

# Metadata for Terminology / KOS Resources

Marcia Lei Zeng

# Registry Types

- Metadata Schemas Registries
  - Elements and refinements, application profiles, schemas in different bindings ...
    - e.g., UKOLN [CORES Registry](#)
- Terminology Registries / Repositories
  - Registries for schemes (metadata) only
  - Registries of the entries of vocabularies (usually accompanied by scheme's metadata)
    - e.g., [OCLC Terminologies Service](#); [BioPortal ontology repository](#)
- Service Registries
  - Terminology services may be listed in a terminology registry or separately hosted in a service registry
- Data Standards Registries (integrated)
  - Registries/repositories of data standards (e.g., data dictionaries, data models, schemas, and code sets)

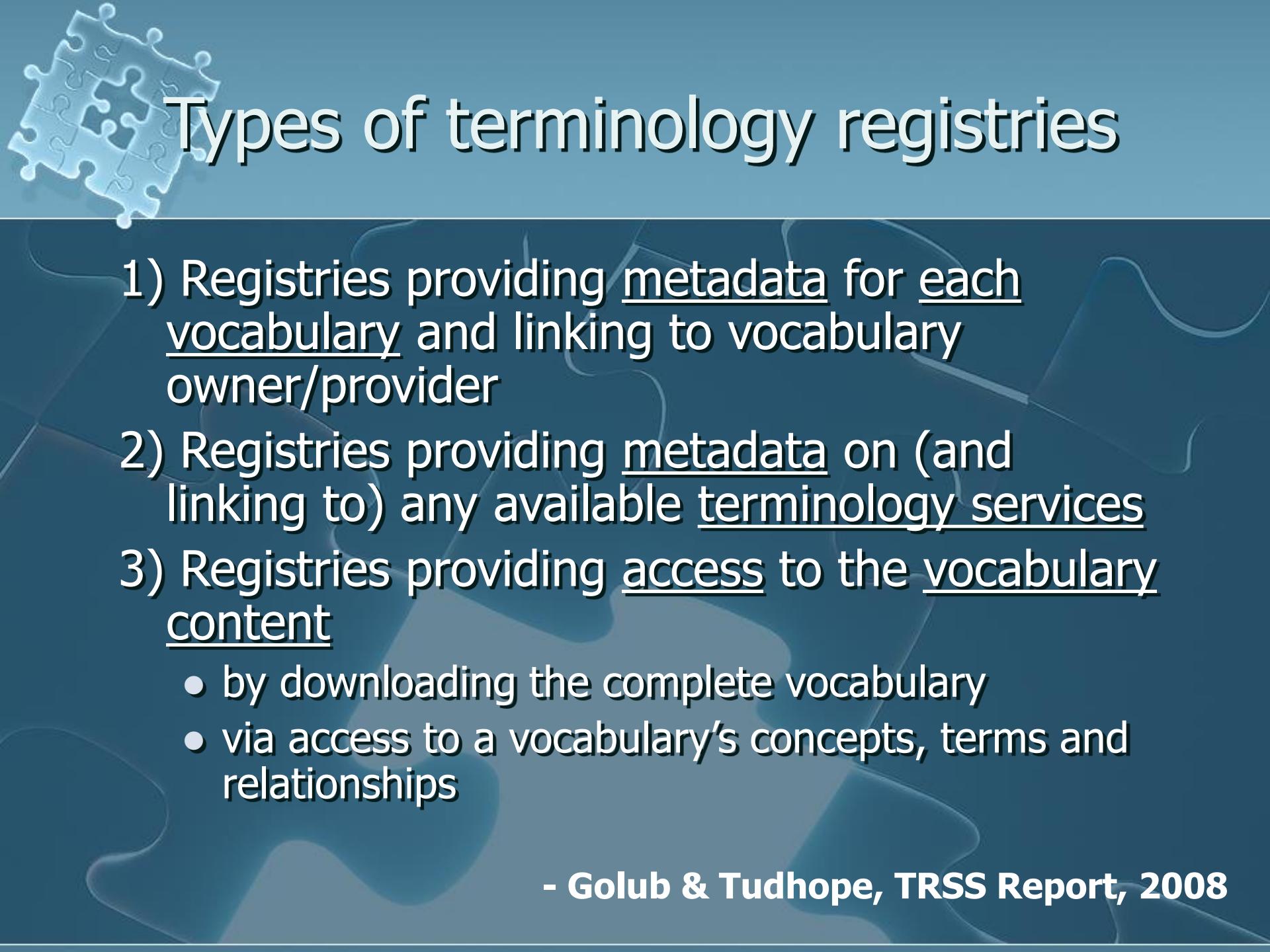
- 
- 1. Why do we need metadata for terminology resources?**
  - 2. What do we need to know about a terminology resource?**
  - 3. Is there a standardized set of metadata elements for terminology resources?**

# **1. Why do we need metadata for terminology resources?**



# Basically, metadata for terminology resources will ...

- record specific characteristics of terminology resources
- facilitate the discovery of terminology resources
- facilitate the evaluation of the terminology resource for a particular application or use
- facilitate sharing, reusing, and collaboration

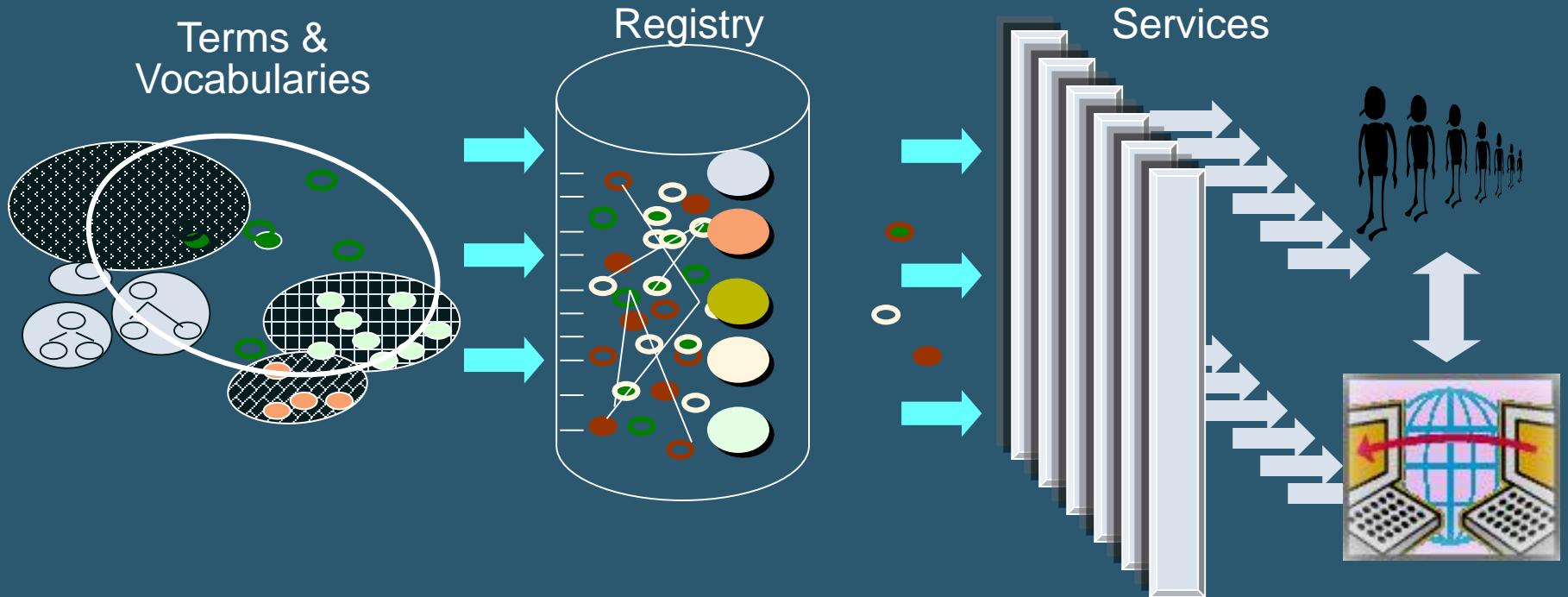


# Types of terminology registries

- 1) Registries providing metadata for each vocabulary and linking to vocabulary owner/provider
- 2) Registries providing metadata on (and linking to) any available terminology services
- 3) Registries providing access to the vocabulary content
  - by downloading the complete vocabulary
  - via access to a vocabulary's concepts, terms and relationships

- Golub & Tudhope, TRSS Report, 2008

# A simplified illustration of Terminology registries and services



- registering machine accessible terminology resources
- mapping among concepts/terms
- making KOS content available in different kinds of tools via terminology (web) services



# Terminology-based Services

- Related to the terminology registries are services, which may also be listed in a terminology registry or separately hosted in a *service registry*.
- These services, based on terminology, are used for automatic classification, term expansion, disambiguation, translation, and semantic reasoning.



# The need of metadata (1)

## -- terminology registries

**Terminology registries need to provide information about:**

- **sources used**
- **creation and revision dates**
- **provenance**
- **trustworthiness of sources**
- **quality assessment metrics for the vocabulary & source materials**
- **licensing, IP limitations**
- **flexibility for integration with other KOS**
- **specific requirements such as**
  - **performance**
  - **security**
  - **Maintainability**



# The need of metadata (2)

## -- service registries

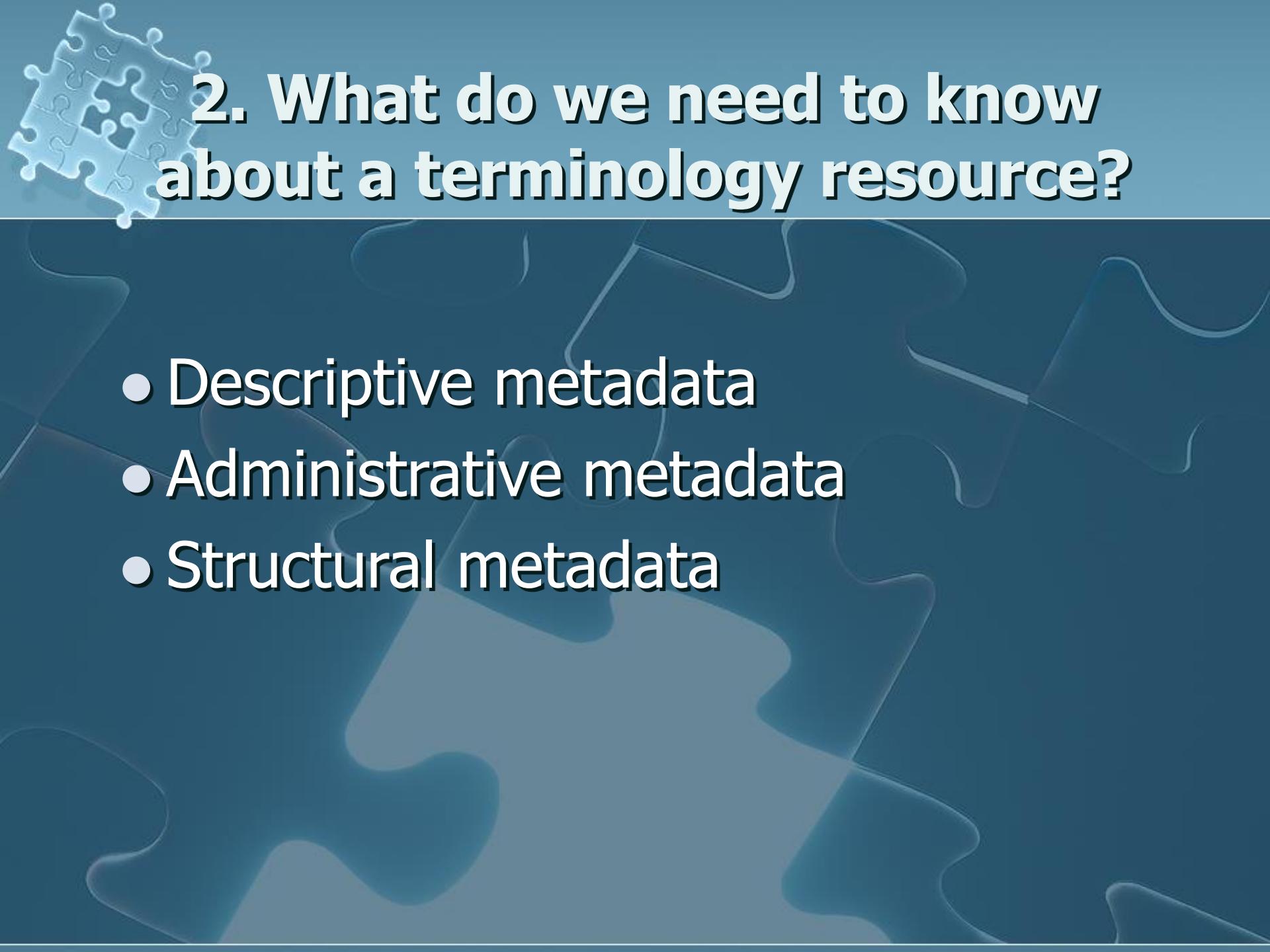
- Service registries need to understand and provide information on:
  - Data models
  - Tool interoperability
  - Protocol
  - Querying and accessibility
  - Affectivity – at what time, location, and/or use is the content applicable or valid
  - Available formats



## The need of metadata (3) -- vocabulary users

Different agents, services, and applications need to *communicate* about KOS data in the form of:

- transferring
- exchange
- transformation
- mediation
- migration
- integration



## **2. What do we need to know about a terminology resource?**

- Descriptive metadata
- Administrative metadata
- Structural metadata

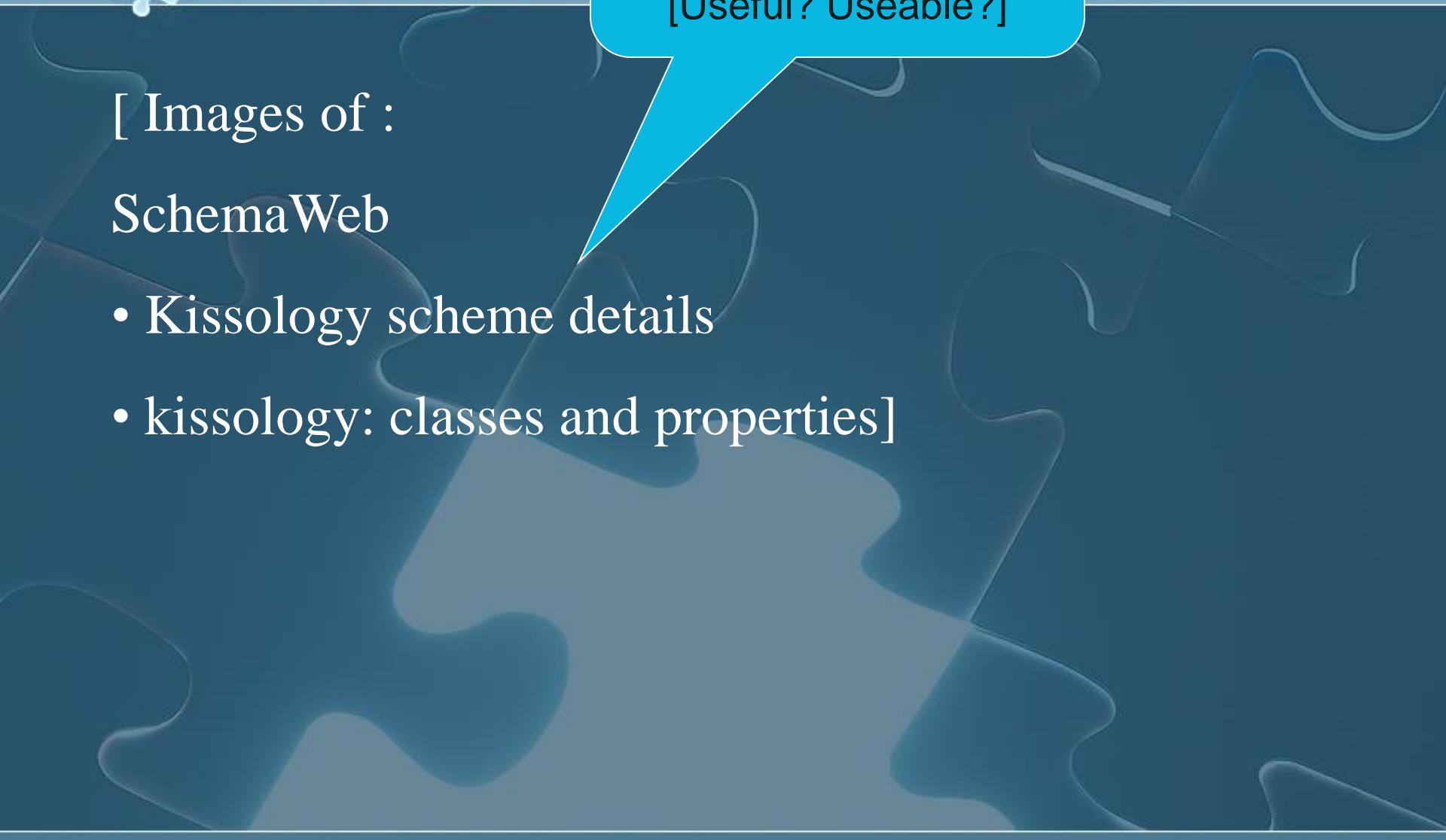


## [ Swoogle's Metadata Scheenshots ]

Creator?  
Domain or subject area?  
Copyright?  
.... ???

*Need more **descriptive**  
metadata*

See Glossary at:  
[http://swoogle.umbc.edu/index.php?option=com\\_swoogle\\_manual&manual=glossary](http://swoogle.umbc.edu/index.php?option=com_swoogle_manual&manual=glossary)

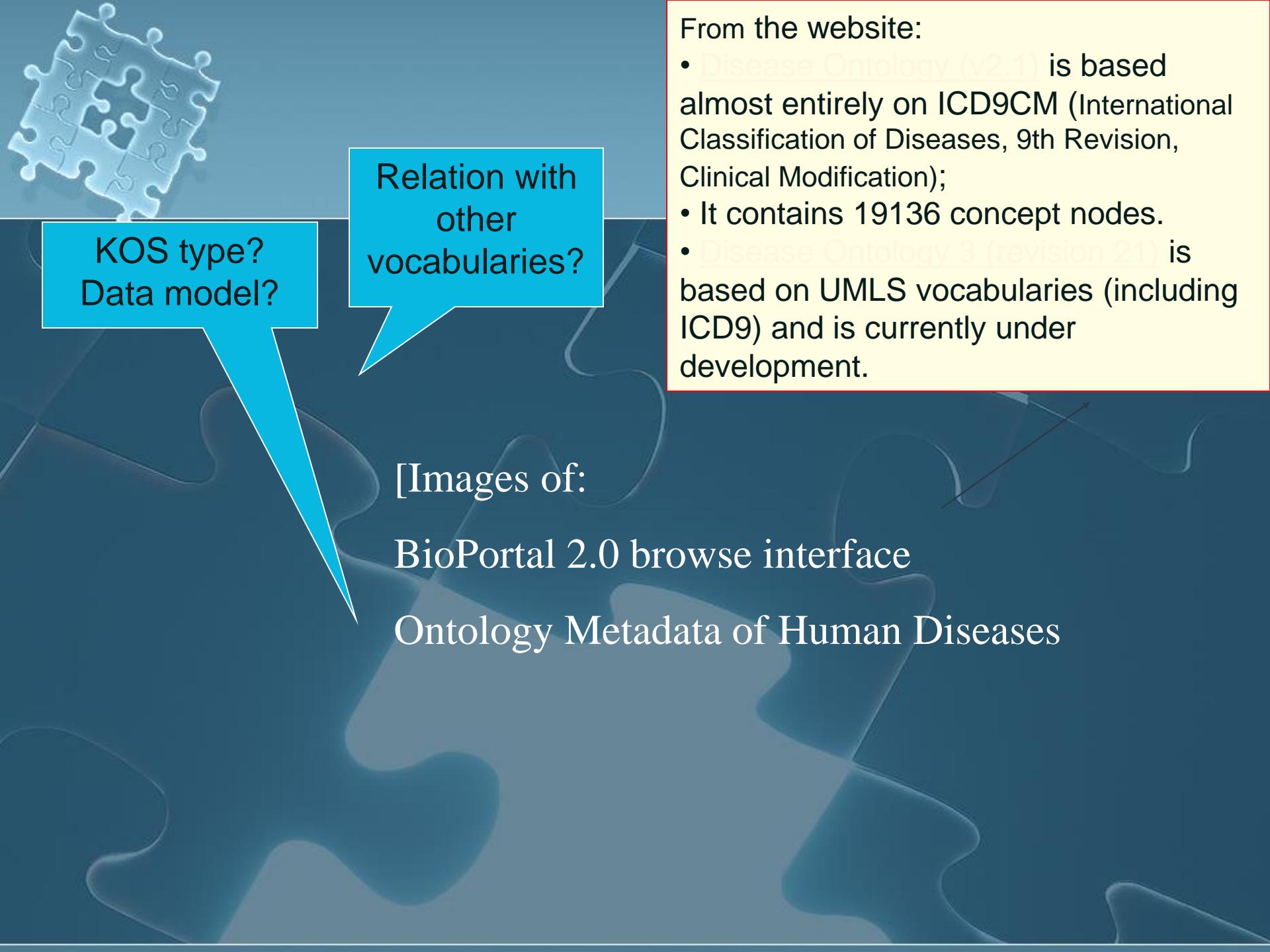


Domain / subject area?  
Number of triples?  
Number of classes?  
[Useful? Useable?]

[ Images of :

SchemaWeb

- Kissology scheme details
- kissology: classes and properties]



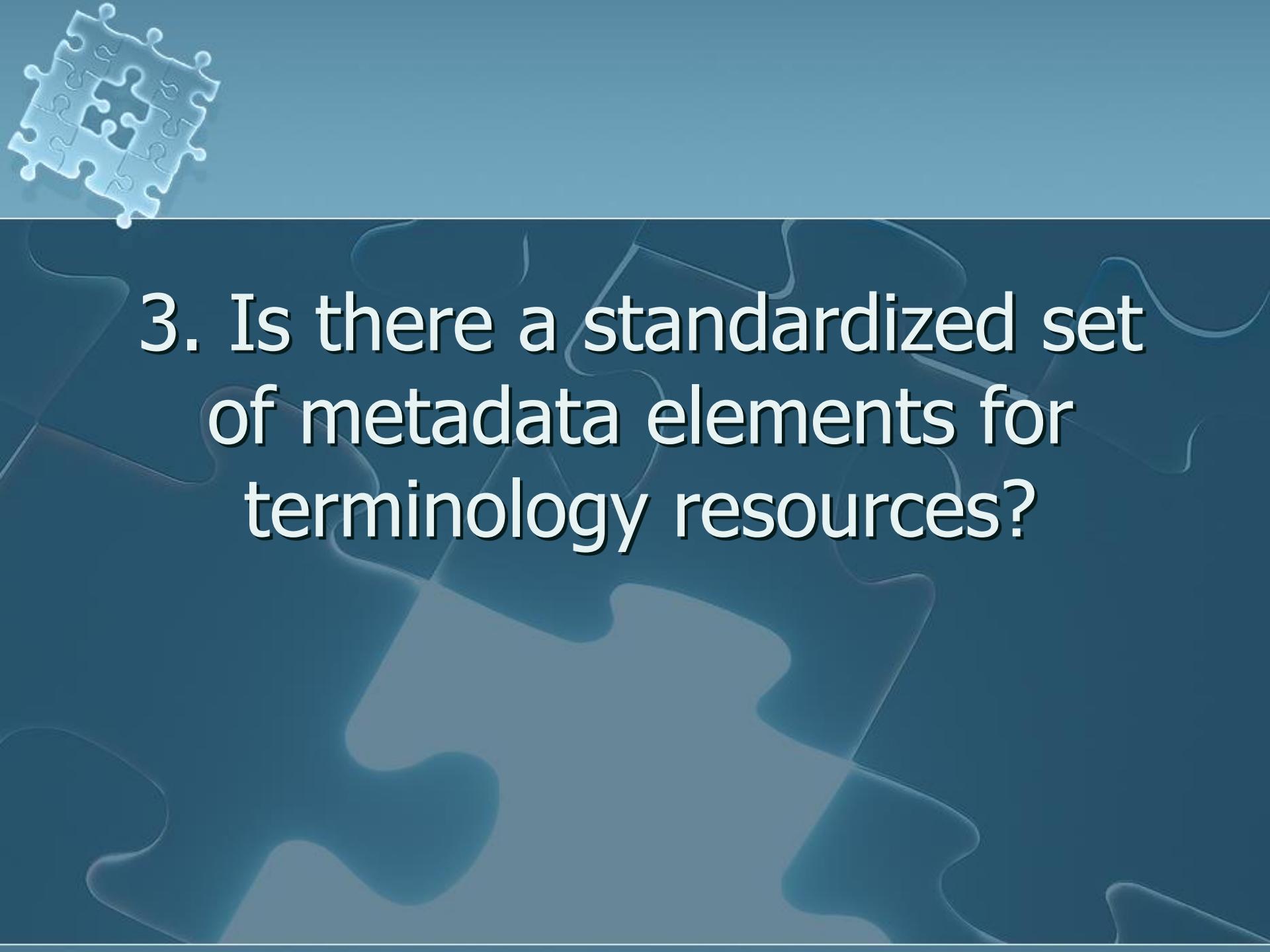
KOS type?  
Data model?

Relation with  
other  
vocabularies?

From the website:

- Disease Ontology (v2.1) is based almost entirely on ICD9CM (International Classification of Diseases, 9th Revision, Clinical Modification);
- It contains 19136 concept nodes.
- Disease Ontology 3 (revision 21) is based on UMLS vocabularies (including ICD9) and is currently under development.

[Images of:  
BioPortal 2.0 browse interface  
Ontology Metadata of Human Diseases



3. Is there a standardized set  
of metadata elements for  
terminology resources?



# International Standard -- ISO 11179

ISO 11179-2 Information Technology --  
**Metadata registries (MDR)**  
- Part 2 Classification\*

(\*CLASSIFICATION schemes include: key words, thesauri, taxonomies, and ontologies.)

<http://metadata-standards.org/11179/>

# Regions of the MDR metamodel containing classification scheme attributes

ISO 11179

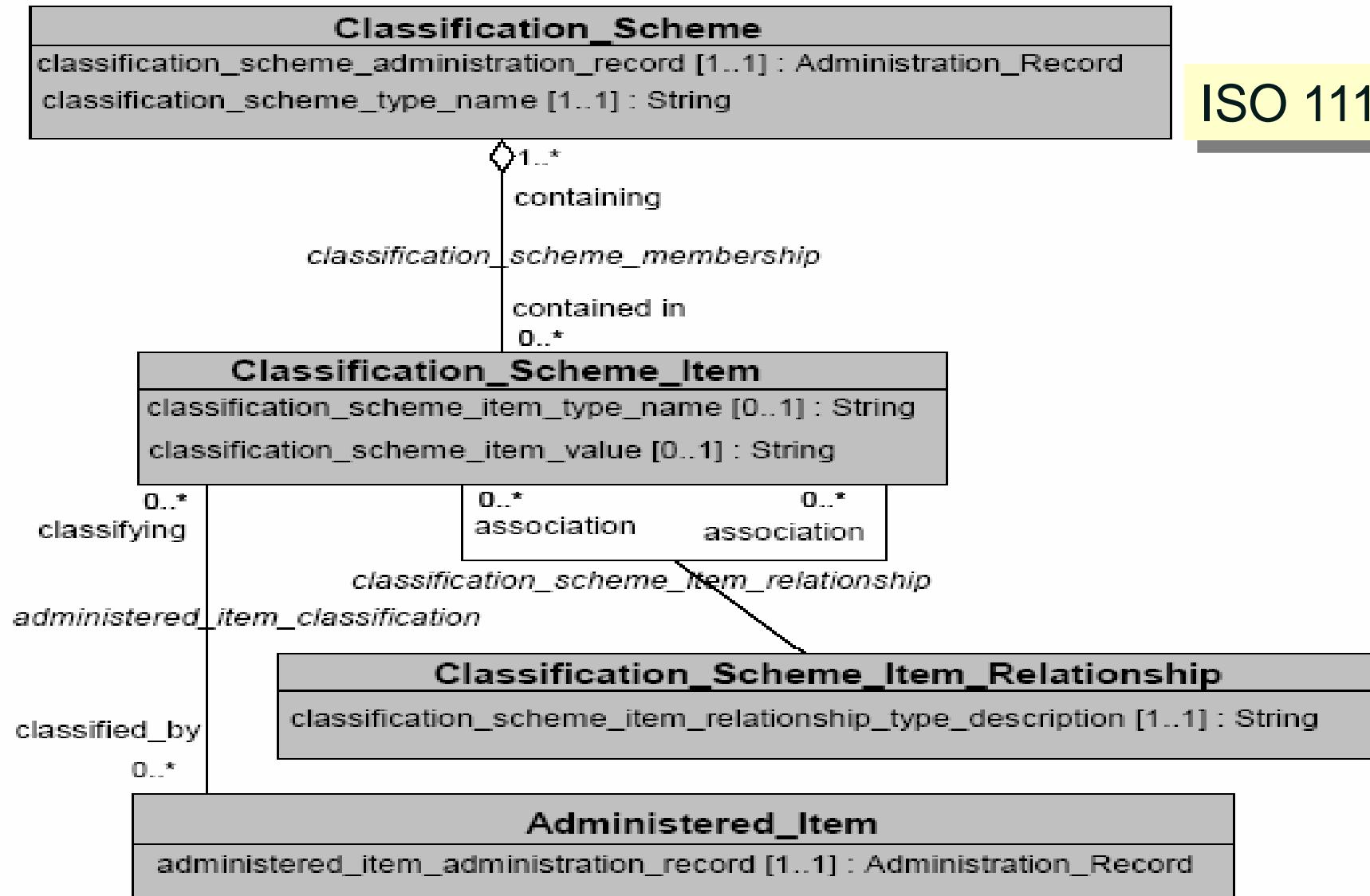


Figure A.1 — Classification metamodel region.



# Attributes of a classification system that may be recorded in an MDR (slide 1)

ISO 11179

## Designation

- *name*
- *preferred designation*
- ***language identifier***

## Definition

- *definition text*
- *preferred definition*
- ***source reference***
- ***language identifier***

## Context

- ***administration record***
- *description*
- ***description language identifier***

## Classification Scheme

- *type name*

## Classification Scheme Item

- *value*
- *type name*

## Classification Scheme Item Relationship

- *type description*

(**boldface**: datatypes containing multiple attribute components)



# Attributes of a classification system that may be recorded in an MDR (slide 2)

ISO 11179

## Administration Record

- *item identifier*
- *registration status*
- *administrative status*
- *creation date*
- *last change date*
- *effective date*
- *until date*
- *change description*
- *administrative note*
- *explanatory comment*
- *unresolved issue*
- *origin*

## Reference Document

- *identifier*
- *type description*
- ***language identifier***
- *title*
- *organization name*
- *organization mail address*



# Attributes of a classification system that may be recorded in an MDR (slide 3)

ISO 11179

## Submission

- *organization name*
- *organization mail address*
- ***contact***

## Stewardship

- *organization name*
- *organization mail address*
- ***contact***

## Registration Authority

- *organization name*
- *organization mail address*
- ***registration authority identifier***
- ***documentation language identifier***

## Registrar

- ***identifier***
- ***contact***



# 11179 Data Element Registries

- [US National Cancer Institute - Cancer Data Standards Repository \(caDSR\)](#)
- [Australian Institute of Health and Welfare - Metadata Online Registry \(METeOR\)](#)
- [US Department of Justice - Global Justice XML Data Model GJXDM](#)
- [US Environmental Protection Agency - Environmental Data Registry](#)
- [US Health Information Knowledgebase \(USHIK\)](#)
- [US National Information Exchange Model NIEM](#)
- [Minnesota Department of Education Metadata Registry \(K-12 Data\)](#)
- [Minnesota Department of Revenue Property Taxation \(Real Estate Transactions\)](#)

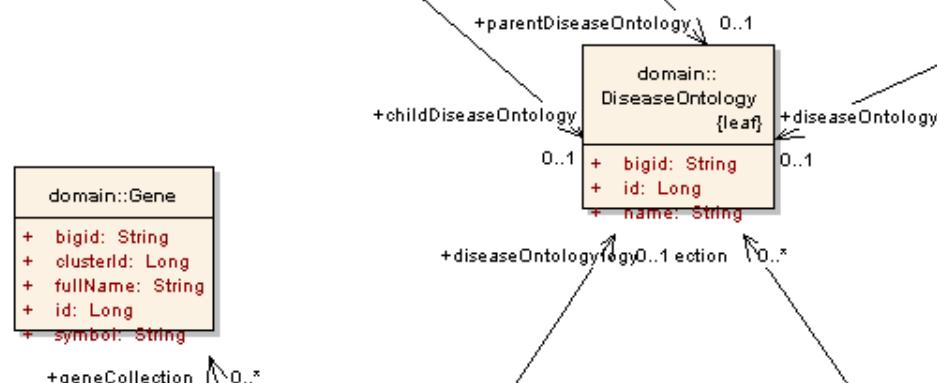
## -- CDE (Common Data Element) Browser

- Important additional items (in addition to “Classification” of ISO 11179)
  - **Form** -- a collection of CDEs (Common Data Elements)
  - **Protocol** -- a collection of Forms.
- For clinical trials applications,
  - Forms correspond to Case Report Forms (CRFs)
  - Protocols correspond to a clinical trial protocol



## Logical View

- + Data Model
- Logical Model
- DIAGRAMS
  - CABIO
  - CADSR
  - COMMON
- gov
- java

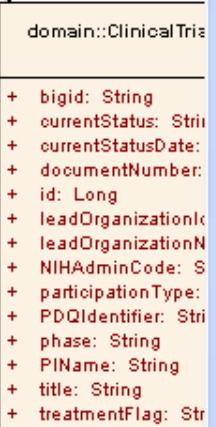
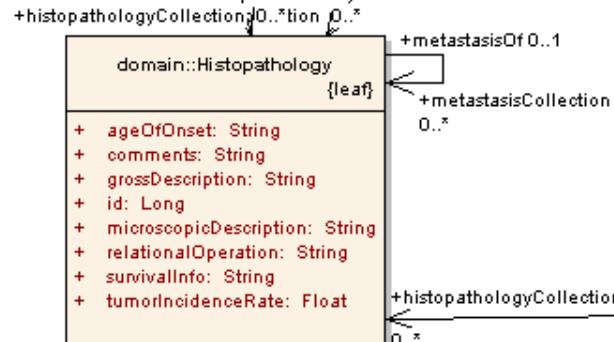


Cancer Data Standards Repository (caDSR)

## -- UML (Unified Modeling Language) Model Browser

### CABIO

- Agent-Anomaly
- CDNA-Object
- CDNA1
- CDNAData
- ClinicalTrialProtocol
- CloneRelativeLocation
- DatabaseCrossReference
- DiseaseOntology
- ExpressionFeature
- Gene
- Gene-ExpressionMeasurement
- HomologousAssociation
- GeneOntology
- GeneRelativeLocation
- GenericReporter
- Histopathology-Disease
- HomologousAssociation



### class diagram: DiseaseOntology

Author:	The Administrator
Project:	Version: 1.0; Locked: false
Dates:	Created: 2/3/2005 1:02:31 PM; Modified: 6/14/2005 12:21:06 PM;
UUID:	{47C124E7-3B2A-4cd3-A215-21CDCE0AAFF}



# NKOS Group's Efforts (1)

## ***NKOS Registry - Draft Set of Thesaurus Attributes, 1999***

(based on Controlled Vocabulary Registry developed by Linda L. Hill and Interconnect Technologies in 1996, with some modification)

[http://nkos.slis.kent.edu/Thesaurus\\_Registry.html](http://nkos.slis.kent.edu/Thesaurus_Registry.html)

## ***Terminology Registry Scoping Study (TRSS), 2008***

(PIs: Kora Golub, Doug Tudhope, Trss Final Report to JISC, UK.)

<http://www.ukoln.ac.uk/projects/trss/>

# TRSS survey report 2008 (draft)

K. Golub, D. Tudhope, Aug, 7, 2008

Product Information	A	B	C	D	E	F	G	H	I	J	K	L
Product Name/Title	+	+	+	+	+	+	+	+	+	+	+	+
Variant Product Name/Title /Acronym	+	+	+	+		+						
Type of Product	+	+	+		+	+		+		+	+	
Product Description	+	+	+*			+	+		+	+	+	+
Auxiliary Lists	+											
Author/Editor	+	+			+	+			+	+		
Current Version/Edition	+						+					+
Date of Current Version	+	+	+							+	+	
Product Update Frequency	+		+									+
Available Format(s) and Size	+	+	+			+						+
Online Availability	+		+		+	+	+**	+	+	+	+	
Notes	+											
URL for Examples	+											

A – NKOS Registry 1998

B – NKOS Registry 2001

C – CENDI

D – Ecoterm (Environmental Terminology and KOS)

E – Food and Agriculture Organization (FAO) of UN

F – Hodge et al. 2007 (10<sup>th</sup> OFMR)

G – National Science Digital Library Registry

H – ISO 11179 (Information Technology – Metadata registries (MDR))

I – OCLC Terminology Services

J – SPECTRUM Terminology Bank

K – Taxonomy Warehouse

L – Vocman (Becta Vocabulary Bank)



# NKOS Registry – Metadata Element Set (slide 1)

## Draft Set of Thesaurus Attributes, 1999

### I. Product Information

- Product Name/Title \*
- Variant Product Name/Title
- Type of Product \*
- Product Description \*
- Auxiliary Lists
- Author/Editor
- Current Version/Edition \*
- Date of Current Version \*
- Product Update Frequency \*
- Available Format(s) and Size \*
- Online Availability
- Notes
- URL for Examples

## TRSS Study, 2008

### Added:

- *Vocabulary type*
- Available terminology services
- Vocabulary identifier



# NKOS Registry – Metadata Element Set (slide 2)

## Draft Set of Thesaurus Attributes, 1999

### II. Scope and Usage

- *Type of NKOS*
- Major Subjects
- Minor Subjects
- Description of User Community and Applications

### Added by TRSS Study, 2008

- Purpose as given by author/publisher
- Used by
- Description of collections where used
- Usage case study
- Use in application profiles
- Rating
- URL to vocabulary users' discussion board
- Change notification details
- Related vocabularies
- Overlap with related vocabularies
- Mappings to other vocabularies
- URL to tutorial for applying vocabulary



# NKOS Registry – Metadata Element Set (slide 3)

## III Detailed Characteristics

Language(s) \*

Type of Terms (e.g. concept terms, geographic names, corporate names, etc.)

\*

Description of Overall Structure \*

Source of New Terminology \*

Number of Preferred Terms or Nodes \*

Number of Non-preferred Terms

Types of Relationships \*

Arrangement of Displays (e.g., alphabetical, hierarchical, graphical)

Depth of Hierarchy (maximum number of levels)

### Added in TRSS, 2008

- Total number of terms\*\*
- Total number of classes\*\*
- \*\*update automatically



# NKOS Registry – Metadata Element Set (slide 5)

## VII. Terms and Conditions

- Subscription Price by Format  
Licensing Availability  
Restrictions (or no-restrictions statement) \*

**Added in TRSS,  
2008**

- Import/download instructions



# NKOS Registry – Metadata Element Set (slide 6)

## VIII. Vendor/Provider Information

... ... [14 elements]

## IX. Contact Information

... ... [5 elements]

## X. Additional Information

- General Note
- Comments to Registry Maintainer

## Simplified in TRSS, 2008

- 6 Vocabulary provider
  - Vocabulary provider name
  - Vocabulary provider URL
  - Vocabulary provider contact details



# NKOS Registry – Metadata Element Set (slide 7)

**NEW -- Added in TRSS, 2008:**

## ***4 Terminology services***

- Available terminology services and their APIs
- Type of terminology service
- If a mappings service, the granularity of the mappings
- If a mappings service, whether mappings derived automatically or manually
- Technical specifications (ways of access etc.)



# NKOS Group's Efforts (2)

***Registry, Version 3 with Reference Document for Data Elements - Draft***

## **For use with Dublin Core**

- core elements only
- consistent with Dublin Core elements and attributes for each element

Draft developed by Diane Vizine-Goetz  
Last updated: June 21, 2001

<http://nkos.slis.kent.edu/registry3.htm>

for facilitating the discovery of KOS resources, (DC-based) :

- KOS Title (R)
- Alternative Title (O)
- Creator (O)
- KOS Subject (R)
- Description (O)
- Publisher (O)
- Date (R)
- KOS Type (R)
- Format (R)
- Identifier (O)
- Language (R)
- KOS Relation (R)
- Rights (O)

for recording specific characteristics, to facilitate the evaluation of the resource for a particular application or use:

- Entity Type (R)
- Entity Value (O)
- Relationships (R)
- Information Given (O)
- Arrangement (R)
- Application (O)
- Minor Subject (O)  
[Should this be a qualifier of KOS Subject?]



# NKOS Group's Efforts (3) KOS Types

**Taxonomy of Knowledge Organization Sources/Systems - Microsoft Internet Expl...** 

File Edit View Favorites Tools Help

Back

Google

Address [http://nkos.slis.kent.edu/KOS\\_taxonomy.htm](http://nkos.slis.kent.edu/KOS_taxonomy.htm) Go Links

## **Taxonomy of Knowledge Organization Sources/Systems (1)**

**Draft June 7, 2000 (revised July 31, 2000)**

**Gail Hodge, 2000**

The descriptions given here are simply to provide an overview of possible sources for the organization of digital libraries.? The descriptions are based on characteristics such as structure and complexity, the relationships between terms, and historical function.? The list is not intended to be comprehensive nor are the definitions specifically based on standards.? The specific types are grouped into general categories -- term lists, which emphasize lists of terms often with definitions; classifications and categories, which emphasize the creation of subject sets; and relationship lists, which emphasize the connections between terms and concepts.? This is extremely draft and was really produced for a different purpose, so there has not been sufficient thought given to describing the differences by how they would need to behave in a networked environment.

## **Term Lists**

### **Authority Files**

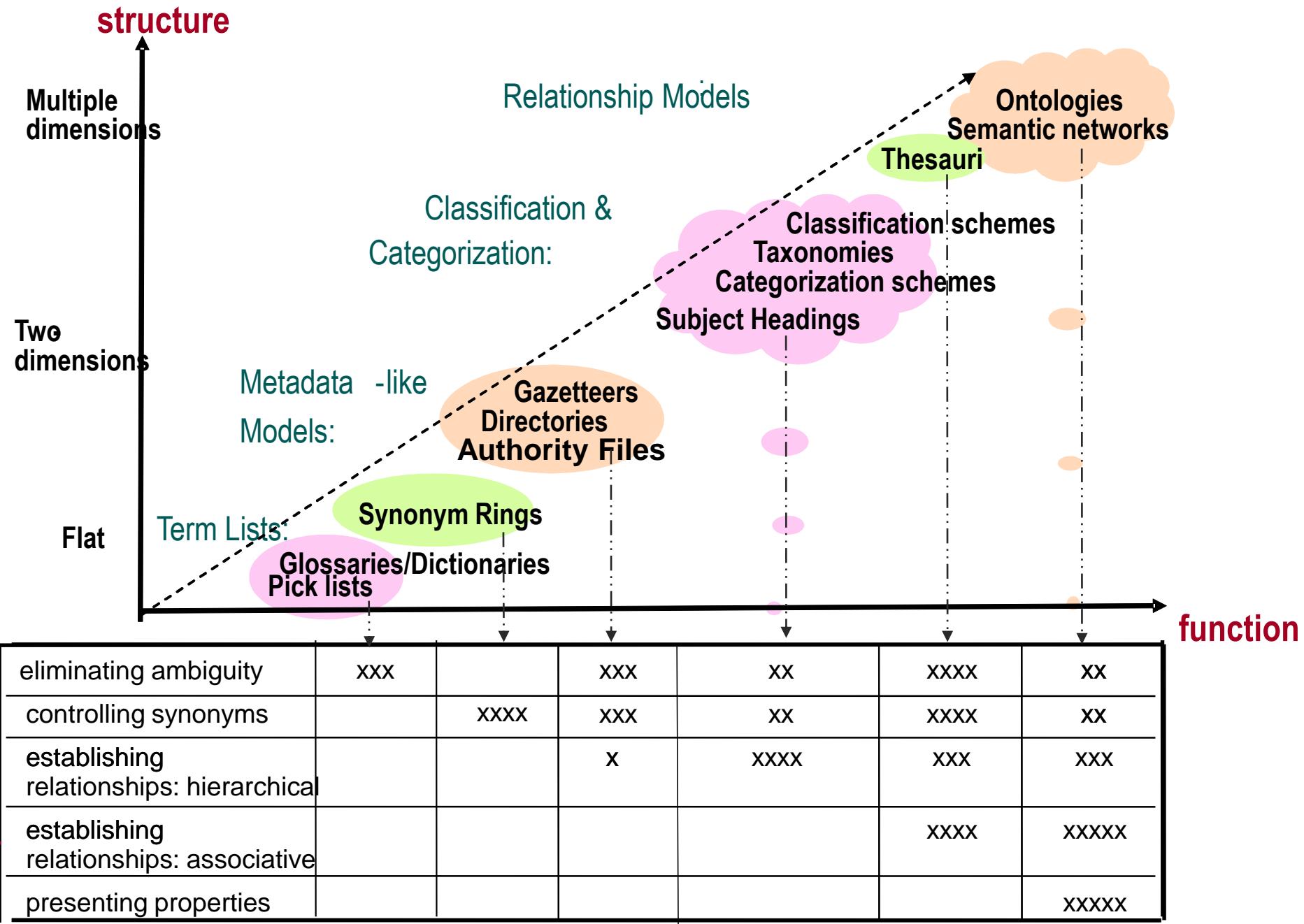
Authority Files are lists of terms that are used to control the variant names for an entity or the domain value for a particular field. Examples include names for countries, individuals, and organizations.? Non-preferred terms may be linked to the preferred versions. This type of KOS generally does not include a deep organization or complex structure.? The presentation may be alphabetical or organized by a shallow classification scheme.? There may be some limited hierarchy applied in order to allow for simple navigation, particularly when the authority file is being accessed manually or is extremely large.? Specific examples of authority files include the Library of Congress Name Authority File and the Getty Geographic Authority File.?

[Classification](#)



4:09 PM

# A Taxonomy of KOS



# Factors governing types of KOS -- **Template**

## **Entities**

Concepts, terms, strings,

Atomic - Composite (attributes)

Enumerative - Synthetic

Low – medium - high degree precombination (coordination in KOS itself)

Size: small – large

Depth: small – medium - large

## **Relationships (internal)**

Types / expressivity of relationships:

low (core set) – medium – high (definable)

concept-concept, concept-term, term-term

monohierarchies - polyhierarchies

Formality: low – medium – high

## **Typical application to objects in domain of interest**

Metadata element: subject, various elements, general

Granularity of application objects: unstructured - complex

## **Relationship applying concepts to objects in domain**

about (fuzzy), instance

Exhaustivity: low - high

Specificity: low - high

Coordination: low - high

expressivity and formality of relationships in coordination (synthesis rules)

**Tudhope,05,NKOS, ECDL2005**

# Factors governing types of KOS -- Thesaurus

## **Entities**

Concepts, terms, strings,  
Atomic - Composite (attributes)  
Enumerative - Synthetic  
Low – medium - high degree precombination (coordination in KOS itself)  
Size: small – large  
Depth: small – medium - large

## **Relationships (internal)**

Types / expressivity of relationships:  
low (core set) – medium – high (definable)  
concept-concept, concept-term, term-term  
monohierarchies - polyhierarchies  
Formality: low – medium – high

## **Typical application to objects in domain of interest**

Metadata element: subject, various elements, general  
Granularity of application objects: unstructured - complex

## **Relationship applying concepts to objects in domain**

about (fuzzy), instance  
Exhaustivity: low - high  
Specificity: low - high  
Coordination: low - high  
expressivity and formality of relationships in coordination (synthesis rules)

# Factors governing types of KOS – [AI] Ontology

## **Entities**

Concepts, terms, strings,

Atomic - Composite (attributes)

Enumerative - Synthetic

Low – medium - high degree precombination (coordination in KOS itself)

Size: small – large

Depth: small – medium - large

## **Relationships (internal)**

Types / expressivity of relationships:

low (core set) – medium – high (definable)

concept-concept, concept-term, term-term

monohierarchies - polyhierarchies

Formality: low – medium – high

## **Typical application to objects in domain of interest**

Metadata element: subject, various elements, general

Granularity of application objects: unstructured - complex

## **Relationship applying concepts to objects in domain**

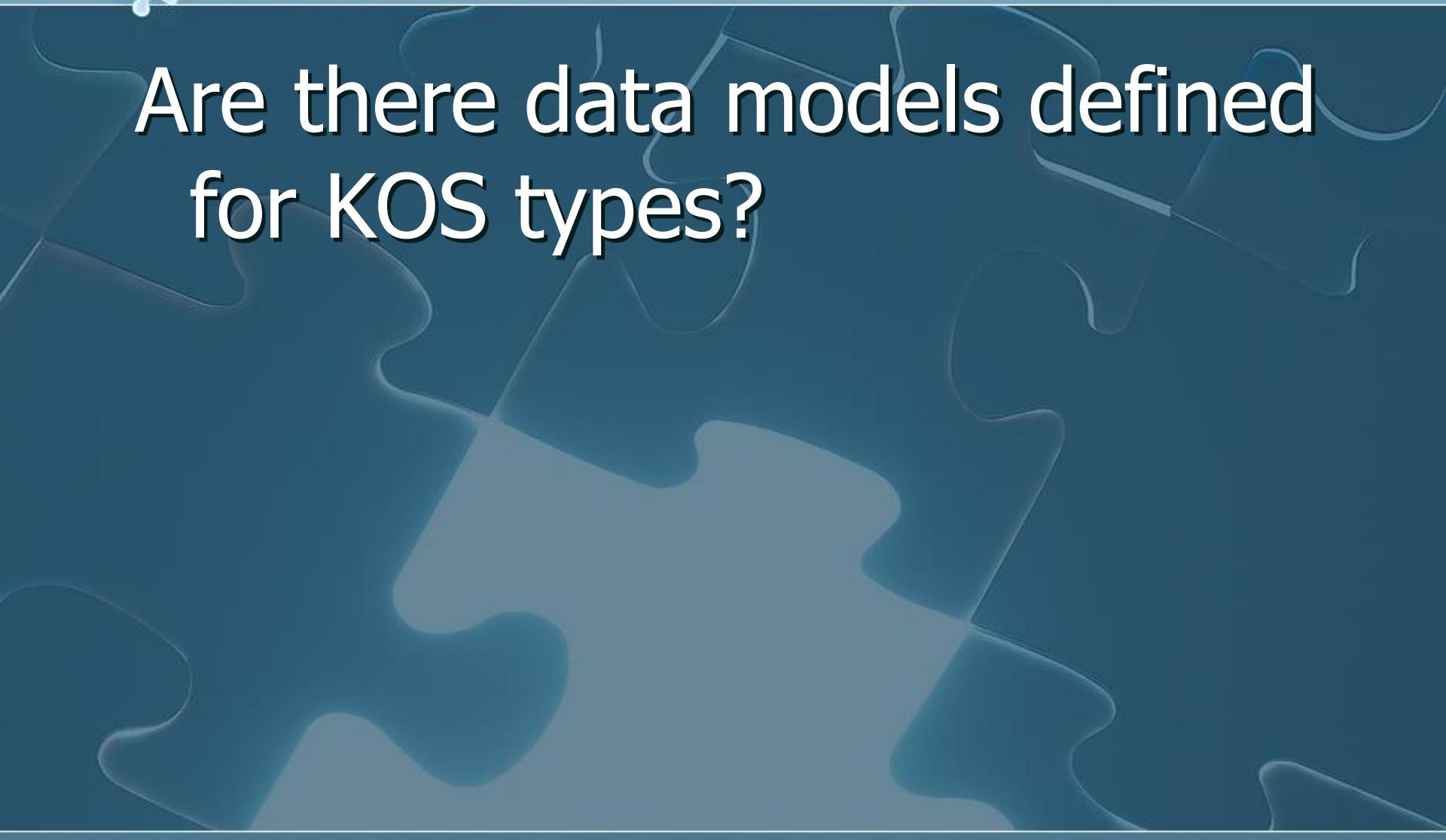
about (fuzzy), instance

Exhaustivity: low - high

Specificity: low - high

Coordination: low - high

expressivity and formality of relationships in coordination (synthesis rules)

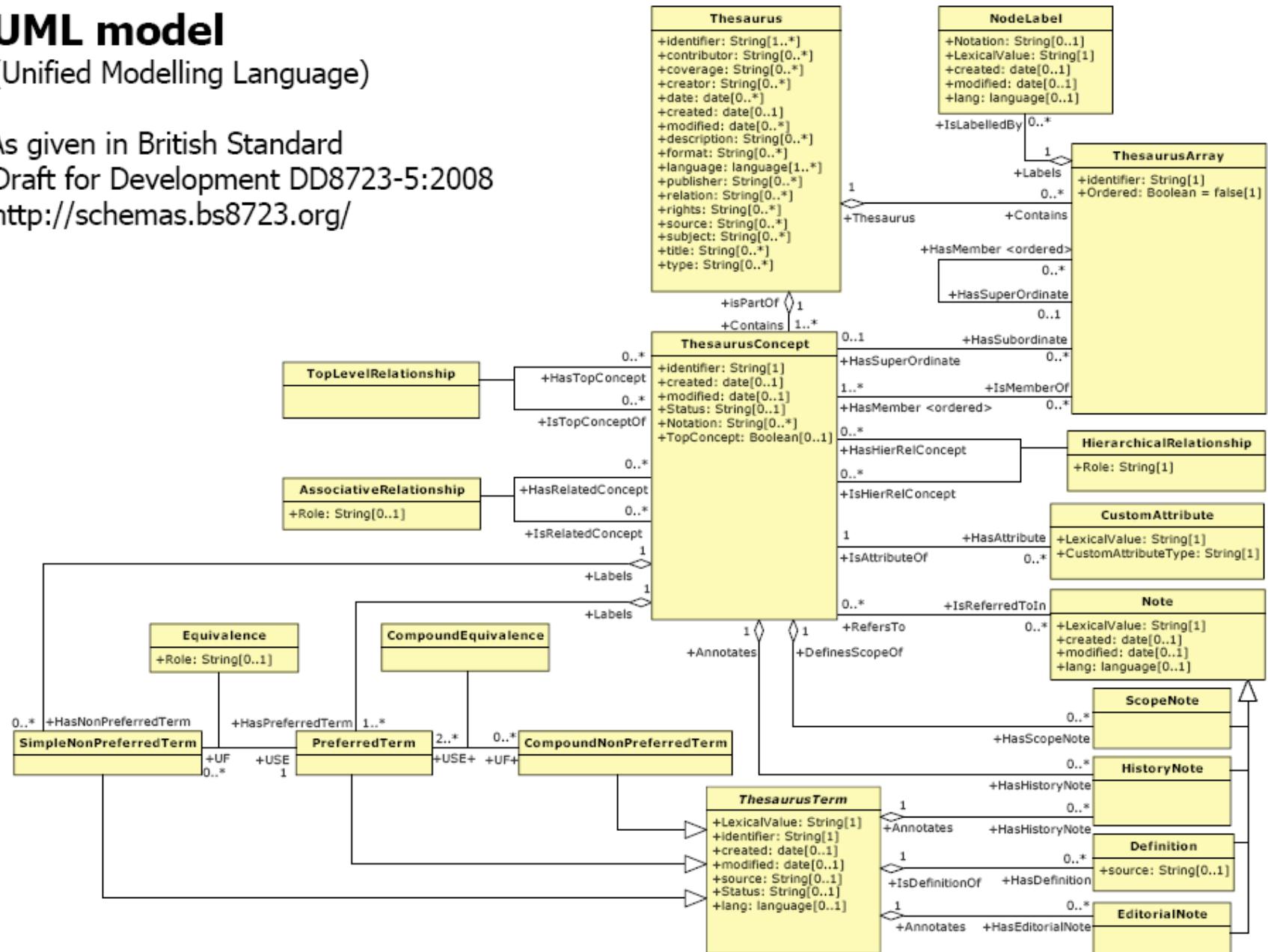


# Are there data models defined for KOS types?

# UML model

(Unified Modelling Language)

As given in British Standard  
 Draft for Development DD8723-5:2008  
<http://schemas.bs8723.org/>





# Summary

- Metadata for KOS resources are important to
  - Terminology registries
  - Service registries
  - Vocabulary users
- Currently there are no standardized metadata element sets
- A KOS typology needs to be implemented
- KOS data models need to be developed and tested



# References

- Hodge, G.; Salokhe, G.; Zolly, L.; Anderson, N. (2007). Terminology Resource Registry: Descriptions for Humans and Computers. Presentation at Integrating Standards in Practice, 10th Open Forum on Metadata Registries, New York City, NY USA, July 9-11, 2007.  
<http://www.metadataopenforum.org/index.php?id=21,74,0,0,1,0>
- ISO/IEC 11179, Information Technology -- Metadata registries (MDR)  
<http://metadata-standards.org/11179/>
- Kendall. E. Metadata Support for OMG's Emerging Ontology & Vocabulary Management Initiative. Joint OOR-OntologySummit2008 Panel Discussion: "Developing an Ontology of Ontologies for OOR"  
[http://ontolog.cim3.net/cgi-bin/wiki.pl?ConferenceCall\\_2008\\_04\\_10](http://ontolog.cim3.net/cgi-bin/wiki.pl?ConferenceCall_2008_04_10)
- Golub, K.; Tudhope, D. TRSS survey report 2008 (draft)  
Aug, 7, 2008



Thank you!

**<http://nkos.slis.kent.edu>**  
(NKOS Website, announcements, listserv,  
workshop materials, etc.)