# IDENTIFYING MANAGEMENT ISSUES IN NETWORKED KOS:
## examples from classification schemes

Aida Slavic
UDC Consortium, The Hague
aida.slavic@udcc.org

Antoine Isaac
Vrije Universiteit Amsterdam
aisaac@few.vu.nl

# OUTLINE

- Presentation focus : classification schemes & SKOS
- Two issues yet to be fully addressed in the vocabulary standards
  - scheme versioning: an issue in vocabulary sharing on the web
  - presentation and identification of pre-coordinated subject statements

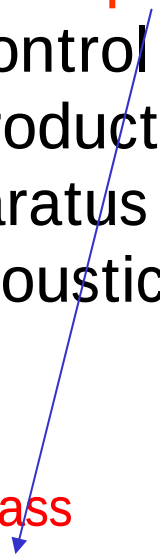
both issues relevant to other KOS as well

# KOS LIFE-CYCLE AND NATURE OF USE

- classification schemes (DDC, UDC, LCC…) – over 100 years of history behind management, development & use
- knowledge changes and causes changing of schemes
- once implemented in resource organization of physical collection classifications tend to be used for at least 20-50 years or longer – they hate updates and amendments
- new and old scheme versions are used side by side causing many issues in managing & mapping changes to facilitate information exchange
- 2 types of each KOS scheme may co-exist published on the web:
  - a standard reference scheme as published by the owner
  - user vocabularies (subject authority files)

# EXAMPLE 1: WHEN NEW KNOWLEDGE EMERGES

681 PRECISION MECHANISMS AND INSTRUMENTS
681.1 Apparatus with wheel or motor mechanisms
681.2 Instrument-making in general. Instrumentation.
681.3 Computers first placed here before 1980s
681.5 Automatic control engineering
681.6 Graphic reproduction machines and equipment
681.7 Optical apparatus and instruments
681.8 Technical acoustics. Musical instruments



Relocated to a new class
004 UDC
004/006 Dewey

# EXAMPLE 2: WHEN 'KNOWLEDGE' IS WRONG

| | |
|---|---|
| **=2** | **Western langauges** |
| =20 | English |
| =3 | Germanic languages |
| =4 | Romance or Neo-Latin languages |
| =50 | Italian |
| =60 | Spanish |
| =690 | Portuguese |
| =7 | Classic languages. Latin and Greek |
| =81 | Slavonic langauges |
| =88 | Baltic languages |
| **=9** | **Oriental, African and other languages** |
| =91 | Various Indo-European languages |
| =92 | Semitic languages |
| =94 | Hamitic languages |

...

Wrong classification of languages - causes wrong classification of:
- peoples
- literatures
- philology

# EXAMPLE 2: CORRECTED 20 YEARS AGO (UDC)

**=2 Western langauges**
=20 English
=3 Germanic languages
=4 Romance or Neo-Latin languages
=50 Italian
=60 Spanish
=690 Portuguese
=7 Cla
=81 Sla
=88 Bal
**=9 Or**
=91 Va
=92 Se
=94 Ha
…

Change to new scientific classification (1980s)

=1/=2 Indo-European languages
=3 Caucasian & other languages. Basque
=4 Afro-Asiatic, Nilo-Saharan, Congo-Kordofanian, Khoisan
=5 Ural-Altaic, Japanese, Korean, Ainu, Palaeo-Siberian, Eskimo-Aleut, Dravidian, Sino-Tibetan
=6 Austro-Asiatic. Austronesian
=7 Indo-Pacific, Australian
=8 American Indian (Amerindian) languages
=9 Artificial languages

# EXAMPLE 3: WRONG….

2 RELIGION. FAITHS

21/28 CHRISTIANITY

21 Natural theology. Theodicy. De Deo
22 The Bible. Holy scripture
23 Dogmatic theology
24 Practical theology
25 Pastoral theology
26 Christian church in general
27 General history of the Christian church
28 Christian churches, sects
29 NON CHRISTIAN RELIGIONS

Another bias inherited from Dewey which was against fundamental principles of UDC – designed to be 'universal' with respect to coverage, culture and application

# EXAMPLE 3: CORRECTED 10 YEARS AGO

2 RELIGION. FAIT

21/28 Christianity
21 Natural theology
22 The Bible. Holy
23 Dogmatic theolo
24 Practical theolog
25 Pastoral theolog
26 Christian church
27 General history
28 Christian church
29 NON CHRISTIAN

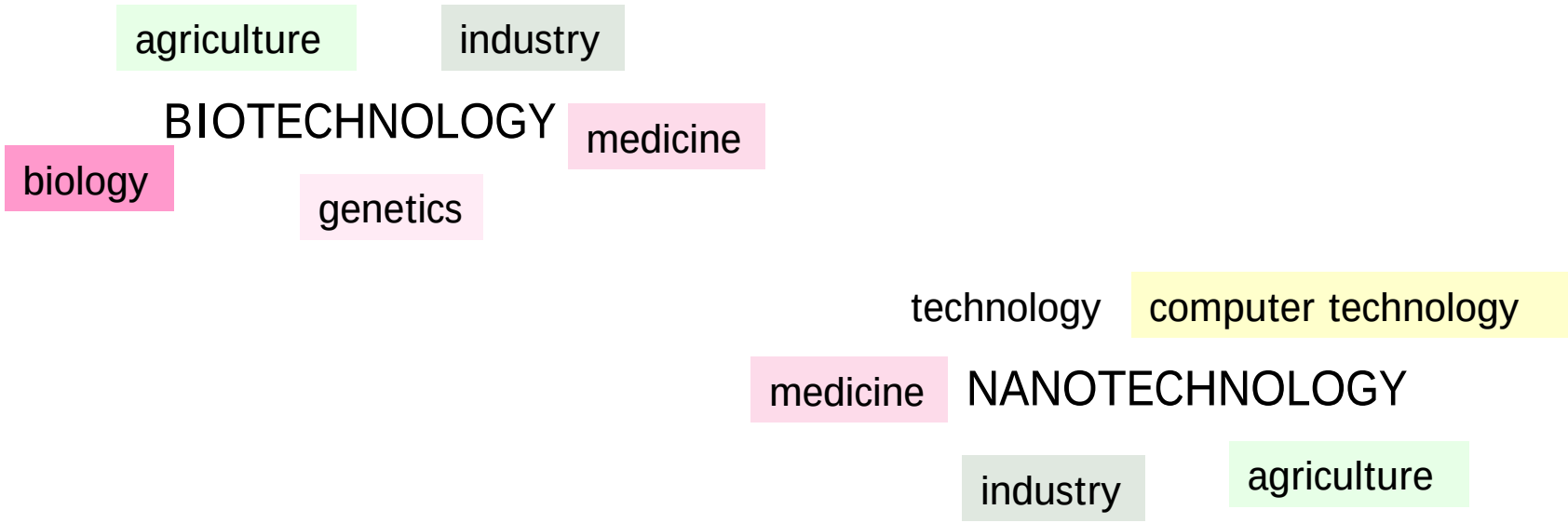
NOW.....

2 RELIGION. FAITHS
21 Prehistoric and primitive religions
22 Religions of the Far East
23 Religions of the Indian subcontinent
24 Buddhism
25 Religions of antiquity
26 Judaism
27 Christianity
28 Islam
29 Modern spiritua

-1 Theory, nature of religion
-2 Evidence of religion
-3 Persons in religion
-4 Religious practice
-5 Worship. Rites. Cult
-6 Processes in religion
-7 Religious organization
-8 Various properties
-9 History of the faith, religion, denomination or church

# IT HAPPENS ALL THE TIME… STARTS AS ONE CONCEPT…

- Finding logical place for new and pervasive concepts




- Problem not specific to DDC or UDC - no KOS is immune

# CHANGES IN SCHEME: ISSUE 1

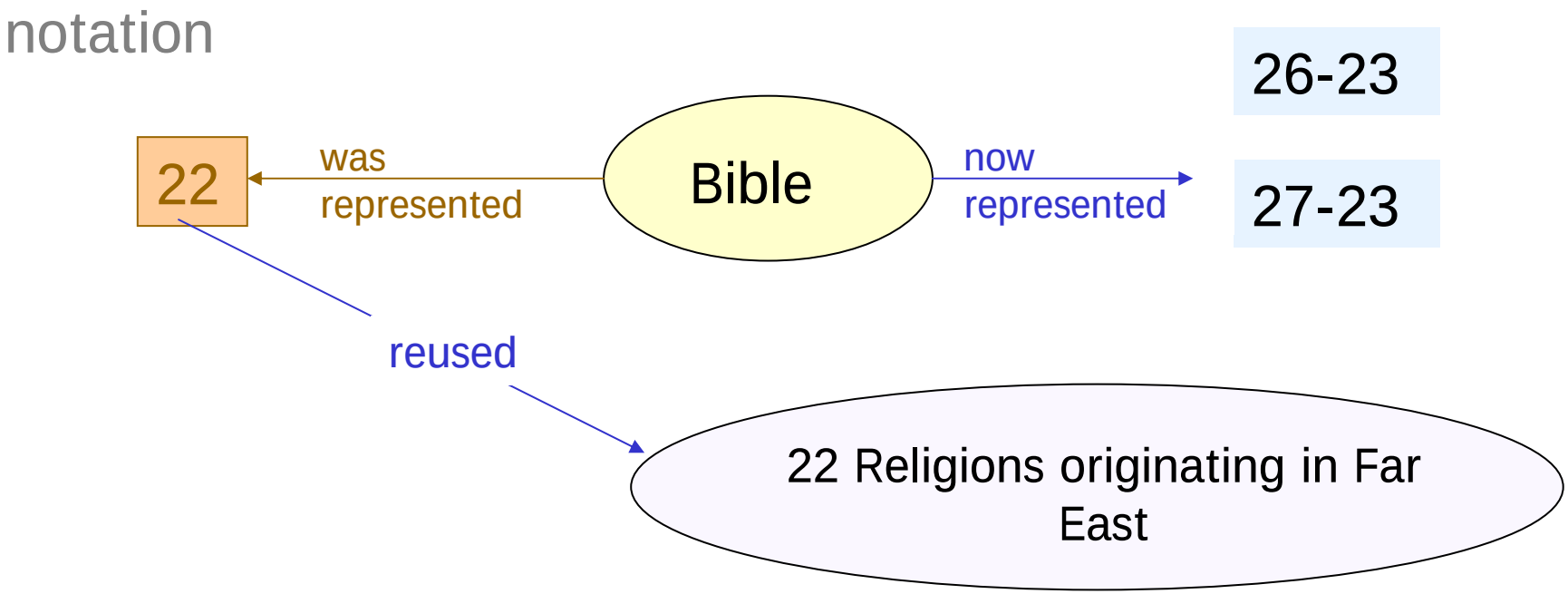
- moving/introducing entire hierarchies from one place of classification structure to another

  e.g. 40% of UDC has changed from 1990-2008

- old notation occasionally has to be re-used in order to preserve the logic of order

- as user continue to use old schemes: co-existence of two different notational representations for the same class/concept in information exchange is a reality

# RE-USE OF THE CANCELLED NOTATION




- unpopular but unavoidable
- can happen 10-50 years apart (desirable) or instantly (to avoid)

# CURRENT CANCELLATION MAPPING

Reports distributed to users: text and database files

http://www.udcc.org/cancellations.htm:

```
UDC CLASS NUMBER: 22
DESCRIPTION: The Bible. Holy scripture
REPLACED BY: 26-23 Judaism – Scriptures
             27-23 Christianity - Scriptures
```

# CANCELLATION MAPPING DATA

Managing notation history in the udc database:

**CLASS ID: 16544**
**NOTATION: 22 CAPTION: Religions originating in the Far East**
**INTRODUCED/DATE: 0012**

**REPLACES ID 15991: 299.1** Religions of Oriental Peoples

**NOTATION HISTORY:** yes **USED FOR: ID:17054:** Bible
**REPLACED BY: ID:17355:** Christian Bible

# URI & NOTATION HISTORY: DEWEY CASE


- Dewey URI example (linked data) - relating notation to an edition http://dewey.info/scheme/e22/ :

  641 Food & drink

- Maybe the same from 15th - 22th edition of Dewey, all still in use by many users

- How do we relate?

  641 from ed. 18

  641 in the ed. 22

# URI & NOTATION, EDITION : UDC CASE

- Standard UDC exist as Master Reference File only (updated and released annually as a file )

  - users purchase it and upload this into their systems
  - publishers world-wide translate and publish national printed editions

- 'time stamp' a weak option. Which date: introduced or db release?

  ..../UDCMRF/1993/22 [Bible]
  ..../UDCMRF/1999/22 [Bible]
  ..../UDCMRF/2000/22 [Religions originating in the Far East]

- More precise solution: notation + database ID

  ..../UDCMRF/22_17054 [Bible]
  .../UDCMRF/22_16554 [Religions originating in the Far East]

# HOW TO REPRESENT THIS IN SKOS?

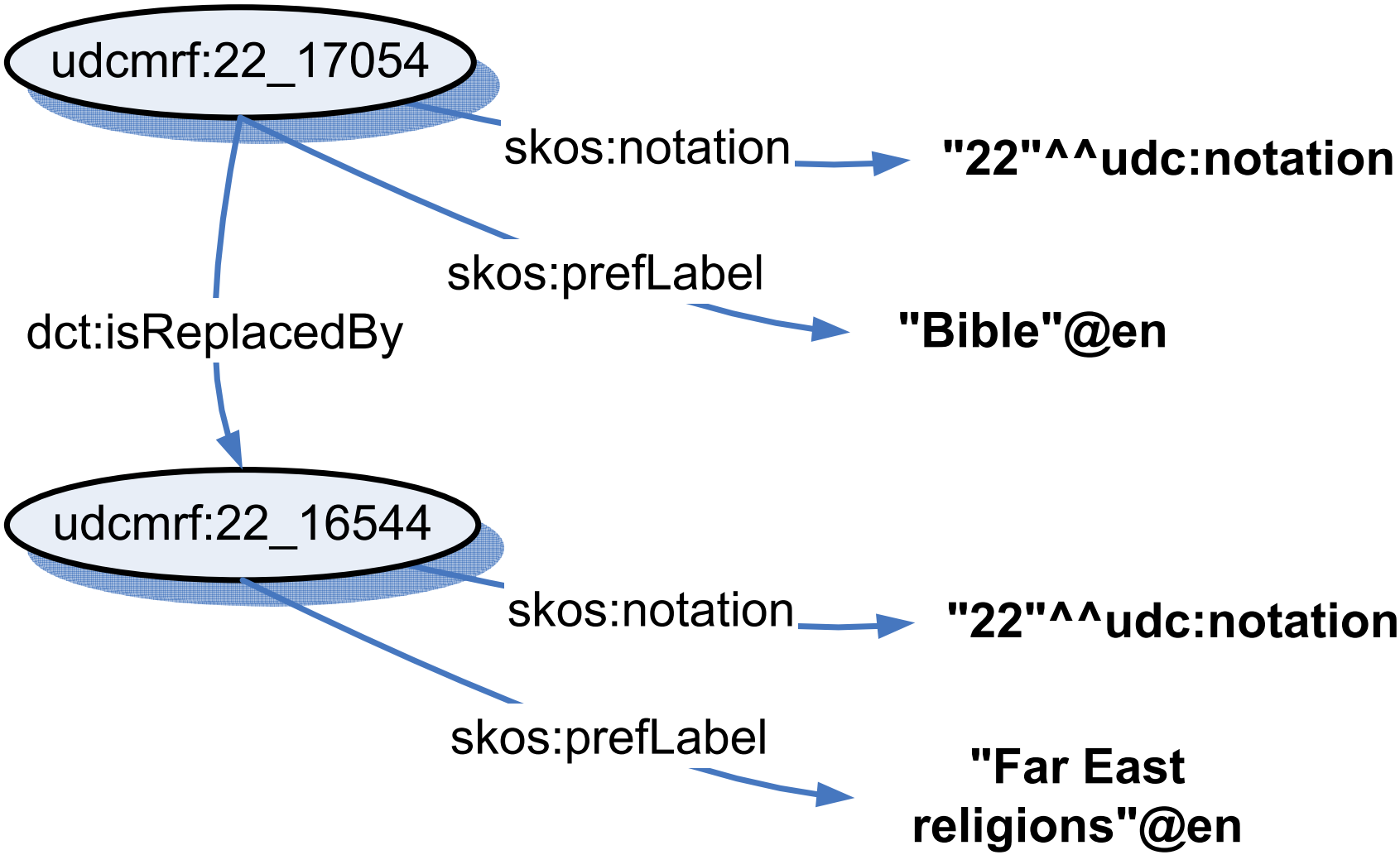
- Notations are first-order entities in classification schemes used as entry points and worth being served as Linked Data

- SKOS does not offer specific solution for concept representation versioning within a single KOS, and the relations between those!

- But in the RDF world, it is possible to use other models in combination with SKOS...
  - Dublin Core for versioning links the two classes using properties in the term namespace : *isReplacedby*; *replaces; isVersionOf*

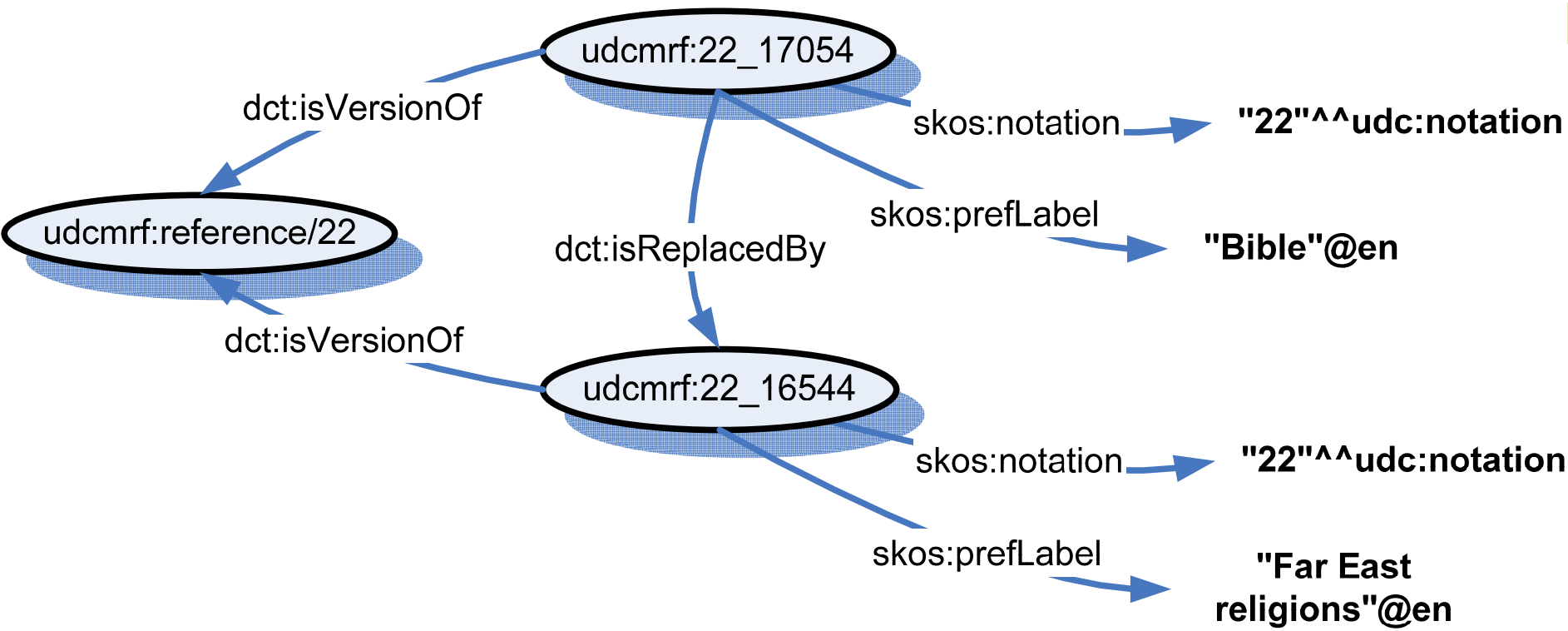
# IF WE FOCUS ON NOTATION... 22

The use of Dublin Core term: isReplacedBy



# NOTATION AS A REFERENCE

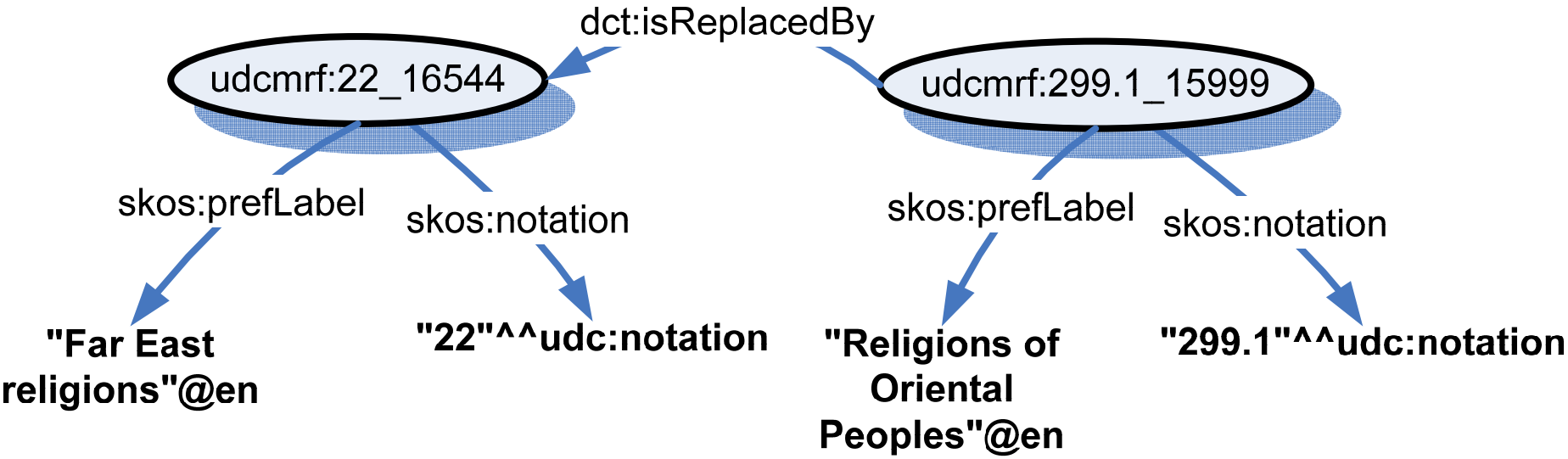
The use of Dublin Core term: isVersionOf



## Slide 18

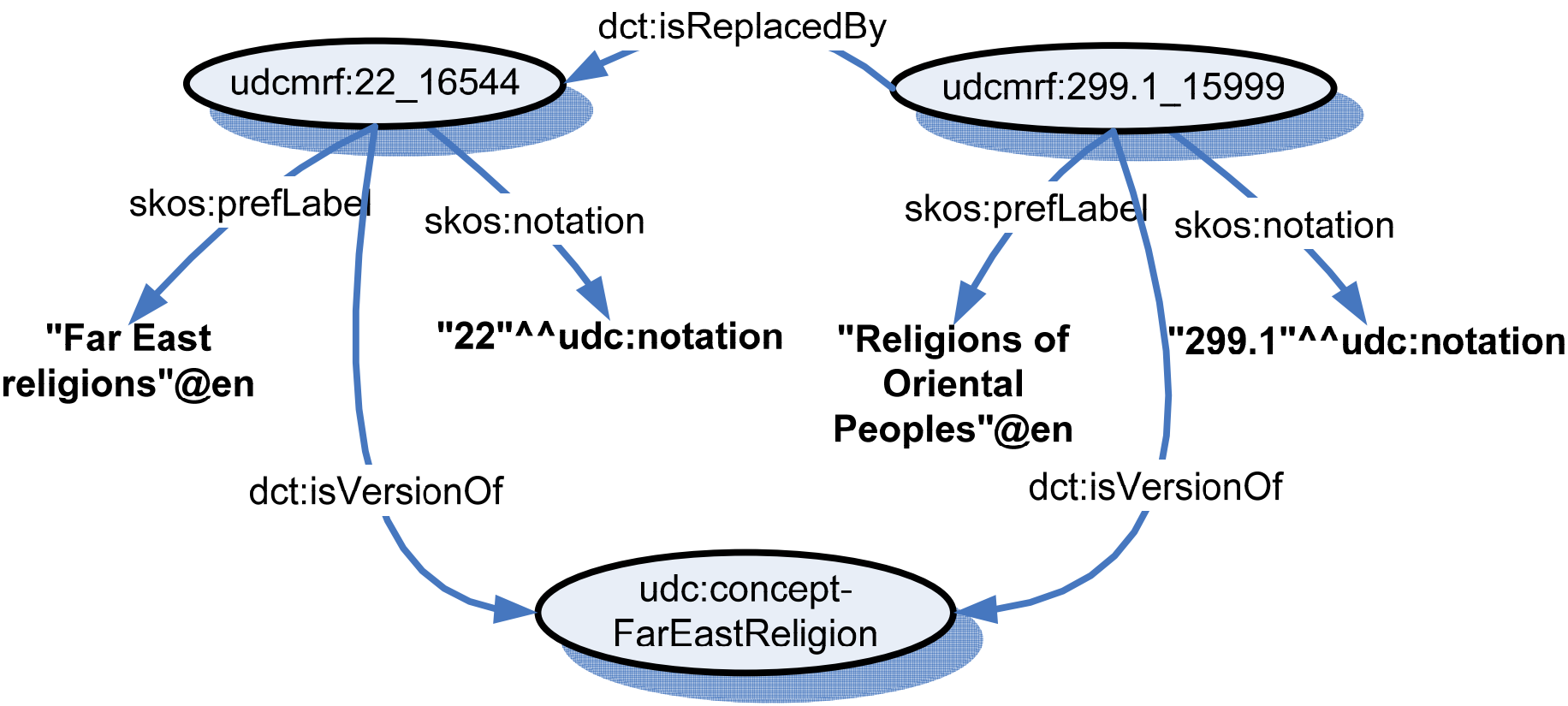
**AI20** I realize that you have removed the discussion on what notations are, but selected the choice that was made before that discussion, namely using "dct:isVersionOf".
I think we should go directly and use ore:aggregates, at least to be in accordance to our final proposal on slide 22.
Antoine, 29/09/2009

# IF WE FOCUS ON CONCEPT: “FarEastReligion”



# TOWARDS UDC:CONCEPT …

AI21

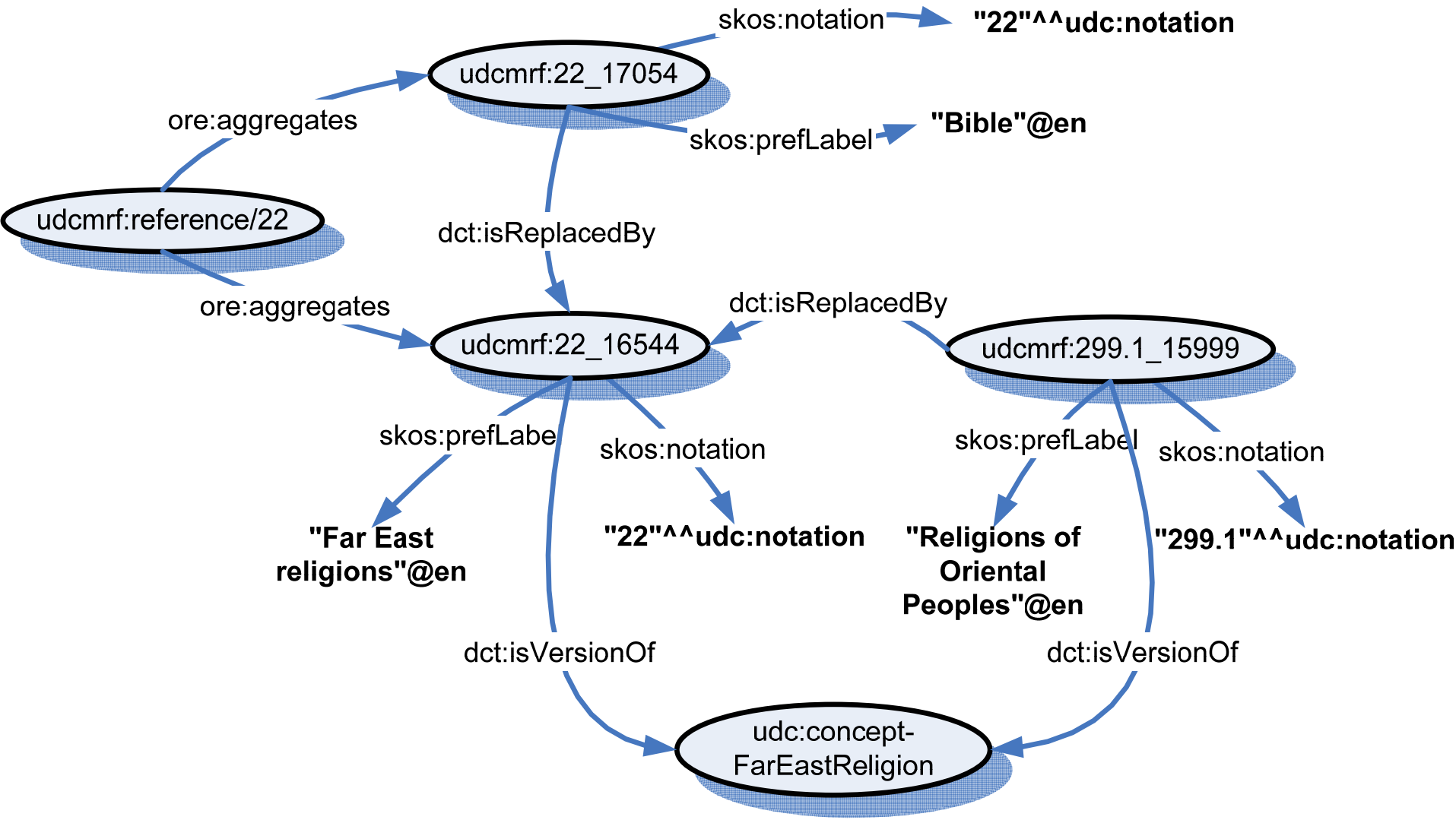


## Slide 20

**AI21** I removed from the graph an arrow that was saying that 299.1_15999 was narrower than 29.
I needed for demonstrating that 2999.1_15999 is of type skos:Concept. But it is not needed anymore as you skept that part of the story -- and rightly so, I guess, even though you should remember it if one asks the question ;-)
Antoine, 29/09/2009

# SAYING IT ALL WITH: SKOS, DC ...+

+ ore:aggregates (Object Reuse and Exchange)



# PART 2
# PROBLEMS IN REPRESENTING PRE-COMBINED SUBJECT STATEMENTS

# SYNTHETIC AND FACETED SCHEMES

- more flexible/modular - widely accepted as the best method of knowledge organization coping with changes while preserving the logic the structure
- instead of enumerating, hierarchical list of all needed/possible/imaginable subjects – we structure knowledge areas as knowledge building blocks: series of mutually exclusive facets
- Fundamental facet categories : broad, high-level categories such as entities, kinds, parts, processes, materials, time, place...

# SIMPLE MODEL: OBJECTS

NOTATION WITH FACET INDICATORS

**1 PRODUCTS**

| | |
|---|---|
| -1 | ACCORDING TO MATERIAL |
| -11 | Cardboard |
| -12 | Wooden |
| -13 | Metal |

| | |
|---|---|
| .0 | ACCORDING TO COLOUR |
| .02 | Red |
| .03 | Blue |
| .04 | Yellow |

| | |
|---|---|
| ’1 | ACCORDING TO SIZE |
| ’12 | small |
| ’13 | medium |
| ’14 | large |

**11 BOXES**
**12 VASES**
**13 SUITCASES**
**14 HANDBAGS**

11-12.03’14 Boxes-wooden-blue-large

13-13.03’14 Suitcases-metal-blue-small

# ADVANTAGES

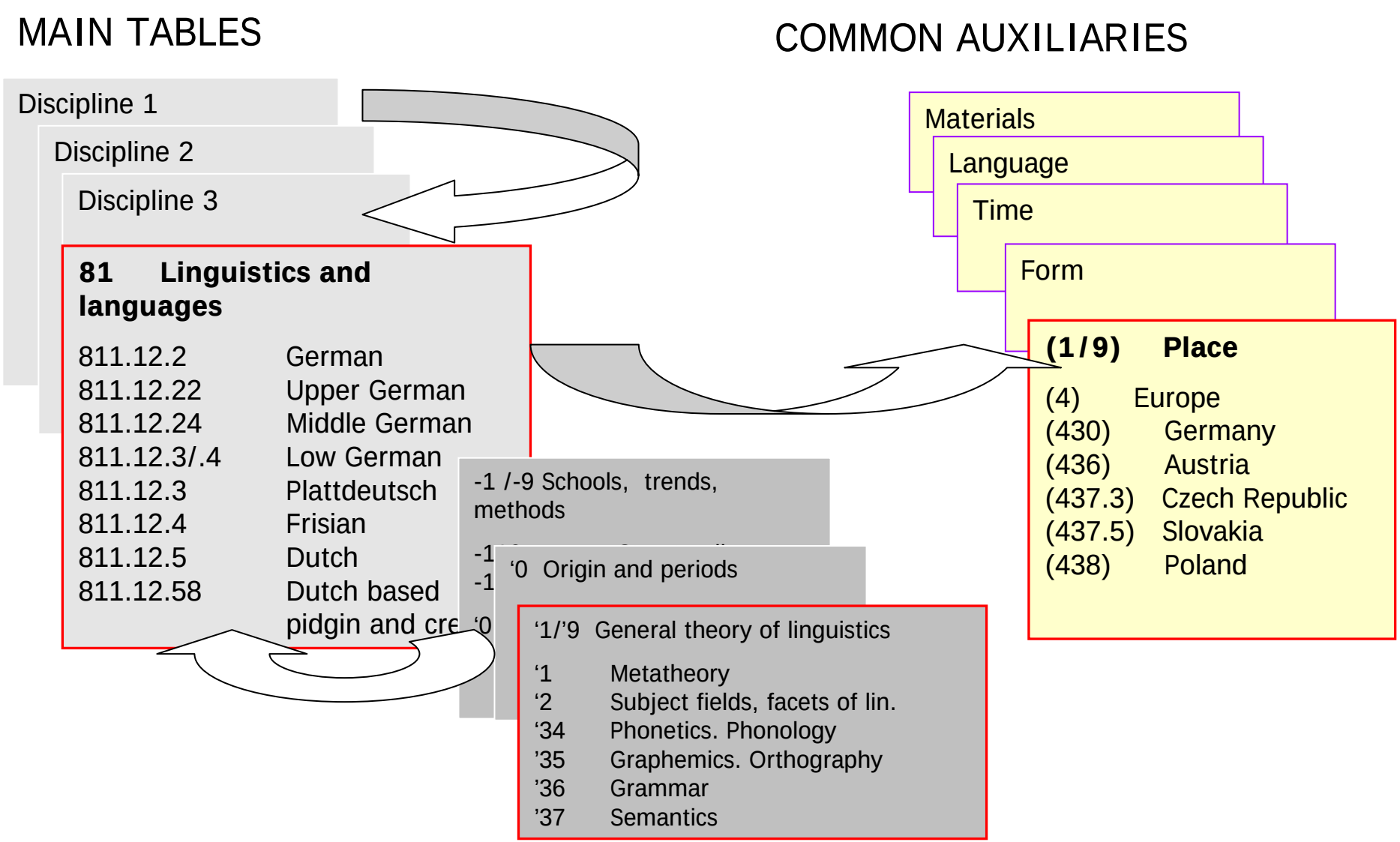
- logical structure – easier to expand, accommodate new knowledge

- smaller classification scheme – larger power in indexing

- no limitation in expressing the information content of the resource

- easier to translate simple notation to verbal expressions: one concept = one notational representation

# KNOWLEDGE: FULLY SYNTHETIC SYSTEM



# PRE-COORDINATION

- schemes contain simple and complex classes i.e. notational representation may appear simple or pre-combined

*special auxiliary numbers for religion*

2-23 Sacred books. Scriptures. Religious texts

| | |
|---|---|
| 2 | Religion |
| 27 | Christianity |
| 27-23 | Bible [Christianity – Sacred books] |
| 28 | Islam |
| 28-23 | Koran [Islam – Sacred books] |

- each simple concept has an URI
- precombined concepts inherit the meaning of their components [URIs of components have to be linked] but combinations create new meanings that may be presented as another URI or blank node in RDF

# SYNTHETIC SCHEMES IN USE

Faceted schemes offer basic vocabulary that is used to expressed complex document content (subject) at the point of indexing.

| | |
|---|---|
| 37:004 | Application of computers in education |
| 32:37 | Relationships between politics and education |

These complex expressions appear in the process of use and may not appear in the original scheme

- Complex notations built in the process of use are being published and shared as authority files independently from scheme owners (new sets of URI related to a standard KOS)

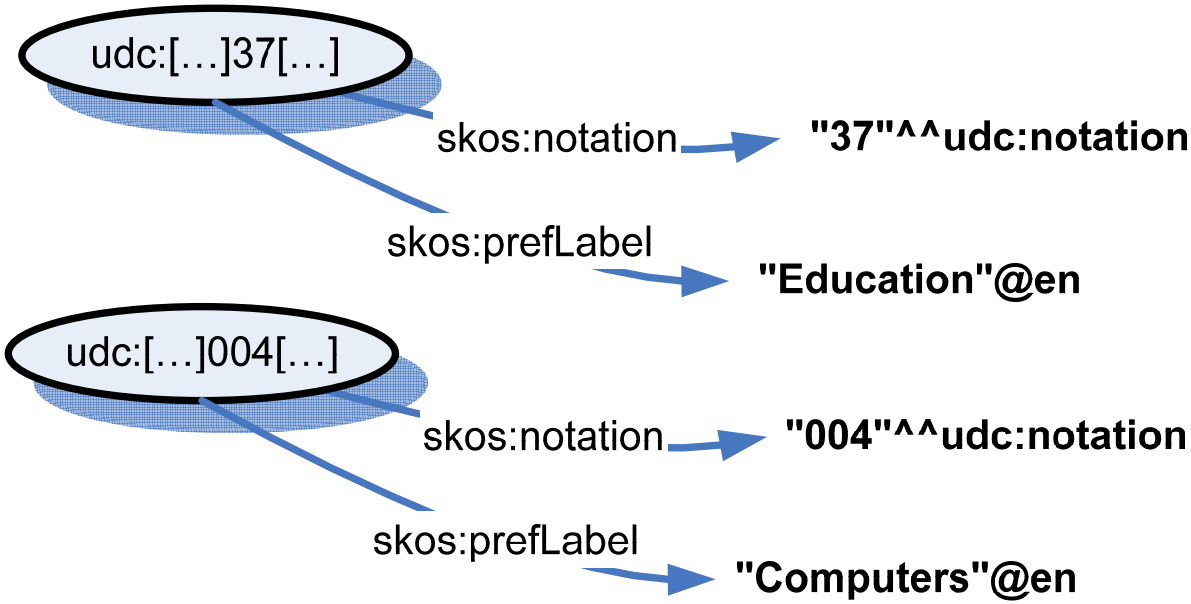
- Issue: standardizing link representations between notations from the original schemes and complex subject expressions developed at the point of indexing

# HOW DOES SKOS ADDRESS THIS?

- Currently no out-of-the box solution represents (pre)coordinations
- Pre(coordination) was acknowledged as an issue - but there were not enough resources to address it
- Some options are available to capture at least part of the information stored in a complex subject statement

# REPRESENTING ELEMENTS AS SEPARATE

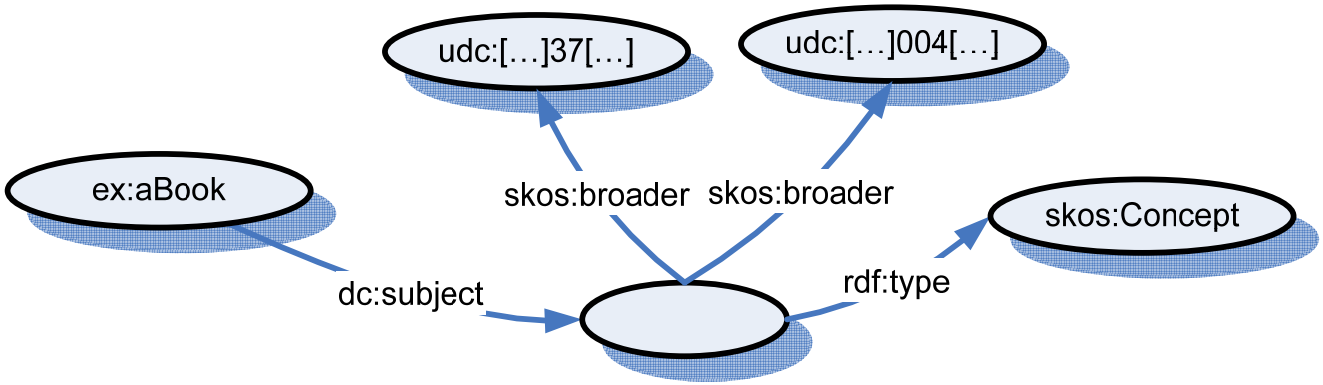
- Starting from 2 classes




- And a "sentence" used to describe a book


*Probably abusing dc:subject...*

# CREATING A NEW SUBJECT NODE



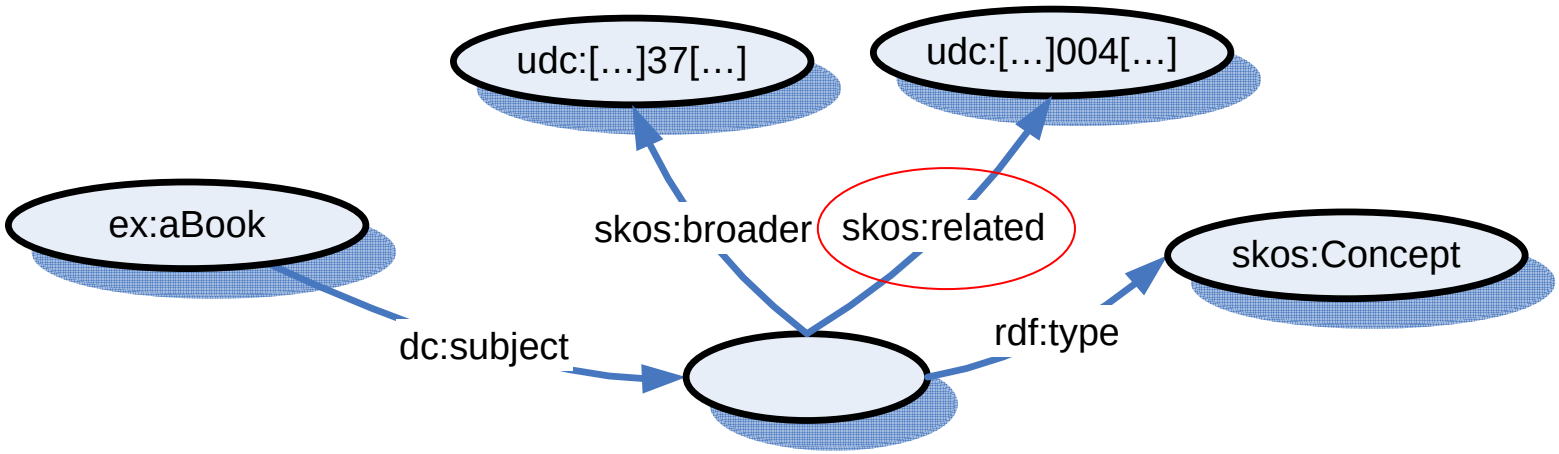

- We can create a new subject node that is connected to the combined concepts
- The node can be represented as anonymous: if information systems do not provide IDs for these collection-specific creations
- It is a skos:Concept (though it has a specific complex flavor)
- We can relate it to the 37 and 004 classes, expressing that the new anonymous class is more specialized
  - It allows to find the book starting from a query combining the skos:Concepts standing for 37 and 004

# PRE-COORDINATION: PHASE RELATIONSHIPS

- Phase relationships are not intersections between two concepts – one concept is the subject of the book the other is a ‘point of view’ from which the subject is presented/treated e.g.

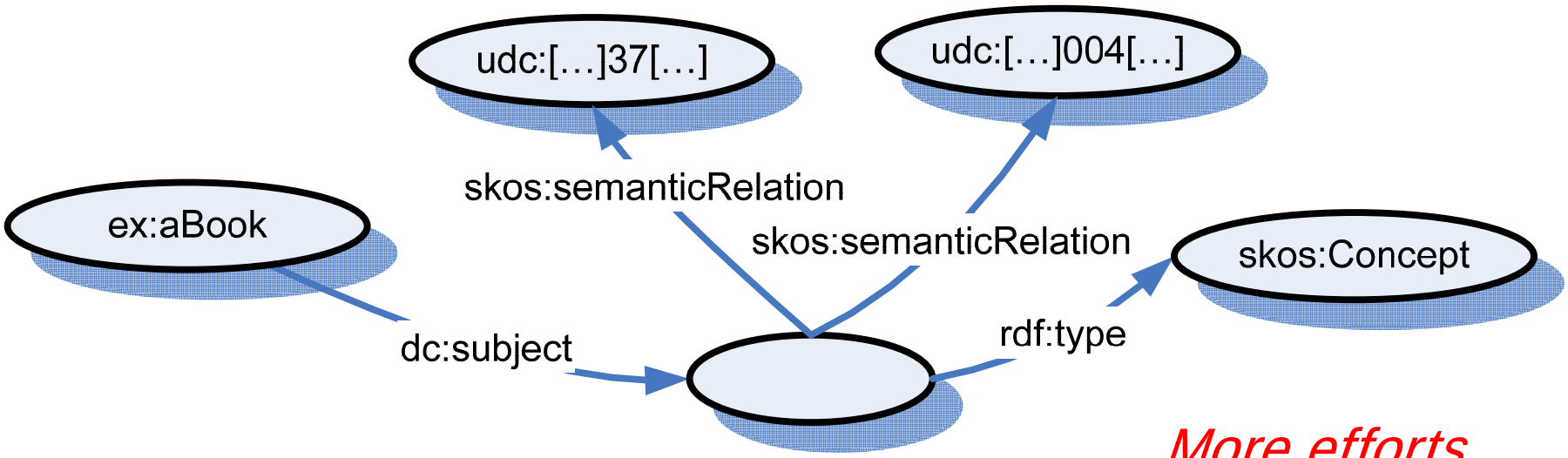
  37:004 Use of computers in education

  this is the subject of pedagogy and 004 is merely a related concept



# ANOTHER OPTION skos: semanticRelation

- specific type depends on:
  - the type of combination and role of concepts used in combination (another discipline?  place, time, process, object, form?)




*More efforts are required!*

# CONCLUSION

- the requirements for classification schemes in the published SKOS standard have not been entirely met

- issue 1: scheme versioning
  - changes in KOS are natural and cannot be avoided
  - the ‘old, obsolete and different views’ of knowledge originating from the same KOS coexist and are used side by side in real life

- issue 2: pre-coordination of terms in subject description
  - are logical requriement in content designation

- the above issues are relevant for other KOS (not only classification) and solutions can be found through shared efforts

# THANK YOU ...

References:

- SKOS Simple Knowledge Organization System: Reference
  http://www.w3.org/TR/skos-reference/
- DCMI Metadata Terms
  http://dublincore.org/usage/terms/
- Dewey as linked data
  http://dewey.info/class/641/2009/08/about.en
- Object Reuse and Exchange
  http://www.openarchives.org/ore/1.0/primer

# YOU ARE CORDIALLY INVITED TO …

http://www.udcc.org/seminar2009/