A Flexible XML-Based Thesaurus Approach for the Federal Government: Version 4

by Ken Sall and Judith Newton
for the XML Community of Practice
April 20, 2005
http://kensall.com/gov/glossary

1

Agenda

- **Candidate Requirements**
- **Generated Search Links Using XSLT** `New!`
- **SKOS = ISO 2788 + W3C + XML + metadata +** `New!` **RDF + Semantic Web + Web Service + Wiki**
- **Recommended Plan of Action**
- **Thesaurus Spreadsheet w/SKOS Subset** `New!`
- **Initial XML Schema**
- **References: Relevant ISO Specifications**
  - **ISO 2788:1986 (oldest)**
  - **ISO 1087:2000**
  - **ISO 704:2000**
  - **ISO 15836:2003 - DCMI Metadata Terms**

Disclaimer

- **The approach discussed herein is not officially endorsed by any specific government agency.**
- **In its latest form, it has not yet been blessed by:**
  - **Chief Architects Forum (CAF)**
  - **DRM Working Group**
  - **Semantic Interoperability Community of Practice (SiCoP)**
- **However, the approach has gained considerable interest and is under active discussion.**
- **CAUTION: Specifics are likely to change.**

Candidate Requirements (1)

- The glossary / lexicon / <u>thesaurus</u> SHOULD use XML syntax with a schema (DTD, XML Schema, or RDF-S) for validation.
- It SHOULD be applicable to any government agency.
- The schema SHOULD be available to any civil servant or citizen. [Should govt be expected to use it?]
- The schema SHOULD not be overly complex.
- The schema SHOULD contain few required elements and many optional and/or repeatable elements.
- It SHOULD be relatively easy to add new terms to the lexicon. Payware SHOULD not be necessary for authoring.

4

Candidate Requirements (2)

- **It SHOULD be relatively easy to combine terms authored by different individuals and different agencies, if desired.**

- **The elements in the schema SHOULD be chosen with ISO standards in mind, to the degree that this does not overly complicate the schema.**

- **It SHOULD be possible to create an XSLT stylesheet based upon the model to display an XML glossary instance document as HTML in modern browsers (IE, Firefox).**

- **It is DESIRABLE that the XSLT generate additional search links not in the source.**

- **Multiple definitions of the same term MUST be permitted, with either same or different context.**

Candidate Requirements (3)

- **The entire approach SHOULD foster a clean separation of collaborative roles:**
  - Developer of schema vs. developer of stylesheets
  - Author/collector of terms and definitions
  - Reviewer/approver of definitions
  - Consumer of results (e.g., agency with custom XSLT)
- **It SHOULD support semantic relationships between terms including related-to and synonyms.**
- **An approval process SHOULD be defined, but it should *not* interfere with contributions. Un-reviewed definitions would still be accessible, but without the "stamp of approval".**
- **It MUST be possible to indicate a term's**
  - Source (agency, author, document, and/or URL)
  - Context
  - Approval status
  - TBD – what else is mandatory?

6

Candidate Requirements (4)

- **Clear authoring conventions SHOULD be established**
  - Case convention (UpperCamelCase, Title Case, lowercase, ?)
  - Pluralization (use singular form)
  - Compound terms (e.g., Data Architecture, Data Class)
  - Placement of acronym/abbreviation (separate element)
  - Placement of source/context/concept (separate element)
  - Citation method (URIs, bibliographical, free form?) [Source could contain child elements for each possible format]
  - TBD others?
- **Usage notes and/or examples are DESIRABLE.**

Sall's XML Glossary Model Strawman

Previous Presentation

Generated with XMLSpy Schema Editor   www.xmlspy.com

8
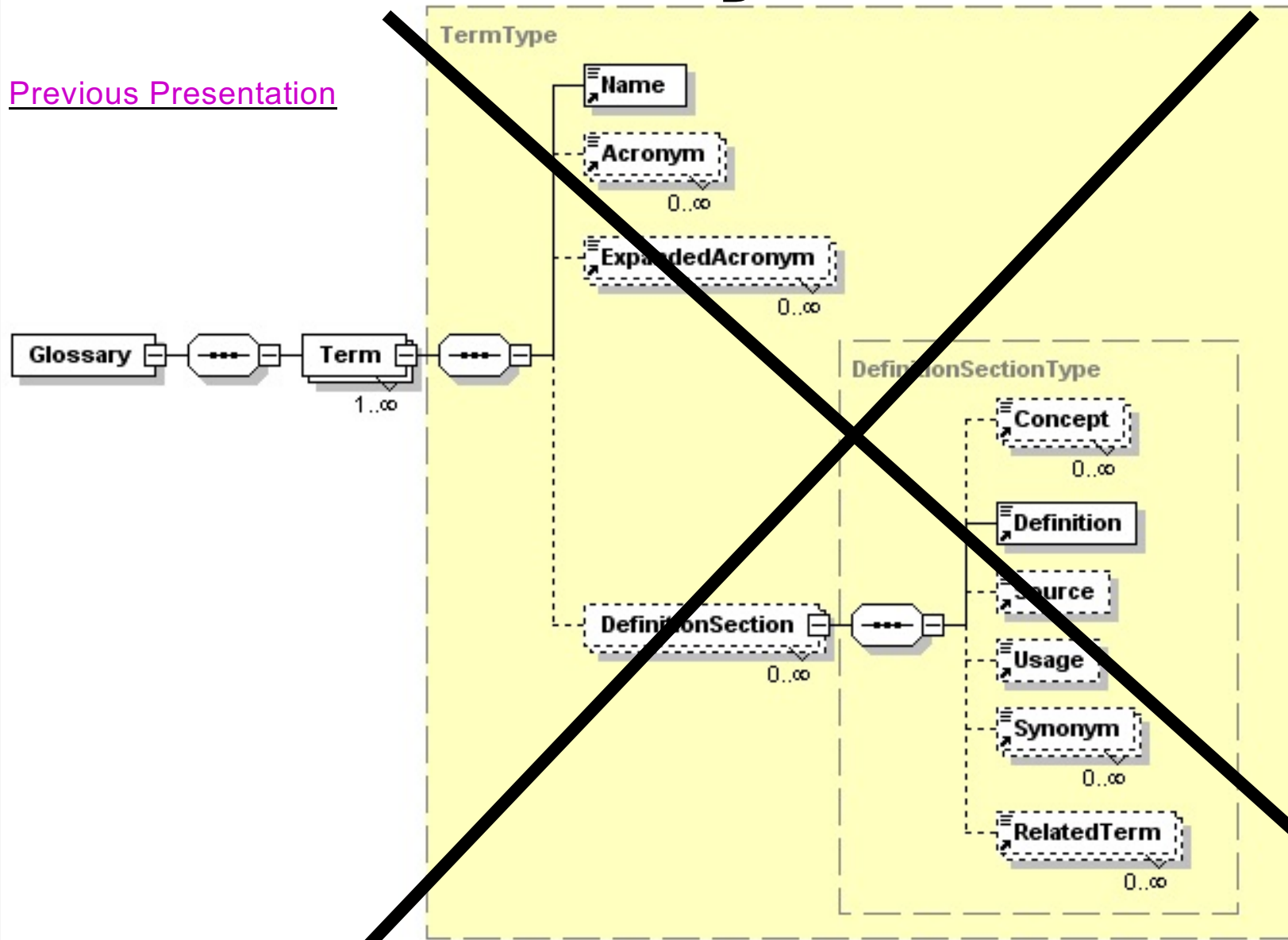
XML Example of One Term

```
<Term id="ontology">
  <Name>ontology</Name>
  <DefinitionSection>
    <Concept>semantic web</Concept>
    <Concept>knowledge management</Concept>
    <Definition>Defines the common words and concepts used to describe and represent
an area of knowledge, and so standardizes the meanings.
    An ontology includes classes in the domains of interest, instances, relationships,
properties and their values,
    functions of and processes involving the objects, and relevant constraints and
rules.</Definition>
    <Source>Daconta, Obrst, Smith</Source>
    <Usage>An onotology can range from the simple notion of a taxonomy to a thesaurus,
to a conceptual model, to a logical theory.
    [Daconta, Obrst, Smith]</Usage>
    <Synonym>classification system</Synonym>
    <RelatedTerm>taxonomy</RelatedTerm>
    <RelatedTerm>OWL</RelatedTerm>
  </DefinitionSection>
  <DefinitionSection>
    <Concept>philosophy</Concept>
    <Definition>[sometimes "Ontology"] the metaphysical study of the nature of being and
existence</Definition>
    <Source>WordNet</Source>
    <Usage>Both the ontology and manner of human existence are of concern to
Existentialism.</Usage>
    <Synonym>metaphysics</Synonym>
  </DefinitionSection>
</Term>
```

9

Search Links Bootstrap:
Based on CDT-FG + CAF Glossary.doc [snapshot]

XSLT-Generated Search Links

- [AcronymFinder](#) – if acronym indicated
- [WikiPedia](#)
- [Clusty](#)
- [Clusty Gov [.gov and .mil]](#)
- [Google Uncle Sam [.gov and .mil]](#)
- [Google Define](#)
- [Google](#)
- [Merriam-Webster](#)
- [W3C](#)
- [W3Schools](#)
- [Webopedia](#)
- [WhatIs](#)
- [WordNet](#)
- [ZVON](#)

ISO Spec Influences: Thesauri

- **ISO 2788:1986 (oldest)**
  - **"Guidelines for the establishment and development of monolingual thesauri"**
  - **How to select terms; how to express relationships**
- **ISO 1087:2000**
  - **"Vocabulary of terminology"**
  - **Subject field, concept, characteristic, extension, etc.**
  - **Hierarchical and associative relations**
- **ISO 704:2000**
  - **"Terminology work – Principles and methods"**
  - **"Concentrate on the essential and delimiting characteristics."**
- **ISO 15836:2003 – Dublin Core Metadata Initiative terms**
  - **Description, Source, Subject, Title, Creator, Date, etc.**
  - **Relation – isPartOf, conformsTo, hasPart, etc.**

**Jump to **

W3C's SKOS [1]

- **Simple Knowledge Organisation System**
- **http://www.w3.org/2004/02/skos/**
- **"SKOS is an open collaboration developing specifications and standards to support the use of knowledge organisation systems (KOS) on the semantic web."**
- **SKOS *Core Vocabulary* (and *Core Guide*) - W3C Working Draft: 3/31/05; work (actively) in progress; subject to backwards incompatible changes!**
- **RDF Schema for thesauri and related knowledge organisation systems**
- **"SKOS Core provides a model for expressing the basic structure and content of concept schemes (thesauri, classification schemes, subject heading lists, taxonomies, terminologies, glossaries and other types of controlled vocabulary)."**

SKOS [2]

- **[Semantic Web Best Practices and Deployment Working Group](#)**
- **SKOS *Core RDF Vocabulary* - for describing thesauri, glossaries, taxonomies, terminologies.**
- **"The SKOS Core Vocabulary is an application of the Resource Description Framework (RDF), that can be used to express a concept scheme as an RDF graph. Using RDF allows data to be linked to and/or merged with other RDF data by semantic web applications."**
- **SKOS *Mapping RDF Vocabulary* - for describing mappings between concept schemes.**
- **SKOS *Web Service API* – WDSL-based**

SKOS [3]

Key

- **Quick Guide to Publishing a Thesaurus on the Semantic Web**
- **W3C Working Draft in Preparation** – **3/30/05**
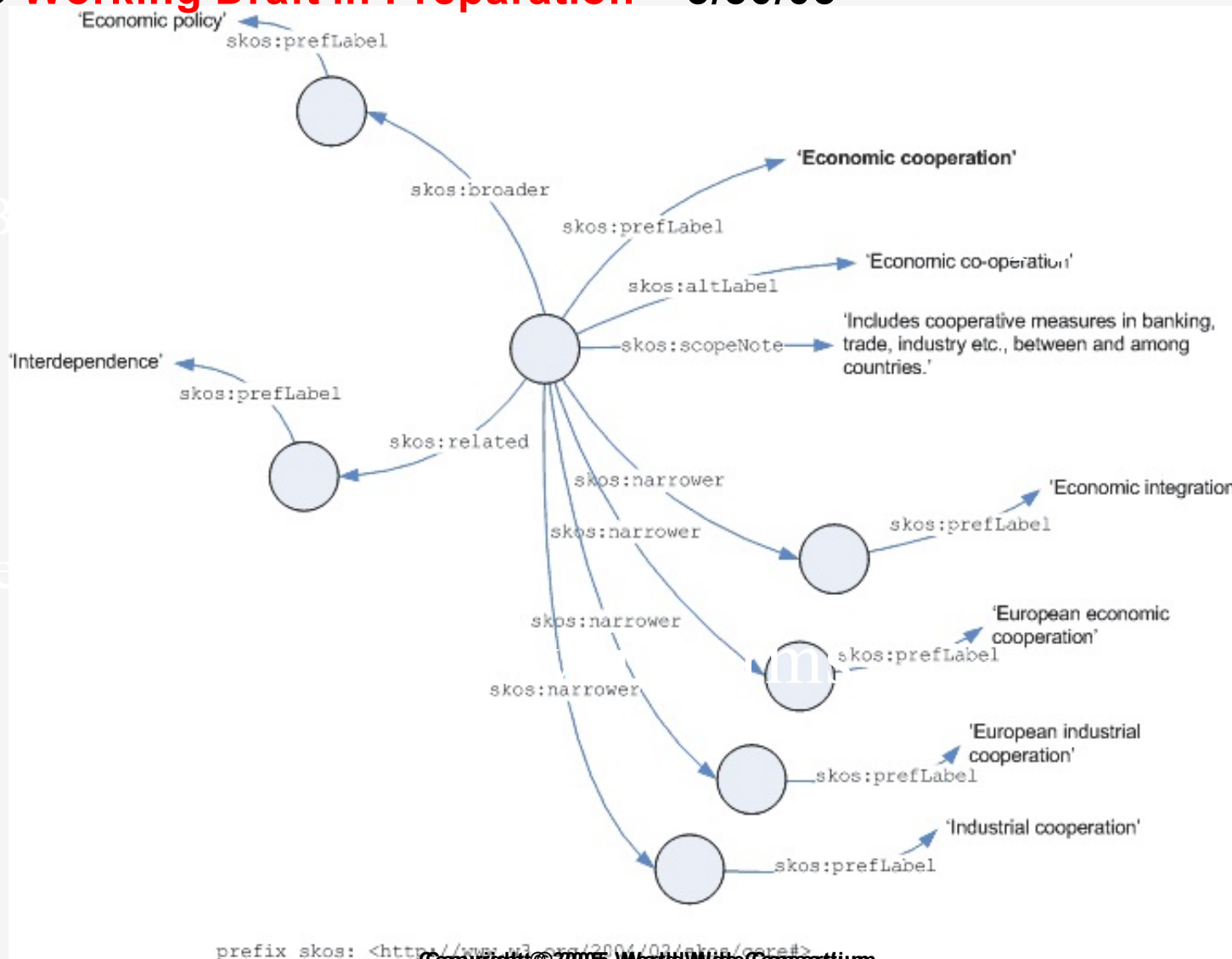
SKOS: RDF Serialization [4]

"Allocating URIs to the concepts in a thesaurus allows anybody to refer to them unambiguously from any context." - <u>**Quick Guide to Publishing a Thesaurus on the Semantic Web**</u>

```xml
<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.,
    <skos:inScheme rdf:resource="http://www.ukat.org.uk/thesaurus"/>
    <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:Concept>

</rdf:RDF>
```

Concept 1750's prefLabel is "Economic cooperation" with broader concept 4382 and narrower concepts 2108, 9505, etc.

SKOS: with Thesaurus Metadata (DCMI) [5]

```xml
<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#"
    xmlns:dc="http://purl.org/dc/elements/1.1/">

    <skos:ConceptScheme rdf:about="http://www.ukat.org.uk/thesaurus">
        <dc:title>The UK Archival Thesaurus</dc:title>
        <dc:description>A subject thesaurus produced to support indexing in the UK arch
        <dc:creator>UK Archival Thesaurus project</dc:creator>
        <dc:date>2004-08-22</dc:date>
        <dc:format>text</dc:format>
        <dc:language>en</dc:language>
        <dc:rights>All rights reserved. Data in the UK Archival Thesaurus may be freely
        <skos:hasTopConcept rdf:resource="http://www.ukat.org.uk/thesaurus/field/1"/>
        <skos:hasTopConcept rdf:resource="http://www.ukat.org.uk/thesaurus/field/2"/>
        <skos:hasTopConcept rdf:resource="http://www.ukat.org.uk/thesaurus/field/3"/>
        <skos:hasTopConcept rdf:resource="http://www.ukat.org.uk/thesaurus/field/4"/>
        <skos:hasTopConcept rdf:resource="http://www.ukat.org.uk/thesaurus/field/5"/>
        <skos:hasTopConcept rdf:resource="http://www.ukat.org.uk/thesaurus/field/6"/>
        <skos:hasTopConcept rdf:resource="http://www.ukat.org.uk/thesaurus/field/8"/>
    </skos:ConceptScheme>

</rdf:RDF>
```

SKOS Complements OWL [6]

- "SKOS-Core is intended as a complement to OWL. It does provide a basic framework for building concept schemes, but it does not carry the strictly defined semantics of OWL. Thus it is ideal for representing those types of KOS, such as thesauri, that cannot be mapped directly to an OWL ontology. SKOS is also easier to use, and harder to misuse than OWL, providing an ideal entry point for those wishing to use the Semantic Web for knowledge organisation. SKOS-Core also provides a framework for linking concepts to the words and phrases that are normally used by people to refer to them. This valuable information, once captured, can be used to support a number of tasks…." – **SKOS Core Guide, 2001** version

- **Latest SKOS Core Guide** – 2/15/05 Working Draft

Subset of SKOS Core Vocabulary [8]

- **Concept** - abstract idea or notion; a unit of thought; • **2** holds term and related terms
- **ConceptScheme** – set of concepts; controlled vocabulary (e.g., what we're developing)
- **prefLabel** – name of term being defined; must be *unique* within a ConceptScheme (e.g., our thesaurus)
- **altLabel** - acronyms, abbreviations, spelling variants, and irregular plural/singular forms
- **related** - concept with which there is an associative semantic relationship
- **broader** - more general in meaning; rendered as parent in a concept hierarchy (tree)
- **narrower** – more specific meaning; child
- **definition**, **example**, **changeNote**, **editorialNote**

SKOS Example [9]

```
<skos:Concept rdf:about="http://my.example.org/GCL/791#concept">
<skos:prefLabel xml:lang="en">Civil Service</skos:prefLabel>
<skos:related rdf:resource="http://my.example.org/GCL/476#concept"/>
</skos:Concept>
<skos:Concept rdf:about="http://my.example.org/GCL/476#concept">
<skos:prefLabel xml:lang="en">Public administration</skos:prefLabel>
<skos:altLabel xml:lang="en">Administration (public)</skos:altLabel>
<skos:altLabel xml:lang="en">Management (public sector)</skos:altLabel>
<skos:related rdf:resource="http://my.example.org/GCL/791#concept"/>
<skos:related rdf:resource="http://my.example.org/GCL/982#concept"/>
</skos:Concept>
<skos:Concept rdf:about="http://my.example.org/GCL/982#concept">
<skos:prefLabel xml:lang="en">Employment relations</skos:prefLabel>
<skos:altLabel xml:lang="en">Conflict (industrial relations)</skos:altLabel>
<skos:altLabel xml:lang="en">Employers' responsibilities</skos:altLabel>
<skos:altLabel xml:lang="en">Industrial disputes</skos:altLabel>
<skos:altLabel xml:lang="en">Industrial relations</skos:altLabel>
<skos:altLabel xml:lang="en">Strikes (labour)</skos:altLabel>
<skos:altLabel xml:lang="en">Trades Unions</skos:altLabel>
<skos:related rdf:resource="http://my.example.org/GCL/474#concept"/>
<skos:related rdf:resource="http://my.example.org/GCL/476#concept"/>
</skos:Concept>
<skos:Concept rdf:about="http://my.example.org/GCL/474#concept">
<skos:prefLabel xml:lang="en">Business management</skos:prefLabel>
<skos:altLabel xml:lang="en">Administration (business)</skos:altLabel>
<skos:altLabel xml:lang="en">Management (business)</skos:altLabel>
<skos:related rdf:resource="http://my.example.org/GCL/982#concept"/>
</skos:Concept>
```

- **2**

SKOS [10]

- **Semantic Web Advanced Development for Europe: <u>SWAD-Europe Thesaurus Activity</u> and <u>SWAD-E home</u>**

- **2**

- **Standards and Best Practises for USING Knowledge Organisation Systems ON THE Semantic Web [<u>PPT from Nov. 2004 </u>conference]**

- **<u>RDF Thesaurus Prototype</u>** – **"thesaurus research prototype demonstrating the SKOS schema by means of the SKOS API web service and a demonstrator containing sample data, some simple clients for using the API, documentation and description of related work."**

- **"Scope of SKOS Core: 'Language-oriented KOS'**
  - **Thesauri**
  - **Glossaries**
  - **Controlled Vocabularies**
  - **Terminologies**
  - **Classification Schemes?**
  - **Taxonomies?**
  - **Web directories … Weblog category schemes … ?"**

> SKOS = ISO 2788 + W3C + XML + metadata + RDF + Semantic Web + Web Service

- **<u>Thesaurus Research Prototype Work Plan</u> : "Refining the existing RDF thesaurus schema to make it compatible with ISO 2788: Guidelines for the establishment and development of monolingual thesauri, will ensure the schema is compatible with most existing thesauri, improving the possibilities of migration."**

- **<u>SKOS Thesaurus Web Service Demonstrations</u>**

- **<u>public-esw-thes@w3.org Mail Archives</u>**

Next Steps - Revised

Update: We may go with SKOS directly.

• **2**

- Determine interested agencies and establish funding.
- Before agencies start authoring, form ad hoc working groups to finalize DTD or XML Schema *using elements that parallel SKOS and ISO 2788*. (Agencies can gather their terms and definitions using an interim schema or using spreadsheets.)
- Determine entry review/approval process and form second team to conduct reviews of submissions.
- Revise initial XSLT to match final Glossary schema.
- Determine repository and submission mechanisms.
  - Could be a good use for <u>CORE.gov</u>?
  - Coordinate with <u>Plans for Derived XML Registry Prototype</u>?
- Write additional XSLT stylesheets for:
  - Merging terms and pulling agency-specific terms
  - Special display requirements
  - Filtering only approved terms
  - Filtering only terms that meet agency-specific criteria

Candidate Review Elements

- **Review – repeatable container element**   • 2
- **ReviewDate – in a standard format a la GJXDM**
- **ReviewerEmail**
- **ReviewerName?**
- **ReviewStatus = {approved, rejected, pending}**
- **ReviewDecision = {primary, secondary, tertiary}**
- **(This idea needs more thought and probably can be deferred.)**

Recommendation: Phased Approach

- Emphasis on ease of implementation and use in the   • 2
  short run, but with expansion path for long run.
- Phase 1:
- a) <u>Developers</u>: Create schema and distribute/post.
- b) <u>Expert</u>: Distill ISO 2788 to 3-4 page authoring guide.
- Phase 2: <u>Authors</u>: Gather terms and definitions.
- Phase 3: <u>Reviewers</u>: Review definitions and approve,
  reject, or defer (tentative approve? Pending?).
- Phase 4: "Publish" Thesaurus version 1.0.
- Phase 5: Iterate Phases 2, 3, and 4 for next version. On-
  going access; can access terms not yet reviewed.
- Phase 6: <u>Developers</u>: Translate schema and Thesaurus
  to SKOS, after evaluating effort. Can be begun after
  Phase 1, but need representative set of terms and
  definitions.

Our Subset: SKOS Core Vocabulary

- *Classes* Collection Concept

- *Properties* altLabel broader (changeNote) definition (editorialNote) example narrower prefLabel related scopeNote subject (semanticRelation)

- 2

- plus 2 more of our own:
  - SOURCE
  - ABBREVIATION_OR_ACRONYM

Borrowed SKOS Properties [1]

2

Concept	Represents an abstract idea or notion; a unit of thought. This is essentially a term with possibly related terms. Column headings in all UPPERCASE represent properties not found in SKOS. All other column headings correspond exactly to "properties" in the W3C SKOS Core Vocabulary Specification Working Draft dated March 3, 2005.
Each Concept (row) is described by 11 columns, each of which is described below.	Note that the only required column is prefLabel, the preferred name for the concept. In practise, we anticipate that most instances will also have a definition, but this is not explicitly required since a true thesaurus does not always define terms.
prefLabel	Preferred name for this concept. This is typically thought of as the name of the term, sometimes called the designator in ISO specifications. NOTE: In SKOS, no two concepts in the same concept scheme may have the same value.
altLabel	Alternate name for this concept. Spelling variants and irregular plural/singular forms may be included among the alternative labels for a concept. NOTE: In a departure from the SKOS Core Vocabulary Specification Working Draft, term authors SHOULD NOT place abbreviations or acronyms here.
ABBREVIATION_OR_ACRONYM	Although the SKOS Core Vocabulary Specification Working Draft suggests that altLabel may include abbreviations and acronyms for the concept, our design calls for placing abbreviations and acronyms in this separate element. We believe this will enable easy filtering or querying to construct abbreviation and acronym lists that often appear in government documents.
definition	A statement or formal explanation of the meaning of a concept. Although this is not required in this XSD, most instances are expected to include a definition.
subject	A concept that is the subject of the current term. This may be one of the "broader" terms. Ex: Subject of Toyota Camry is automobiles
scopeNote	A note that helps to clarify the meaning of a concept. This could be contextual information, a particular domain, or some other constraint to bound the scope of the concept, perhaps to distinguish it from the same term (prefLabel) used in a different scope. While ISO 2788 allows definitions to appear in the scopeNote, we are following the SKOS approach of using the separate definition element for that purpose.

Borrowed SKOS Properties [2]

- 2

scopeNote	A note that helps to clarify the meaning of a concept. This could be contextual information, a particular domain, or some other constraint to bound the scope of the concept, perhaps to distinguish it from the same term (prefLabel) used in a different scope. While ISO 2788 allows definitions to appear in the scopeNote, we are following the SKOS approach of using the separate definition element for that purpose.
SOURCE	Source of the definition. Official document names and URLs are preferred, but specific names of people or agencies are acceptable. Please follow the biblographic citation conventions in the Guidelines. NOTE: This element does not appear in the SKOS Core Vocabulary Specification Working Draft.
example	An illustration of the use of the concept, such as in a sentence.
narrower	A concept that is more specific in meaning. Narrower concepts are typically rendered as children in a concept hierarchy (tree). Ex: Bald eagle is narrower than eagle; eagle is narrower than bird.
broader	A concept that is more general in meaning. Broader concepts are typically rendered as parents in a concept hierarchy (tree). Ex: Bird is broader than eagle; eagle is broader than bald eagle.
related	A concept with which there is an associative semantic relationship. Ex: Nest, flying, and wings are all concepts that are related to the concept bird. However, these terms are neither "narrower" nor "broader". They reflect an association based on our expereince with the concept.

Initial XML Schema - Main

- **2**

CollectionType

ConceptType

prefLabel

altLabel
0..∞

ABBREVIATION_OR_ACRONYM
0..∞

definition
0..∞

subject
0..∞

scopeNote

SOURCE
0..∞

example
0..∞

narrower
0..∞

Collection — Concept — broader
0..∞

Root element.   1..∞

related
0..∞
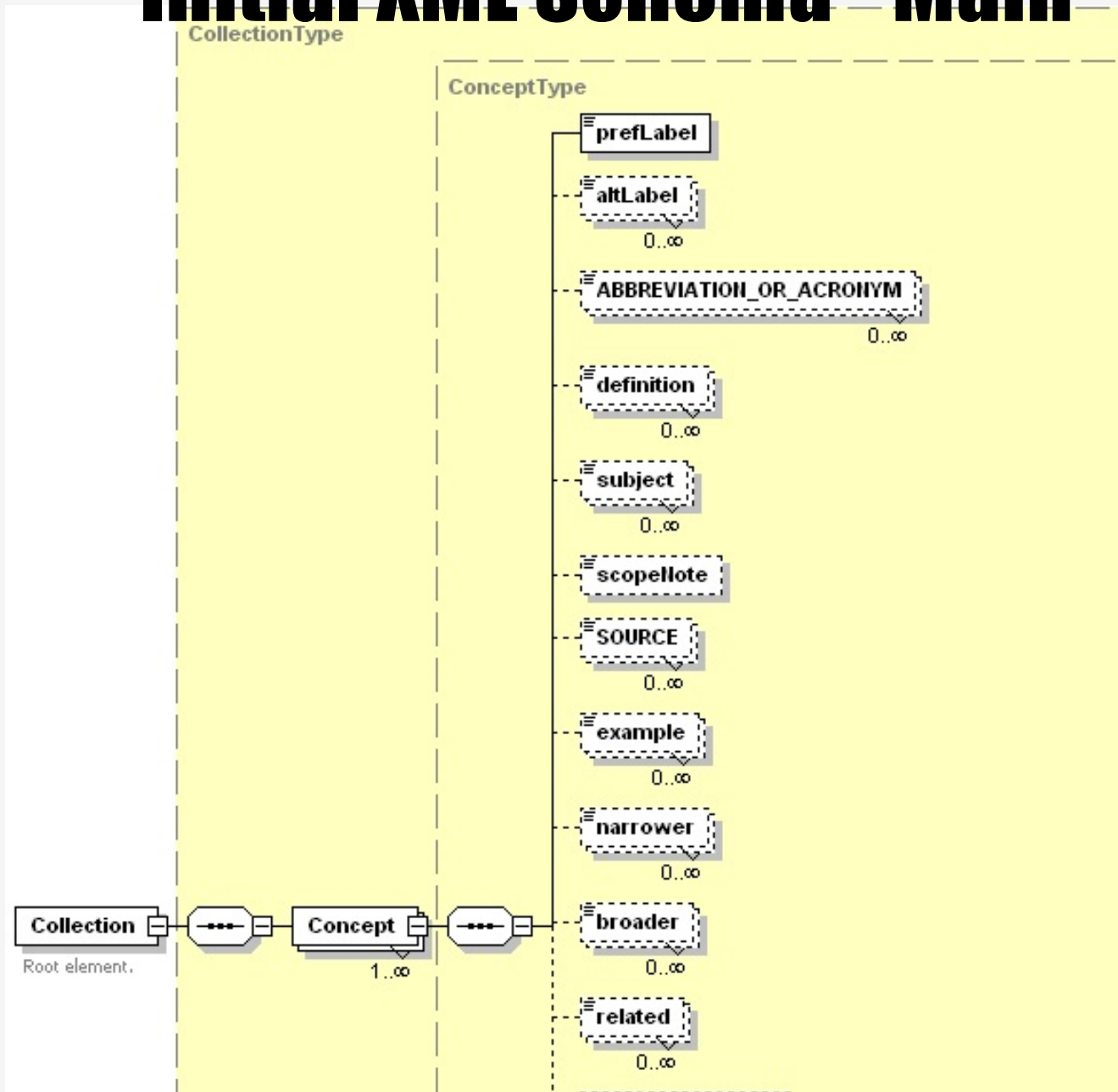
Initial XML Schema - Ancillary

- 3
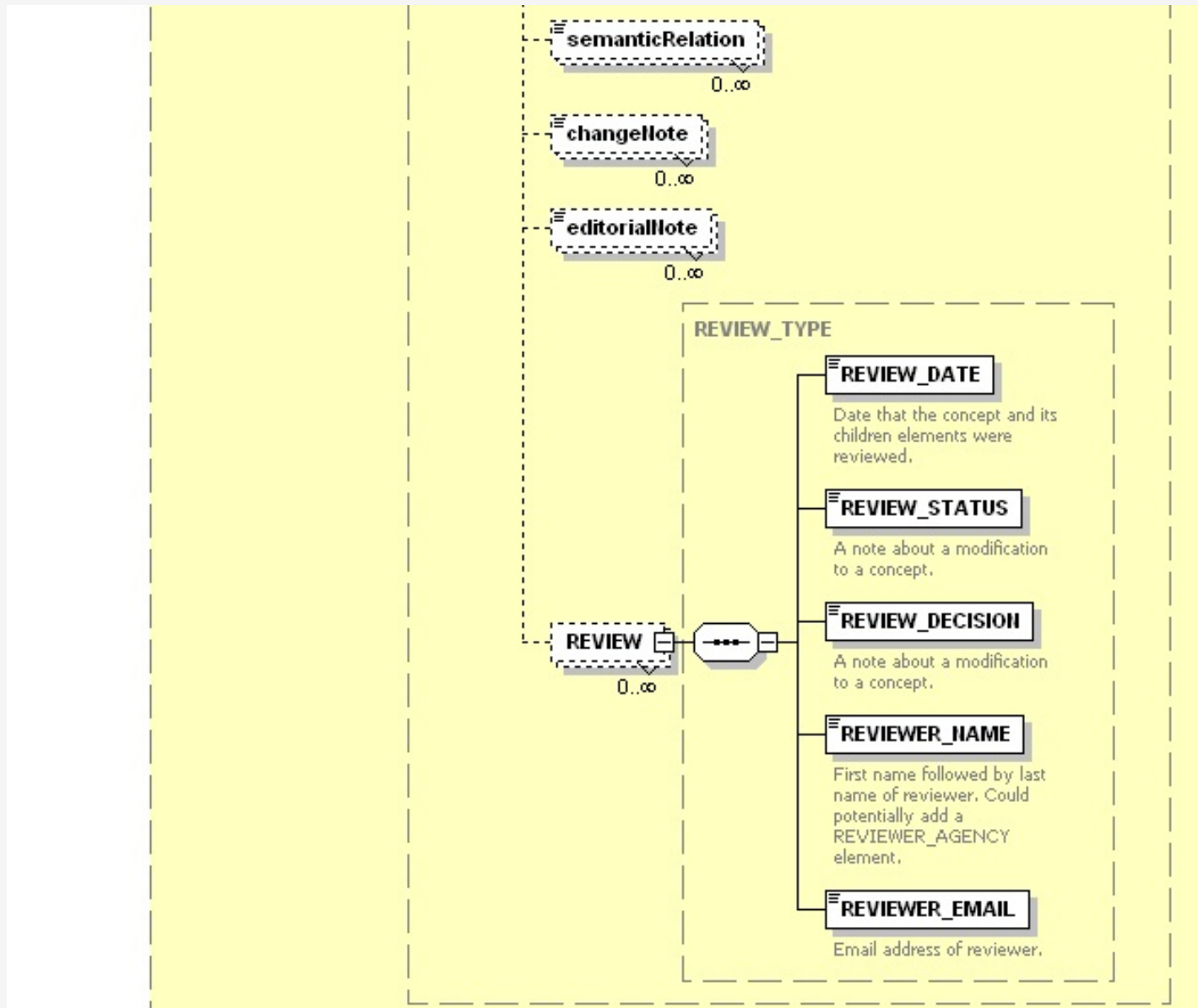
GAO Thesaurus Excerpt in Our .xls

- 3

prefLabel	altLabel	ABBREVIATION_OR_ACRONYM	definition	subject	scopeNote	SOURCE	example	narrower	broader	related
computer	computer equipment; computer hardware							embedded computer system; mainframe computer; personal computer; supercomputer	electronic equipment	computer crime; computer equipment contract; computer equipment industry; computer equipment management; computer network; information technology
computer network								Internet; intranet; local area network; wide area network	computer; telecommunication	data transmission; computer network protocol; computer bulletin board
computer security	computer security measures		Measures to protect a computer's hardware, software and data.						security	computer crimes; computer viruses; data encryption; data integrity; facility security; hackers; identity verification; information resources management
computer software	computer programs; software		Instructions that cause computers to function. Includes data bases and programs.					application software; computer virus; data base; operating system	information technology; programming languages	computer software documentation; computer software industry; computer software reuse; computer software verification and validation; computer-aided software engineering
computer virus			computer software that copies itself by attaching to other software and						computer software	computer crime; computer security; hacker

GAO.xml - validated • **3**

C:\XML\CDT-FG\Development\GAO.xml - Microsoft Internet Explorer

File   Edit   View   Favorites   Tools   Help

Address   C:\XML\CDT-FG\Development\GAO.xml   Go   Links

```xml
<?xml version="1.0" encoding="UTF-8" ?>
- <Collection xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:noNamespaceSchemaLocation="thesaurus.xsd">
  - <Concept>
      <prefLabel>computer</prefLabel>
      <altLabel>computer equipment; computer hardware</altLabel>
      <narrower>embedded computer system; mainframe computer; personal computer; supercomputer</narrower>
      <broader>electronic equipment</broader>
      <related>computer crime; computer equipment contract; computer equipment industry; computer equipment
        management; computer network; information technology</related>
    </Concept>
  - <Concept>
      <prefLabel>computer network</prefLabel>
      <narrower>Internet; intranet; local area network; wide area network</narrower>
      <broader>computer; telecommunication</broader>
      <related>data transmission; computer network protocol; computer bulletin board</related>
    </Concept>
  - <Concept>
      <prefLabel>computer security</prefLabel>
      <altLabel>computer security measures</altLabel>
      <definition>Measures to protect a compute's hardware, software and data.</definition>
      <narrower>computer crimes; computer viruses; data encryption; data integrity; facility security; hackers;
        identity verification; information resources management</narrower>
      <broader>security</broader>
    </Concept>
  - <Concept>
      <prefLabel>computer software</prefLabel>
      <altLabel>computer programs; software</altLabel>
      <definition>Instructions that cause computers to function. Includes data bases and programs.</definition>
      <narrower>application software; computer virus; data base; operating system</narrower>
      <broader>information technology; programming languages</broader>
      <related>computer software documentation; computer software industry; computer software reuse; computer
        software verification and validation; computer-aided software engineering</related>
    </Concept>
  - <Concept>
      <prefLabel>computer virus</prefLabel>
      <definition>computer software that copies itself by attaching to other software and executing unwanted
        operations</definition>
      <broader>computer software</broader>
      <related>computer crime; computer security; hacker</related>
    </Concept>
</Collection>
```

Done   My Computer

SUPPORTING MATERIAL:

Thesaurus-Related ISO Specifications

ISO 2788:1986 [1]

- 3

- "Documentation – Guidelines for the establishment and development of monolingual thesauri"; replaces ISO 2788:1974

- From Technical Committee ISO/TC **46**, **Documentation**

- Guidelines for:
  - Selecting terms for inclusion in *thesaurus*
  - Expressing *relationships* between the selected terms
  - Could serve as our guidelines for term selection and definition concepts

- preferred term – descriptor (main entry point)

- non-preferred term - synonym

ISO 2788:1986 [2]

- 3

ISO 2788:1986 [3]

- **3**

SN Scope note; a note attached to a term to indicate its meaning within an indexing language

USE The term that follows the symbol is the preferred term when a choice between synonyms or quasi-synonyms exists

UF Use for; the term that follows the symbol is a non-preferred synonym or quasi-synonym

TT Top term; the term that follows the symbol is the name of the broadest class to which the specific concept belongs; sometimes used in the alphabetical section of a thesaurus

BT Broader term; the term that follows the symbol represents a concept having a wider meaning

BTG Broader term (generic)

BTP Broader term (partitive)

Judy Newton has offered to create an "executive summary" of ISO 2788.

NT Narrower term; the term that follows the symbol refers to a concept with a more specific meaning

NTG Narrower term (generic)

NTP Narrower term (partitive)

RT Related term; the term that follows the symbol is associated, but is not a synonym, a quasi-synonym, a broader term or a narrower term

ISO 1087-1:2000 [1]

- 3

- **1990: "Vocabulary of terminology"**

- **2000: "TERMINOLOGY WORK — VOCABULARY — Part 1: Theory and application"**

- **Mainly vocabulary (normative)**

- **Concept diagrams (informative)**

ISO 1087-1:2000 [2]

- **3**

abbreviation ....
acronym ....
admitted term ....
alphabetical arrangement ....
alphabetical order (admitted) ....
antonymy ....
appellation ....
associative relation ....
base list ....
blend...
borrowed term ....
broader concept (admitted) ....
causal relation ....
characteristic ....
clipped term ....
complex term ....
comprehensive concept ....
concept ....
concept diagram ....
concept field ....
concept harmonization ....
context ....
concept system ....
coordinate concept ....
corpus ....
country identifier ....
definition ....
delimiting characteristic ....
deprecated term ....
designation ....
designator (admitted) ....
domain (admitted) ....
entry term ....
equivalence ....
essential characteristic ....
extension ....
extensional definition ....
general concept ....
generic concept ....
generic relation ....
genus – species relation (admitted) ....
glossary ....
grammatical label ....
hierarchical relation ....
homonymy ....
individual concept ....
initialism ....
intension ....
intensional definition ....
language for special purposes (admitted) ......

mononymy ....
monosemy ....
name (admitted) ....
narrower concept (admitted) ....
neologism (admitted) ....
neoterm ....
nomenclature ....
note   ....
object   ....
obsolete term ....
partitive concept...
partitive relation ....
part – whole relation (admitted) ....
polysemy ....
pragmatic relation (admitted) ....
preferred term ....
sequential relation ....
simple term ....
source identifier ....
special language ....
specific concept ....
subject field ....
subject label ....
subordinate concept ....
superordinate concept ....
synonymy ....
system of concepts (admitted) ....
systematic arrangement ....
systematic order (admitted) ....
technical dictionary (admitted) ....
temporal relation ....
term   ....
term bank ....
term excerption ....
term harmonization ....
term identification ....
terminography ....
term acceptability rating ....
terminological concordance ....
terminological data ....
terminological database ....
terminological data bank (admitted) ....
terminological dictionary ....
terminological entry ....
terminological format ....
terminologization ....
terminology ....
terminology planning ....
terminology processing ....
terminology science (admitted) ....
terminology work ....
thematic arrangement ....
thematic order (admitted) ....
type of characteristics ....
vocabulary...

ISO 1087-1:2000   [3]

- *Subject field* (domain) – field of special knowledge
- *Concept* – unit of knowledge created by a unique combination of characteristics
- *Characteristic* – abstraction of a property of an object or of a set of objects
- *Extension* – set of objects to which concept corresponds
- *Intension* – set of characteristics which make up the concept

ISO 1087-1:2000   [4]

- **Hierarchical Relation**
  - *Generic Relation:* vehicle and car
  - *Partitive Relation:* week and day
- *Associative Relation:* baking and oven
- *Extensional definition* = enumerating all subordinate concepts under one criterion of subdivision (e.g., noble gases = {helium, neon, argon, crypton, xenon, or radon})

- 4

ISO 1087-1:2000   [5]

- *Terminology work* has 3 types of *Designators*   • 4 (representation of a concept by a sign that denotes it)
  - *Symbol*
  - *Appellation* – verbal designation of individual concept
  - *Term* - verbal designation of a general concept in a specific subject field; may have variants (i.e., alternate spellings)

ISO 1087-1:2000   [6]

- **Kinds of Terms (sample)**   • 4
  - *Simple* – **one root**
  - *Complex* – **two or more roots (e.g., bookmaker, fault tolerance)**
  - *Clipped term* – **abbreviation formed by truncating part of a simple term (e.g., flu for influenza, vet for veterinarian)**
  - *Blend* – **formed by clipping and combining two separate terms (e.g., infomercial = information + commercial)**
  - *Preferred term* – **rated as the primary term for a given concept; usually the *entry term***

ISO 1087-1:2000   [7]

- *Polysemy* ⁻ one designation represents two or more concepts <u>sharing</u> certain characteristics
  - e.g., bridge: structure to carry traffic over a gap; dental plate
- *Homonymy* - one designation represents two or more <u>unrelated</u> concepts
  - e.g., bark: sound made by dog; sailing vessel; outer surface of a tree
  -
- The more common *terminological data* include:
- entry term, definition, note, grammatical label, subject label, language identifier, country identifier, and source identifier.

- 4

ISO 1087-1:2000   [8]

- *Terminological dictionary* - collection of terminological • 4 entries presenting information related to concepts or designations from one or more specific subject fields
- *Vocabulary* - terminological dictionary which contains designations and definitions from one or more specific subject fields
- *Glossary* - terminological dictionary which contains a list of designations from a subject field, together with equivalents in one or more languages [In English common language usage glossary can refer to a unilingual list of designations and definitions in a particular subject field.]

ISO 704:2000 [1]

- "Terminology work — Principles and methods" • 4
- Replaces ISO 704:1987.
- Technical Committee ISO/TC **37**, **Terminology**
- Establishes basic principles and methods for preparing and compiling terminologies.
- Describes the links between objects, concepts, and their representations through the use of terminologies.
- Borrows terms from ISO 1087-1:2000 (i.e., object, concept, characteristic, intension, extension, etc.)

ISO 704:2000 [2]

- *Essential vs. non-essential characteristics*   • 4
  - Graphite is encased in wood?
  - One end may be sharpened to a point?
  - Is it indispensable to understanding a concept?
  - Property may be essential characteristic of a concept in one subject field but non-essential in another.
- *Delimiting characteristics* – essential characteristic that distinguishes one concept from another.
- "When modeling a concept system, one shall concentrate on the essential and delimiting characteristics."

ISO 704:2000 [3]

- *Hierarchical relations* – see ISO 1087 slides   • 4
- *Associative relations* – thematic connection between concepts based on experience
  - Pencil case : pencil :: container : contained
  - Writing : pencil :: activity : tool

ISO 704:2000 [4]

- **Terminology isn't a random collection of terms.• 4**
- **"The *terminology* of a *subject field* is the collection of *designations* attributed to *concepts* making up the knowledge structure of the field."**
- **Concept systems:**
  - **"model concept structures based on specialized knowledge of a field;**
  - **clarify the relations between concepts;**
  - **form the basis for a uniform and standardized terminology;**
  - **facilitate the comparative analysis of concepts and designations across languages;**
  - **facilitate the writing of definitions."**

DCMI Metadata [1]

- Dublin Core Metadata Initiative: http://dublincore.org/
- Terms: http://dublincore.org/documents/dcmi-terms/
- Type vocabulary: http://dublincore.org/documents/dcmi-type-vocabulary/
- Browse Dublin Core Metadata Registry
- ISO 15836:2003(E). Information and documentation — The Dublin Core metadata element set
- Element list from Users Guide: 16 (or 18?)

- 4

Content	Intellectual Property	Instantiation
Coverage	Contributor	Date
Description	Creator	Format
Type	Publisher	Identifier
Relation	Rights	Language
Source		
Subject		
Title		
Audience		

DCMI Metadata [2]

- xmlns:dc="http://purl.org/dc/elements/1.1/"
- 5
- Creator="Internal Revenue Service. Customer Complaints Unit" (a person, an organization, or a service). See also Contributor.
- Date="1998-02-16"
- Relation "is Refined by": conformsTo hasFormat hasPart hasVersion isFormatOf isPartOf isReferencedBy isReplacedBy isRequiredBy isVersionOf references replaces requires
- Identifier – would be desirable if registry could assign this automatically as a UID
- Audience
- Title == Term
- Subject == Context