

# SKOS and Its Application in Transferring Traditional Thesauri into Networked Knowledge Organization Systems



**Marcia Lei Zeng**

Kent State University

**Wei Fan**

National Science Library  
Chinese Academy of Sciences

Acknowledgement:  
**Dongbo Wang** and **Shuqing Bu**  
National Library of China

OCLC/ISKO-NA Conference, Aug. 2008, Université de Montréal



# Outline

## I. Background

- Chinese Classified Thesaurus (CCT)

## II. SKOS -- a quick review

## III. Challenges and Issues -- when applying SKOS to CCT

- SKOS for [enumerative] classification systems
- SKOS for mapped vocabularies

## IV. Future Direction:

From machine-readable to machine-processable

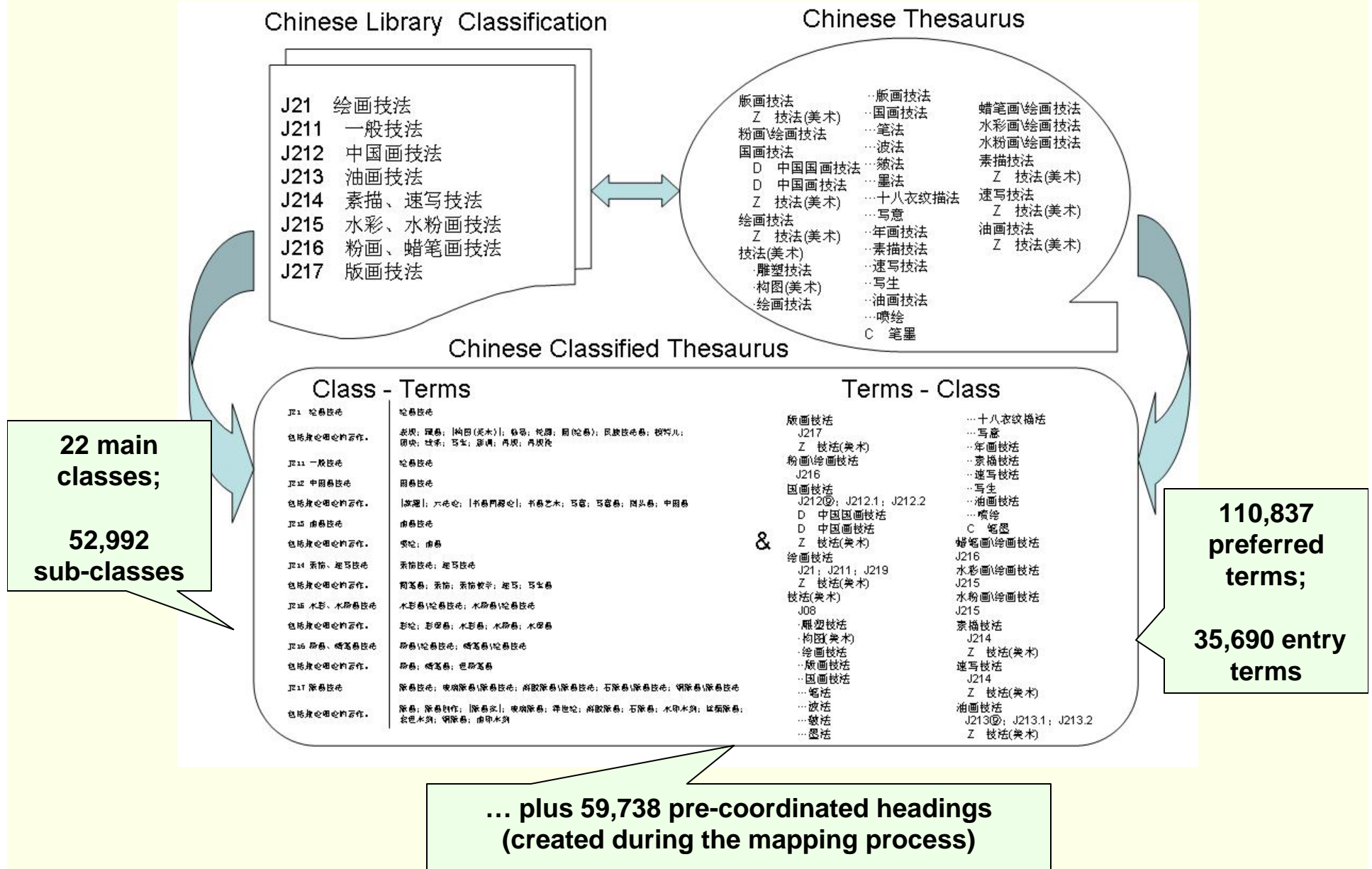
SKOS = Simple Knowledge Organisation System



# I. Background

***Chinese Classified Thesaurus (CCT)***

## CCT: The Vocabulary



# CCT: The Electronic Version (1. browsing)

Notation-to-Term		Term-to-Notation
notation + caption + notes	Term(s)	term entry + notation(s)
<b>R51 传染病</b> 总论传染病的预防入R183。	<b>传染病</b> 传染病学；动物源性疾病	a b c d e f g h i j k l m n o p q r s t w 其他
R511 病毒传染病 病毒性肺炎入R563.1；病毒性皮肤病入R752；腹股沟肉芽肿入R759.4；沙眼入R777.32。	<b>病毒病</b>	叮丁啉 Y 氮杂环丁烷
R511.1 麻疹	<b>麻疹</b>	<b>Non-preferred term</b> Y = Use  <b>Preferred term</b> English term notations D = Used For Z = Top Term C = Related Term S = Broader Term F = Narrower Term
R511.2 风疹	<b>风疹</b>	
R511.3 天花 种痘入R186+.1。	<b>天花</b>	叮啉 Acridine 0626.32+6 D 氮杂蒽 Z 杂环化合物 C 叮啉生物碱 C 阿的平
R511.5 水痘	<b>水痘</b> 痒疹	叮啉\化工生产 TQ253.26
R511.6 感冒（伤风）	<b>感冒（伤风）</b>	叮啉\衍生物 0626.32+6
R511.7 流行性感冒	<b>流行性感冒</b>	叮啉\衍生物\化工生产 TQ253.26
R511.8 腺病毒感染	<b>腺病毒\感染</b>	叮啉染料 Acridine dyes TQ613.9 Z 染料
R512.1 流行性腮腺炎	<b>流行性腮腺炎</b>	叮啉染料\化工生产 TQ613.9
R512.3 脑炎及脑脊髓膜炎	<b>流行性脑炎；脑炎 脑脊髓炎</b>	叮啉生物碱
R512.31 甲型脑炎（尼克诺莫氏脑炎）	<b>流行性脑炎\类型</b>	
R512.32 乙型脑炎（日本脑炎）	<b>流行性乙型脑炎</b>	
R512.33 丙型脑炎（美国圣路易脑炎）	<b>流行性脑炎\类型</b>	
R512.34 森林脑炎（苏联脑炎）	<b>森林脑炎</b>	
R512.39 其他 淋巴球性脉络丛脑膜炎、澳大利亚第十管睡性脑炎、无菌性脑脊髓膜炎入此。	<b>流行性脑炎</b>	
R512.4 脊髓灰质炎（小儿麻痹症）	<b>脊髓灰质炎</b>	

**multiple terms (;)** points to R511.8 and R512.1

**Term \ term coordination** points to R511.8, R512.1, R512.3, R512.31, R512.32, R512.33, R512.34, R512.39, and R512.4

Source: <http://clc.nlc.gov.cn/ztfzfbdzj.jsp>

## CCT: The Electronic Version (2. searching)

The screenshot displays the CCT (China Classification Table) software interface, which is used for searching and classifying terms. The interface is divided into several panes and sections, each with specific functions:

- Class Hierarchy:** Located on the left, it shows a tree structure of classification categories. The selected path is R5 (内科学) > R51 (传染病学) > R511 (病毒传染病).
- Class Notation, Caption, Notes + Thesaurus Terms + Record #:** This section at the bottom left provides detailed information for the selected term. It includes the notation R51, the term 传染病 (Infectious diseases), a note about the prevention of infectious diseases (总论传染病的预防入R183), and the record number C019527.
- Term Hierarchical Display:** This pane shows the hierarchical structure of the selected term. It lists various sub-terms under 传染病, such as 传染病防治 (Infectious disease prevention), 传染病学 (Infectious diseases), 传染病医院 (Infectious disease hospital), 传染病媒介 (Infectious disease vector), 传染期 (Infectious period), 传染性单核细胞增多症 (Infectious mononucleosis), 传染性坏死性肝炎 (Infectious necrotizing hepatitis), 传染性胃肠炎 (Infectious gastroenteritis), and 传染源 (Infectious source).
- Term Entry with Notations:** This pane shows the term entry for 传染病, including its English translation (Infectious diseases), the notation R51@, R519.9, and a list of related terms like 传染性法氏囊病 (Infectious bursal disease), 法定传染病 (Notifiable infectious diseases), 自然疫源性传染病 (Zoonotic infectious diseases), 病毒病 (Viral diseases), 立克次氏体病 (Rickettsial diseases), and 螺旋体感染 (Spiral body infection).
- Top Term Index:** This pane provides a list of top-level terms and their corresponding classification numbers. The list includes R511, R512.99, R51, R519.9, R183, R183.9, S855, R197.5, R18, R373.1, and R155.3+4.
- Term Index:** This pane shows a list of terms and their corresponding classification numbers. The list includes 传染病 (Infectious diseases), 病毒病 (Viral diseases), 立克次氏体病 (Rickettsial diseases), 螺旋体感染 (Spiral body infection), 细菌病 (Bacterial diseases), 衣原体感染 (Chlamydia infection), 真菌病 (Fungal diseases), and 支原体病 (Mycoplasma diseases).



# CCT: The Format – a combination of ...

## *CNMARC Format for Classification Data*

(= a.k.a. CLCMARC )

Based on UNIMARC Classification Format

&

## *CNMARC for Authorities*

– Subject Authorities

Based on UNIMARC Authorities format

### (四) 数据区 : IV. Data

- 0-- 标识块 Identification Block
- 1-- 编码信息块 Coded Information Block
- 2-- 类目标识块 Heading Block
- 3-- 附注块 Notes block
- 4-- 单纯参照块 See Reference Block
- 5-- 相关参照块 See Also Reference Block
- 66- 类号组配说明块 Number Building Block
- 700-754 索引款目块 Index Term Block
- 8-- 信息来源块 Source Information Block
- 9-- 国内使用块 National Use Block

- 0-- 标识块 Identification Block
- 1-- 编码信息块 Coded Information Block
- 2-- 标目块 Heading Block
- 3-- 标目附注块 Information Note Block
- 4-- 单纯参照根查块 See Reference Tracing Block
- 5-- 相关参照根查块 See Also Reference Tracing Block
- 6-- 分类号块 Classification Number Block
- 7-- 连接标目块 Linking Heading Block
- 8-- 来源信息块 Source Information Block
- 9-- 国内使用块

Mapping levels: Major mapping, minor mapping, overlapping

[Link to CCT's [format summary](#) ]

2008-08-05

©zeng

7

Source: Based on CLC FAQ, <http://clc.nlc.gov.cn/qa.do?action=listAll&page=1>



## II. SKOS – a quick review





# SKOS

SKOS = Simple Knowledge Organization System

- a common data model for sharing and linking knowledge organization systems (KOS) via the Semantic Web.
  - KOS examples: thesauri, taxonomies, classification schemes, subject heading systems ... ..



# SKOS

- Does not aim to replace the original objects in their initial context of use
- Allows to **port them to a shared space** based on a **simplified model**
- Enables wider re-use and better interoperability



# SKOS

- Current version  
*SKOS 2008*
  - SKOS Reference (W3C Working Draft 9 June 2008)
    - Replaces SKOS Core Guide (i.e. *SKOS 2005*)
    - SKOS Primer (W3C Working Draft 21 February 2008)
    - SKOS RDF Schema (June 3rd 2008 Edition)
    - SKOS eXtension for Labels (XL) RDF Schemas (June 3rd 2008 Edition)

# SKOS Synopsis (1)

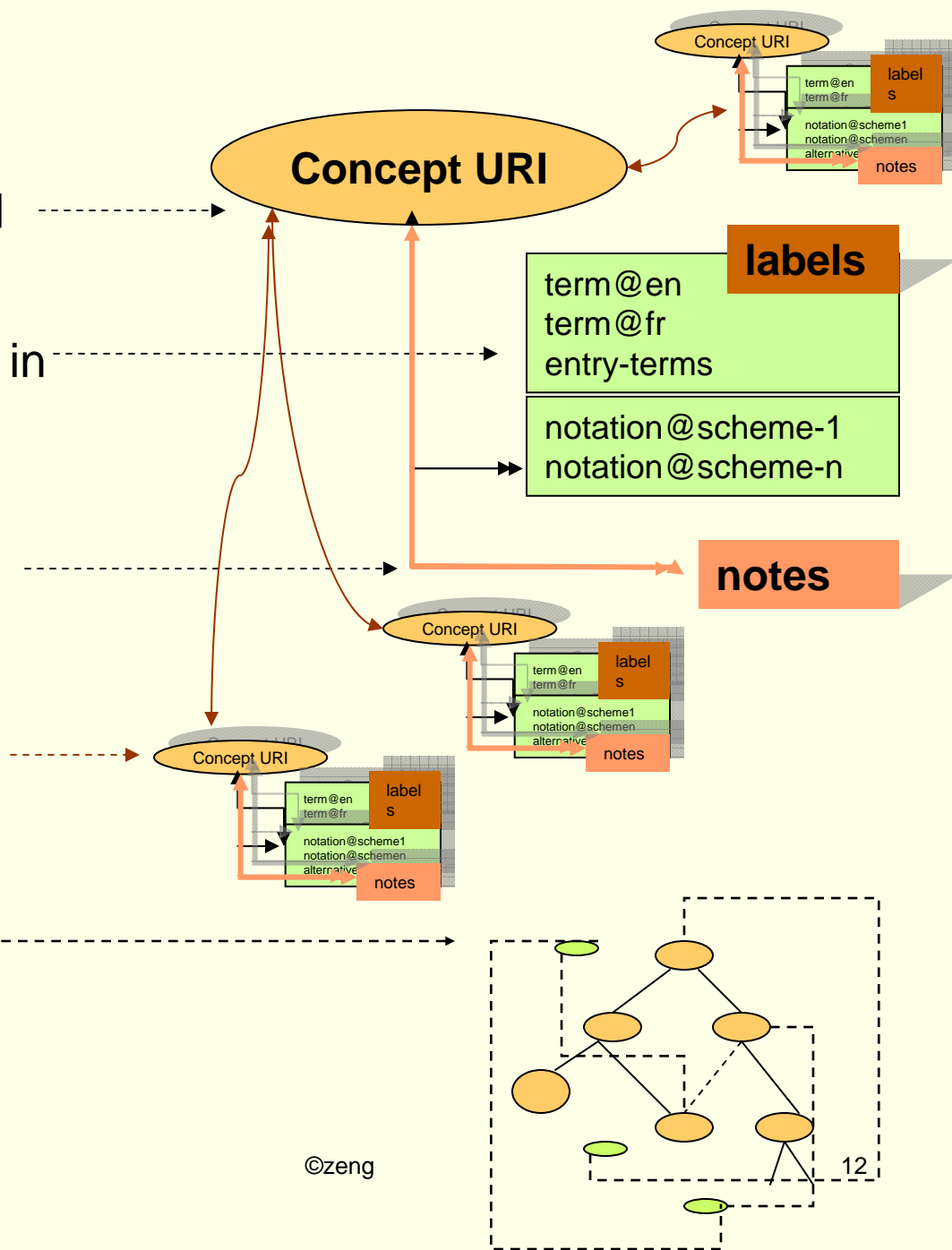
Using SKOS --

- **concepts** can be identified using URIs
  - **labeled** with lexical strings in one or more natural languages, and assigned **notations** (lexical codes)
  - **documented** with various types of notes
  - **linked** to other concepts
- and

organized into informal hierarchies and association networks

(to be continued →)

2008-08-05



©zeng

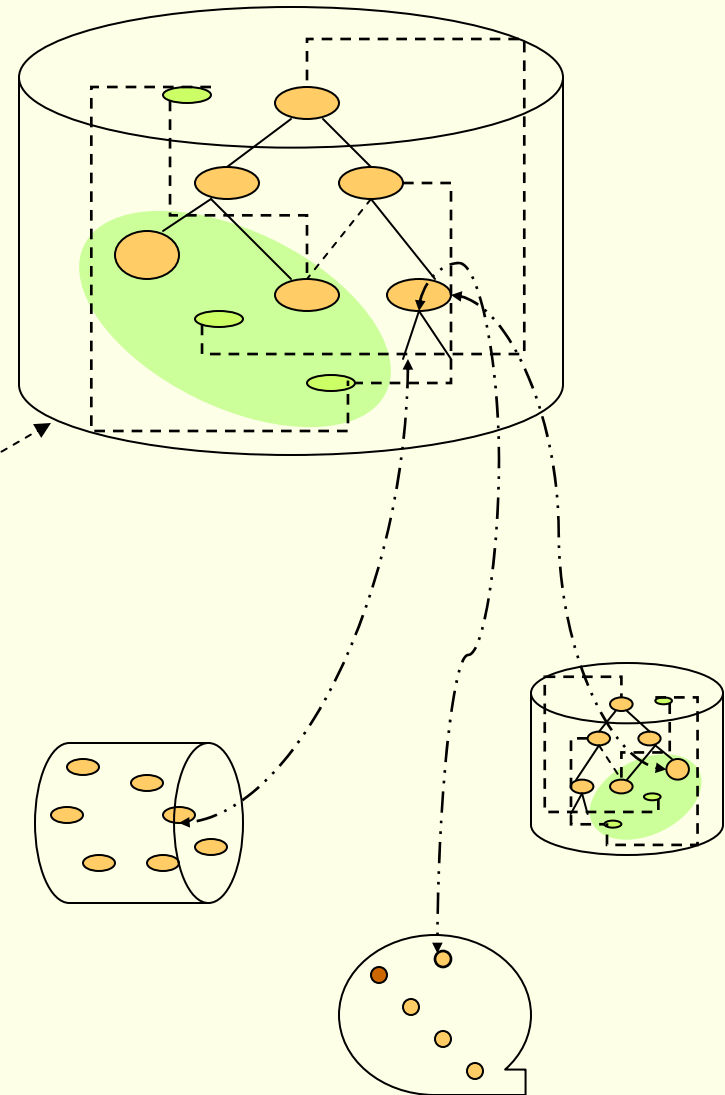
## SKOS Synopsis (2)

(continued from previous page)

- aggregated into **concept schemes**,

- [grouped into labeled and/or ordered **collections**, ]

- and **mapped** to concepts in other schemes.

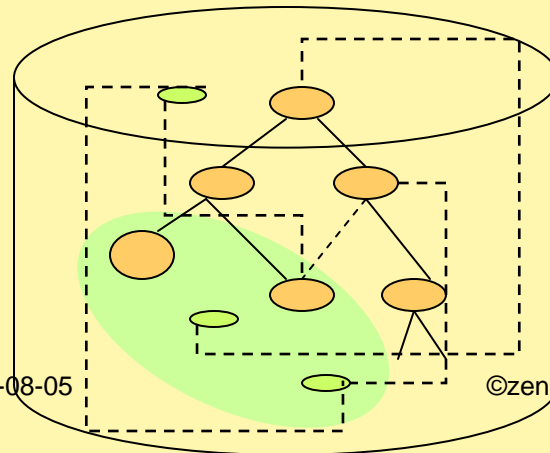




## SKOS: A Standard Set of...

- Resource types (Classes)
- Link types (Properties)

Classes	Properties	
<ul style="list-style-type: none"><li>• ConceptScheme</li><li>• Concept</li></ul>	<ul style="list-style-type: none"><li>• semantic relations (broader, narrower, related, etc.)</li></ul>	<ul style="list-style-type: none"><li>• Lexical properties (prefLabel, altLabel)</li></ul>
<ul style="list-style-type: none"><li>• Collection</li><li>• OrderedCollection</li></ul>	<ul style="list-style-type: none"><li>• membership relations (member, memberList)</li></ul>	<ul style="list-style-type: none"><li>• Document notes (scopeNote, definition, etc.)</li></ul>



2008-08-05

©zeng



# A typical thesaurus entry

[concept]

Term: Economic cooperation

Used For:  
Economic co-operation

Broader terms:  
Economic policy

Narrower terms:  
Economic integration  
European economic cooperation  
European industrial cooperation  
Industrial cooperation

Related terms:  
Interdependence

Scope Note:  
Includes cooperative measures in  
banking, trade, industry etc.,  
between and among countries.

**Preferred label**

**Alternative label**

**Broader concept**

**Narrower concept**

**Related concept**

**Scope note on concept**

2008-08-05

Example of the entry is from:  
Alistair Miles, SKOS Core Tutorial, DC-2005 Madrid



## An RDF/XML serialization of the RDF description of the 'Economic cooperation' concept

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:skos="http://www.w3.org/2004/02/skos/core#">

  <skos:Concept rdf:about="http://www.ukat.org.uk/thesaurus/concept/1750">
    <skos:prefLabel>Economic cooperation</skos:prefLabel>
    <skos:altLabel>Economic co-operation</skos:altLabel>
    <skos:scopeNote>Includes cooperative measures in banking, trade, industry etc.,
      between and among countries.</skos:scopeNote>
    <skos:broader rdf:resource="http://www.ukat.org.uk/thesaurus/concept/4382"/>
    <skos:narrower rdf:resource="http://www.ukat.org.uk/thesaurus/concept/2108"/>
    <skos:narrower rdf:resource="http://www.ukat.org.uk/thesaurus/concept/9505"/>
    <skos:narrower rdf:resource="http://www.ukat.org.uk/thesaurus/concept/15053"/>
    <skos:narrower rdf:resource="http://www.ukat.org.uk/thesaurus/concept/18987"/>
    <skos:related rdf:resource="http://www.ukat.org.uk/thesaurus/concept/3250"/>
    <skos:inScheme rdf:resource="http://www.ukat.org.uk/thesaurus"/>
  </skos:Concept>

</rdf:RDF>
```



# SKOS eXtension for Labels (XL)

- an OPTIONAL extension of SKOS
- provides additional support for identifying, describing and linking lexical entities
- Defines:

## **class**

**xl:Label**

## **labeling properties**

**xl:prefLabel**

**xl:altLabel**

**xl:hiddenLabel**

## **relation property**

**xl:labelRelation**

### **SKOS 2008 June version changes:**

- A new section on notations
- A new appendix "SKOS eXtension for Labels"
- The section on label relations was dropped. This is replaced by the xl:labelRelation property in the eXtension for Labels appendix.

2008-08-05

©zeng

17



# III. Challenges and Issues

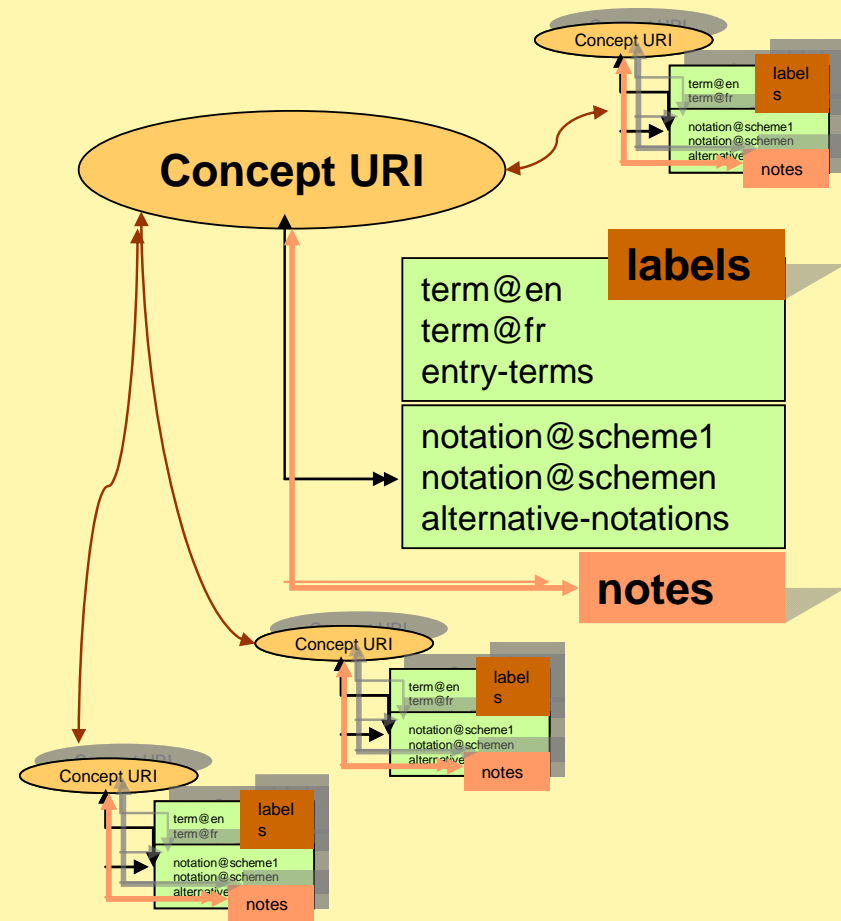
-- when applying SKOS to CCT



# 1. SKOS for [enumerative] Classification Systems

# SKOS is very good for thesauri (1)

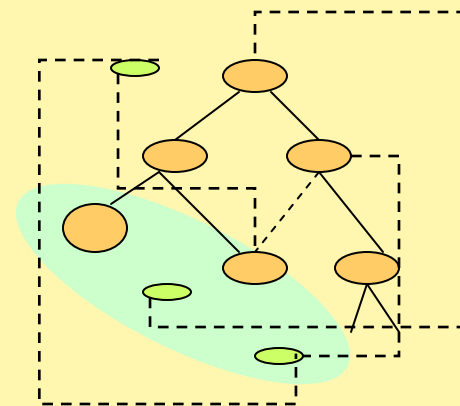
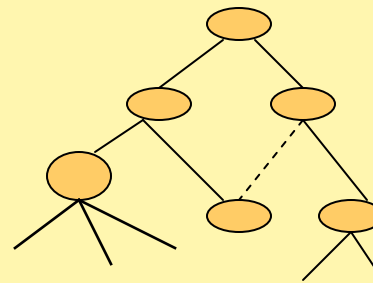
- For concepts:
  - Clear representation of:
    - a concept's properties
    - and the relationships among or between the concepts
  - (hierarchical and associated)





## SKOS is very good for thesauri (2)

- For hierarchical displays:
  - `skos:hasTopConcept`
  - `skos:broader`
  - `skos:narrower`
- For non-hierarchically related concepts
  - `skos:related`
- For groups under node labels
  - `skos:Collection`
  - `skos:member`



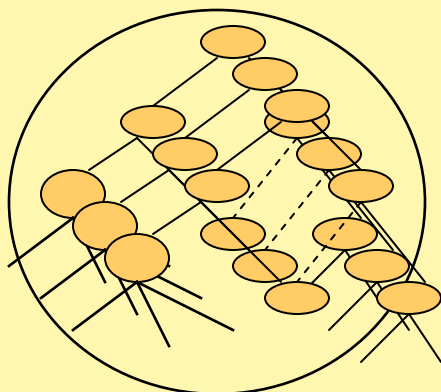


Is SKOS <sup>good?</sup>  
<sup>weak?</sup>  
<sup>not ?</sup>  
for [enumerative] Classification Systems ?

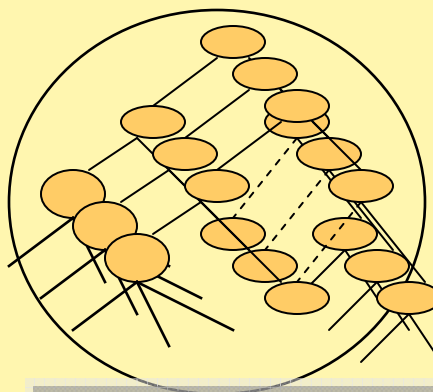
- Good in showing hierarchical relationships
- But ...

# 1. What about the order of main classes/ schedules?

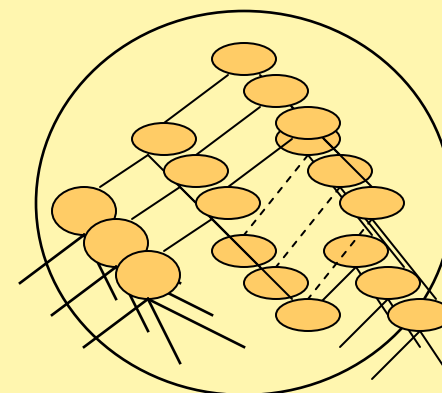
Main class B



Main class F



Main class Y



2008-08-05

Yes, it is still fine. It is just not perfect ...





(Cont.) 1. What about the order of main classes/schedules?

## Options discussed :

- Treat main classes as independent schemes  
(Main classes of CLC have been issued as specialized classification schemes).
- Use **skos:OrderedCollection** to include main classes and used **skos:memberList** to show the members in an order.



## 2. Alternative classification notations

- cross-listed notations
- e.g., **[Q89]** environmental biology  
Preferred class: X17

### N 自然科学总论

**N0** 自然科学理论与方法论  
**N1** 自然科学现状及发展  
**N2** 自然科学机关、团体、会议  
**N3** 自然科学研究方法  
**N4** 自然科学教育与普及  
**N5** 自然科学丛书、文集、连续性出版物  
**N6** 自然科学参考工具书  
**[N7]** 自然科学文献检索工具  
**N79** 自然科学非书资料、视听资料  
**N8** 自然科学调查、考察  
**N91** 自然研究、自然历史  
**N94** 系统科学  
**[N99]** 情报学、情报工作

### O 数理科学和化学

**O1** 数学  
**O3** 力学  
**O4** 物理学  
**O6** 化学  
**O7** 晶体学

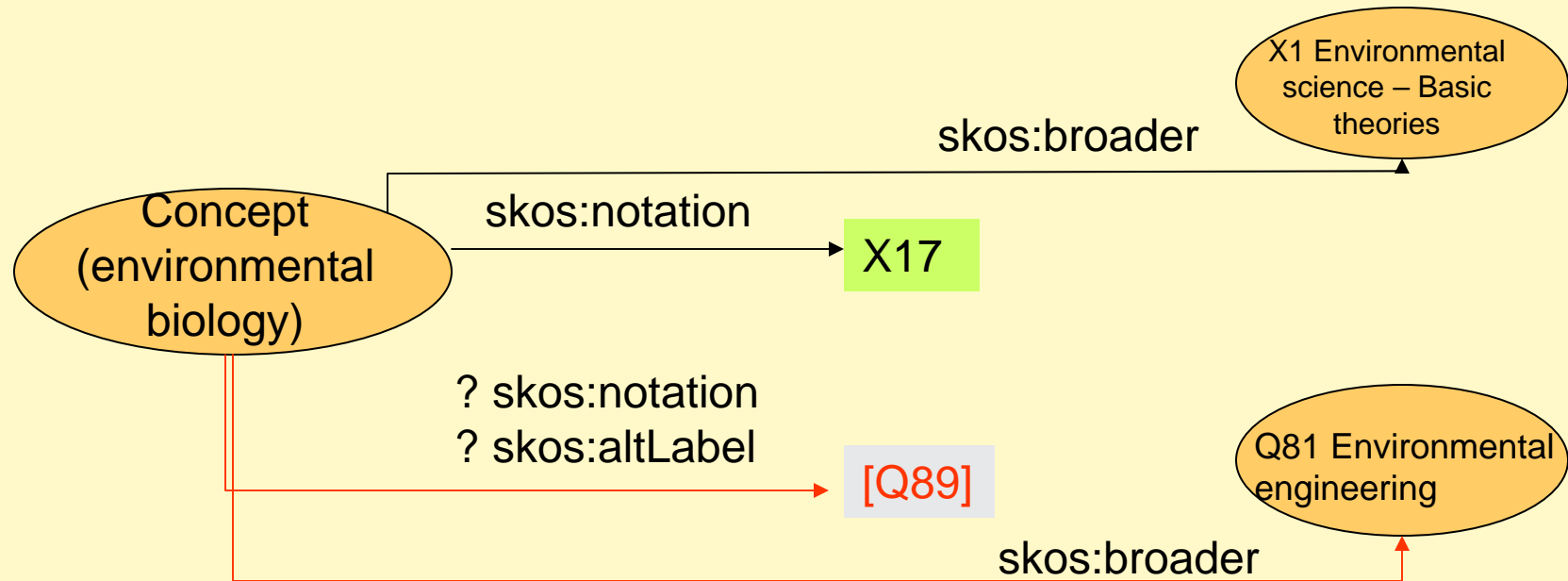
### P 天文学、地球科学

**P1** 天文学  
**P2** 测绘学  
**P3** 地球物理学  
**P4** 大气科学(气象学)  
**P5** 地质学  
**P7** 海洋学  
**P9** 自然地理学

### Q 生物科学

**Q1** 普通生物学  
**Q2** 细胞生物学  
**Q3** 遗传学  
**Q4** 生理学  
**Q5** 生物化学  
**Q6** 生物物理学  
**Q7** 分子生物学  
**Q81** 生物工程学(生物技术)  
**[Q89]** 环境生物学  
**Q91** 古生物学  
**Q93** 微生物学  
**Q94** 植物学  
**Q95** 动物学  
**Q96** 昆虫学  
**Q98** 人类学

**[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.



### 3. Top Concept types...for auxiliary tables, etc.

- **skos:hasTopConcept** is for classes in the main schedule
- What about those in the auxiliary tables
  - Auxiliary table concepts should be handled as **skos:Concept**.
- Discussion of options:
  - Add local attributes  
e.g., “hasTop**Table**Concept”

```
and ...  
my:hasTopFamilyConcept  
my:hasTopClassConcept  
my:hasTopTableConcept  
... ..
```



## 4. Notations are constructed in various ways...

- The **skos:notation** element doesn't record how a classification notation is built
- CCT Format field 260:

### 260 分类号（必备，不可重复）

指示符 1: 类号的编辑制度

- 0 层累制
- 1 八分法
- 2 双位制
- 3 借上级类号
- 4 借下级类号
- 5 借同级类号
- 6 借 0 编号
- 7 使用“-”列举编号
- 8 顺序制编号(例如:A\B\C\D...)
- n 以上都不适用的

### 260 Notation (Mandatory, non-repeatable)

Indicator 1: Notation forming rule

- 0 hierarchical notation
- 1 expanding at number 9
- 2 double-digits
- 3 borrowing a super-ordinate class' notation
- 4 borrowing a sub-ordinate class' notation
- 5 borrowing a coordinate class' notation
- 6 borrowing “0” for a notation
- 7 using “-” for summary number span
- 8 sequential notations
- n other rules

### Discussion:

**skos:notation** may need to be extended,  
or local types to be added

2008-08-05

©zeng



## Other notation building issues

- **synthesized numbers (and terms)**
  - A number that has been made appropriately specific by adding or appending numbers from a table or from other parts of the schedule  
(*MARC Glossary*, <http://www.loc.gov/marc/classification/cdapndxd.html>)
- **add or divide like instructions**
  - Instructions to the classifier to construct a classification number by adding numbers from other parts of the schedule, from a table, or by basing it on a pattern defined in another part of the schedule. (*MARC Glossary*, <http://www.loc.gov/marc/classification/cdapndxd.html>)
- **parallel schedules**
  - E.g., 'law' class has two parallel schedules
- **faceted scheme features**
  - E.g., 'bridges' class has three facets, subgroups are listed under node labels.



## 2. SKOS for Mapped Vocabularies





- One important feature of SKOS is that it is possible for a same concept to be linked to several concept schemes, using the **skos:inScheme** property.



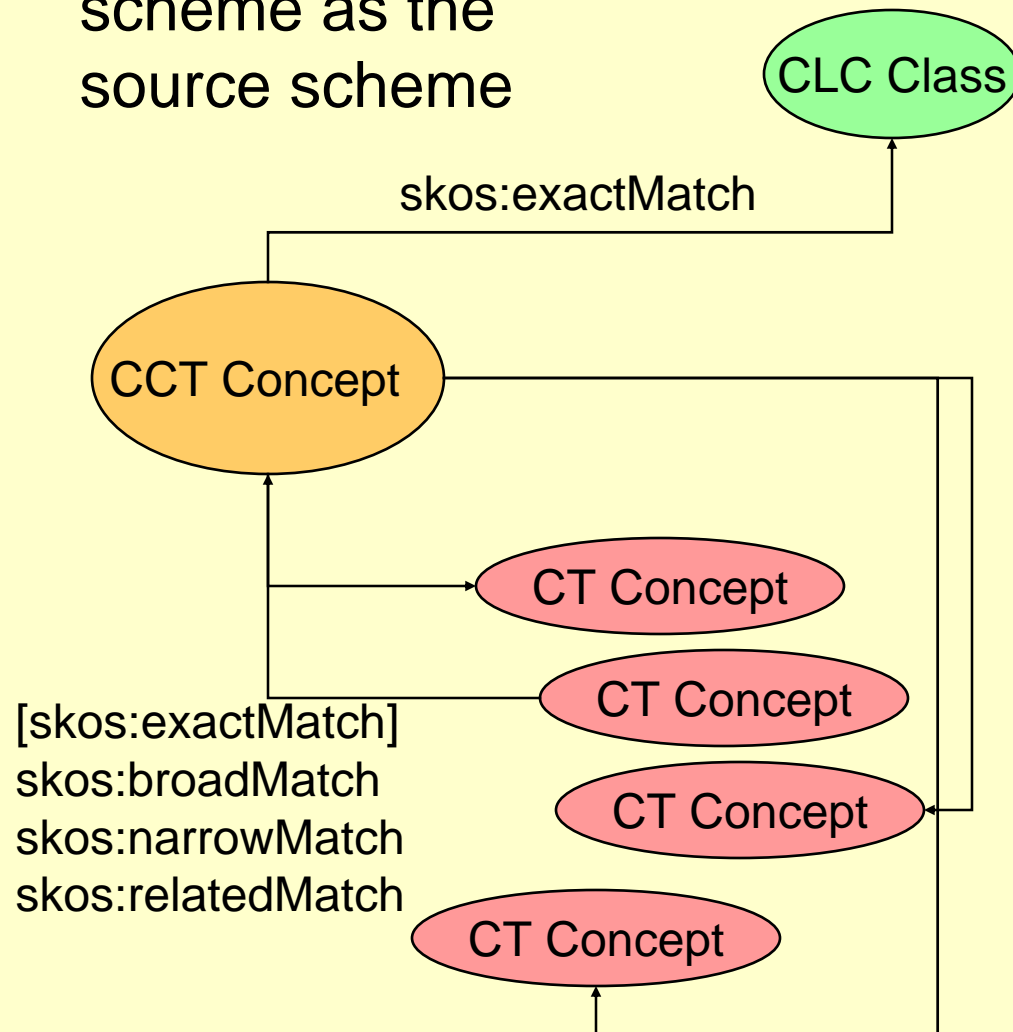


## Option 1. Treat CCT as a mapping result

- Advantages:
  - Semantic relationships are clear
  - Avoids semantic conflicts in applications
- Disadvantages:
  - Complicated
  - Time-consuming

## Two kinds of possible situations

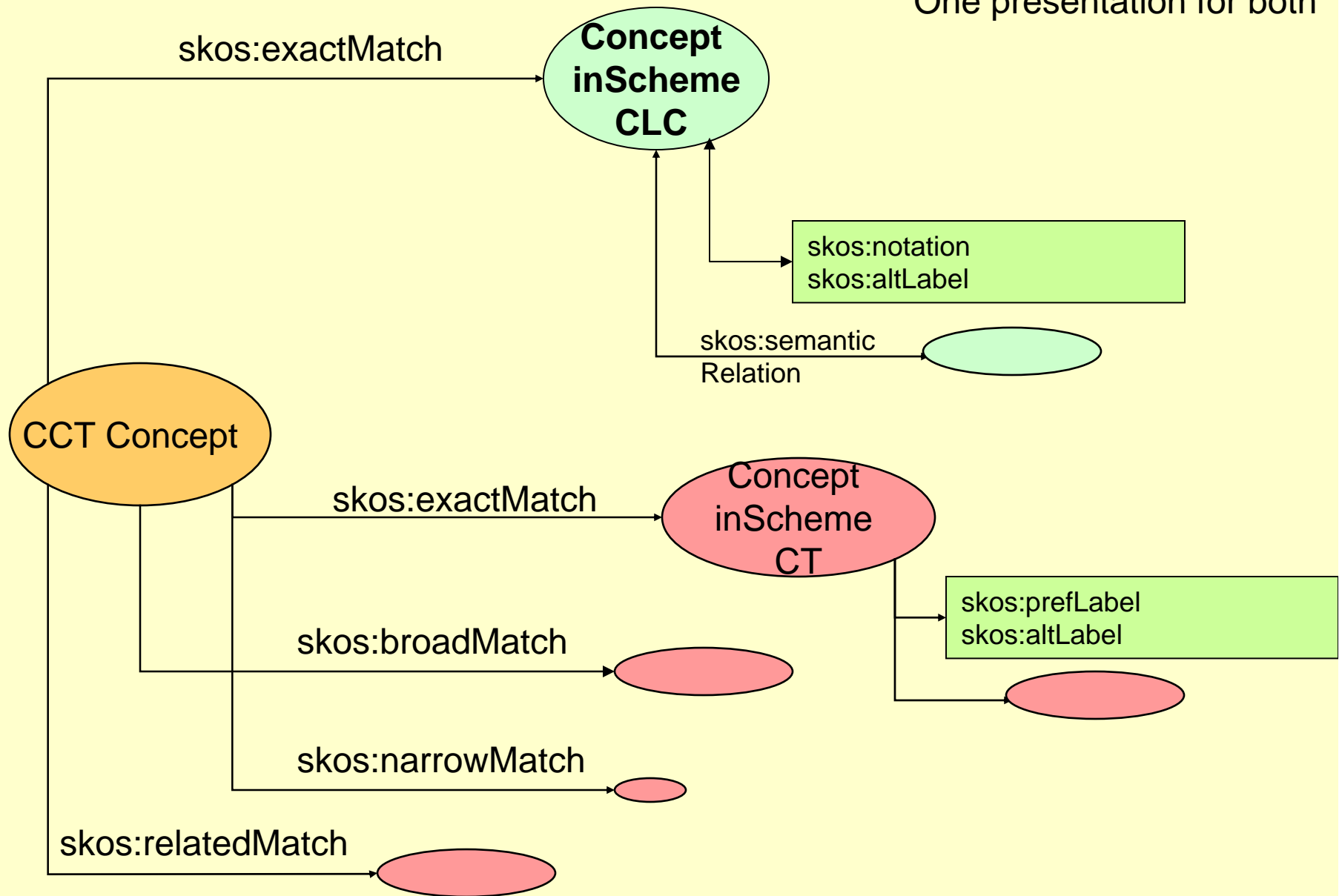
- Classification scheme as the source scheme



- Thesaurus as the source scheme

Same situation

One presentation for both





## Sub-issue 1. Many-to-one mapping

- *“No two concepts in the same concept scheme are given the same notation.”*
- *“No two concepts have the same preferred lexical label in a given language when they belong to a same concept scheme.”*

-- SKOS Primer & SKOS reference

## Notation-to-Terms

## Term-to-Notations

notation + caption + notes

term

term entry + notations

R511 病毒传染病  
病毒性肺炎入R563.1；病毒性皮肤病入R752；腹股沟肉芽肿入R759.4；沙眼入R777.32。

病毒病

R511.1 麻疹

麻疹

R511.2 风疹

风疹

R511.3 天花  
种痘入R186+.1。

天花

R511.5 水痘

水痘  
痒疹

R511.6 感冒（伤风）

感冒（伤风）

R511.7 流行性感冒

流行性感冒

R511.8 腺病毒感染

腺病毒\感染

R512.1 流行性腮腺炎

流行性腮腺炎

R512.3 脑炎及脑脊髓膜炎

流行性脑炎；脑炎  
脑脊髓炎

R512.31 甲型脑炎（尼克诺莫氏脑炎）

流行性脑炎\类型

R512.32 乙型脑炎（日本脑炎）

流行性乙型脑炎

R512.33 丙型脑炎（美国圣路易脑炎）

流行性脑炎\类型

R512.34 森林脑炎（苏联脑炎）

森林脑炎

R512.39 其他  
淋巴球性脉络丛脑膜炎、澳大利亚第十管睡性脑炎、无菌性脑脊髓膜炎入此。

流行性脑炎

R512.4 脊髓灰质炎（小儿麻痹症）

脊髓灰质炎

term / term  
coordination  
(\)

multiple  
terms (;)  
for one  
class

其他

吡啶

Y 氮杂环丁烷

吡啶

Acridine

0626.32+6

D 氮杂蒽

Z 杂环化合物

C 吡啶生物碱

C 阿的平

吡啶\化工生产

TQ253.26

吡啶\衍生物

0626.32+6

吡啶\衍生物\化工生产

TQ253.26

吡啶染料

Acridine dyes

TQ613.9

Z 染料

吡啶染料\化工生产

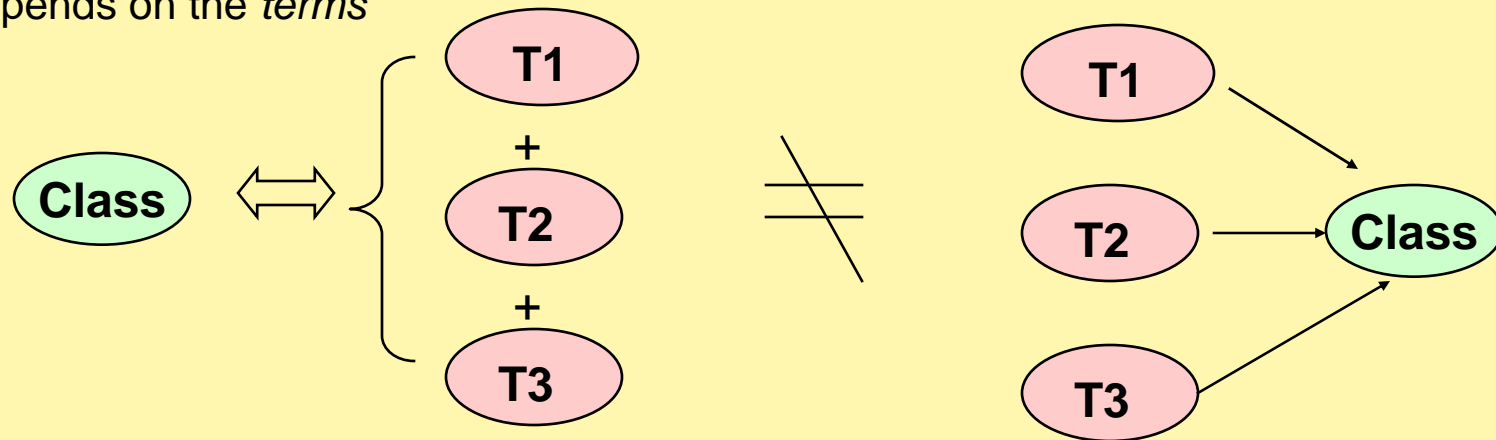
TQ613.9

吡啶生物碱





- Although the mapping is *concept*-based, the representation of concepts depends on the *terms*



The result could be: more than one term (each represents a different concept) point to the same notation

**Option discussed:** deal with a combined term as a **string** only, each representing a concept that did not exist in the original source schemes.

There might be problems for these new strings to be semantically linked with the concepts that each unit of the string originally represents.

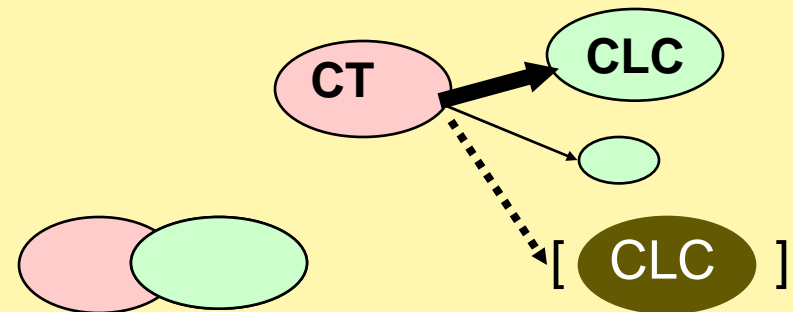


## Sub-issue 2: One-to-many mapping

### Sub-issue 3: Degree of mapping

CCT specifies degrees of mapping:

- Major mapping
- Minor mapping
- Alternative mapping
- Overlapping

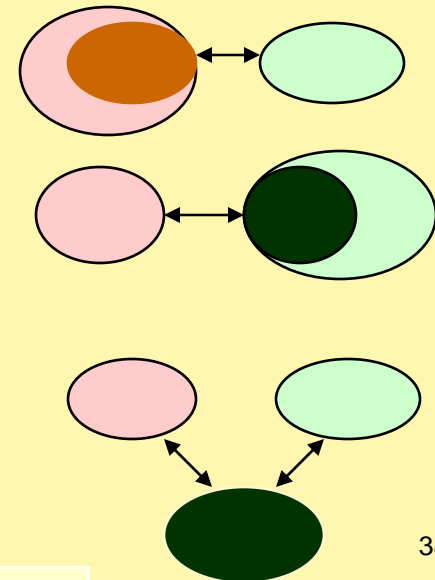


**How to indicate the degrees of mapping?**

SKOS' 'mapping relation' properties :

- broadMatch
- narrowMatch
- relatedMatch

These are not the same as CCT's mapping relations.



**Option discussed:** Define a local set of relation properties (?)

**52,992**  
**sub-classes**

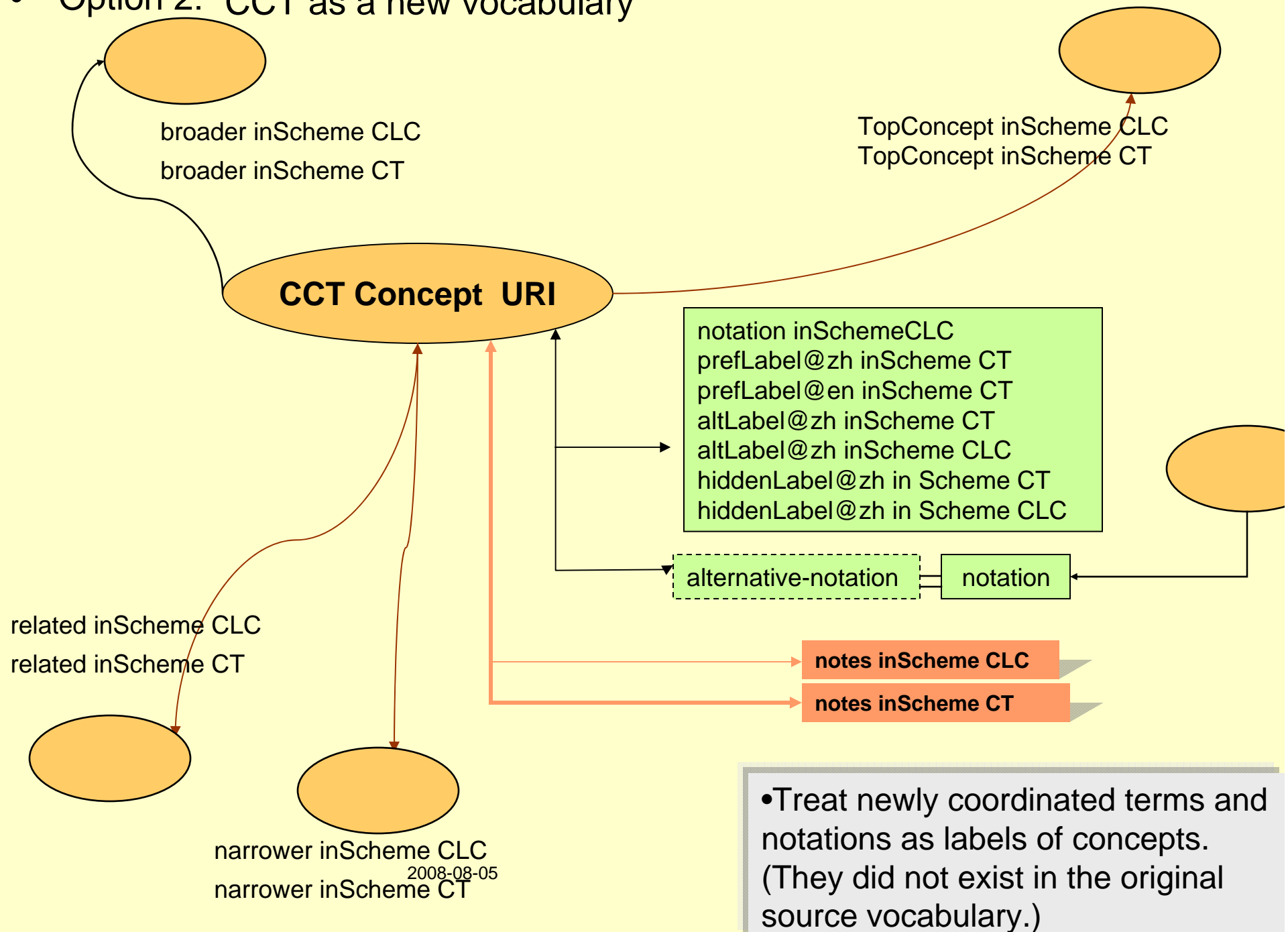
第21 唎易技法	唎易技法
包括健康心理的示作。	表现：碌易；[响田(无木)]；响易；陀陀；屈(唎易)；反肢唎易易；拽呀儿； 响快：纹字；马字；添清；再皮；再皮反
第22 一原技法	唎易技法
第23 中屈易技法	屈易技法
包括健康心理的示作。	[响田]；六心；[有易再屈心]；有易之木；马雷；马雷易；图从易；中屈易
第24 市易技法	市易技法
包括健康心理的示作。	求陀；市易
第25 系指、超马技法	系指技法；超马技法
包括健康心理的示作。	系指易；系指；系指般若；超马；马生易
第26 水形、水即易技法	水形易；唎易技法；水即易；唎易技法
包括健康心理的示作。	彭陀；彭受易；水形易；水即易；水即易
第27 形易、喊易易技法	形易；唎易技法；喊易易；唎易技法
包括健康心理的示作。	形易；喊易易；世即易易
第28 形易技法	形易技法；吸吸形易；形易技法；高散形易；形易技法；石形易；形易技法；响形易；形易技法
包括健康心理的示作。	形易；形易技法；[形易实]；吸吸形易；弹世陀；高散形易；石形易；水即水即；喊形易易 念世实；响形易；由取木实



- 版画技法 J217
  - Z 技法(美术)
- 粉画(绘画)技法 J216
- 国画技法 J212②; J212.1; J212.2
  - D 中国国画技法
  - D 中国画技法
  - Z 技法(美术)
- 绘画技法 J211; J211.1; J219
  - Z 技法(美术)
- 技法(美术) J08
  - 雕塑技法
  - 构图(美术)
  - 绘画技法
  - 版画技法
  - 国画技法
  - 笔法
  - 波法
  - 斜法
  - 墨法
- 十八衣纹描法 …十八衣纹描法
- 写意 …写意
- 年画技法 …年画技法
- 景稿技法 …景稿技法
- 速写技法 …速写技法
- 写生 …写生
- 油画技法 …油画技法
- 喷绘 …喷绘
- C 笔墨
- 蜡笔画(绘画)技法 J216
- 水彩画(绘画)技法 J215
- 水粉画(绘画)技法 J215
- 景稿技法 J214
- Z 技法(美术)
- 速写技法 J214
- Z 技法(美术)
- 油画技法 J213②; J213.1; J213.2
- Z 技法(美术)

**35,690 entry terms**

**... plus 59,738 pre-coordinated headings  
(created during the mapping process)**

- Option 2: CCT as a new vocabulary





## 4. Future Direction:

From machine-readable to  
machine-processable

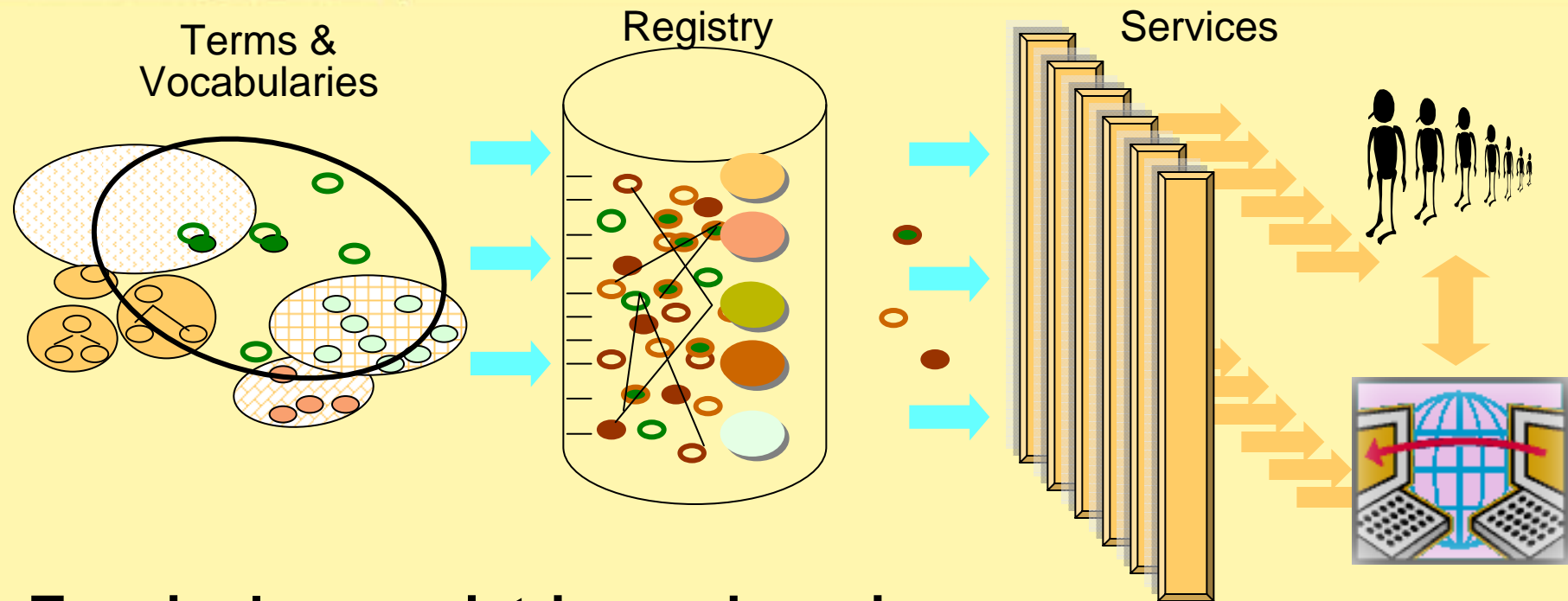


# Shareable Vocabularies

- SKOS enables KOS Vocabularies to become machine-processable
  - Share the vocabularies
  - Re-use and re-purpose vocabularies
    - Derive specialized schemes
      - e.g., Chinese culture related schemes
      - multilingual vocabularies
    - Integrate into new vocabularies
      - e.g., some structures to be used in ontologies
    - Use for learning, reasoning, data mining ...



# Effective Web Services



## Terminology registries and services

- registering machine accessible KOS
  - mapping among concepts/terms
  - making KOS content available in different kinds of tools
- via

terminology (web) services

2008-08-05

©zeng

43



Thank you!





# Acknowledgements

Special thanks for those who shared experiences and provided advise:

- Diane Vizine-Goetz, OCLC
- Andrew Houghton, OCLC
- Doug Tudhope, University of Glamorgan, U.K. and
  - STRA (Semantic Technologies for Archaeology Resources) project