaurus versions and versioned data sets (proposal)



Want a copy of ISO 25964-1 ?

- Download it from ISO at

  http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=53657

- Order it from your national standards body (e.g. BSI, DIN, ANSI, AFNOR)

- Some public/academic reference libraries may stock it

- The XML schema (http://www.niso.org/schemas/iso25964/schema-intro/) and the SKOS/SKOS-XL mapping (http://www.niso.org/schemas/iso25964/#schema) for exchange of thesaurus data is in an Annex which is available online without charge or password control.

  Go to http://www.niso.org/schemas/iso25964/

References

- DD8723-5, Data Model for BS 8723. In: *Structured vocabularies for information retrieval. Part 5. Exchange formats and protocols for interoperability*. London: British Standards Institution. Available at: http://schemas.bs8723.org/Model.aspx
- ISO 25964-1:2011, *Thesauri and interoperability with other vocabularies. Part 1: Thesauri for information retrieval*. Geneva: International Organization for Standards, August 8, 2011.
- *ISO 25964-1 Schema and Data Model*. http://www.niso.org/schemas/iso25964/#schema
- W3C Recommendation, *SKOS Simple Knowledge Organization System Reference.* W3C Recommendation, August 18, 2009. Alistair Miles and Sean Bechhofer, eds. Available at: http://www.w3.org/TR/skos-reference/
- W3C Recommendation. *SKOS eXtension for Labels (SKOS-XL)*. In: SKOS Simple Knowledge Organization System Reference, Appendix B. W3C, August 18, 2009. Available at: http://www.w3.org/TR/2009/REC-skos-reference-20090818/#xl
- Correspondences between ISO-2788/5964 and SKOS constructs. In: SKOS Simple Knowledge Organization System Primer. W3C Working Group Note 18 August 2009. Antoine Isaac, A. and Ed Summers, eds. Available at: http://www.w3.org/TR/skos-primer/#seccorrespondencesISO
- Thesaurus UML Model introduction by Leonard Will: http://www.willpowerinfo.co.uk/Will_ISKO2012_paper.doc and http://www.willpowerinfo.co.uk/LWill-ISKO2012.pdf for an introduction.
- The "GND-Systematik" - Deutsche Nationalbibliothek (Leipzig, Frankfurt am Main), (ISBN 978-3-941113-33-6), available at http://d-nb.info/1018626042/34.

- Detailed documentation about the "Correspondence between ISO 25964, SKOS/SKOS-XL and MADS Models": http://www.niso.org/schemas/iso25964/#schema
- This slide pack: Conference proceeding