Update on ISO 25964: Thesauri and interoperability with other vocabularies

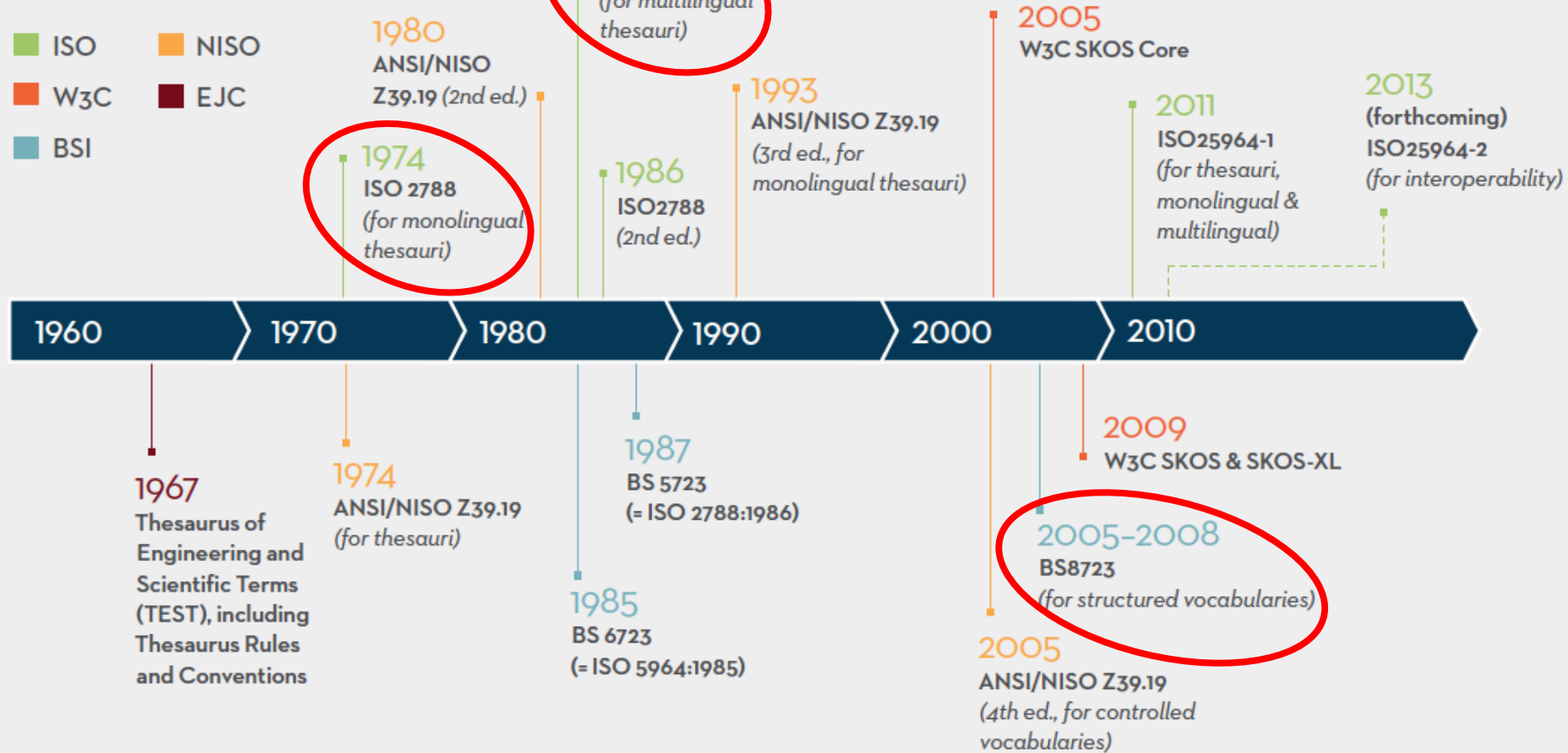## Reported by Doug Tudhope

TPDL – 2012-09-27 – NKOS workshop

What is ISO 25964?

ISO 25964: Thesauri and interoperability with other vocabularies

- Part 1: Thesauri for information retrieval
- Part 2: Interoperability with other vocabularies

- It updates ISO 2788 and ISO 5964
- based on BS 8723, with much reworking
- Part 1, **published in August 2011**, covers monolingual and multilingual thesauri
- Part 2, to be published in 2012, covers mapping between thesauri and other types of vocabulary
- information retrieval seen as main application; mapping applies to index terms or to search terms

Timeline of Landmark Thesaurus Standards in the English Language

Legend: ISO · NISO · W3C · EJC · BSI

1980
ANSI/NISO Z39.19 (2nd ed.)

1985
ISO 5964
(for multilingual thesauri)

1974
ISO 2788
(for monolingual thesauri)

1986
ISO2788
(2nd ed.)

1993
ANSI/NISO Z39.19
(3rd ed., for monolingual thesauri)

2005
W3C SKOS Core

2011
ISO25964-1
(for thesauri, monolingual & multilingual)

2013
(forthcoming)
ISO25964-2
(for interoperability)

Timeline axis: 1960 · 1970 · 1980 · 1990 · 2000 · 2010

1967
Thesaurus of Engineering and Scientific Terms (TEST), including Thesaurus Rules and Conventions

1974
ANSI/NISO Z39.19
(for thesauri)

1987
BS 5723
(= ISO 2788:1986)

1985
BS 6723
(= ISO 5964:1985)

2009
W3C SKOS & SKOS-XL

2005–2008
BS8723
(for structured vocabularies)

2005
ANSI/NISO Z39.19
(4th ed., for controlled vocabularies)

Who is involved in developing the standard?

- A Working Group (WG8), under the ISO subcommittee known as ISO TC46/SC9, has drafted the standard.

- WG8 has members from 15 countries.

- The WG8 Secretariat is provided by NISO in the USA

- Currently active members of WG8 include:

Johan De Smedt	Marianne Lykke
Stella Dextre Clarke (Leader)	Esther Scheven
Michèle Hudon	Douglas Tudhope
Daniel Kless	Leonard Will
Jutta Lindenthal	Marcia Lei Zeng

What does "interoperability" mean?

Definition: ability of two or more systems or components to exchange information and to use the information that has been exchanged.

In the case of thesauri and other KOS, broadly speaking interoperability applies at 2 levels:
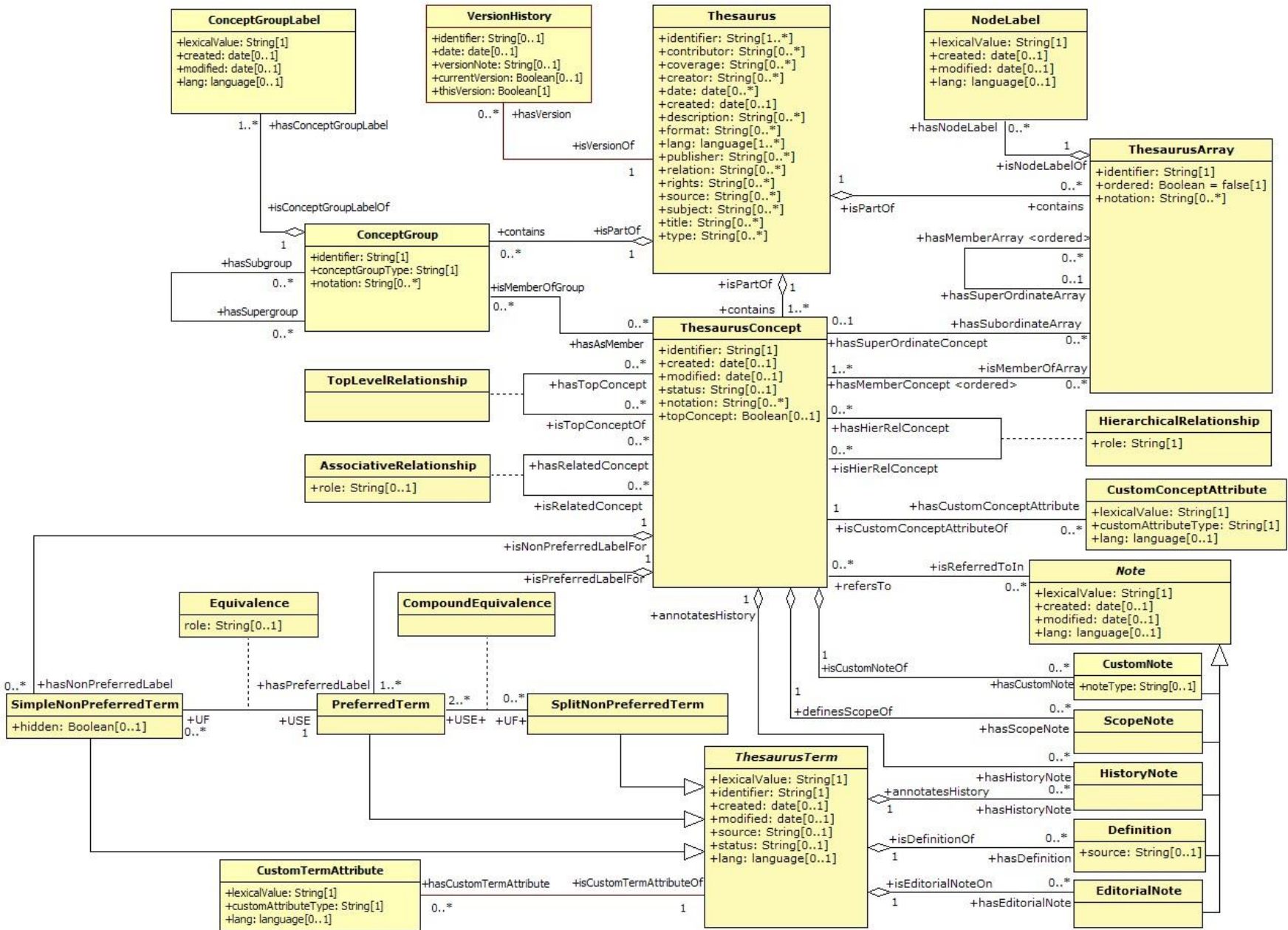
- presenting data in a standard way to enable import and use in other systems   (ISO 25964 Part 1)
- providing mappings between the terms/concepts of one KOS and those of another (ISO 25964 Part 2)

What's in Part 1?

All that was in ISO 2788 and ISO 5964, revised and extended to include:

- thesaurus content and construction, mono- or multi-lingual.
- guidance on applying facet analysis to thesauri
- guidance on managing thesaurus development and maintenance
- functional requirements for software to manage thesauri
- a data model and derived XML schema

ConceptGroupLabel
+lexicalValue: String[1]
+created: date[0..1]
+modified: date[0..1]
+lang: language[0..1]

VersionHistory
+identifier: String[0..1]
+date: date[0..1]
+versionNote: String[0..1]
+currentVersion: Boolean[0..1]
+thisVersion: Boolean[1]

Thesaurus
+identifier: String[1..*]
+contributor: String[0..*]
+coverage: String[0..*]
+creator: String[0..*]
+date: date[0..*]
+created: date[0..1]
+description: String[0..*]
+format: String[0..*]
+lang: language[1..*]
+publisher: String[0..*]
+relation: String[0..*]
+rights: String[0..*]
+source: String[0..*]
+subject: String[0..*]
+title: String[0..*]
+type: String[0..*]

NodeLabel
+lexicalValue: String[1]
+created: date[0..1]
+modified: date[0..1]
+lang: language[0..1]

ThesaurusArray
+identifier: String[1]
+ordered: Boolean = false[1]
+notation: String[0..*]

ConceptGroup
+identifier: String[1]
+conceptGroupType: String[1]
+notation: String[0..*]

ThesaurusConcept
+identifier: String[1]
+created: date[0..1]
+modified: date[0..1]
+status: String[0..1]
+notation: String[0..*]
+topConcept: Boolean[0..1]

TopLevelRelationship

HierarchicalRelationship
+role: String[1]

AssociativeRelationship
+role: String[0..1]

CustomConceptAttribute
+lexicalValue: String[1]
+customAttributeType: String[1]
+lang: language[0..1]

Note
+lexicalValue: String[1]
+created: date[0..1]
+modified: date[0..1]
+lang: language[0..1]

Equivalence
role: String[0..1]

CompoundEquivalence

CustomNote
+noteType: String[0..1]

ScopeNote

HistoryNote

SimpleNonPreferredTerm
+hidden: Boolean[0..1]

PreferredTerm

SplitNonPreferredTerm

ThesaurusTerm
+lexicalValue: String[1]
+identifier: String[1]
+created: date[0..1]
+modified: date[0..1]
+source: String[0..1]
+status: String[0..1]
+lang: language[0..1]

Definition
+source: String[0..1]

CustomTermAttribute
+lexicalValue: String[1]
+customAttributeType: String[1]
+lang: language[0..1]

EditorialNote

Relationship labels:
1..* +hasConceptGroupLabel
+isConceptGroupLabelOf 1
+hasSubgroup 0..*
+hasSupergroup 0..*
0..* +hasVersion +isVersionOf 1
+contains +isPartOf 0..* 1
+isMemberOfGroup 0..*
+hasAsMember 0..*
+hasNodeLabel 0..*
+isNodeLabelOf 1 0..*
+isPartOf 1 +contains
+hasMemberArray <ordered> 0..*
+hasSuperOrdinateArray 0..1
+hasSubordinateArray 0..*
+isPartOf 1 +contains 1..*
+hasSuperOrdinateConcept 0..1
+isMemberOfArray 1..*
+hasMemberConcept <ordered> 0..*
+hasTopConcept 0..*
+isTopConceptOf 0..*
+hasHierRelConcept 0..*
+isHierRelConcept 0..*
+hasRelatedConcept 0..*
+isRelatedConcept 0..*
+hasCustomConceptAttribute 1
+isCustomConceptAttributeOf 0..*
+isNonPreferredLabelFor 1
+isPreferredLabelFor 1
+isReferredToIn 0..*
+refersTo 0..*
+annotatesHistory 1
+isCustomNoteOf 1
+hasCustomNote 0..*
+definesScopeOf 1
+hasScopeNote 0..*
+hasHistoryNote 0..*
+annotatesHistory 1
+hasHistoryNote
+isDefinitionOf 0..*
+hasDefinition 1
+isEditorialNoteOn 0..*
+hasEditorialNote 1
+hasNonPreferredLabel 0..*
+hasPreferredLabel 1..*
+UF 0..*
+USE 1
+USE+ 2..* +UF+ 0..*
+hasCustomTermAttribute +isCustomTermAttributeOf
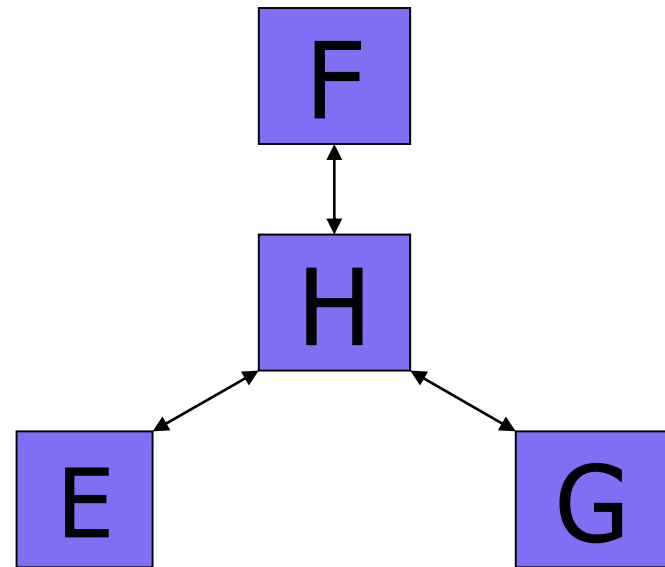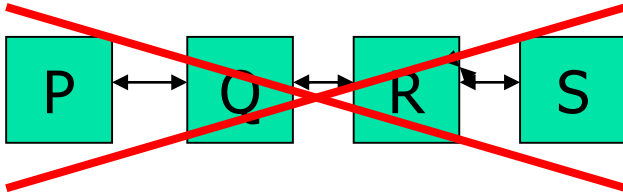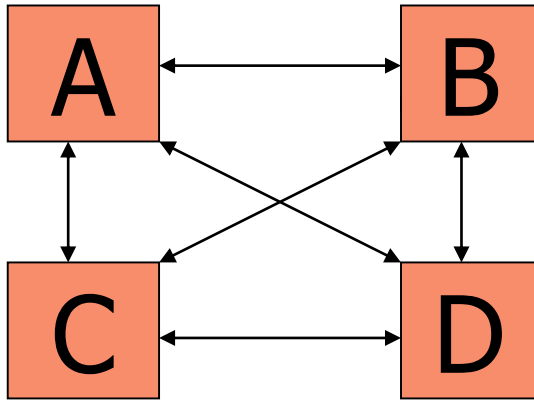0..* 1

7

ISO 25964 data model <-&->SKOS

+ The ISO 25964 data model is broadly compatible with SKOS

+ Greater compatibility awaits SKOS extension to allow for:

   - version history

   - compound equivalence among terms

   - use of concept group and of array

   - custom extensions

   - specific properties of concepts, terms, notes and thesaurus that are not covered in SKOS

+ Johan De Smedt will report more on this aspect.

What's in Part 2?

+ Models for mapping
+ Guidelines for mapping
    + Recommendations on mapping types
    + How to handle pre-coordination
    + Mapping to vocabularies other than thesauri:
        + classification schemes
        + file plans (Classification schemes used for records management)
        + taxonomies
        + subject heading schemes
        + ontologies
        + terminologies
        + name authority lists
        + synonym rings
+ Brief guidance on handling mappings data

Recommend "Models for mapping"

How does Part 2 support interoperability?

- mapping = "process of establishing relationships between the **concepts** of one vocabulary and those of another"

- Recommended types of mapping are based on the standard internal relationship types, basically: **equivalence, hierarchical and associative**

- Greater differentiation of mapping types is allowed, but is **optional**, to avoid complexity in simple applications

Vocabularies other than thesauri

- ISO 25964 is a standard for thesauri; it does not attempt to standardize other types of KOS. It guides only on **interoperability** between thesauri and other types of KOS.

- The clause on each KOS type presents:
    - Key characteristics of the KOS (non-normative)
    - Semantic components/relationships (non-normative)
    - Recommendations for mapping between the KOS and a thesaurus (normative)

Vocabularies other than thesauri – Classification Schemes (eg)

- Selected classes from a classification scheme (UDC)
  with a notation that uses numbers and symbols

(084)	**pictorial documents**
(084.12)	**photographs**
5	**mathematics and natural science**
59	**zoology**
591.1	**animal physiology**
591.132	**digestion**
599	**mammals**
599.74	**carnivores**
599.742.71	**"big cats"; lions; tigers**

599.742.71:591.132 digestion in lions
599.742.71(084.12) photographs of lions

Vocabularies other than thesauri – Classification Schemes (eg)

- Thesaurus relationships only hierarchical if generic, partitive or instantial but Classification scheme allows discipline based hierarchies, eg chemical elements under chemistry
- when classes of a classification scheme mapped to concepts of a thesaurus, even exact equivalence mapping, commonly the hierarchical structure between the classes is not the same as the hierarchical relationships between the thesaurus concepts
- scope of a class should be determined by examining its caption(s), superordinate and subordinate classes, any descriptive notes and corresponding index entries

Vocabularies other than thesauri – Terminologies

- A terminology is a "set of designations belonging to one special language" (ISO 1087-1), each designation representing a concept by means of a sign, symbol, term or appellation.

- Terminological data can be presented eg as term banks, termbases, glossaries or other publications.

- Traditional aim to improve clarity of communication in spoken or written discourse. Eg guide a translator or a writer find a term for a precise meaning, understand meaning of a term or find different terms for a particular concept in the same or a different language.

- Information retrieval not prime motivation. Nonetheless, the terms in a terminology can be useful in searching full text, sometimes using natural language processing tools.

Vocabularies other than thesauri – Terminologies

- Two types of concept distinguished in ISO 704

- An "individual concept" corresponds to only one object, and is broadly comparable to the type of entity named in a name authority list.

- A "general concept" corresponds to two or more "objects" (including abstract concepts as well as physical objects).

- Most of the concepts in a thesaurus would be considered "general concepts" in a terminology, although a thesaurus may also admit "individual concepts".

- Because terminologies not used for indexing, mapping does not have the same goals as most of the vocabularies described in this standard. However some use cases that call for selective mapping.

- Extending a thesaurus in a domain covered by a terminology
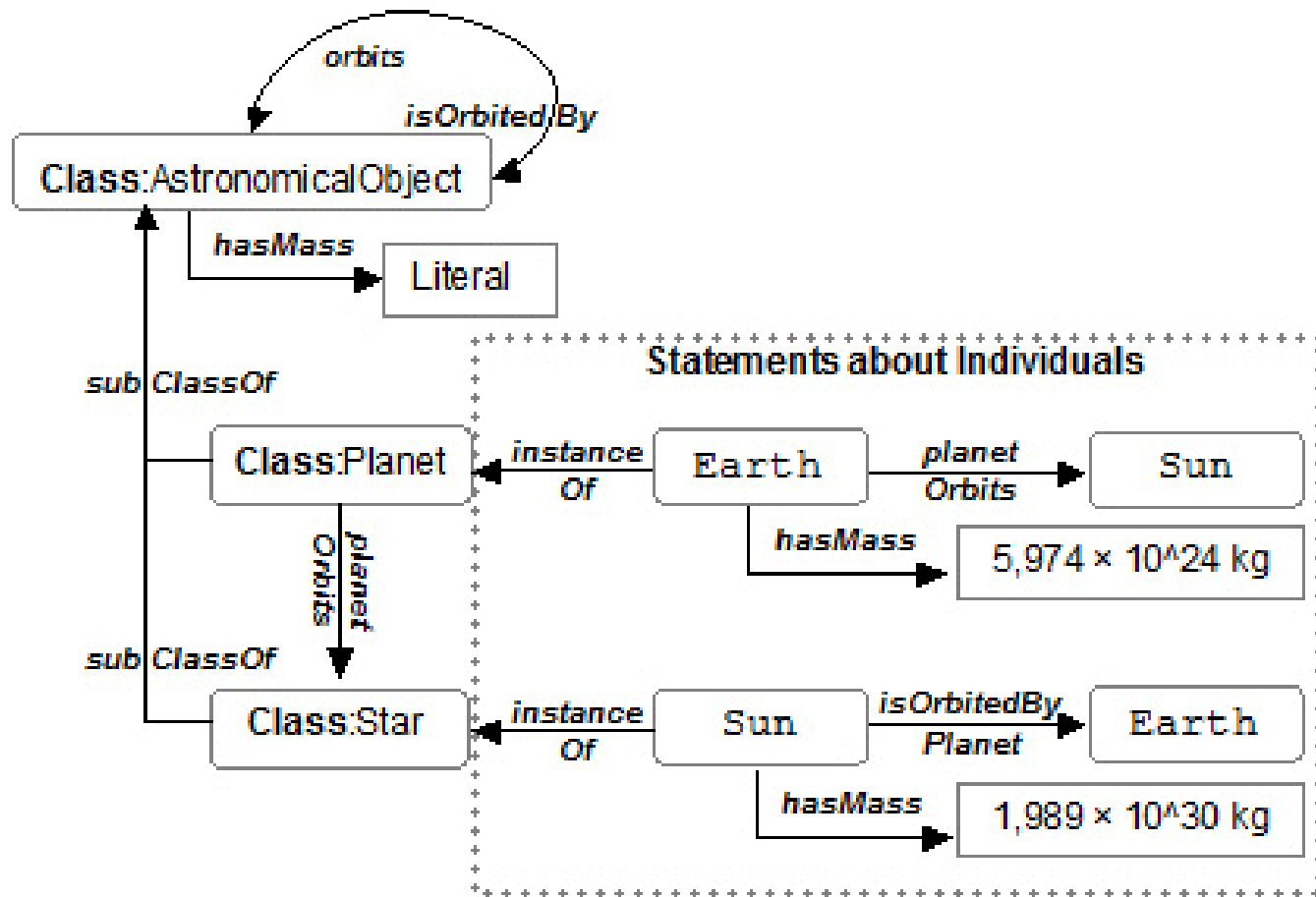
- Complementing a thesaurus in retrieval applications

Ontologies are unlike other types of KOS …

- Definition of ontology excludes "lightweight" examples such as thesauri and classification schemes

- The Gruber/Studer definition is adopted, and interpreted broadly enough to admit OWL-based examples such as ORE and FOAF.

- Mapping between ontologies and thesauri is not recommended.

- Interoperability recommendations focus on use cases such as reengineering a thesaurus as an ontology, and complementary use of thesaurus with ontology.

Simple ontology illustration
(credit: Jutta Lindenthal)

Structural comparison

- The illustration is used to draw out key similarities and differences between ontologies and thesauri.

- The aim is to encourage emerging applications in which thesauri and ontologies can usefully interoperate.

Progress with ISO 25964-2

- A draft (ISO DIS 25964-2) was issued in early 2012, for public comment
- All the feedback has been incorporated in a revised version, submitted to the NISO Secretariat in August 2012
- Publication by ISO is expected later in 2012.

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
- It is not cheap to purchase
- However, the XML schema 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/

Want a copy of ISO 25964-2 ?

- Send us an email and we'll alert you when it is published.   stella@lukehouse.org
- You can then obtain it from ISO or from your national standards body.

PS Members of WG8 (all unpaid volunteers) greatly regret the copyright restrictions and prices charged by ISO and the standards bodies. No other way of releasing the work has been found.

ISO 2788 (1986) + ISO 5964 (1985) + New content, adapted from BS8723

extensive revision

New content, adapted from BS8723

ISO 25964 Part 1 (published 2011)

ISO 25964 Part 2 (expected 2012)