

A Flexible XML-Based Glossary Approach for the Federal Government: *The Next Generation*

by [Ken Sall](#)

for the DHS Core Data Types Focus Group

February 25, 2005

<http://kensall.com/gov/glossary>

Agenda

- **Candidate Requirements**
- **Original Sall Strawman Approach – Simplicity But...**
- **Relevant ISO Specifications**
 - ISO 2788 (oldest)
 - ISO 1087
 - ISO 704
 - Others?
- **DCMI Metadata Terms**
- **XML Glossary/Thesauri/Lexicon Implementations**
 - GlossXML
 - SALT, MARTIF, OLIF, XLT, CLS....
 - SKOS
- **Recommended Plan of Action**

Candidate Requirements (1)

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

Candidate Requirements (2)

7. It **SHOULD** be relatively easy to combine terms authored by different individuals and different agencies, if desired.
8. The elements in the schema **SHOULD** be chosen with ISO standards in mind, to the degree that this does not overly complicate the schema.
9. 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).
10. It is **DESIRABLE** that the XSLT generate additional search links not in the source.
11. Multiple definitions of the same term **MUST** be permitted, with either same or different context.

Candidate Requirements (3)

12. The entire approach **SHOULD** foster a clean separation of collaborative roles:
 - a) Developer of schema vs. developer of stylesheets
 - b) Author/collector of terms and definitions
 - c) Reviewer/approver of definitions
 - d) Consumer of results (e.g., agency with custom XSLT)
13. It **SHOULD** support semantic relationships between terms including related-to and synonyms.
14. 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”.
15. It **MUST** be possible to indicate a term’s
 - a) Source (agency, author, document, and/or URL)
 - b) Context
 - c) Approval status
 - d) TBD – what else is mandatory?

Candidate Requirements (4)

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

Vote by requirement # to: xml@kensall.com ; subject "glossary". Comments optional.

+ = in favor (desirable)

++ = change SHOULD to MUST (mandatory)

-- = not a requirement

0 = no opinion

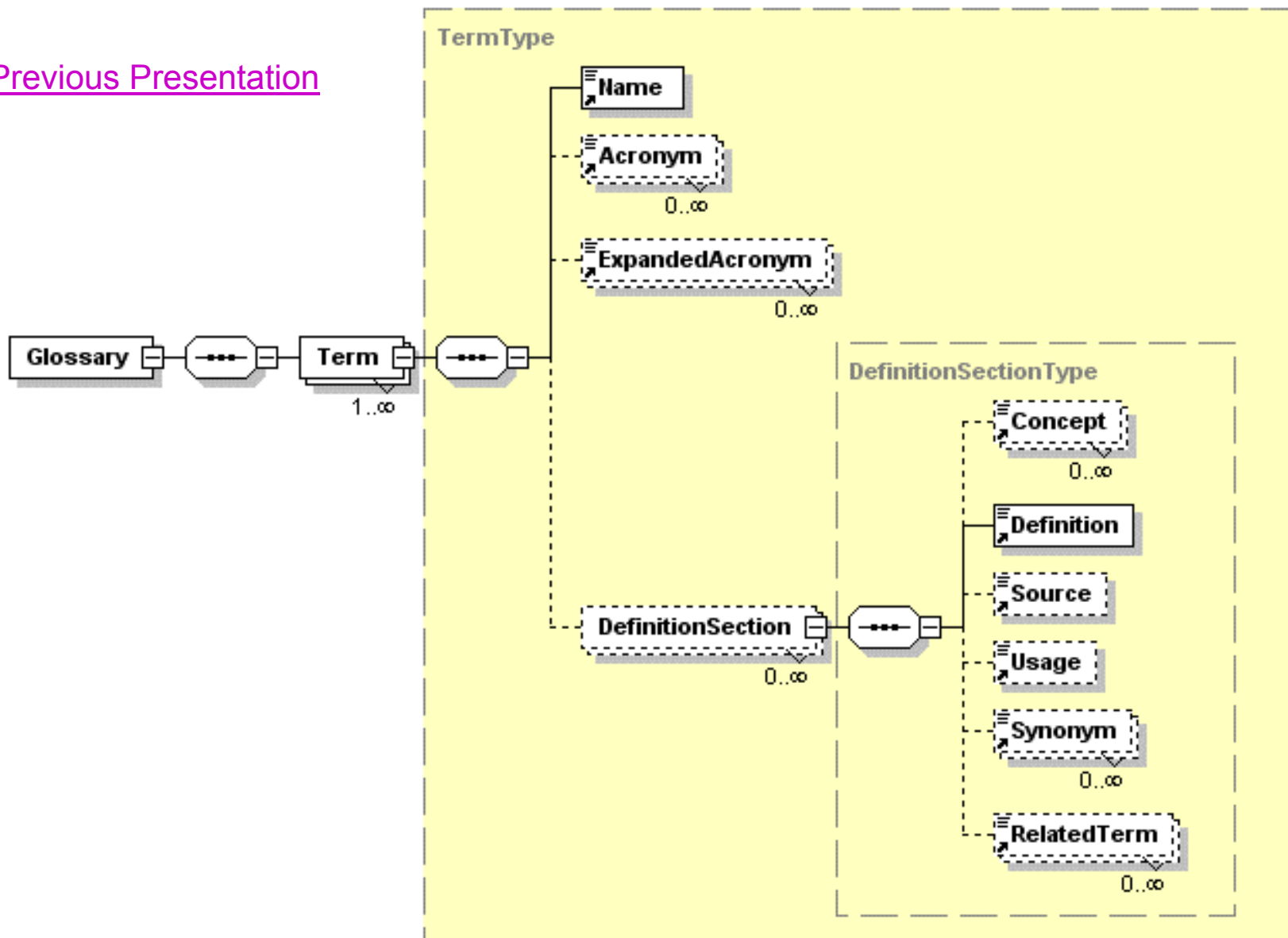
CAF Glossary Considerations

“The following were brainstormed in no particular order and represent a dump, not a culled consensus:

- 1. Need to add terms like transition strategy, transition plan, sequencing plan, and risk management plan.**
- 2. Need some basic rules for how to create the definitions- like how to treat compound worded entities.**
- 3. Use the RMs as taxonomy if necessary not NIST.**
- 4. Need to tell the “to-do” aspect in the term. (i.e. How is it used? What for? Why?)**
- 5. Need clearly defined validation criteria. (Correct, Cohesive, Complete and Context - what do these mean and how evaluate?)**
- 6. Examples could add value in context and understanding.**
- 7. When two terms are similar - either clearly distinguish or make synonyms.**
- 8. Need basic terms first then use to define others..**
- 9. Ontological rules could/should be used/followed.**
- 10. Capture the dialogue on terms first - then vote on validation.**
- 11. Listing currently includes mostly all nouns. We need verbs also to describe a story to business people.**
- 12. Many of the definitions could use Antedotes. Definitions are dry by themselves.**
- 13. Context of OMB official understanding and official linkages should be established up front. (Use legislation and regulation- A-11, A-119, A-130, FEAPMO Reference Models, so forth). (Issue- what to do about poor or incorrect definitions in these?)**
- 14. Multiple presentation views - alphabetical, reference models, taxonometric, other - Dummies Guide**
- 15. How to relate to other disciplines and their terminologies?**
- 16. There are experts that know how to do this...perhaps we should consult or engage with them.**
- 17. Only use the Reference Models to organize the glossary.**
- 18. Some of the words in the glossary are too technical or too much in the weeds. These do not belong in this glossary...diffuses use and purpose. These fit elsewhere - in the RMs for example.”**

Sall's XML Glossary Model Strawman

[Previous Presentation](#)



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>
```

XML Example: XSLT Details

The screenshot shows a Mozilla Firefox browser window titled "Strawman Glossary Example - Mozilla Firefox". The address bar shows the URL <http://kensall.com/gov/glossary/glossary-ex1.xml>. The browser displays a glossary page with two entries. The first entry is for "ontology" and the second is for "philosophy".

ontology

Concept(s): semantic web; knowledge management

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.

Source: Daconta, Obrst, Smith

Usage Example: An ontology can range from the simple notion of a taxonomy to a thesaurus, to a conceptual model, to a logical theory. [Daconta, Obrst, Smith]

Synonym(s): classification system

RelatedTerm(s): taxonomy; OWL

philosophy

Concept(s): philosophy

Definition: [sometimes "Ontology"] the metaphysical study of the nature of being and existence

Source: WordNet

Usage Example: Both the ontology and manner of human existence are of concern to Existentialism.

Synonym(s): metaphysics

Generated Searches for ontology: [Google Define](#); [WordNet](#); [Google Search](#).

Callouts in the image:

- ontology** (points to the term)
- CSS Styling** (points to the styling of the first entry)
- DefinitionSection based on Concept** (points to the definition of "ontology")
- Optional and Repeatable Elements** (points to the "Usage Example" of "ontology")
- New DefinitionSection based on 2nd Concept** (points to the definition of "philosophy")
- Auto-generated Search Links** (points to the search links for "ontology")

Previous Strawman Terminology

Unused ISO 1087 Terms

- Characteristic
- Designation
- Dictionary
- Nomenclature
- Object
- PreferredTerm
- Terminological Dictionary / technical dictionary
- Terminological Record
- Terminological Database
- Terminological Dictionary
- Terminology Work
- Vocabulary

ISO 1087 Terms Used:

- Concept
- Definition
- Term

Used but not ISO 1087:

- Glossary
- Synonym
- RelatedTerm

Additional Terms by Sall (next slide):

- Name
- Acronym
- ExpandedAcronym
- DefinitionSection
- Source
- Usage

Additional {Sall} Terminology

- Glossary – change to **Dictionary**, **Vocabulary**, **Technical Dictionary** or Terminology?
- Name – added only to allow Term to be a container; could change Term to **Entry** and Name to Term?
- Acronym – necessary option for technical terms
- ExpandedAcronym – ditto
- DefinitionSection - added simply as a repeatable container to encompass all aspects pertaining to a specific definition of a term
- Source - useful for traceability and credibility
- Usage – useful to have an optional example sentence for a given definition (use in context)

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

The screenshot shows a Microsoft Internet Explorer window titled "Strawman Glossary Example - Microsoft Internet Explorer". The browser's address bar is empty, and the menu bar includes File, Edit, View, Favorites, Tools, and Help. The toolbar contains icons for Back, Forward, Stop, Home, Search, Favorites, Media, and other standard browser functions. The main content area displays a glossary entry for "Class".

Class

Concept(s):
Definition: A description of a set of objects that share the same attributes, operations, methods, relationships, and semantics
Source: ISO 11179, Metadata Registries (MDR) ? Part 1: Framework, 20 May 2003 (Draft)

Generated Searches: [Google Define](#), [WordNet](#), [Merriam-Webster](#), [Whats](#), [W3C](#), [W3Schools](#), [Webopedia](#), [ZVON](#), [Google Uncle Sam \[gov and mil\]](#), [Google Search](#)

Community of Practice (COI)

Concept(s):
Definition: An affinity group. An informal network or forum where tips are exchanged and ideas generated. A group of professionals informally bound to one another through exposure to a common class of problems, common pursuit of solutions, and thereby themselves embodying a store of knowledge
Source: [Thomas A. Stewart][McKinsey & Co.]

Generated Searches: [Google Define](#), [WordNet](#), [Merriam-Webster](#), [Whats](#), [W3C](#), [W3Schools](#), [Webopedia](#), [ZVON](#), [Google Uncle Sam \[gov and mil\]](#), [Google Search](#)

Conceptual Data Model (CDM)

Concept(s):
Definition: A data model that defines the real world entities, and the relationships between these entities, in a business context. A CDM is typically constructed as an ERD, e.g., UML class diagram or ERwin model.

Generated Searches: [Google Define](#), [WordNet](#), [Merriam-Webster](#), [Whats](#), [W3C](#), [W3Schools](#), [Webopedia](#), [ZVON](#), [Google Uncle Sam \[gov and mil\]](#), [Google Search](#)

Conceptual Data Model [11179]

Concept(s):
Definition: A data model that represents an abstract view of the real world

The search links in the "Generated Searches" sections are circled in pink. The browser's status bar at the bottom shows "My Computer".

Search Links Bootstrap: Based on CAF-index.html {?}

The screenshot shows a Microsoft Internet Explorer browser window titled "Strawman Glossary Example - Microsoft Internet Explorer". The address bar is empty. The menu bar includes File, Edit, View, Favorites, Tools, and Help. The toolbar contains icons for Back, Forward, Stop, Home, Search, Favorites, Media, and other navigation functions. The main content area displays a glossary with the following entries:

- BPEL4WS**
Generated Searches: [Google Define](#); [WordNet](#); [Merriam-Webster](#); [WhatIs](#); [W3C](#); [W3Schools](#); [Webopedia](#); [ZVON](#); [Google Uncle Sam \[.gov and .mil\]](#); [Google Search](#).
- Broadband**
Generated Searches: [Google Define](#); [WordNet](#); [Merriam-Webster](#); [WhatIs](#); [W3C](#); [W3Schools](#); [Webopedia](#); [ZVON](#); [Google Uncle Sam \[.gov and .mil\]](#); [Google Search](#).
- Builder's View (Technology Model)**
Generated Searches: [Google Define](#); [WordNet](#); [Merriam-Webster](#); [WhatIs](#); [W3C](#); [W3Schools](#); [Webopedia](#); [ZVON](#); [Google Uncle Sam \[.gov and .mil\]](#); [Google Search](#).
- Business**
Generated Searches: [Google Define](#); [WordNet](#); [Merriam-Webster](#); [WhatIs](#); [W3C](#); [W3Schools](#); [Webopedia](#); [ZVON](#); [Google Uncle Sam \[.gov and .mil\]](#); [Google Search](#).
- Business Alignment and Assessment**
Generated Searches: [Google Define](#); [WordNet](#); [Merriam-Webster](#); [WhatIs](#); [W3C](#); [W3Schools](#); [Webopedia](#); [ZVON](#); [Google Uncle Sam \[.gov and .mil\]](#); [Google Search](#).

The status bar at the bottom shows "My Computer".

ISO Mania

- ISO 639:1988. Code for the representation of names of languages.
- ISO 690:1987. Bibliographic references - Content, form and structure
- **ISO 704:2000. Terminology work — Principles and methods**
- **ISO 1087-1:2000. Terminology work — Vocabulary — Part 1: Theory and application**
- ISO/1087-2:2000, Terminology work – Vocabulary – Part 2: Computer applications.
- ISO 1951:1973. Lexicographical symbols and typographical conventions for use in terminography
- **ISO 2788:1986. Documentation – Guidelines for the establishment and development of monolingual thesauri**
- ISO 9115:1987. Documentation - Bibliographic identification (biblid) of contributions in serials and books.
- ISO 10241:1992. International terminology standards - Preparation and layout.
- ISO **12200:1999**, Computer applications in terminology – Machine-readable terminology interchange format (**MARTIF**) – Negotiated interchange.
- ISO 12620, Terminology – Computer applications – Data categories.
- **ISO 15836:2003(E). Information and documentation — The Dublin Core metadata element set**

ISO / TC 37 Terminology (principles and coordination)

- From Håvard Hjulstad, June 2000
 - 639 ♦ language codes
 - **704** ♦ principles and methods
 - 860 ♦ harmonization of concepts and terms
 - **1087** ♦ vocabulary of terminology
 - 12199 ♦ alphabetical ordering
 - **12200** ♦ machine-readable terminology interchange format (MARTIF)
 - 12616 ♦ translations-oriented terminography
 - 12618 ♦ creation and use of terminological databases and text corpora
 - **12620** ♦ data categories

ISO 2788:1986 [1]

- “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]

6	Indexing terms
6.1	General
6.2	Forms of terms
6.3	Choice of singular or plural forms
6.4	Homographs or polysemes
6.5	Choice of terms
6.6	Scope notes and definitions
7	Compound terms
7.1	General
7.2	Terms that should be retained as compounds
7.3	Terms that should be syntactically factored
7.4	Order of words in compound terms
8	Basic relationships in a thesaurus
8.1	General
8.2	The equivalence relationship
8.3	The hierarchical relationship
8.4	The associative relationship
9	Display of terms and their relationships
9.1	General
9.2	Alphabetical display
9.3	Systematic display
9.4	Graphic display
10	Management aspects of thesaurus construction
10.1	Methods of compilation

ISO 2788:1986 [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)

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

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

ISO 1087-1:2000 [1]

- **1990: “Vocabulary of terminology”**
- **2000: “TERMINOLOGY WORK — VOCABULARY — Part 1: Theory and application”**
- **Mainly vocabulary (normative)**
- **Concept diagrams (informative)**

ISO 1087-1:2000 [2]

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, krypton, xenon, or radon})

ISO 1087-1:2000 [5]

- ***Terminology work*** has 3 types of ***Designators*** (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)**
 - *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 sharing certain characteristics (e.g., bridge: structure to carry traffic over a gap; dental plate)
- ***Homonymy*** - one designation represents two or more unrelated concepts (e.g., bark: sound made by dog; sailing vessel)
- The more common *terminological data* include:
 - entry term, definition, note, grammatical label, subject label, language identifier, country identifier, and source identifier.

ISO 1087-1:2000 [8]

- ***Terminological dictionary*** - collection of terminological 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”**
- **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***
 - 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
- ***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.
- “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?)

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

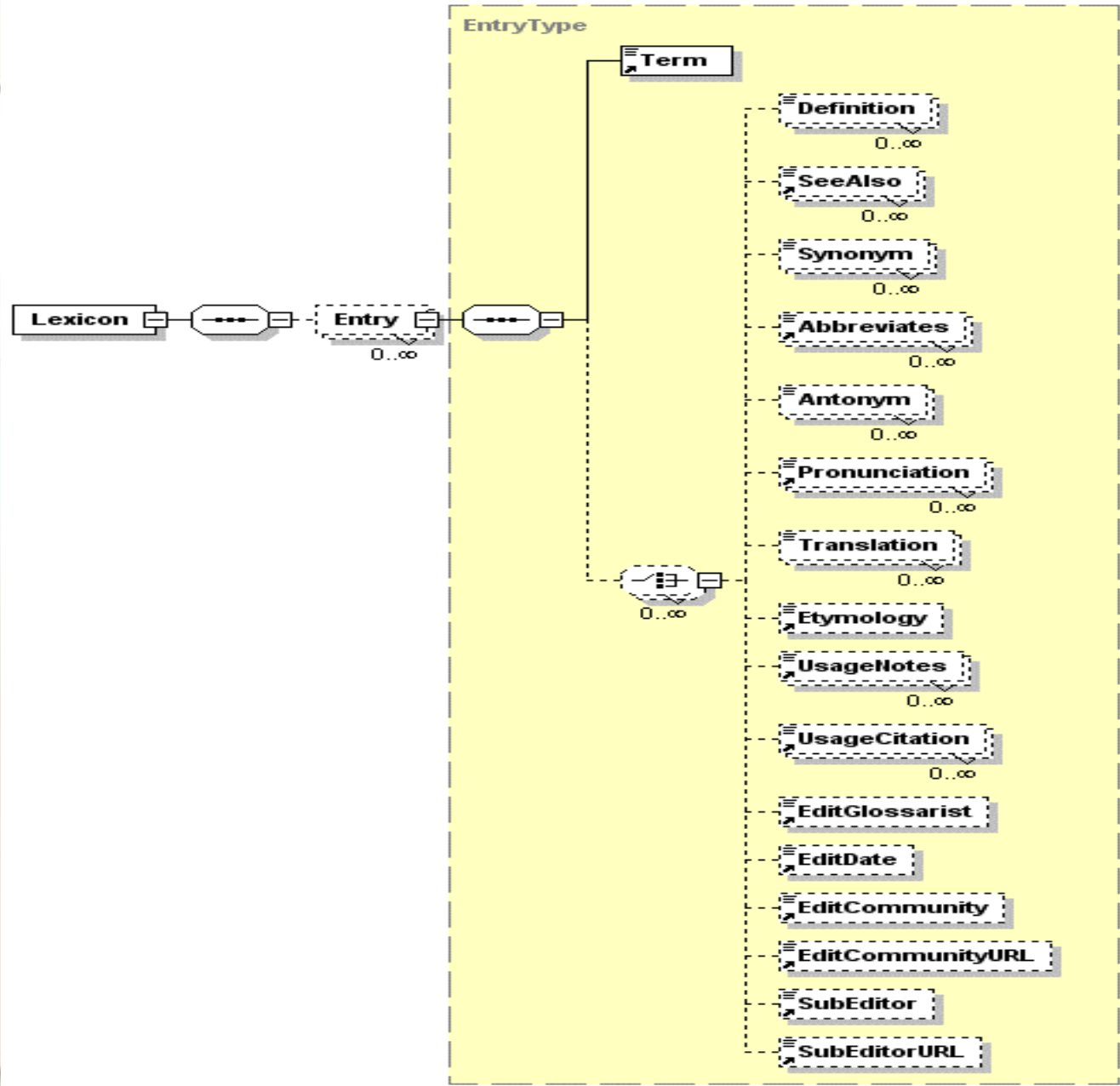
DCMI Metadata [2]

- `xmlns:dc="http://purl.org/dc/elements/1.1/"`
- `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](#) [isFormat Of](#) [isPartOf](#) [isReferencedBy](#) [isReplacedBy](#) [isRequired By](#) [isVersionOf](#) [references](#) [replaces](#) [requires](#)
- `Identifier` – would be desirable if registry could assign this automatically as a UID
- `Audience`
- `Title == Term`
- `Subject == Context`

GlossXML

- **Proposed XML Format for Glossaries**
- **<http://www.creativyst.com/Prod/Glossary/Doc/XMLOut.htm>**
- **GlossXML DTD**
- **Related standards: MARTIF, SALT, CLS, etc.**
- **“In some cases the existing standard was simply too broad or ambitious to be useable for this simple application, since concise interoperability across multiple independent implementations is required and may be diluted in broadly defined standards. In other cases, the functionality may have been defined in platform specific language or targeted toward proprietary systems.”**

GlossXML DTD



OLIF

- **Open Lexicon Interchange Format**
- **<http://www.olif.net/>**
- **“Designed for users of language technology, OLIF is an open, XML-compliant standard that can streamline the exchange of terminological and lexical data. With its flexible design and representative array of terminological and linguistic features, OLIF can help the user address language data management needs ranging from basic terminology exchange to managing lexicons for natural language processing (NLP) systems, such as machine translation.”**
- **[DTD \(2002\)](#) and [description of elements and attributes](#)**
- **Opinion: Powerful but complex; too many elements.**

MARTIF

- **MAchine-Readable Terminology Interchange Format, also known as ISO (FDIS) 12200**
- **MARTIF DTD (1997)**
- **150 data categories are described for MARTIF in ISO (FDIS) 12620**
- **Does not match the needs of non-concept-oriented approaches to terminology, i.e. lexicographic and NLP approaches, because MARTIF presupposes a *concept* orientation rather than a *word* orientation.**
- **See also <http://coral.lili.uni-bielefeld.de/~ttrippel/terminology/terminology.html>**

XLT

- XML representation of Lexicons and Terminologies
- <http://www.ttt.org/oscar/xlt/DXLT.html>
- See also the [CLS Framework](#) (Concept-oriented with Links and Shared references)

SALT

- **Standards-based Access to multilingual Lexicons and Terminologies**
- <http://www.loria.fr/projets/SALT/saltsite.html>
- “The SALT project combines two recently finalized interchange formats:
 - **OLIF** (Open Lexicon Interchange Format), which focuses on the interchange of data among lexbase resources from various machine translation systems, (Thurmaier et al. 1999), and
 - **MARTIF** (ISO 12200:1999, MACHine-Readable Terminology Interchange Format), which facilitates the interchange of termbase resources with conceptual data models ranging from simple to sophisticated.”
- “The goal of SALT is to integrate lexbase and termbase resources into a new kind of database, a lex/term-base called **XLT** (eXchange format for Lex/Term-data). XLT is based on XML (Xtensible Markup Language), which is a data format for structured document interchange on the Web and is under development by the World Wide Web Consortium (XML 1999a).”

SKOS [1]

- **Simple Knowledge Organisation System**
- **W3C Working Draft: 2/10/05; work in progress; subject to backwards incompatible changes!**
- **RDF Schema for thesauri and related knowledge organisation systems**
- **<http://www.w3.org/2004/02/skos/>**
- **“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).”**

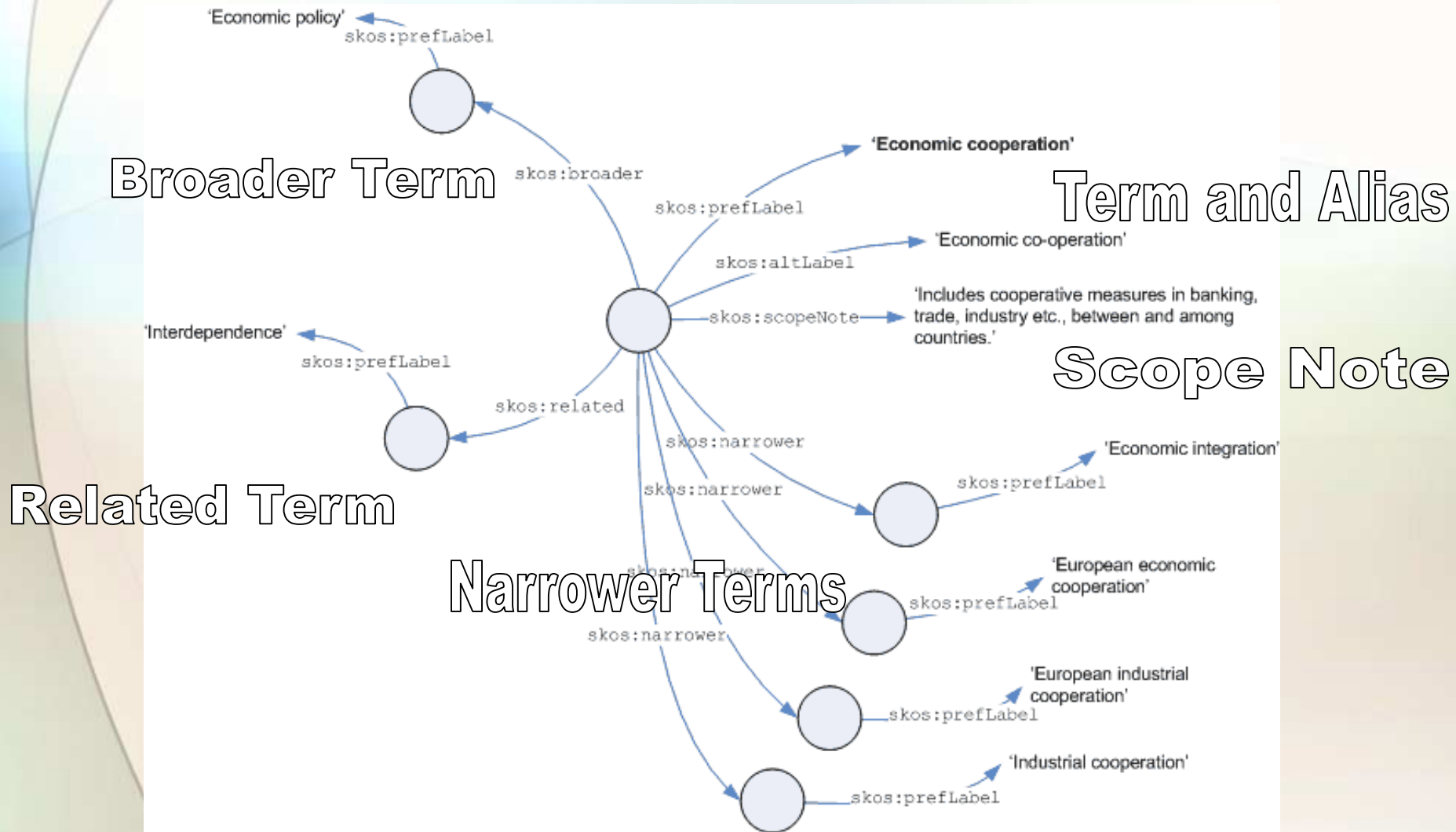
SKOS [2]

- **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
- **[Semantic Web Best Practices and Deployment Working Group](#)**

SKOS [3]

Key Slide

- Quick Guide to Publishing a Thesaurus on the Semantic Web
- **W3C Working Draft in Preparation – 2/8/05**



SKOS: RDF Serialization [4]

```
<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>
```

SKOS: with Thesaurus Metadata (DCMI) [5]

```
<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

SKOS Core Vocabulary [7]

Key Slide

- **Classes**

CollectableProperty

Collection

Concept

ConceptScheme

OrderedCollection

- **Properties**

altLabel

altSymbol

broader

changeNote

definition

editorialNote

example

hasTopConcept

hiddenLabel

historyNote

inScheme

isPrimarySubjectOf

isSubjectOf

member

memberList

narrower

prefLabel

prefSymbol

primarySubject

privateNote

publicNote

related

scopeNote

semanticRelation

subject

subjectIndicator

Subset of SKOS Core Vocabulary [8]

- **Concept** - abstract idea or notion; a unit of thought; 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, historyNote**

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>
```


Next Steps - Revised

- 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 [CORE.gov](#)?
 - Coordinate with [Plans for Derived XML Registry Prototype](#)?
- 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
- **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 short run, but with expansion path for long run.

Phase 1:

- a) Developers: Create schema and distribute/post.
- b) Expert: Distill ISO 2788 to 3-4 page authoring guide.

Phase 2: Authors: Gather terms and definitions.

Phase 3: Reviewers: 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: Developers: Translate schema and Thesaurus to SKOS, after evaluating effort. Can be begun after Phase 1, but need representative set of terms and definitions.