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)

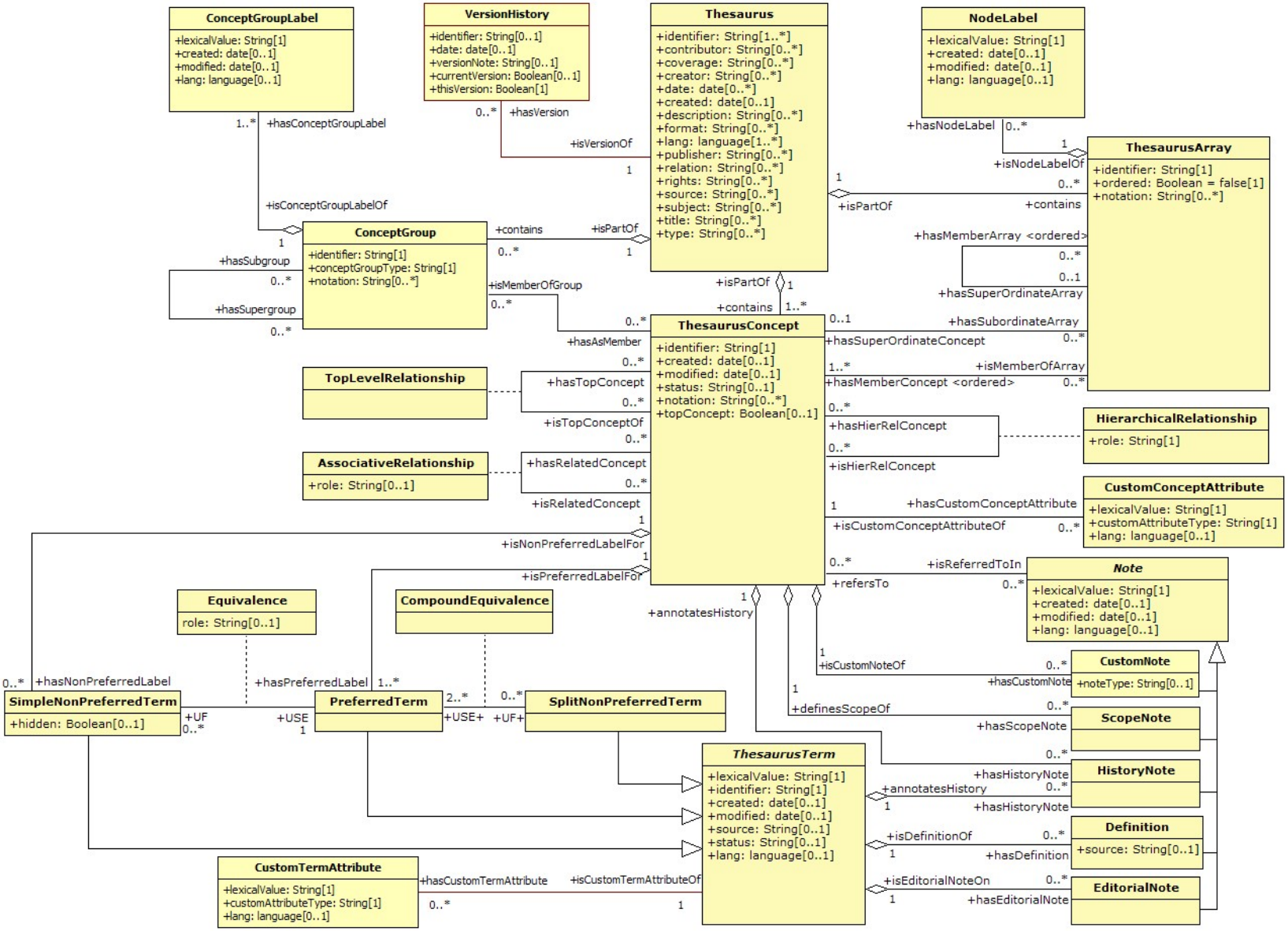
- Make correction and/or update to the Appendix "Correspondences between ISO-2788/5964 and SKOS constructs" of the SKOS Simple Knowledge Organization System Primer
  - The update is needed because ISO 25964-1:2011 has been published, replacing the earlier ISO standards ISO 2788:1986 and ISO 5964:1985
- 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.

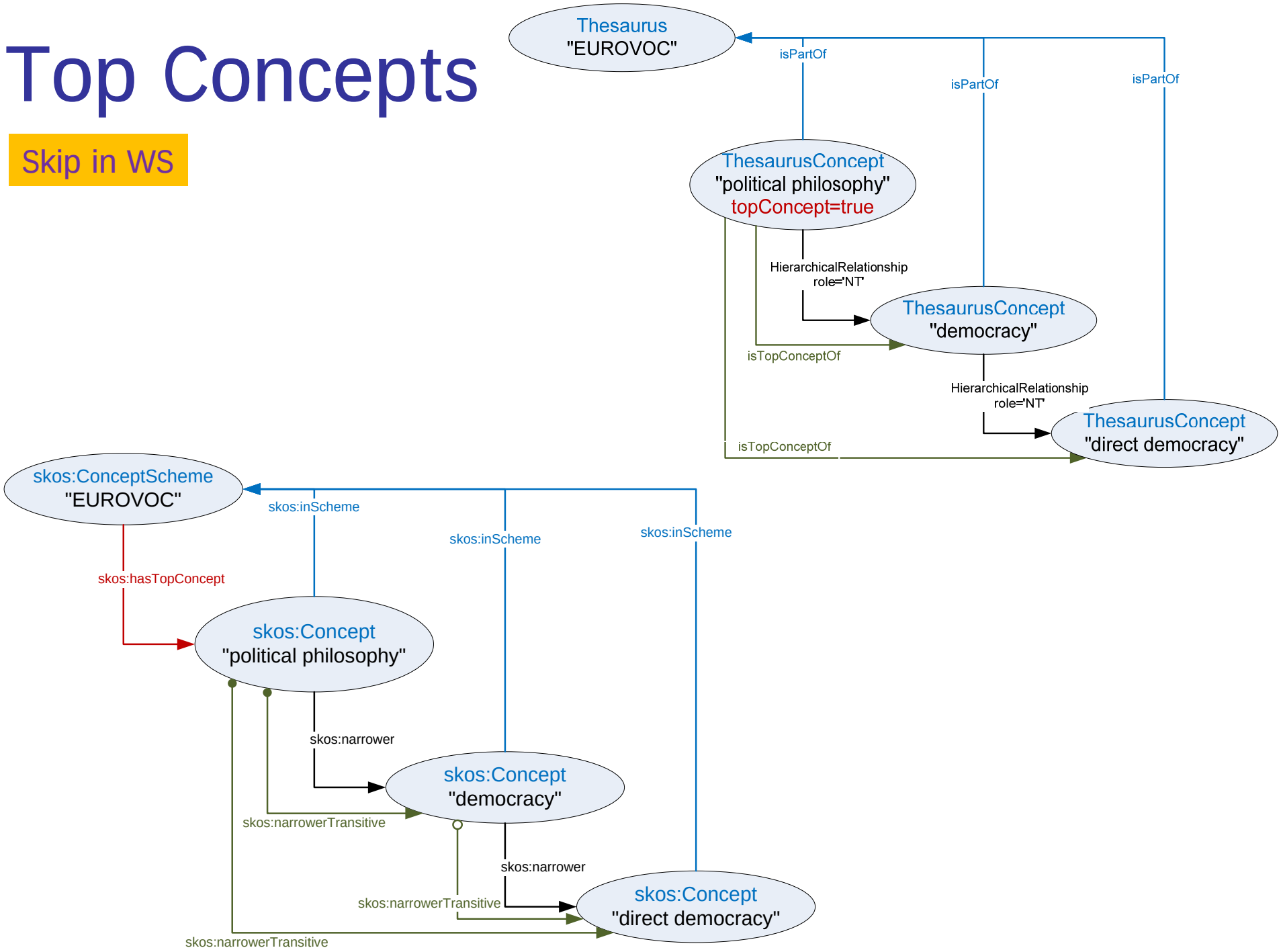
Skip in WS

Thesaurus

ISO-25964	SKOS/SKOS-XL/extension	Remark
Thesaurus	skos:ConceptScheme	
isPartOf (Thesaurus)	skos:inScheme	
- hasTopConcept (TopLevelRelationship)	Not explicitly mapped. The relationship may be derived (see remark)	restriction to the sub-properties of skos:broaderTransitive having as range the domain of skos:topConceptOf
Example		

Top Concepts

Skip in WS



Skip in WS

ThesaurusConcept (1/2)

ISO-25964	SKOS / SKOS-XL / extension	Remark
ThesaurusConcept	skos:Concept	
- status	iso-thes:status	Proposed extension
- isPartOf (Thesaurus)	skos:inScheme	
- notation	skos:notation	The datatype of the notation range is set to distinguish different types of notations
ThesaurusConcept[topConcept=true]		Has special attribute mapping (see next 2 rows)
- isPartOf (Thesaurus)	skos:topConceptOf	
- isTopConceptOf (TopLevelRelationship)	Not explicitly mapped. The relationship may be derived (see remark)	sub-properties of skos:narrowerTransitive having the same domain as skos:topConceptOf

Skip in WS

ThesaurusConcept (2/2)

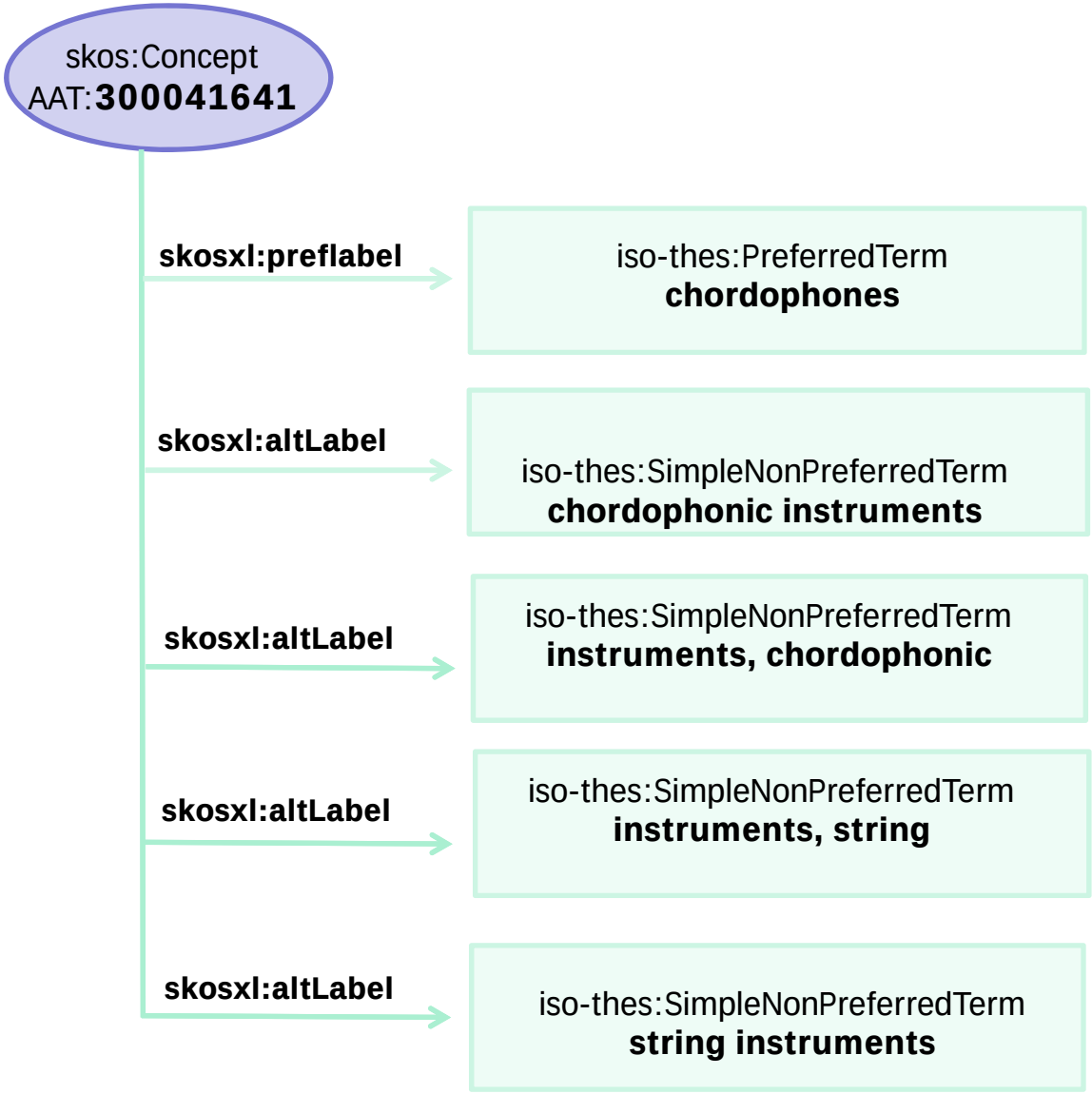
ISO-25964	SKOS/SKOS-XL/extension	Remark
HierarchicalRelationship [role]	skos:broader skos:narrower	sub-properties may be needed to model different hierarchical relationships identified by the attribute "role" (e.g. BTP/NTP, BTI/NTI, BTG/NTG)
AssociativeRelationship [role]	skos:related	sub-properties may be needed to model different associative relationships identified by the attribute "role"
CustomConceptAttribute [customAttributeType]	custom RDF property	The property name depends on the customAttributeType

Skip in WS

ThesaurusTerm (1/8)

ISO-25964	SKOS/SKOS-XL/extension	Remark
ThesaurusTerm	- rdf:PlainLiteral - xl:Label	- preferred simple mapping - extended mapping to handle label relationships
PreferredTerm	iso-thes:PreferredTerm	The restriction of xl:Label to the range of xl:prefLabel is only required for expressing label relations (see ConpoundEquivalence)
(ThesaurusConcept) hasPreferredLabel (PreferredTerm)	- skos:prefLabel - xl:prefLabel	
CustomTermAttribute [customAttributeType]	requires xl:Label mapping custom RDF property	The property name depends on the customAttributeType

ThesaurusTerm (2/8 - equivalence example)



(Art & Architecture Thesaurus)

Skip in WS

ThesaurusTerm (3/8)

ISO-25964	SKOS/SKOS-XL/extension	Remark
SimpleNonPreferredTerm	iso-thes: SimpleNonPreferredTerm	A sub-class of xl:Label and of the union of: the range of xl:altLabel the range of xl:hiddenLabel
ThesaurusConcept - hasNonPreferredLabel (SimpleNonPreferredTerm [hidden=false])	- skos:altLabel - xl:altLabel	The inverse mapping requires the distinction - SimpleNonPreferredTerm - SplitNonPreferredTerm (see CompoundEquivalence)
ThesaurusConcept - hasNonPreferredLabel (SimpleNonPreferredTerm [hidden=true])	- skos:hiddenLabel - xl:hiddenLabel	

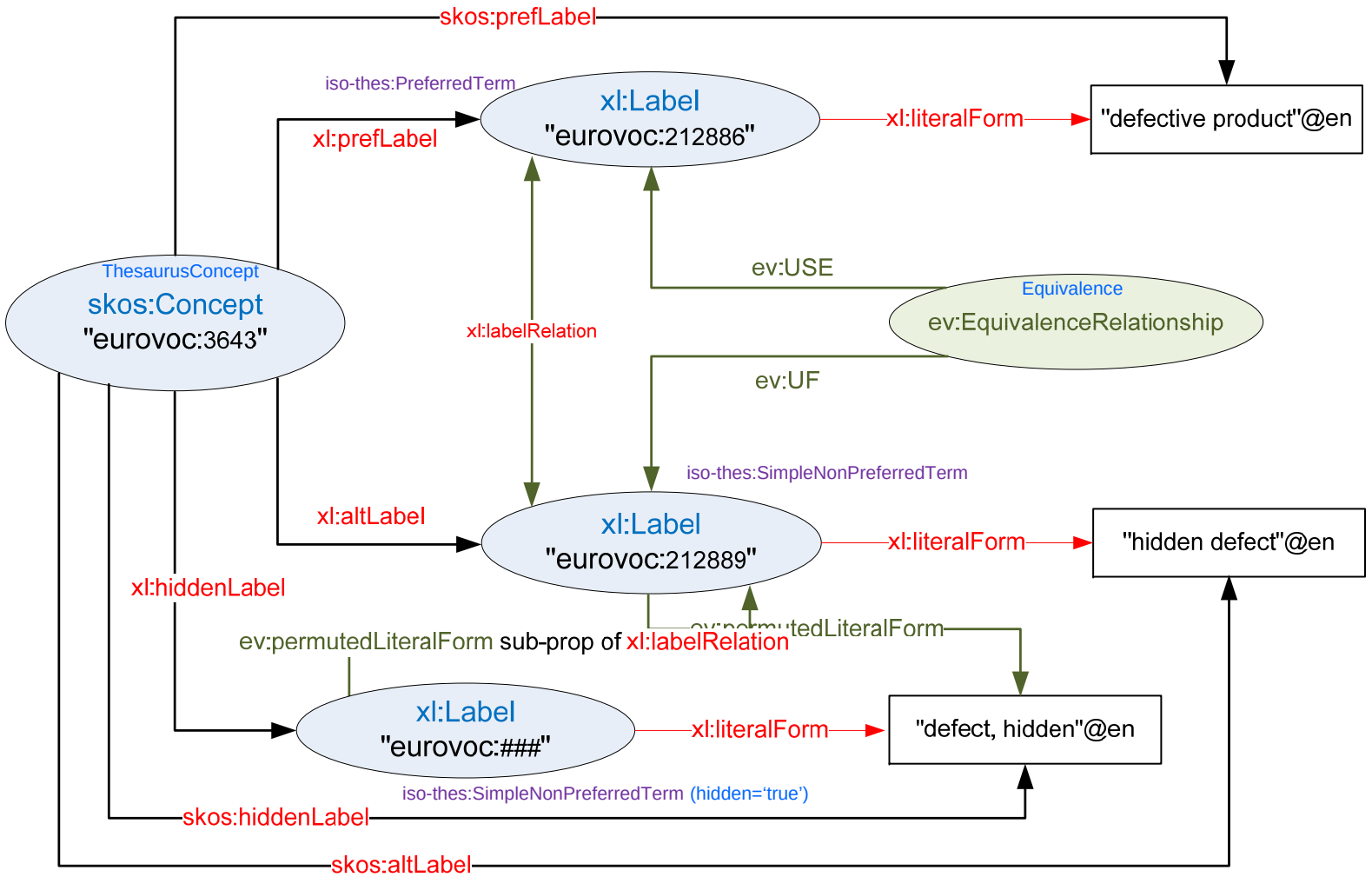
ISSUE 1: equivalence as inferred relation

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

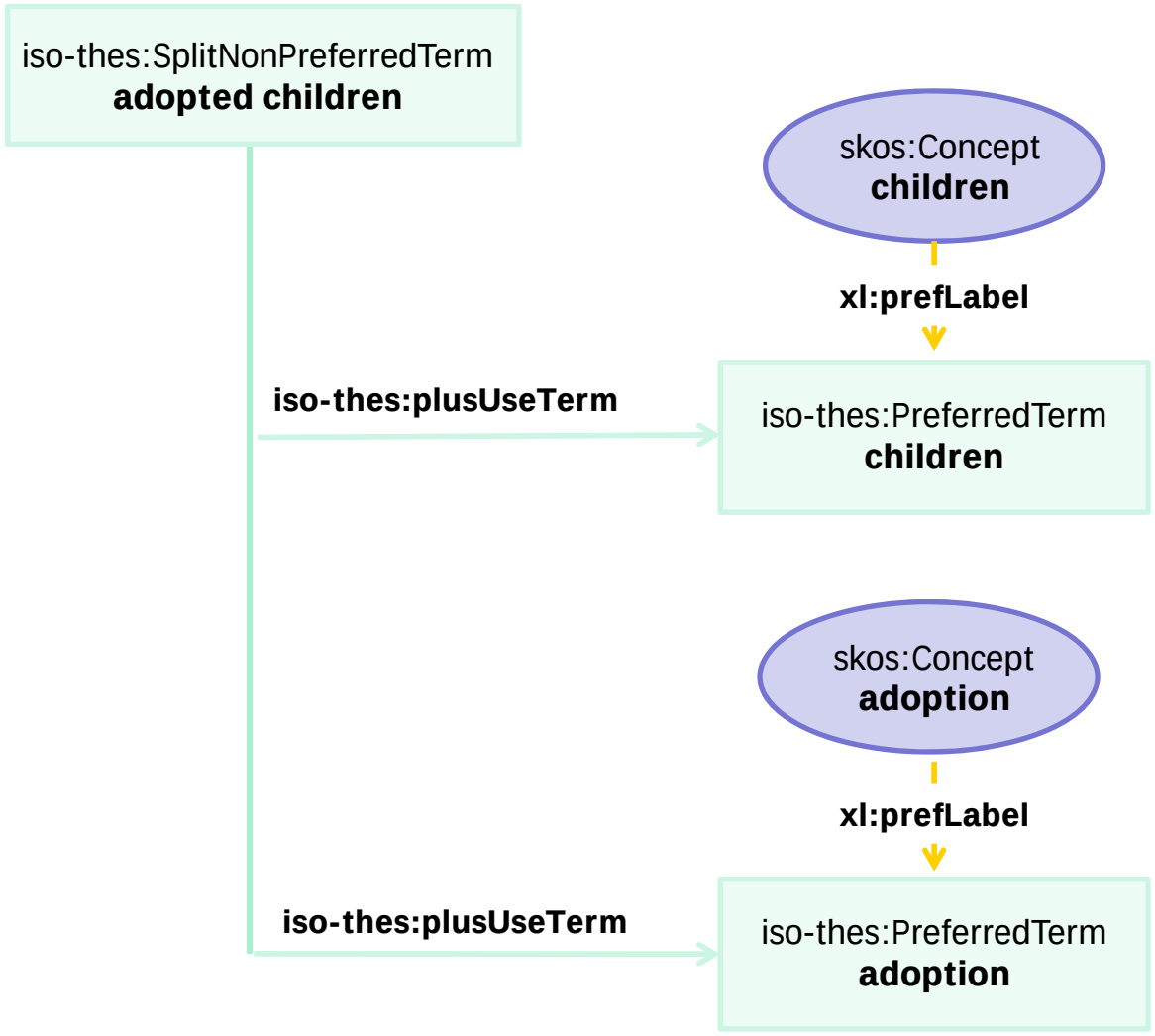


Skip in WS

ThesaurusTerm (5/8)

ISO-25964	SKOS/SKOS-XL/extension	Remark
Equivalence	implicit in SKOS	All skos:altLabel and skos:hiddenLabel having the same language and belonging to the same skos:Concept are equivalent to the skos:prefLabel with the same language of the skos:Concept.
- USE		the skos:prefLabel has role: USE
- UF		the skos:altLabel and skos:hiddenLabel have role: UF
example		

ThesaurusTerm (6/8 -compound equivalence example)

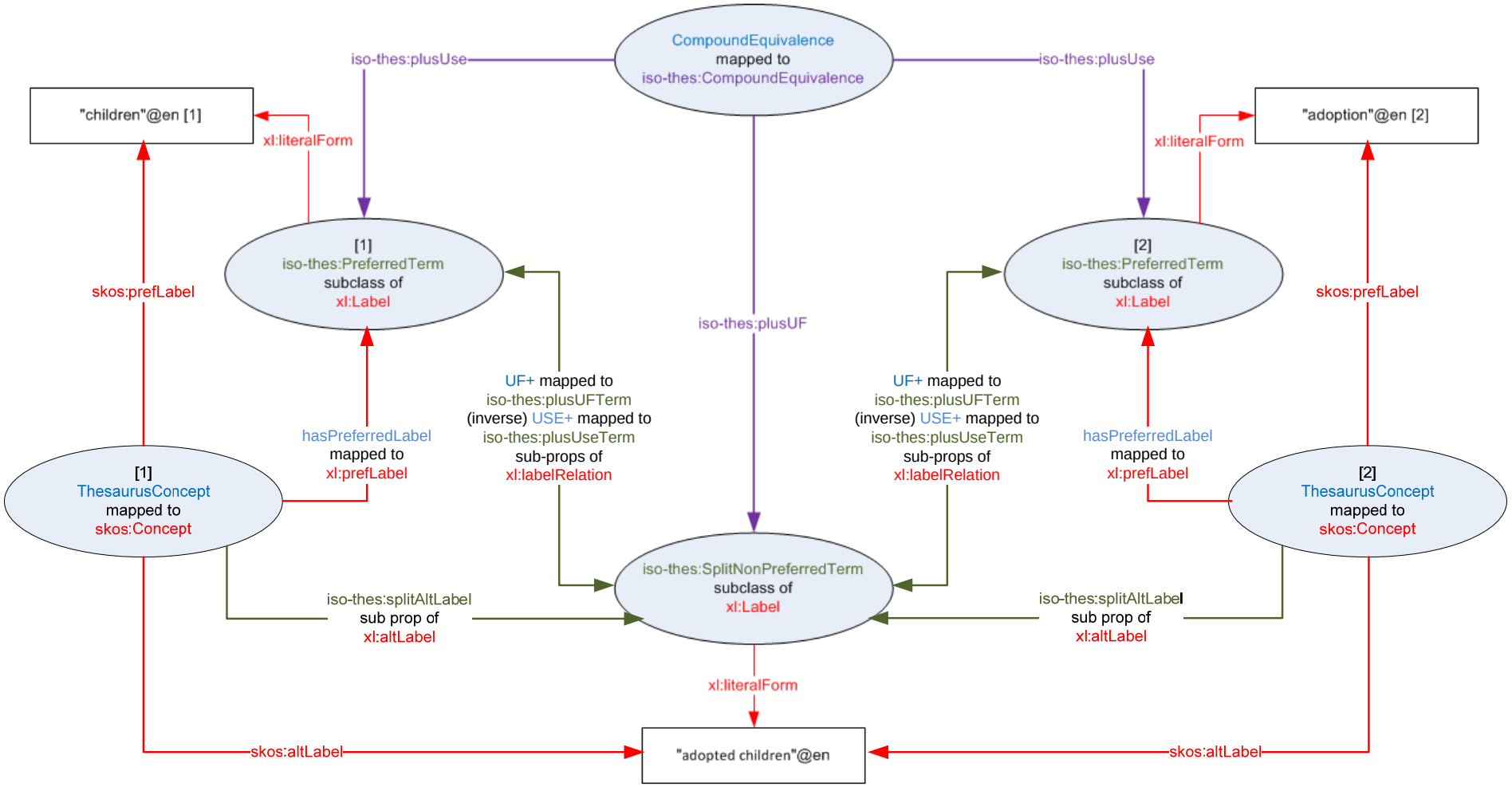


Example: Thesaurus for the Social Sciences
(Thesaurus Sozialwissenschaften)
http://www.gesis.org/sowiport/suche/thesaurus.html

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)

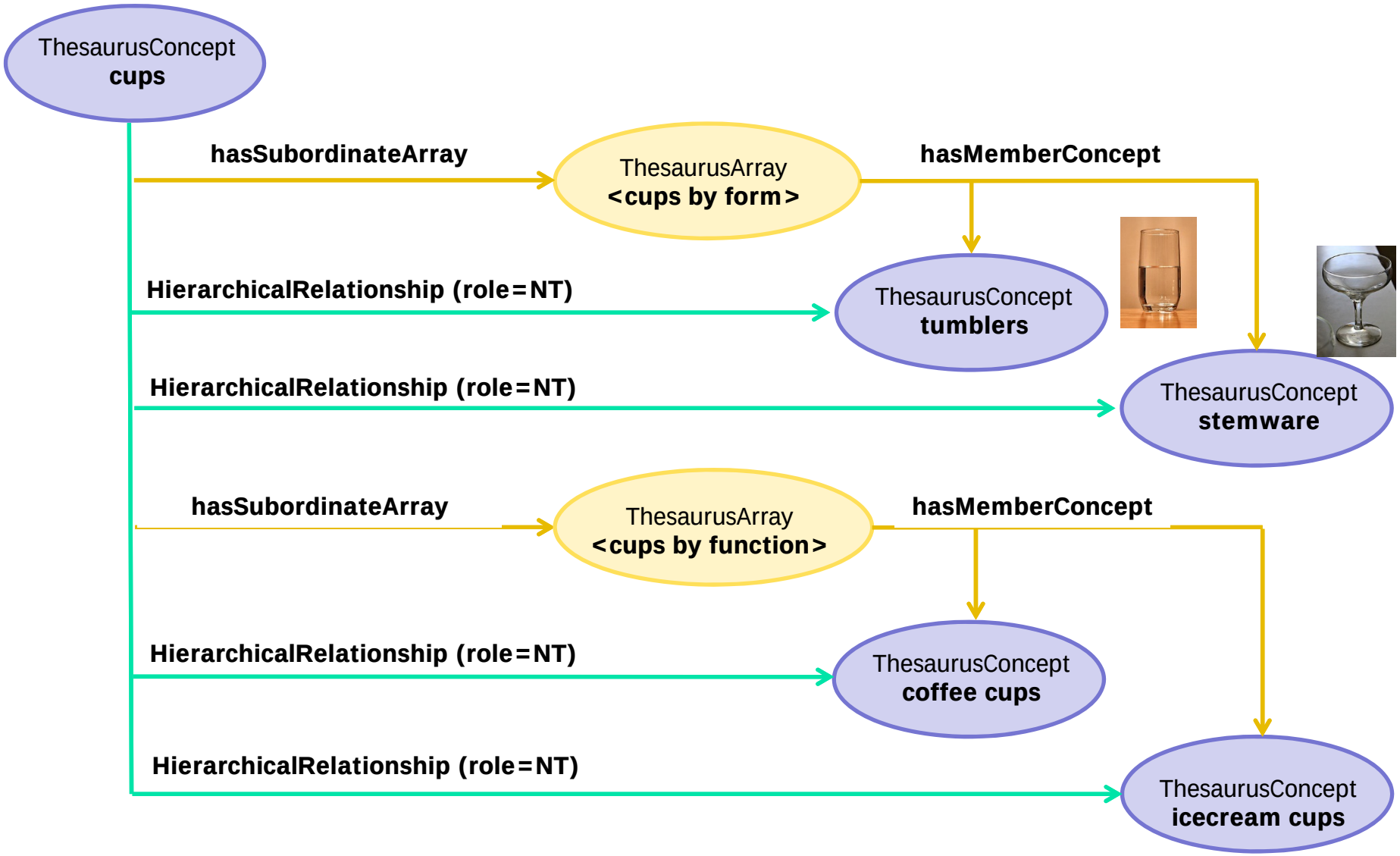


Skip in WS

ThesaurusTerm (8/8)

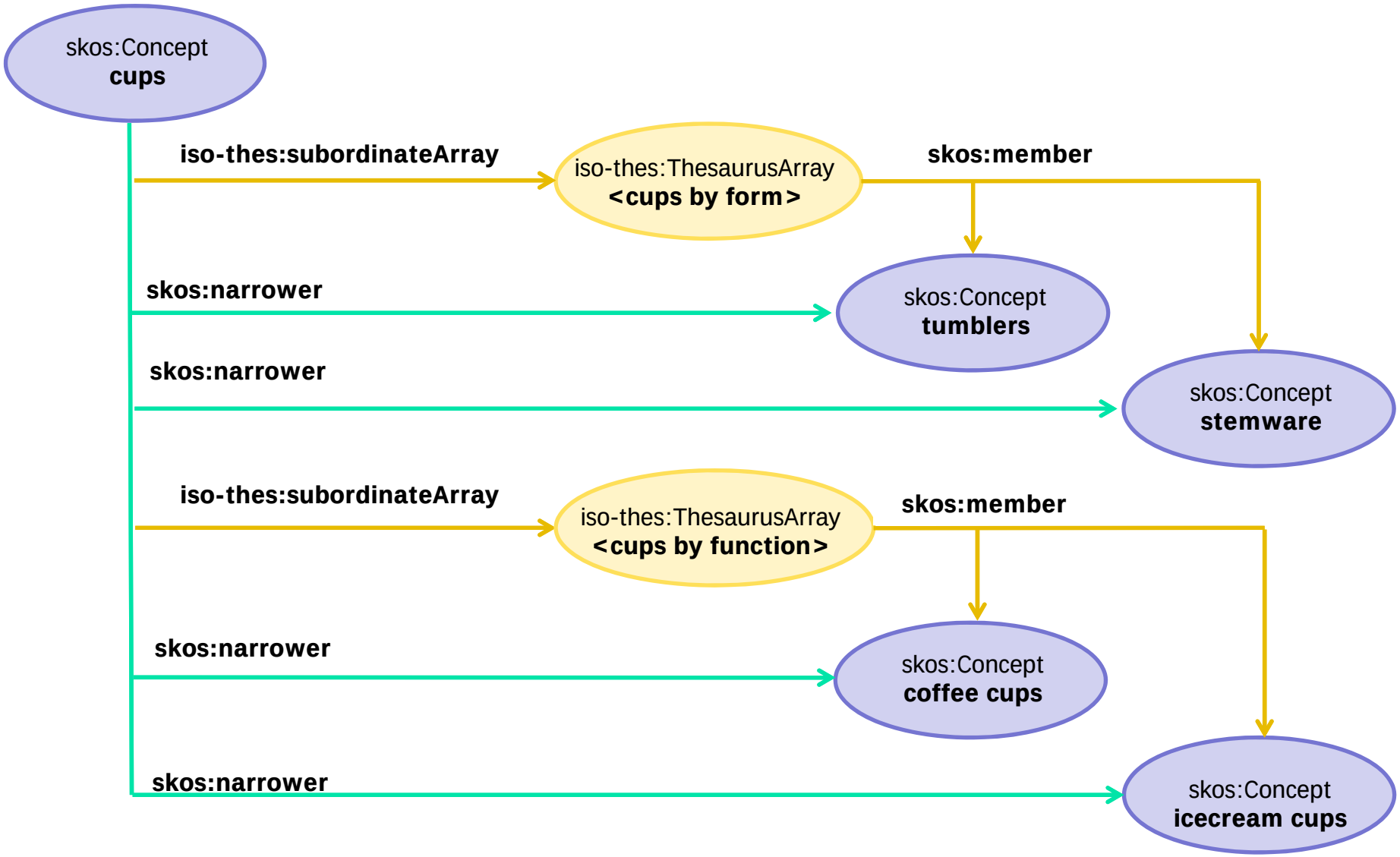
ISO-25964	SKOS/SKOS-XL/extension	Remark
SplitNonPreferredTerm	iso-thes: SplitNonPreferredTerm	The restriction of xl:Label to a subset of the union of: the range of xl:altLabel the range of xl:hiddenLabel
CompoundEquivalence	iso-thes:CompoundEquivalence - iso-thes:plusUF - iso-thes:plusUse (2+)	One instance per equivalence; - the split compound term; - two or more preferred terms
- USE+	iso-thes:plusUseTerm	A sub-property of xl:labelRelation
- UF+	iso-thes:plusUFTerm	A sub-property of xl:labelRelation
example		See above Thesaurus for the Social Sciences (Thesaurus Sozialwissenschaften) http://www.gesis.org/sowiport/suche/thesaurus.html

ThesaurusArray (1/4 – not ordered)

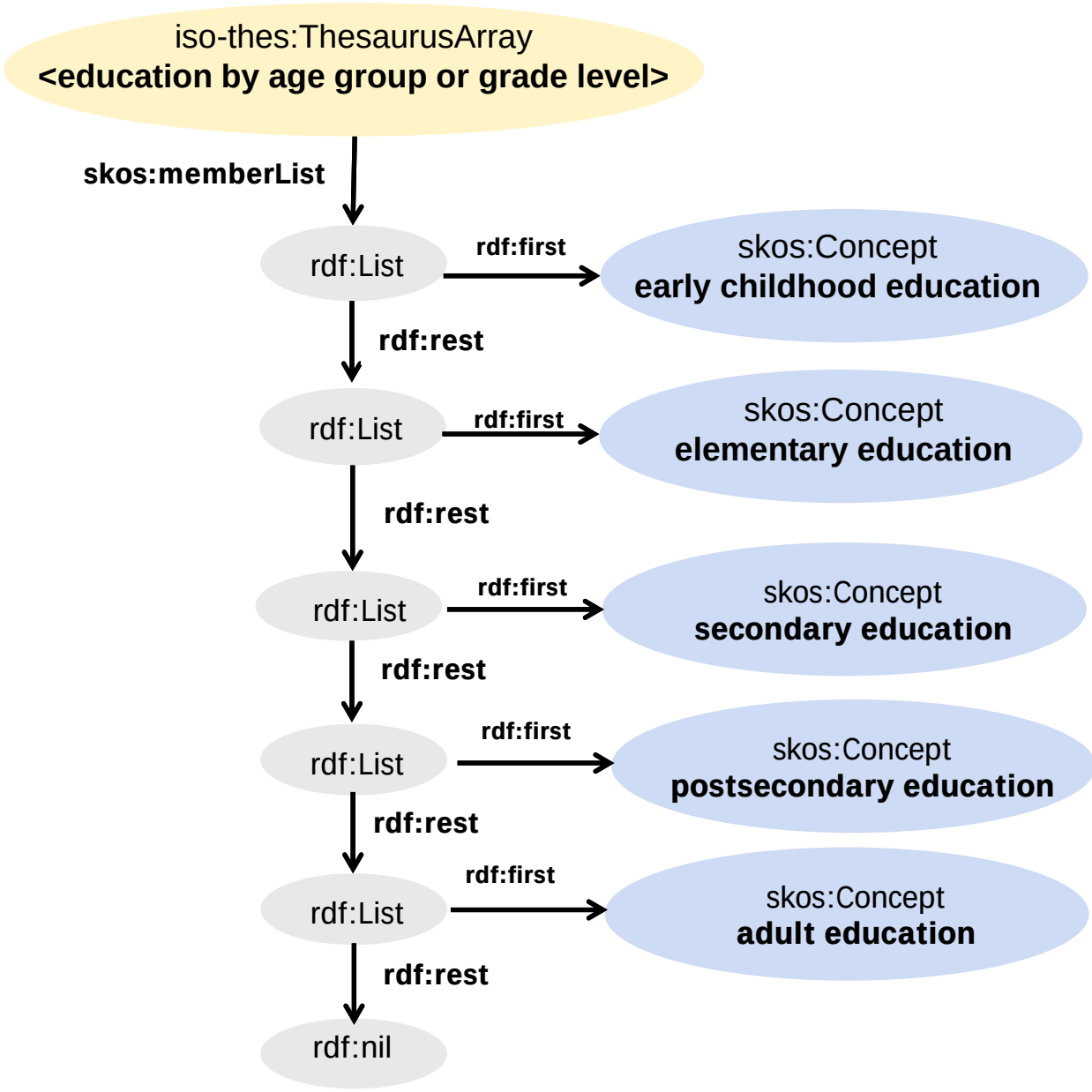


images from wikipedia

ThesaurusArray (1/4 – not ordered)

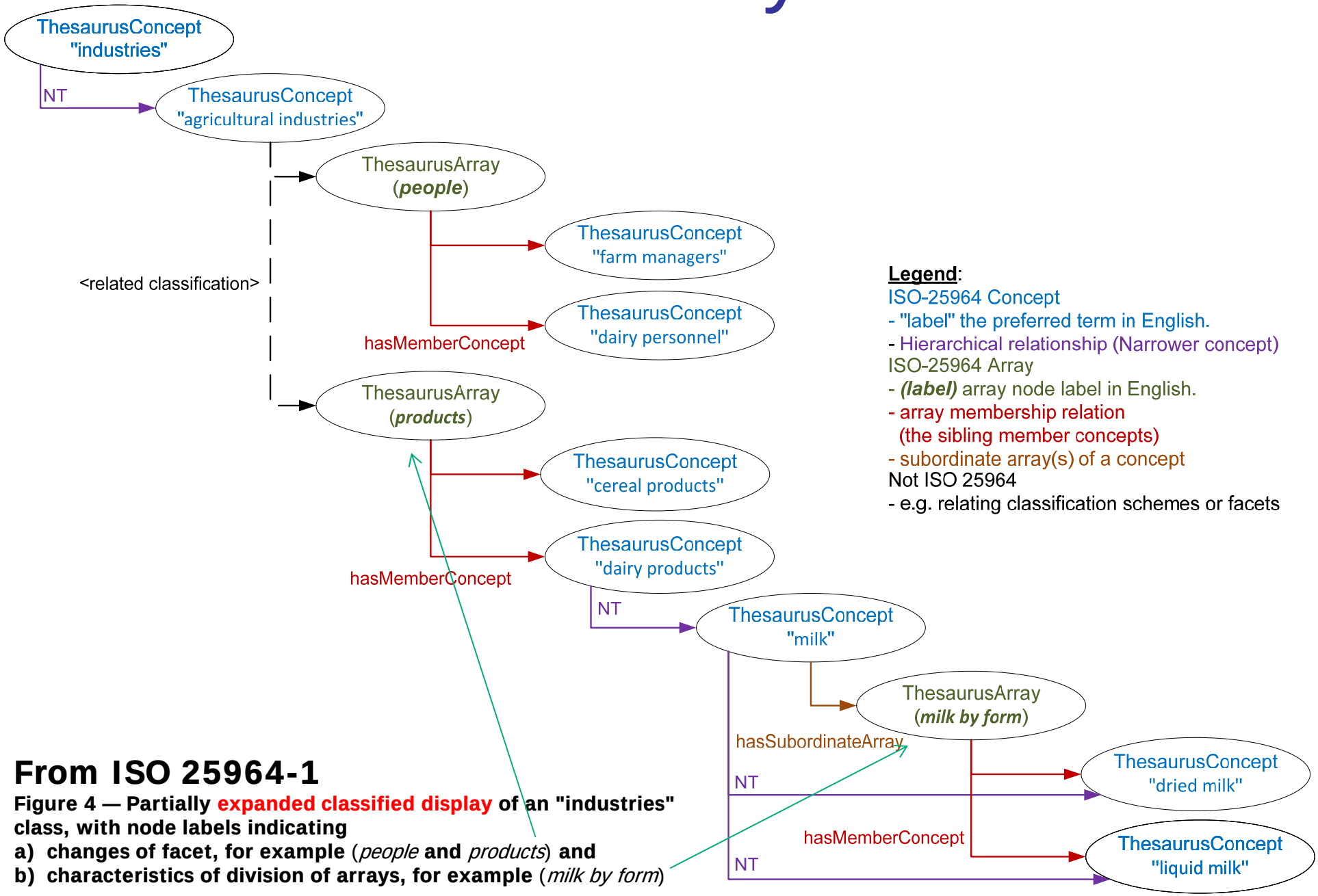


ThesaurusArray (2/4 – ordered)



(Art & Architecture Thesaurus)

ThesaurusArray 3/4 – facet vs narrower



Legend:
ISO-25964 Concept
- "label" the preferred term in English.
- Hierarchical relationship (Narrower concept)

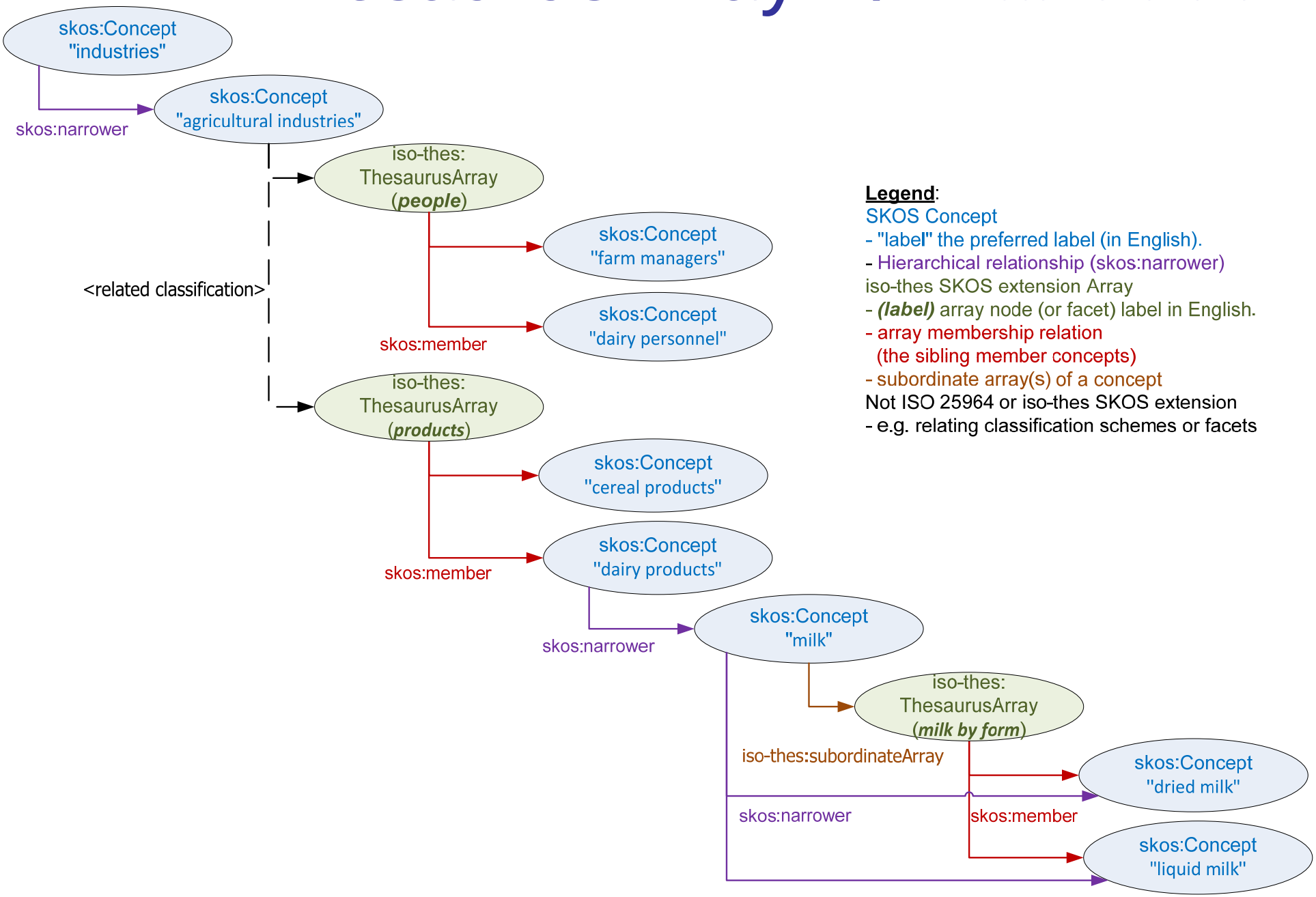
ISO-25964 Array
- ***(label)*** array node label in English.
- array membership relation
  (the sibling member concepts)
- subordinate array(s) of a concept

Not ISO 25964
- e.g. relating classification schemes or facets

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



ThesaurusArray (5/5)

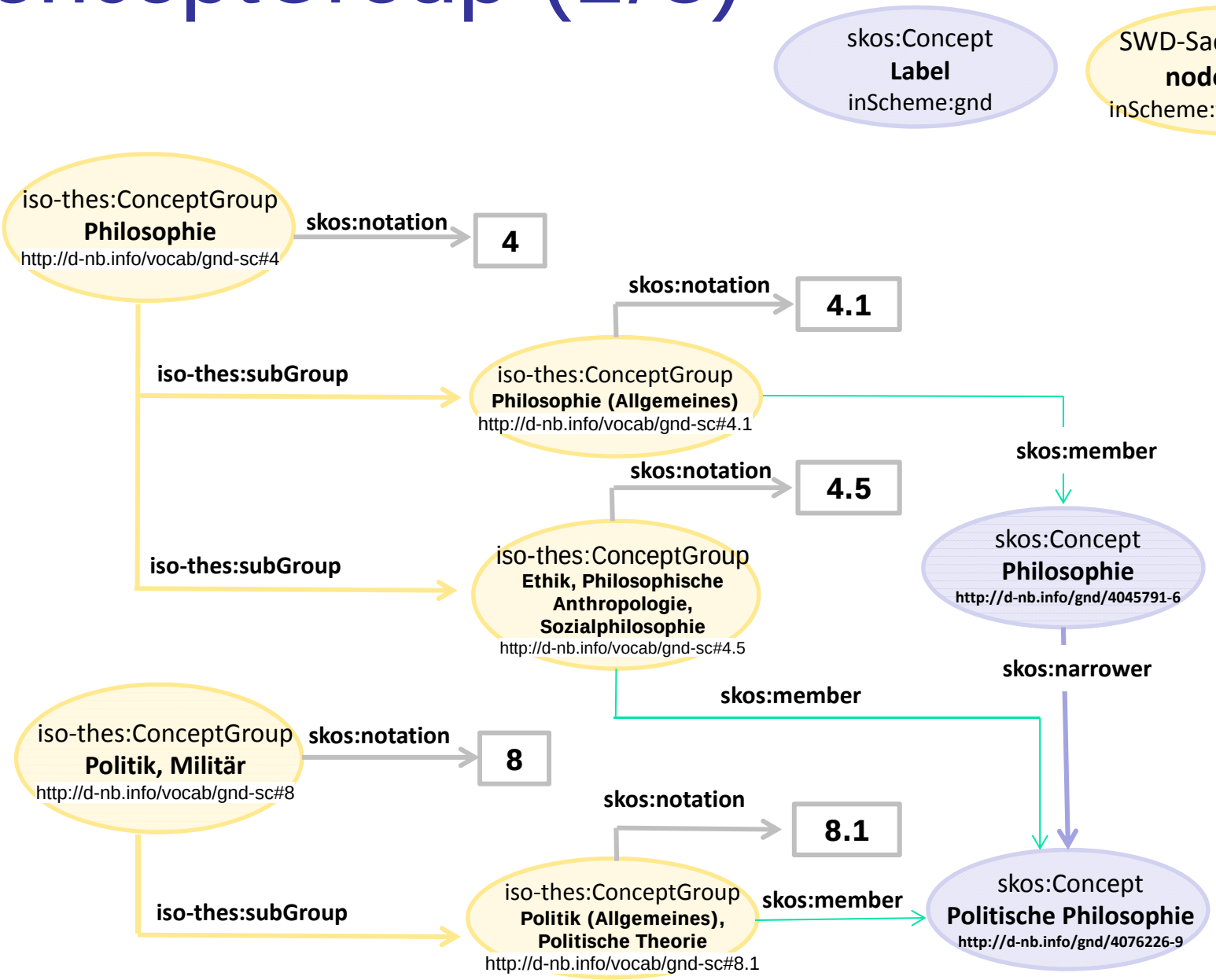
Skip in WS

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

ISO-25964	SKOS/SKOS-XL/extension	Remark
ThesaurusArray[ordered=false]	iso-thes:ThesaurusArray and skos:Collection	
ThesaurusArray[ordered=true]	iso-thes:ThesaurusArray and skos:OrderedCollection	
- hasMemberConcept - hasMemberArray	skos:member	Not ordered array (skos:Collection)
- hasMemberConcept<order> - hasMemberArray<order>	skos:memberList (rdf:List, rdf:first, rdf:rest, rdf:nil)	Ordered array (skos:OrderedCollection)
- isPartOf	skos:inScheme	
- notation	skos:notation	
- hasSubordinateArray	iso-thes:subordinateArray	Domain = skos:Concept Range = skos:Collection
- hasSuperOrdinateConcept	iso-thes:superOrdinate	Inverse of iso-thes:subordinateArray

ConceptGroup (1/3)

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



Skip in WS

ConceptGroup (2/3)

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:gnd="http://d-nb.info/gnd#">
    <gnd:SubjectHeadingSensoStricto rdf:about="http://d-nb.info/gnd/4076226-9">
        <gnd:variantNameForTheSubjectHeading>Staatsauffassung &lt;Politische Philosophie&gt;</gnd:variantNameForTheSubjectHeading>
        <gnd:broderTermGeneral rdf:resource="http://d-nb.info/gnd/4045791-6"/>
        <gnd:relatedTerm rdf:resource="http://d-nb.info/gnd/4115590-7"/>
        <gnd:variantNameForTheSubjectHeading>Staatsgedanke &lt;Politische Philosophie&gt;</gnd:variantNameForTheSubjectHeading>
        <gnd:variantNameForTheSubjectHeading>Staatsphilosophie</gnd:variantNameForTheSubjectHeading>
        <gnd:relatedTerm rdf:resource="http://d-nb.info/gnd/4077784-4"/>
        <gnd:relatedDdcWithDegreeOfDeterminacy3 rdf:resource="http://d-nb.info/ddc/class/320.101"/>
        <gnd:preferredNameForTheSubjectHeading>Politische Philosophie</gnd:preferredNameForTheSubjectHeading>
        <gnd:gndSubjectCategory rdf:resource="http://d-nb.info/vocab/gnd-sc#4.5"/>
        <gnd:relatedTerm rdf:resource="http://d-nb.info/gnd/4055876-9"/>
        <gnd:gndIdentifier>4076226-9</gnd:gndIdentifier>
        <gnd:relatedDdcWithDegreeOfDeterminacy3 rdf:resource="http://d-nb.info/ddc/class/320.01"/>
        <gnd:gndSubjectCategory rdf:resource="http://d-nb.info/vocab/gnd-sc#8.1"/>
        <gnd:relatedTerm rdf:resource="http://d-nb.info/gnd/4046563-9"/>
        <gnd:oldAuthorityNumber>(DE-588c)4076226-9</gnd:oldAuthorityNumber>
    </gnd:SubjectHeadingSensoStricto>
</rdf:RDF>
```

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)

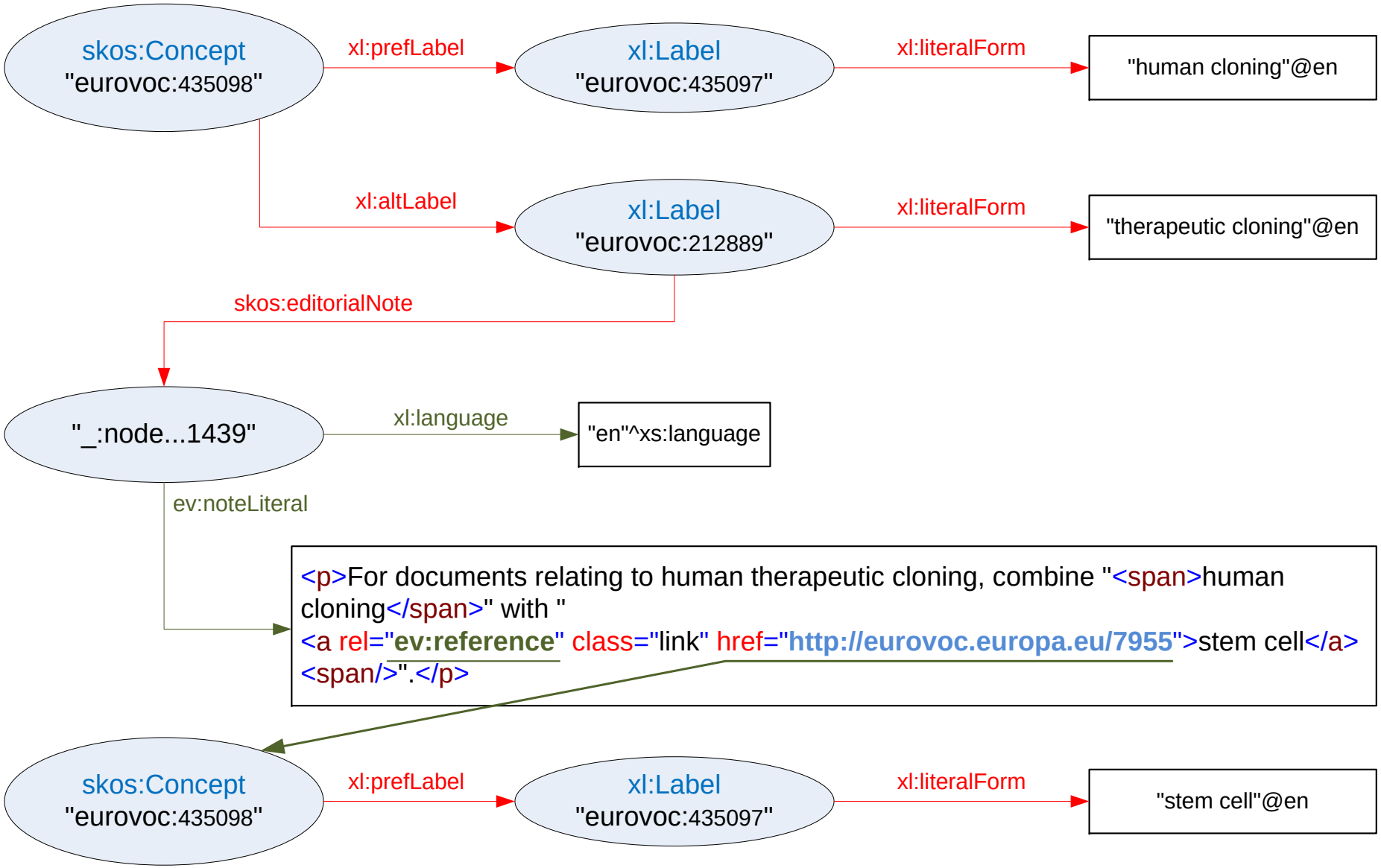
Example 1 Neubert (2009) http://events.linkeddata.org/ldow2009/papers/ldow2009_paper7.pdf
Example 2 EuroVoc
Work ongoing in the RDF group

Skip in WS

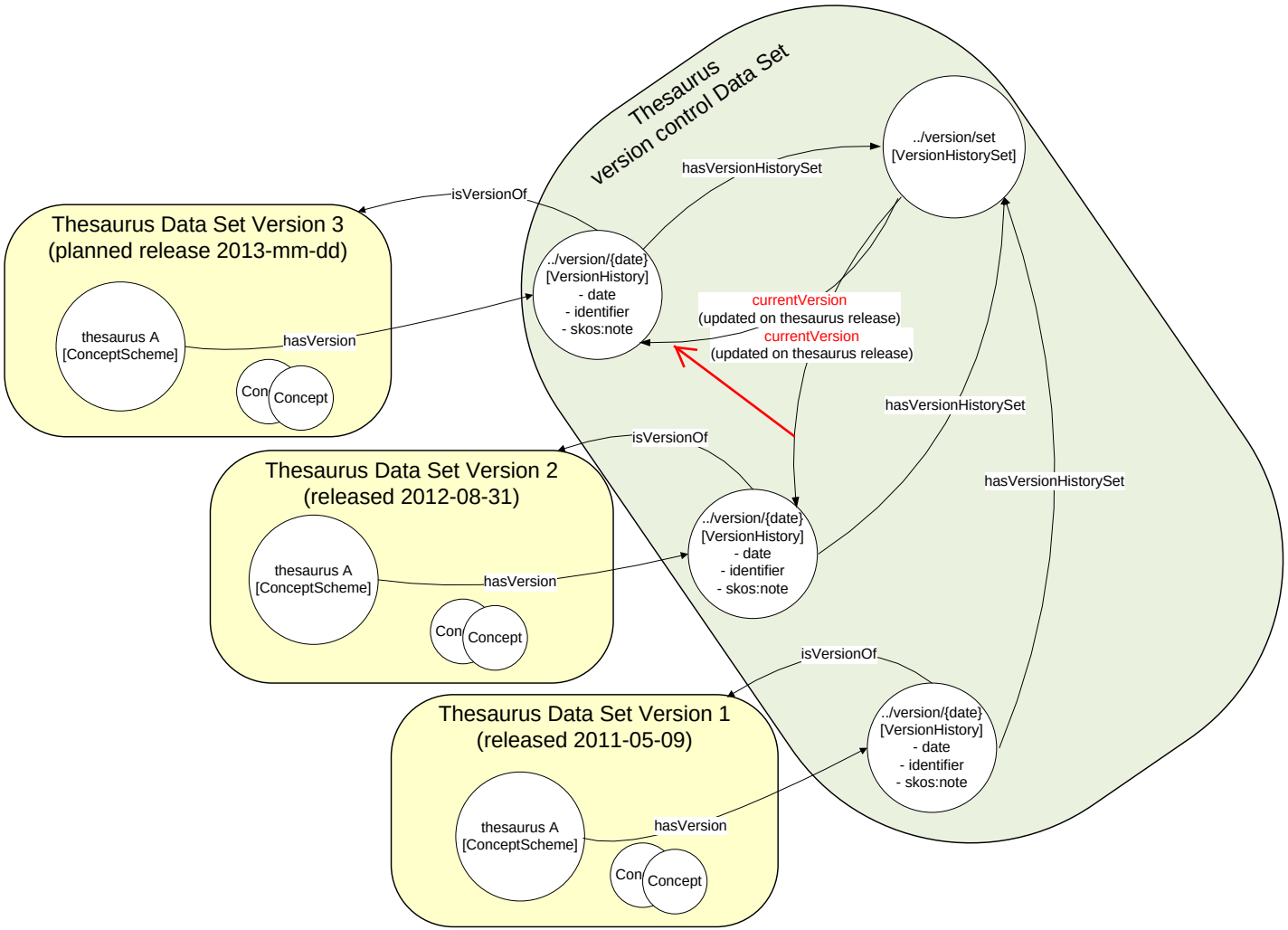
Notes (1/2)

ISO-25964	SKOS / SKOS-XL / extension	Remark
Note	range of skos:note	
- refersTo	tagged embedded hyperlinks	work is on going to in the RDF group to type the content explicitly as HTML or XML In RDF1.1: http://dvcs.w3.org/hg/rdf/raw-file/default/rdf-concepts/index.html#section-html
hasScopeNote	skos:scopeNote	ISO 25964 is more restrictive than SKOS
hasHistoryNote	skos:historyNote	ISO 25964 is more restrictive than SKOS
hasEditorialNote	skos:editorialNote skos:changeNote	ISO 25964 is more restrictive than SKOS
hasDefinition	skos:definition	ISO 25964 is more restrictive than SKOS
hasCustomNote[noteType]	custom sub property of skos:note or skos:example	depending noteType a custom property may be defined

Notes (2/2 – eurovoc example)



Thesaurus versions and versioned data sets (proposal)



Skip in WS

VersionHistory (2/2)

ISO-25964	SKOS/SKOS-XL/extension	Remark
VersionHistory	iso-thes:VersionHistorySet	The typical unique instance is a set aggregating all “historic versions or variants” of a thesaurus.
	iso-thes:VersionHistory	Each instance details a “historic version or variant” of the thesaurus
	iso-thes:currentVersion	Current “historic version or variant” of a thesaurus
	iso-thes:isVersionOf	Identifies the thesaurus data set of a “historic version or variant”
	iso-thes:hasVersion	Identifies the “historic version or variant” instance providing the historic or variant identifiers of a thesaurus publication (data set)

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 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

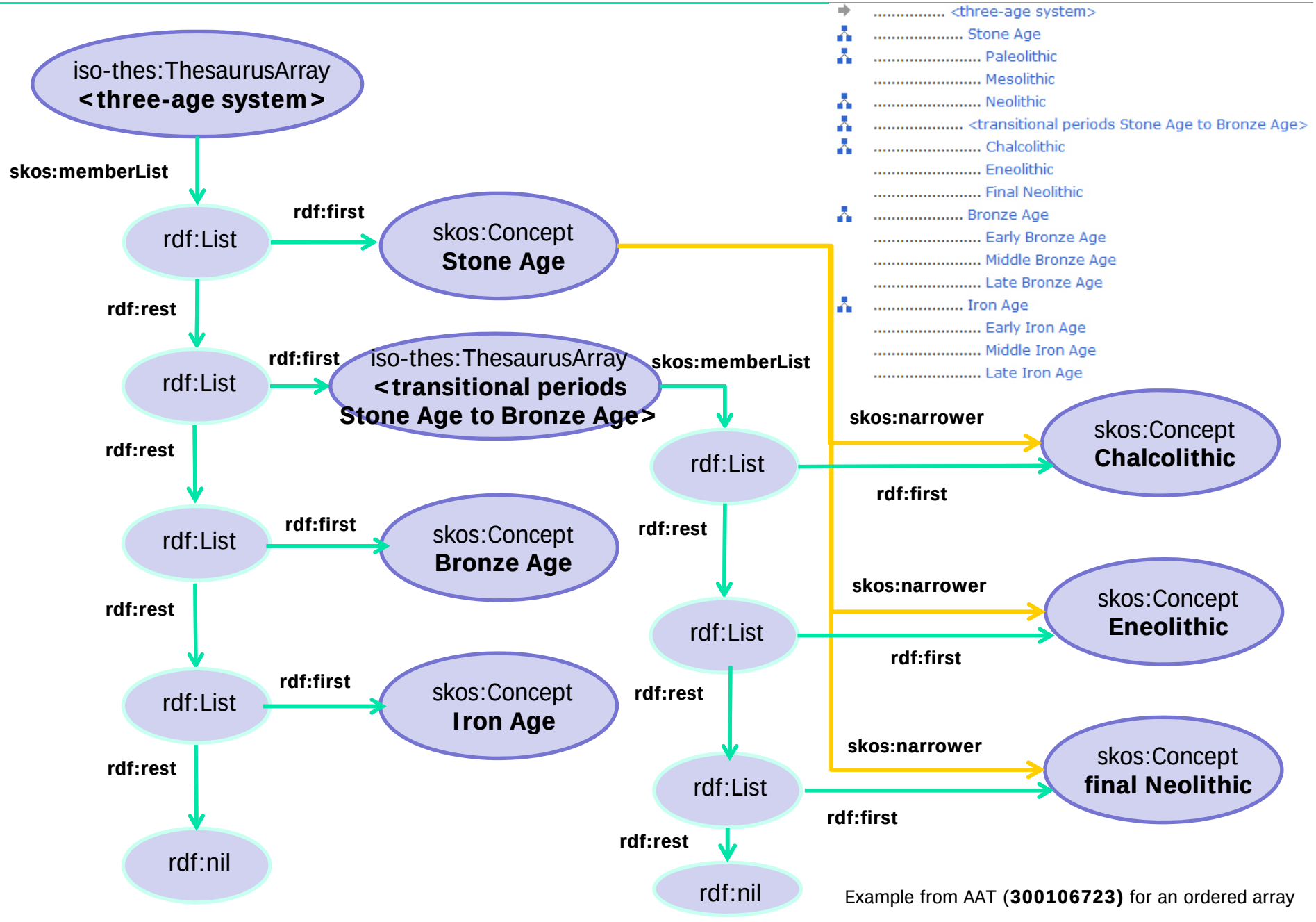
- DD8723-5, Data Model for BS 8723. In: *Structured vocabularies for information retrieval. Part 5. Exchange formats and protocols for interoperability*. London: British Standards Institution. Available at: http://schemas.bs8723.org/Model.aspx
- ISO 25964-1:2011, *Thesauri and interoperability with other 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

back-up - hidden

ThesaurusArray <ordered> (2/4)

(Arts and Architecture Thesaurus)



Example from AAT (**300106723)** for an ordered array