Modeling Classification Systems in SKOS:

Some Challenges and Best-Practice Recommendations

Michael Panzer, OCLC
Marcia Lei Zeng, Kent State University

DC-2009 “Semantic Interoperability of Linked Data”, Seoul

1. Representational issues

1. Special types of concepts
2. Index terms
3. Class–topic relationships
4. Internal structure of notes
5. Alternative classification notations
6. Orders/sequences of coordinate classes

2.1. Special types of concepts

- Non-assignable concepts
  - DDC example: “Centered entries”
  - Expressible as `skos:Collection`?
- Concepts in auxiliary tables
  - DDC example: “Table 2: Areas, Periods, Persons”
  - Expressible as `skos:Collection`?

See reference

—485 Sweden

Class here *Baltic Sea Region

For divisions of Sweden, see —486–488

Centered entry

See also —16334 for Baltic Sea; also —479 for Baltic States

> —486–488 Divisions of Sweden

Class comprehensive works in —485

—486 Southern Sweden (Götaland)

Including Blekinge, Gotland, Halland, Jönköping, Kalmar, Kronoberg, Malmöhus, Östergötland, Skåne, Västra Götaland counties (län); former Älvsborg, Göteborg och Bohus, Kristianstad, Skaraborg counties (län)

Including Öland Island

See also —16334 for Baltic Sea, Kattegat

—487 Central Sweden (Svealand)

Including Dalarna (Kopparberg), Örebro, Södermanland, Uppsala, Värmland, Västmanland counties (län)

Gävleborg county (län) relocated to —488

—487 3 Stockholm county (Stockholms län)

Class here Stockholm

—488 Northern Sweden (Norrland)

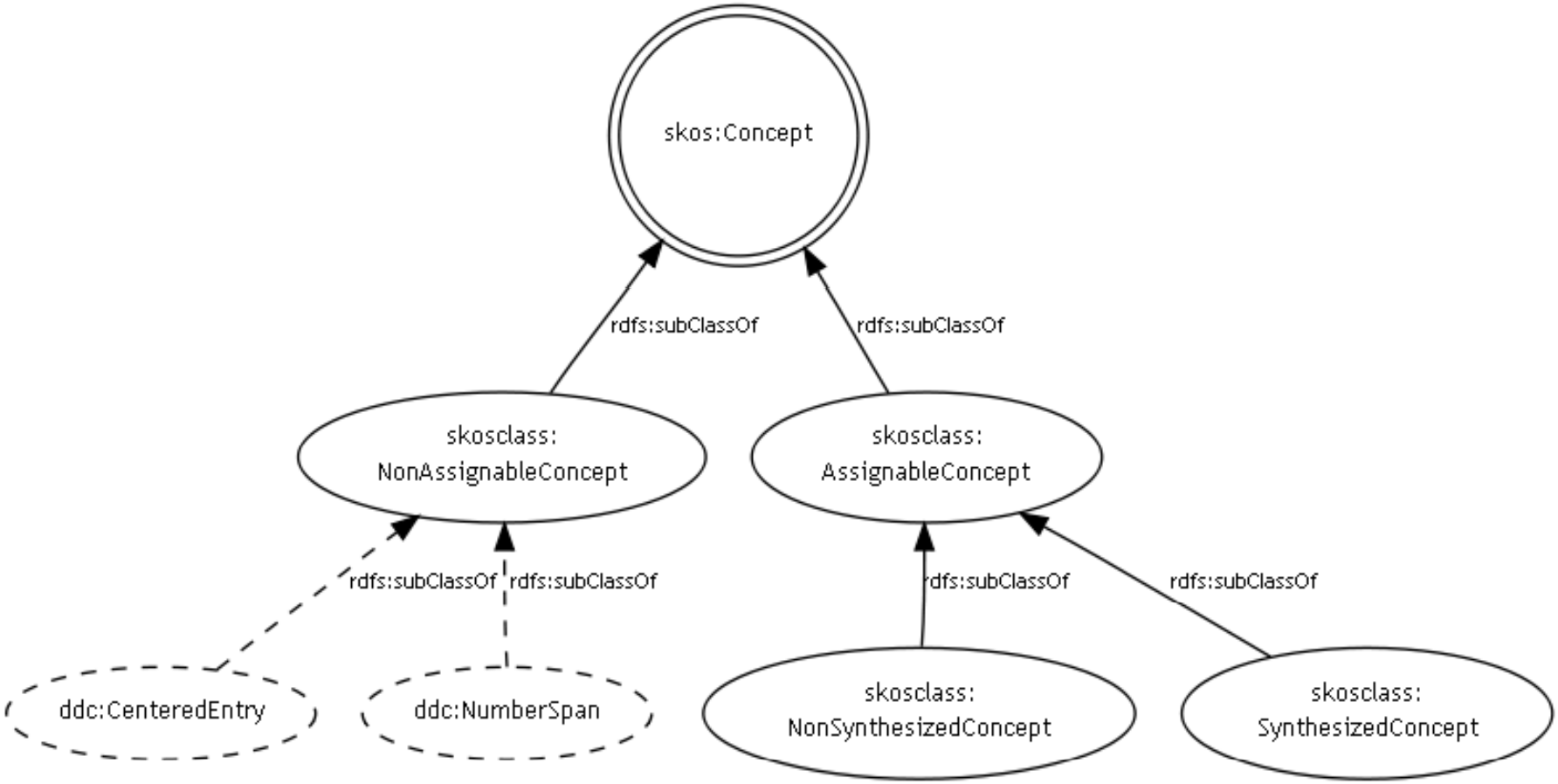
Including Gävleborg county (län) [*formerly* —487]; Jämtland, Norrbotten, Västerbotten, Västernorrland counties (län)

See also —16334 for Gulf of Bothnia

Centered Entries vs. `skos:Collection`

- Centered entries
  - Part of hierarchy
  - Establish hierarchical array
  - Grouping function, but do not establish a collection of concepts
- `skos:Collection`
  - Disjoint with `skos:Concept`
  - Not in domain/range of semantic relationships
- Alternative: Interoperable definition of subclasses (refinement) of `skos:Concept`

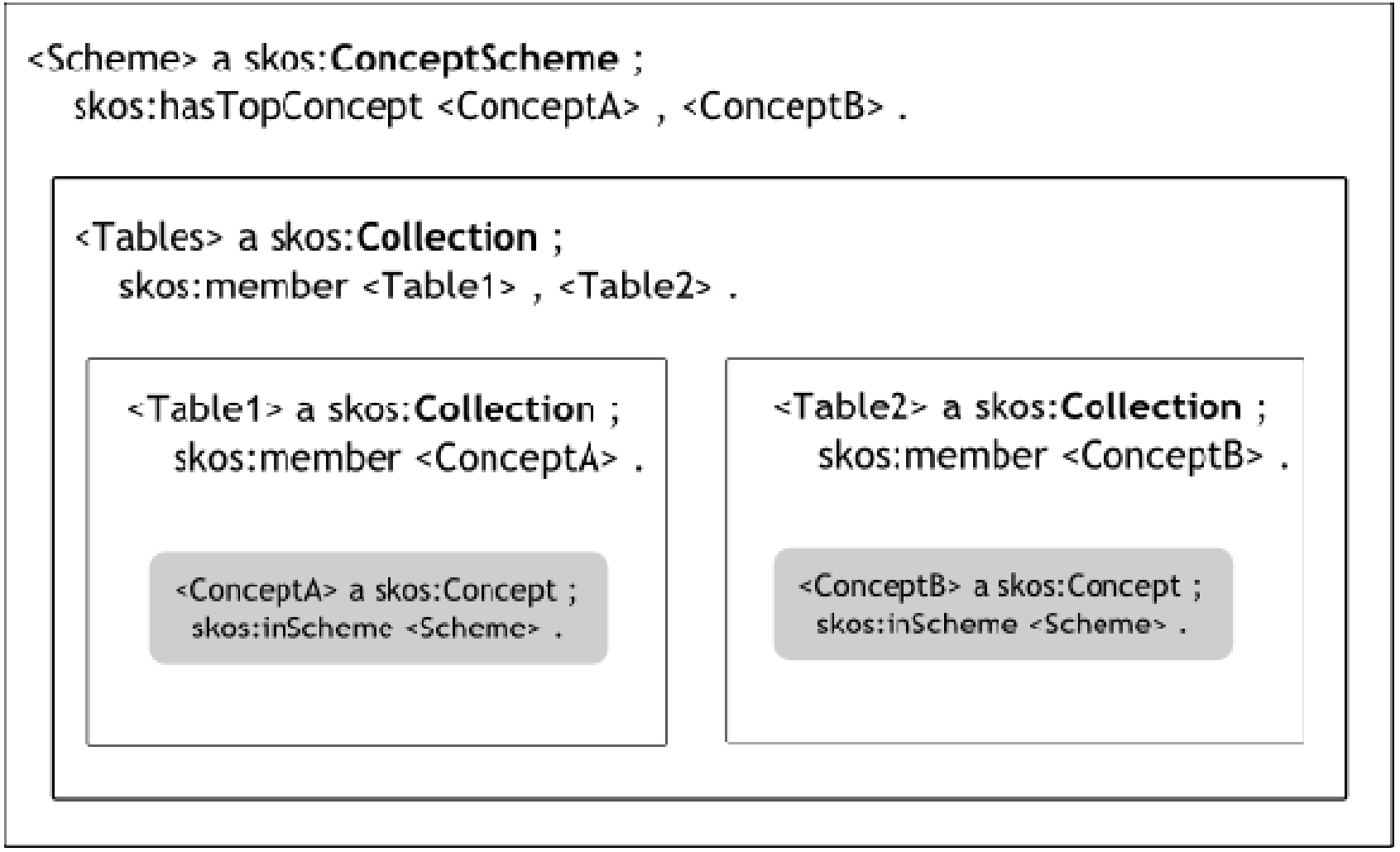
Refining skos:Concept



Auxiliary tables vs. `skos:Collection`

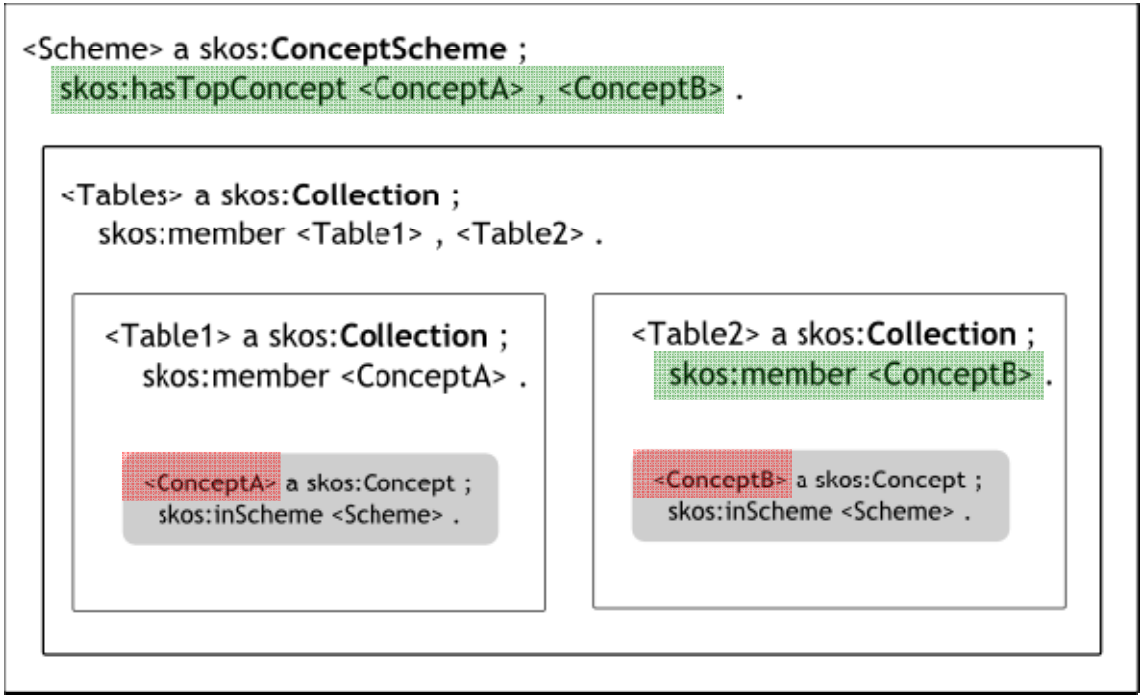
- Auxiliary tables
  - Partitioning of concepts into separate lists for synthesis
  - Form separate hierarchies
  - Have top concepts separate from main schedules
- `skos:Collection`
  - Can have documenation properties (labels, notes)
  - Can have a notation
  - Disjoint with `skos:ConceptScheme`
  - Not in domain of `skos:hasTopConcept`
- Alternative: “Nested” `skos:Collections`

Tables as “nested” collections



Retrieving top table concepts

```
SELECT ?concept WHERE {
  <Scheme> skos:hasTopConcept ?concept .
  <Table2> skos:member ?concept .
}
```



Tables as “nested” collections

- Drawbacks:
  - `skos:Collection` can never explicitly be part of a `skos:ConceptScheme`
  - `skos:Collection` orginally intended for a specific use case (“node labels”)
  - Implicit “nesting” with schemes blurs the line between concept schemes and concept collections
- Alternatives:
  - Declare subclasses of `skos:Concept` for table concepts?
  - Use of other relations to indicate membership of concepts in specific subgroupings (`dct:partOf`)?

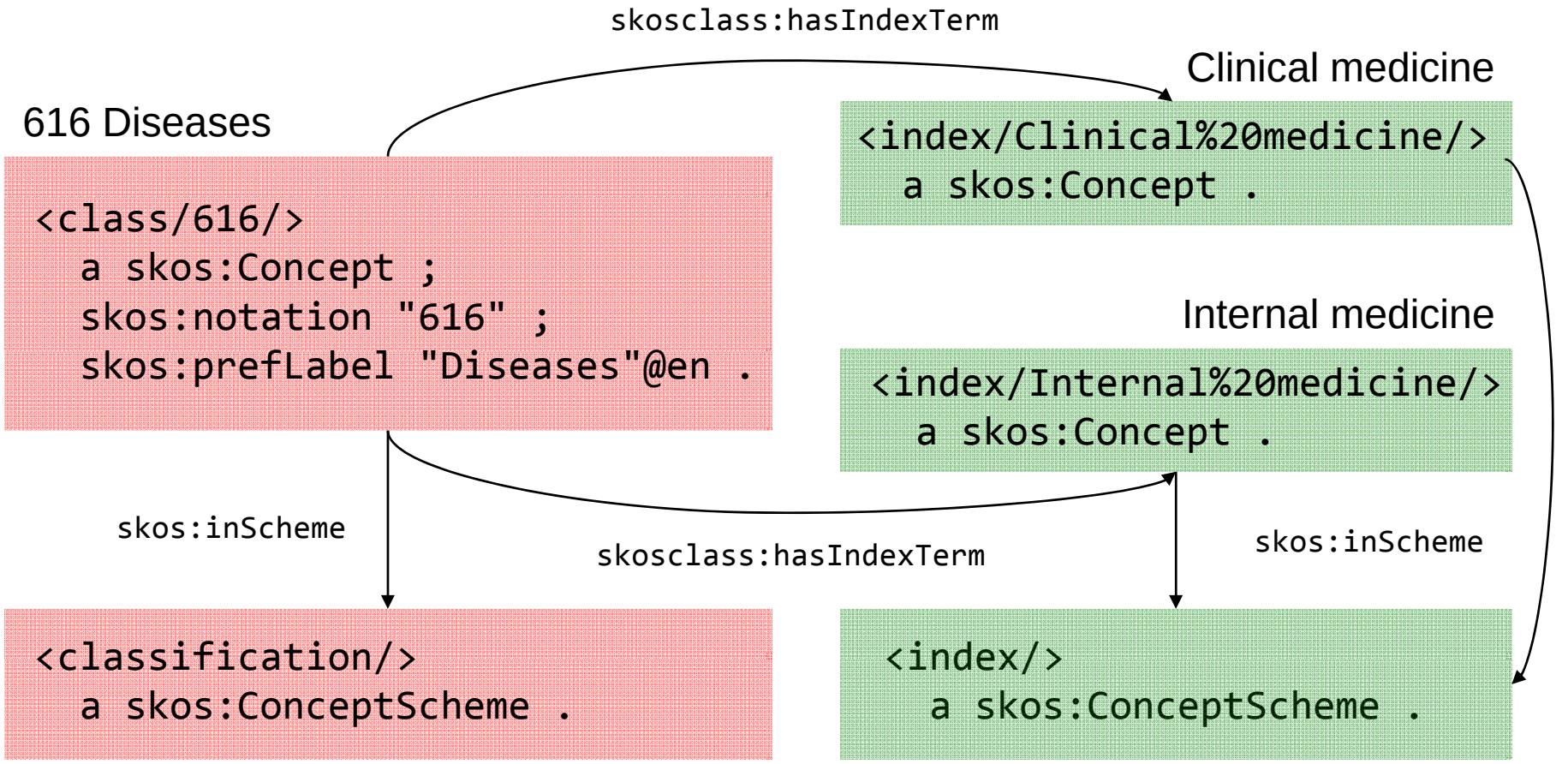
2.2. Index terms

- Index is important part of many classification systems
- DDC Example: Relative Index terms of 616 Diseases:

Clinical medicine
Diseases—humans—-medicine
Illness—medicine
Internal medicine
Physical illness—medicine
Sickness—medicine

Index terms

- Treat index as seperate concept scheme



2.3. Class–topic relationships

- Index terms are just one example of the association of classes with topics
- Various relationships have to be modeled to describe the topical neighborhood of a class
- Basic distinction between classes and topics and class–class and class–topic relationships necessary
- Not easy to make that basic distinction with SKOS
- Alternative: Expressive power of OWL?

2.3. Class–topic relationships

Topical neighborhood shaped by (among other things):

Inclusion of topics

Caption	**370.113 Vocational education**
Class-here note	**Class here career education, occupational training, vocational schools**

Exclusion of topics

Class-elsewhere note	**Class on-the-job training, vocational training provided by industry in 331.2592**
See reference	***For vocational education at secondary level, see 373.246; for adult vocational education, see 374.013***

2.4. Internal structure of notes

- Notes carry different types of information in classification systems
- `skos:note` can serve as an extension point for some of them
- DDC example: history notes

2.4. Internal structure of notes

MARC 21:

```
685  41  $tBiochemistry$iformerly located in
         $b574.192$d1996$221
```

SKOS:

```
<class/572/> ddc:formerlyNote [
   dct:issued "1996-01-01"^^<http://purl.org/dc/terms/W3CDTF> ;
   dct:isPartOf <scheme/e21/> ;
   dct:description "Class immediately reused"@en ;
   ddc:previousNumber "574.192"^^<schema-terms/Notation> ;
   rdf:value "Biochemistry formerly located in 574.192"@en ;
   ddc:topic "Biochemistry"@en .
]
```

2.5 Alternative classification notations

Q Biological sciences

… …

[Q89] 环境生物学
宜入X17。

[Q89] Environmental biology
Preferred class: X17

X17 环境生物学
环境生物工程入此。

X171 生态系统与污染生态学
环境生态学入此。
参见S181。

X171.1 生态系统与生态环境

[X171.3] 生态农业
宜入S181。

X171.4 生态建设与生态恢复
生态工程的研究入此。

X171.5 污染生态学
生态毒理学入此。

X172 环境微生物学
微生物降解和转化污染物的规律入此。

X173 环境植物学
植物毒性、树木园林与环境保护的关系入此。

X174 环境动物学
动物毒性研究入此。

X176 生物多样性保护
参见Q16。

X1 Environmental Sciences – Basic Theory

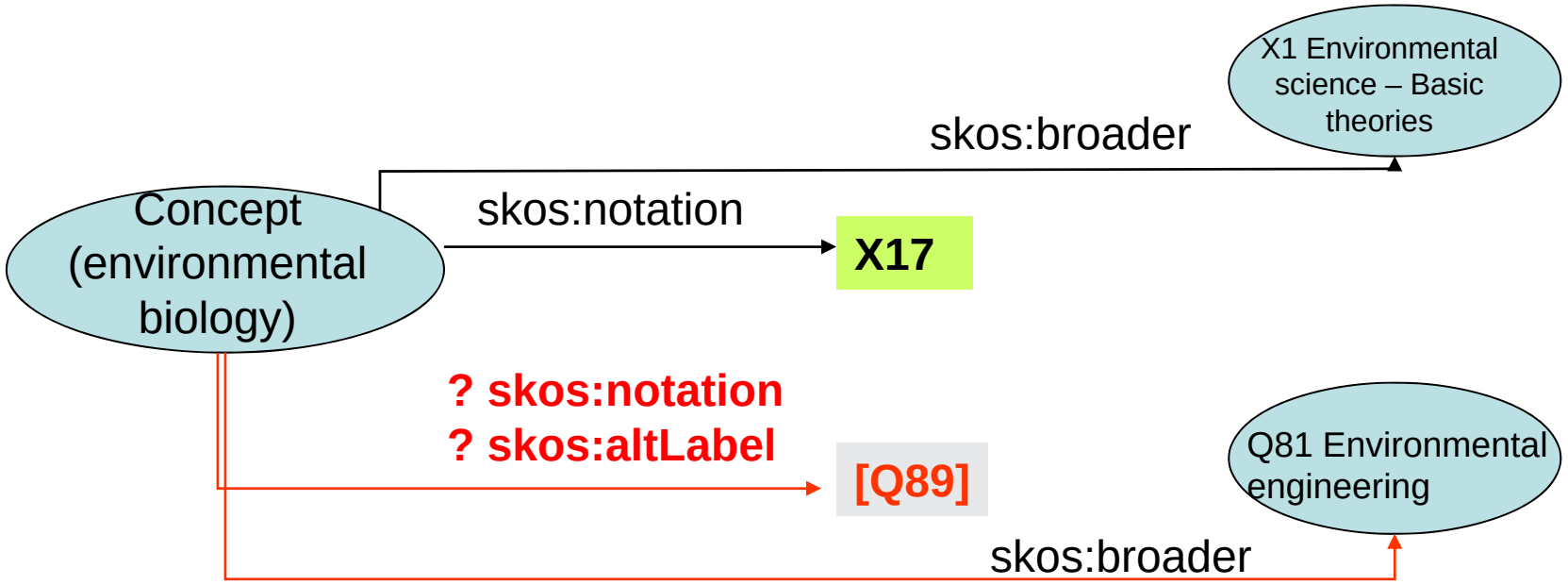
… …

X17 Environmental biology

If Q89 is chosen as the preferred class:

- X17 becomes alternative class;
- The bracket moves to X17;
- Subdivisions under Q89 will be formed following X17 class.

[Q89] environmental biology
preferred class: X17



Discussion of options:

? skos:notation

Yes – if SKOS allows for more than one notation for one concept.
But how can you show which one is preferred and which one is not?

? skos:altLabel

No. This is not an alternative label. It is a concept, with its own semantic relations.

[Q89] environmental biology
preferred class: X17

Suggested solution

(following the SKOS extension for labels)

- transform the alternative notations into resources (instead of literals)
  - As resources, they can possess different semantic relationships:

```
<X17> skos:broader <Environmental Science_Basic Theory>
<Q89> skos:broader <BiologicalSciences>
<Environmental Biology> clc:altNotation <X17>
```

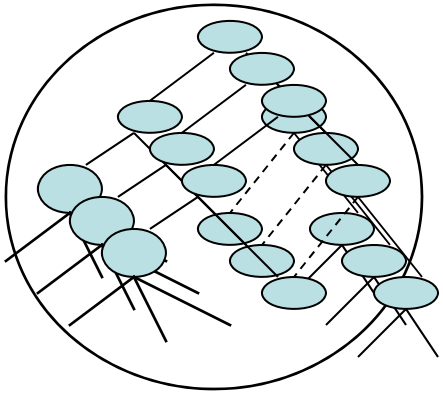
- The connection back to the original resource is established by the definition of a property chain axiom.
  - works only if we define “X17” to be an instance of skos:Concept

```
skos:broader owl:propertyChainAxiom ( clc:altNotation
  skos:broader )
```

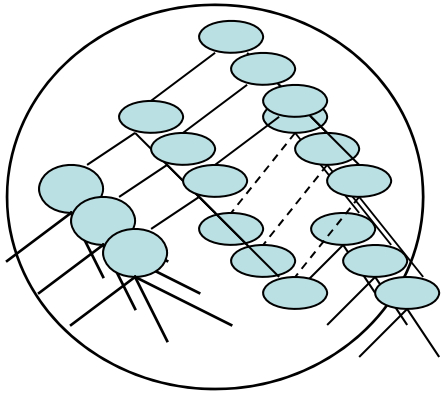
2.6 Orders/sequences of coordinate classes

"The creation of a meaningful order is equally as important in information organization as the grouping of documents into classes" (Svenonius: *The intellectual foundation of information organization,* 2000, p. 191)

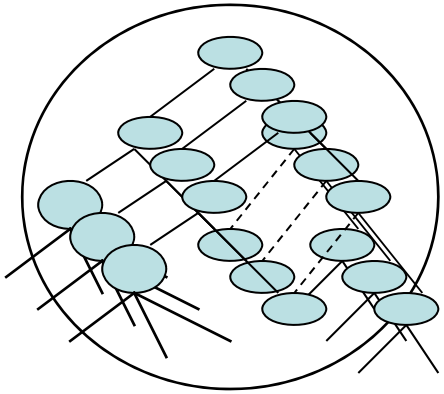
Main schedule B



Main schedule F



Main schedule Y



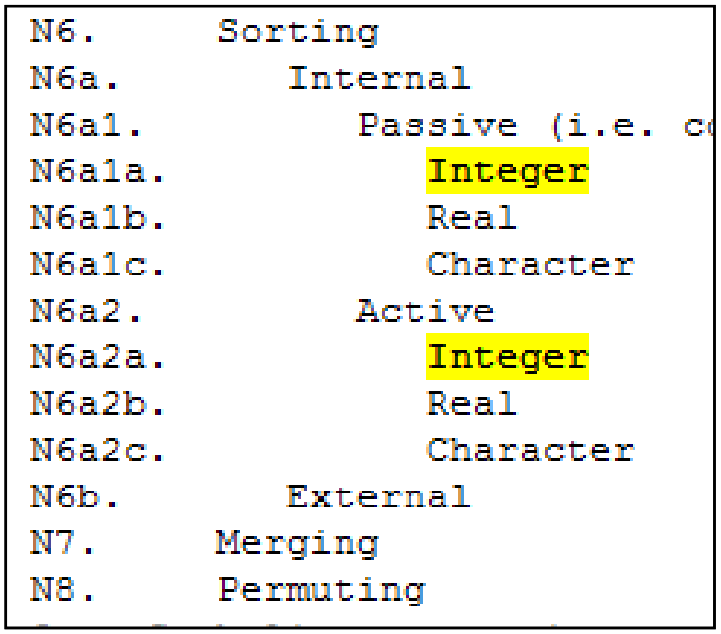
See example on the cover slide

Semantically meaningful orders in a classification system

- Common arrangements of coordinate classes may be based on one or more of the following principles:
  - stages in a process (e.g., brewing processes, packaging of product processes);
  - time or evolutionary sequence (e.g., ancient Greek sculptures, paleontology, stars);
  - degree of complexity (e.g., geometric figures),
  - size (e.g., town, cities, metropolis, and other administrative unites)
- According to Literary Warrant principle (e.g., arrange literature according to publication amount)
- According to User Warrant principle (e.g., arrange services and products according to popularity)

A notation has its *semantic value* and an *ordinal value*

- The semantic value of a classification number is the subject or concept it stands for.
  - handled through *skos:notation* for a given concept
- The ordinal value of a number or code places the subject into its determined rank in the scheme.
  - Question: how to handle?



Examples extracted from GAMS Problem Taxonomy

Options discussed for the issue of orders/sequences of coordinate classes

Options discussed :

- Use *skos:OrderedCollection* to include main schedules and use *skos:memberList* to show the member in an order.
- Combine nested ordered and unordered collections.
- Use of expressive notation that allows reliable sequencing.
- Other approaches discussed for previous issues.