




# Building, Enhancing, and Integrating Taxonomies

ENDORSE

The European Data Conference on Reference Data and Semantics  
17 March 2021

The background of the slide is a light blue-grey color. It features a top-down view of a person's hands using a tablet. The tablet screen displays a colorful Venn diagram with four overlapping circles in green, yellow, red, and blue, each containing a number. To the right of the tablet, there is a complex network diagram consisting of white lines connecting various circular nodes. Some nodes contain icons, such as a bar chart, a gear, and a globe with a 'T' on it. The overall aesthetic is clean and professional, suggesting a focus on data and technology.

**Heather Hedden**  
Data and Knowledge Engineer,  
Semantic Web Company

**Helmut Nagy**  
COO, Semantic Web Company

# About the Speakers

---

## Helmut Nagy

COO

Semantic Web Company

Helmut has been in the field of knowledge management for around 20 years and has been working as senior consultant in lots of projects introducing knowledge graphs to industry and public administration since joining SWC in 2010.

## Heather Hedden

Data and Knowledge Engineer

Semantic Web Company

Heather has over 25 years of experience in developing and managing taxonomies, metadata, and other knowledge organization systems for various organizations and applications. She provides training in taxonomy creation and is author of the book *The Accidental Taxonomist*.

**Semantic Web Company** is a leading provider of graph-based metadata, search, and analytic solutions and the developer of PoolParty Semantic Suite software.

## Part 1: Introduction to taxonomies and other knowledge organisation systems (KOSs)

- ▶ Types of KOSs
- ▶ Standards for KOSs
- ▶ Comparisons of KOSs
- ▶ Uses of KOSs
- ▶ Combinations of KOSs

## Part 2: How to start building or enhancing a taxonomy or other KOS

- ▶ Gathering users' needs and input
  - ▷ Focus example: card sorting
- ▶ Gathering terms from the content

## Part 3: Integrating existing taxonomies

- ▶ Linking taxonomies
- ▶ Mapping taxonomies
- ▶ Merging taxonomies

**Part 1:**  
**Introduction to taxonomies  
and other knowledge  
organisation systems**

**Types**  
**Standards**  
**Comparisons**  
**Uses**  
**Combinations**

## Knowledge organisation system (KOS)

- ▶ Any system of terms, terminology, classification, etc.
- ▶ to organise, define, manage, and/or retrieve information.
- ▶ Not any method to organise knowledge directly, but rather a *scheme* to organise concepts for organising, classifying, defining, tagging, or retrieving information.
- ▶ Broader, includes more than just “controlled vocabularies”

KOS types:

term lists  
synonym rings  
name authorities  
taxonomies  
thesauri  
glossaries  
dictionaries  
gazetteers  
terminologies  
categorisation schemes  
classification systems  
subject heading schemes  
semantic networks  
ontologies



Controlled  
Vocabularies  
for information  
retrieval

## Common types of controlled vocabularies

- ▶ Term list
- ▶ Synonym ring
- ▶ Name authority
- ▶ Taxonomy
  - ▷ Hierarchical taxonomy
  - ▷ Faceted taxonomy
- ▶ Thesaurus

“Taxonomy” sometimes means any controlled vocabulary.

## Other common kind of knowledge organisation system

- ▶ Ontology

# Knowledge Organisation System Types

## Term List

- ▶ A simple list of terms
- ▶ Usually alphabetical, but could be in other logical order
- ▶ Lacking synonyms, it is usually short enough for quick browsing
- ▶ Can appear in drop-down scroll boxes
- ▶ May be used for various metadata values, facets, concept schemes
- ▶ Part of a larger set of controlled vocabularies; part of a KOS



Country of publication



Language

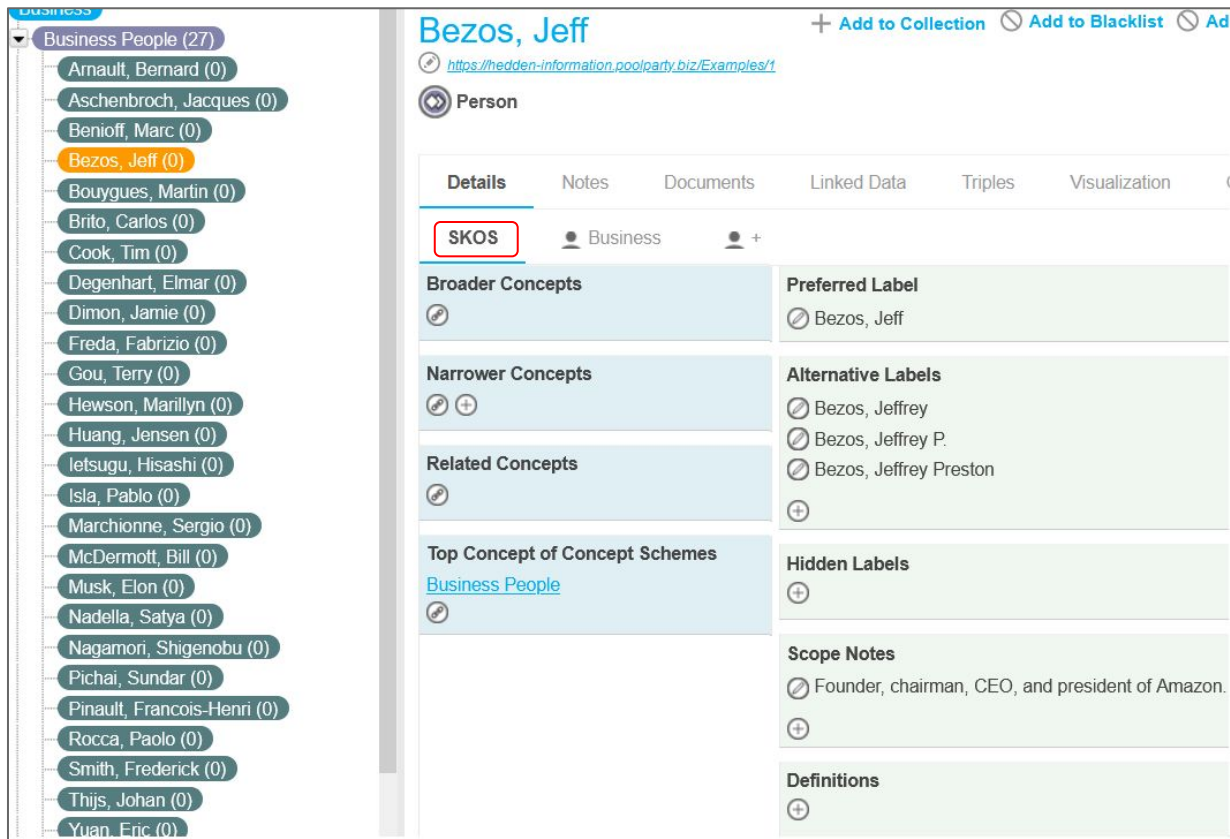


Format

# Knowledge Organisation System Types

## Name authority

- ▶ For named entities, concrete entities, proper nouns
- ▶ A controlled vocabulary with preferred names and variant/alternative names.
- ▶ May or may not have hierarchical relationships between named entities.
- ▶ Usually has additional information/attributes (metadata) for each named entity.



The screenshot displays a Knowledge Organisation System (KOS) interface. On the left, a list of 'Business People (27)' is shown, with 'Bezos, Jeff (0)' highlighted. On the right, the detailed view for 'Bezos, Jeff' is shown, including a URL, a 'Person' type, and various concept relationships.

**Business People (27)**

- Arnault, Bernard (0)
- Aschenbroch, Jacques (0)
- Benioff, Marc (0)
- Bezos, Jeff (0)**
- Bouygues, Martin (0)
- Brito, Carlos (0)
- Cook, Tim (0)
- Degenhart, Elmar (0)
- Dimon, Jamie (0)
- Freda, Fabrizio (0)
- Gou, Terry (0)
- Hewson, Marilyn (0)
- Huang, Jensen (0)
- Ietsugu, Hisashi (0)
- Isla, Pablo (0)
- Marchionne, Sergio (0)
- McDermott, Bill (0)
- Musk, Elon (0)
- Nadella, Satya (0)
- Nagamori, Shigenobu (0)
- Pichai, Sundar (0)
- Pinault, Francois-Henri (0)
- Rocca, Paolo (0)
- Smith, Frederick (0)
- Thijs, Johan (0)
- Yuan, Eric (0)

**Bezos, Jeff** + Add to Collection ⊗ Add to Blacklist ⊗ Ad

<https://hidden-information.poolparty.biz/Examples/>

**Person**

**Details** Notes Documents Linked Data Triples Visualization

**SKOS** Business +

**Broader Concepts**

- [Business People](#)

**Narrower Concepts**

- 

**Related Concepts**

- 

**Top Concept of Concept Schemes**

- [Business People](#)

**Preferred Label**

- Bezos, Jeff

**Alternative Labels**

- Bezos, Jeffrey
- Bezos, Jeffrey P.
- Bezos, Jeffrey Preston
- 

**Hidden Labels**

- 

**Scope Notes**

- Founder, chairman, CEO, and president of Amazon.
- 

**Definitions**

-



# Knowledge Organisation System Types

## Name authority

- ▶ The SKOS model can be extended by custom ontologies to support extended attributes, often desired for named entities.

The screenshot displays a knowledge organization system interface. On the left, a sidebar shows a list of 'Business People (27)' with names in rounded rectangles, including Arnault, Bernard (0), Aschenbroch, Jacques (0), Benioff, Marc (0), **Bezos, Jeff (0)**, Bouygues, Martin (0), Brito, Carlos (0), Cook, Tim (0), Degenhart, Elmar (0), Dimon, Jamie (0), Freda, Fabrizio (0), Gou, Terry (0), Hewson, Marilyn (0), Huang, Jensen (0), Ietsugu, Hisashi (0), Isla, Pablo (0), Marchionne, Sergio (0), McDermott, Bill (0), Musk, Elon (0), Nadella, Satya (0), Nagamori, Shigenobu (0), Pichai, Sundar (0), Pinault, Francois-Henri (0), Rocca, Paolo (0), Smith, Frederick (0), Thijs, Johan (0), Yuan, Eric (0), and Zuckerberg, Mark (0). The main area shows the profile for 'Bezos, Jeff' with a URL, a 'Person' icon, and tabs for 'Details', 'Notes', 'Documents', 'Linked Data', 'Triples', and 'Visualization'. The 'SKOS' tab is active, showing a 'Business' ontology and 'No relations defined'. The right panel lists attributes: Birthdate (12.01.1964), Birthplace (Albuquerque, New Mexico, U.S.A.), Education degree (Princeton University (BSE)), Job title (Founder, CEO, President, and Chairman), and Nationality.

# Knowledge Organisation System Types

## Taxonomy

- ▶ A KOS with broader/narrower relationships that includes all concepts to create a hierarchical structure.
- ▶ Has a focus on categorising and organising concepts.
- ▶ May or may not have “synonyms” to point to the correct, preferred terms/labels.
- ▶ May comprise several hierarchies, concept schemes, or facets.
- ▶ (A facet can be considered as a hierarchy.)
- ▶ “Taxonomy” sometimes refers to any kind of controlled vocabulary (term list, authority file, classification scheme, thesaurus, etc.)

Leisure and culture

- . Arts and entertainment venues
  - . Museums and galleries
- . Children's activities
- . Culture and creativity
  - . Architecture
  - . Crafts
  - . Heritage
  - . Literature
  - . Music
  - . Performing arts
  - . Visual arts
- . Entertainment and events
- . Gambling and lotteries
- . Hobbies and interests
- . Parks and gardens
- . Sports and recreation
  - . Team sports
    - . Cricket
    - . Football
    - . Rugby
  - . Water sports
  - . Winter sports
- . Sports and recreation facilities
- . Tourism
  - . Passports and visas
- . Young people's activities

Hierarchical taxonomy

### Career Level

Student  
Entry Level  
Experienced  
Manager  
Director  
Executive

Faceted  
taxonomy

### Function

Customer Service & Support  
Delivery  
Engineering  
Finance  
General Management  
Legal & Regulatory Affairs  
Marketing & Advertising  
[more]

### Industry

Agriculture  
Apparel & Fashion  
Automotive  
Aviation & Aerospace  
Banking  
Biotechnology  
Broadcast Media  
Chemicals  
[more]

# Knowledge Organisation System Types

## Hierarchical taxonomy

Concepts have broader-concept and/or narrower-concept relationships to other concepts.

The screenshot displays a knowledge organization system interface. On the left, a hierarchical taxonomy of recipes is shown. The 'Desserts (4)' category is highlighted with a red box, and its sub-category 'Cakes (4)' is highlighted with an orange box. Under 'Cakes (4)', four sub-concepts are listed: 'Cheese cakes (0)', 'Chocolate cakes (0)', 'Fruit cakes (0)', and 'Layer cakes (0)', which are enclosed in a purple box. On the right, the 'Cakes' concept is detailed. It includes a URL, action buttons for 'Add to Collection', 'Add to Blacklist', 'Add to ExactMatch', and 'Delete Concept', and a 'Dish' icon. Below this, there are tabs for 'Details', 'Notes', 'Documents', 'Linked Data', 'Triples', and 'Visualization'. The 'Details' tab is active, showing 'Quality Management' and 'History' sections. The 'SKOS' section is expanded, showing 'Recipe-Scheme' and a '+' button. The 'Broader Concepts' section is highlighted with a red box and contains the link 'Desserts'. The 'Narrower Concepts' section is highlighted with a purple box and contains links for 'Cheese cakes', 'Layer cakes', 'Chocolate cakes', and 'Fruit cakes'. To the right of these sections are 'Preferred Label' (with 'Cakes' selected and an 'en' language indicator), 'Alternative Labels', and 'Hidden Labels' sections, each with a '+' button.

## Thesaurus

- ▶ A controlled vocabulary that has standard structured relationships between “terms” (concepts)
  - ▷ **Hierarchical**: broader term/narrower term (BT/NT)
  - ▷ **Associative**: related terms (RT)
  - ▷ **Equivalence**: preferred term (“use for” or “used for”)/ non-preferred term (use) (USE/UF)
- ▶ Created in accordance with standards:
  - ▷ **ISO 25964** (2011, 2013) *Thesauri and Interoperability with Other Vocabularies*
  - ▷ **ANSI/NISO Z39.19** (2005, renewed 2010) *Guidelines for Construction, Format, and Management of Monolingual Controlled Vocabularies*  
[www.niso.org/publications/ansiniso-z3919-2005-r2010](http://www.niso.org/publications/ansiniso-z3919-2005-r2010)
- ▶ “Thesaurus” is usually the kind of controlled vocabulary used in indexed articles databases, such as accessed through libraries.

materials	acquisitions
UF	acquisitions (of materials) library acquisitions
BT	collection development
NT	accessions approval plans gifts and exchanges materials claims materials orders subscriptions
RT	book vendors jobbers subscription agencies subscription cancellations

ASIS&T thesaurus

# Knowledge Organisation System Types

## Thesaurus

ANSI/NISO or ISO  
thesaurus model  
and SKOS model  
compared

**Church music**

**Thesaurus model**

- UF** [Pastoral music \(Sacred\) \(Subjects\)](#)
- UF** [Sacred music \(Subjects\)](#)
- BT** [Religious music \(Subjects\)](#)
- NT** [Mass \(Music\) \(Subjects\)](#)
- NT** [Oratorios \(Subjects\)](#)
- NT** [Requiems \(Subjects\)](#)
- NT** [Sacred vocal music \(Subjects\)](#)
- RT** [Carillons \(Subjects\)](#)
- RT** [Choirs \(Music\) \(Subjects\)](#)
- RT** [Christmas music \(Subjects\)](#)
- RT** [Church \(Subjects\)](#)
- RT** [Church musicians \(Subjects\)](#)
- RT** [Classical music \(Subjects\)](#)
- RT** [Contemporary Christian music \(Subjects\)](#)
- RT** [Devotional exercises \(Subjects\)](#)
- RT** [Easter music \(Subjects\)](#)
- RT** [Liturgics \(Subjects\)](#)
- RT** [Organ music \(Subjects\)](#)

Gale Subject Thesaurus

**Church music**

**SKOS model**

<https://hedden-information.poolparty.biz/Examples/29>

Details Notes Documents Linked Data Triples Visualization

Quality Management History

**SKOS**

**Broader Concepts**

- [Religious music](#)

**Narrower Concepts**

- [Mass \(music\)](#)
- [Oratorios](#)
- [Requiems](#)
- [Sacred vocal music](#)

**Related Concepts**

- [Carillons](#)
- [Choirs \(Music\)](#)
- [Christmas music](#)
- [Church](#)
- [Church musicians](#)

**Preferred Label**

Church music en

**Alternative Labels**

- Pastoral music (Sacred) en
- Sacred music

**Hidden Labels**

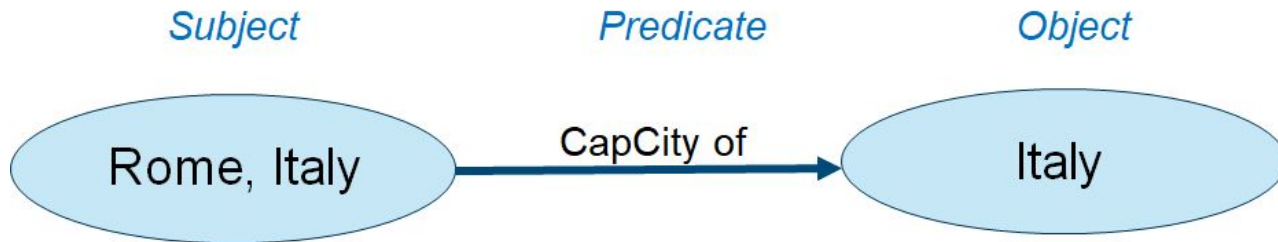
**Scope Notes**

**Definitions**

## Ontology



- ▶ The most complex or semantically rich kind of KOS.
- ▶ A more abstract layer in describing a KOS (taxonomy, thesaurus, etc.)
- ▶ A formal naming and definition of the types, properties and interrelationships of entities in a particular domain.
- ▶ Relations contain meaning, are “semantic.”
- ▶ Common standards provided by W3C: Web Ontology Language (OWL) and RDF-Schema.
- ▶ Comprises classes, relations, and attributes, which are linked in triples.





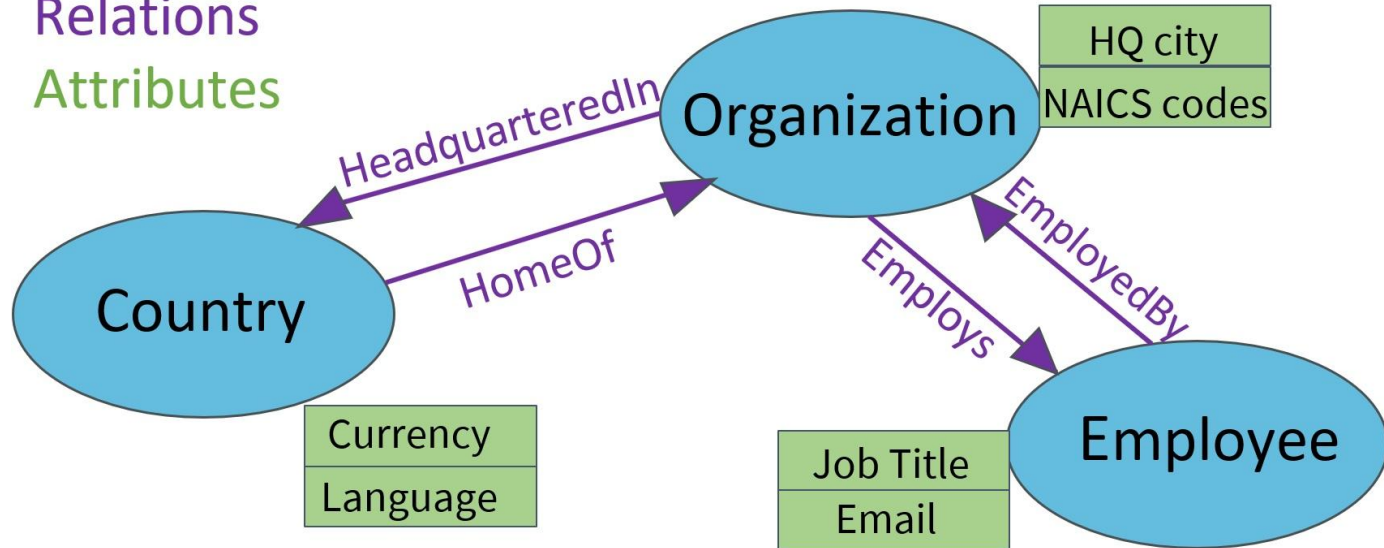
# Knowledge Organisation System Types

## Ontology

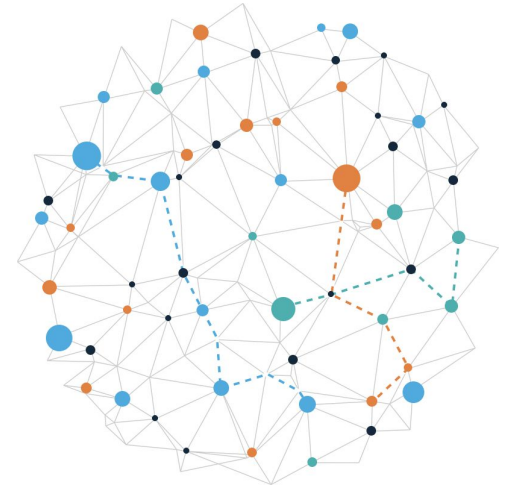
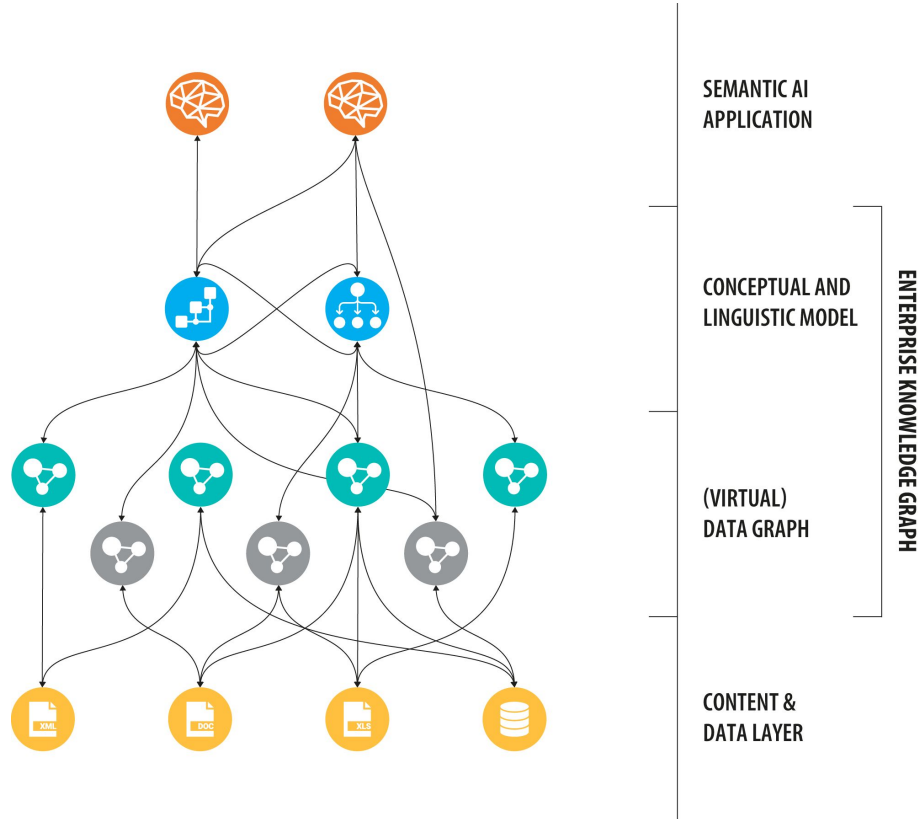
Classes

Relations

Attributes



# What is an Enterprise Knowledge Graph (EKG)?



An Enterprise Knowledge Graph (EKG) contains business objects and topics that are closely linked, classified, semantically enriched, and connected to existing data and documents.



## Types of standards:

1. **Standards for design**
  - Supports an expected experience and results by varied users without training.
2. **Standards for specifications** (measurements, protocols, coding, etc.)
  - Supports exchange and interoperability.

## Standards for knowledge organisation systems of each type:

1. **Standards for design**

ISO 25964 (2011 and 2013) *Thesauri and Interoperability with Other Vocabularies*  
ANSI/NISO Z39.19-2005 *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies* [www.niso.org/publications/ansiniso-z3919-2005-r2010](http://www.niso.org/publications/ansiniso-z3919-2005-r2010)
2. **Standards for specifications and interoperability**

Dublin Core, MARC, ZThes, DD 8723-5, SKOS, RDF, RDFS, and OWL

## ISO 25964 and ANSI/NISO Z39.19

- ▶ Does not have to be machine-readable.
- ▶ Standards first published in 1974. Thesauri have existed since the 1960s.
- ▶ Principles can be followed also within a SKOS model.

## Examples from guidelines

- ▶ Concepts are things: nouns or noun phrases.
- ▶ No duplicates: Concept labels must be unique.
- ▶ No relationship clashes: A pair of concepts can be either hierarchically or associatively related to each other, but not both.
- ▶ No circular relationships: hierarchical relationship logic extends:
  - ▷ Concept A is narrower to Concept B, and
  - ▷ Concept B is narrower to Concept C,
  - ▷ Concept C cannot be narrower to Concept A.



## SKOS (Simple Knowledge Organisation System)

- ▶ A data model to represent knowledge organisation systems.
- ▶ A World Wide Web (W3C) recommendation (initial version 2004 - revised 2009)
- ▶ “A common data model for sharing and linking knowledge organisation systems via the Web”  
<https://www.w3.org/TR/skos-reference/>
- ▶ A KOS built on SKOS is machine-readable and interchangeable.
- ▶ Encoded using XML and RDF (Resource Description Framework).
- ▶ To enable easy publication and use of such vocabularies as linked data.
- ▶ Different KOS types (name authority, thesaurus, taxonomy, ontology) can all be built on the SKOS standard (although ontologies are usually based on the OWL standard instead).



## SKOS principles

- ▶ A KOS is a group of **concepts** identified with URIs and
- ▶ Concepts can be grouped hierarchically into a **concept scheme**.
- ▶ Concepts can be grouped into **collections**, which can be labeled and/or ordered.
- ▶ Concept can be labeled with any number of lexical strings (**labels**) in any natural language.
- ▶ Concepts can have one **prefLabel** in any natural language.
- ▶ Concepts can be documented with **notes** of various types: scope notes, definitions, editorial notes, etc.
- ▶ Concepts can be linked to each other using hierarchical and associative semantic **relations**.
- ▶ Concepts of different concept schemes can be mapped using four basic types of **mapping relations**.



## SKOS elements

Concept Scheme & Collection	Concepts	Labels & Notation	Documentation	Semantic Relations	Mapping Relations
ConceptScheme	Concept	prefLabel	scopeNote	broader	exactMatch
inScheme	hasTopConcept	altLabel	definition	narrower	closeMatch
Collection	topConceptOf	hiddenLabel	example	related	broaderMatch
orderedCollection		notation	changeNote		narrowerMatch
member			editorialNote		relatedMatch
memberList			historyNote		

## RDF (Resource Description Framework)

- ▶ A World Wide Web (W3C) recommendation [www.w3.org/TR/rdf11-concepts](http://www.w3.org/TR/rdf11-concepts)
- ▶ “A standard model for data interchange on the Web”
- ▶ Requires the use of URIs to specify things and to specify relations.
- ▶ Models information as **subject – predicate – object** triples.



## RDFS (RDF-Schema)

- ▶ A W3C recommendation [www.w3.org/2001/sw/wiki/RDFS](http://www.w3.org/2001/sw/wiki/RDFS)
- ▶ Published as part of the RDF Specification Suite Recommendations in 2004.
- ▶ “A general-purpose language for representing simple RDF vocabularies on the Web”
- ▶ Goes beyond RDF to designate classes and properties of RDF resources.

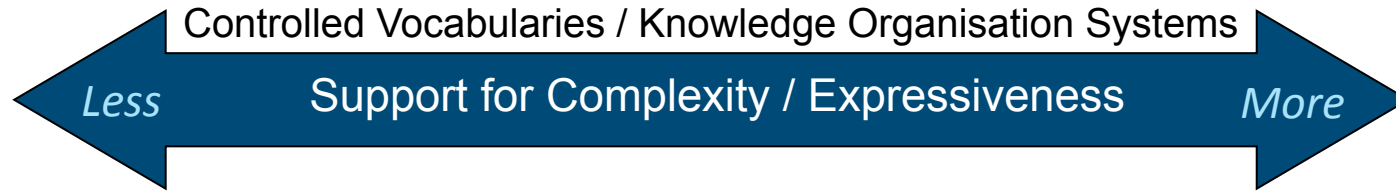


## OWL (Web Ontology Language)

- ▶ A W3C specification [www.w3.org/OWL](http://www.w3.org/OWL)
- ▶ “A Semantic Web language designed to represent rich and complex knowledge about things, groups of things, and relations between things”
- ▶ Based on RDF and RDFS; OWL is W3C’s attempt to extend RDFS.



# Comparison of Knowledge Organisation Systems



Term List	Name Authority	Taxonomy	Thesaurus	Ontology
Ambiguity control	Ambiguity control Synonym control (Attributes)	Ambiguity control (Synonym control) Hierarchical relationships	Ambiguity control Synonym control Hierarchical relationships Associative relationships	Semantic relationships Classes Attributes

## Taxonomies

---

- > All concepts belong to a limited number of major hierarchies (or facets).
- > Support classification, categorisation, concept organisation. (Like Linnaean taxonomy)
- > Do not strictly follow ISO standards.
- > Approach is a top-down, drilling down browse navigation.
- > Especially serving end-users when browsing, non-expert users who benefit from guidance.
- > For a subject area with a defined scope.
- > For relatively small collections of concepts (100s).

## Thesauri

---

- > All concepts have relationships, but “hierarchies” may be as few as 2 terms.
- > Support concept scoping, disambiguation, and relationships with similar concepts. (Like Roget’s)
- > Follow ISO thesaurus standards.
- > Approach is term-centered and what terms are linked to/from it.
- > Especially serving indexers/indexing and users who are subject matter experts, looking for specifics.
- > For a broad, undefined, or unlimited topic area.
- > For large or constantly growing vocabulary.



## Taxonomies are suited for

- ▶ Content and concepts that can naturally be hierarchically categorised
- ▶ A subject area with defined scope and limits
- ▶ Categories and subcategories
- ▶ Non-expert users, who benefit from guidance of hierarchies
- ▶ Relatively small collections of concepts (10s, 100s)
- ▶ Browsing, filtering, sorting

## Thesauri are suited for

- ▶ Concepts that are not easily categorised into hierarchies or facets
- ▶ Multiple, overlapping subject areas
- ▶ Highly specific concepts for detailed indexing
- ▶ Subject-matter experts and those who likely look for specific concepts
- ▶ Vocabulary that is large and/or constantly growing
- ▶ Searching and retrieving

## Ontologies are suited for

- ▶ Concepts in a domain of knowledge, but not necessarily in hierarchies
- ▶ A defined domain with many aspects
- ▶ Broad or specific concepts for knowledge modeling
- ▶ Both expert and non-expert end-users
- ▶ Controlled vocabularies of any size
- ▶ Discovery, recommendation

## “Taxonomy”

Any kind of controlled vocabulary, in a/an...

- ▶ enterprise, corporate setting
- ▶ content management system
- ▶ website navigation (e.g. ecommerce site)

## “Thesaurus”

Any kind of controlled vocabulary...

- ▶ for indexing articles / literature retrieval databases
- ▶ used by librarians, indexers, or other information professionals
- ▶ that includes synonyms/alternative labels (synonym rings)

## “Ontology”

Any kind of controlled vocabulary...

- ▶ with customised, semantic relationships

## Benefits of taxonomies/controlled vocabularies

### 1. Controlled vocabulary



Brings together different wordings (synonyms) for the same concept

- ▷ Helps people search for information by different names

### 2. Classification and structure



Organises information into a logical structure

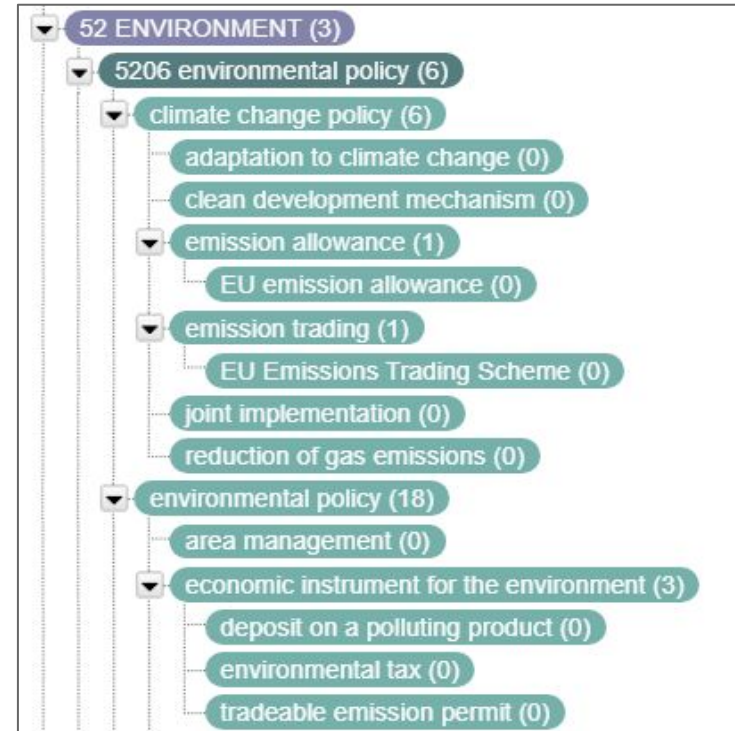
- ▷ Helps people browse or navigate for information
- ▷ Provides context and meaning for concepts for indexing and retrieval

## Multiple purposes and uses of controlled vocabularies

- ▶ Consistent tagging/indexing
- ▶ Topic/category browsing
- ▶ Search (matching search strings to concepts)
- ▶ Discovery (related concept links, or content sharing the same concepts)
- ▶ Filtering results
- ▶ Sorting results
- ▶ Content management workflow (rights, audience, retention, etc.)
- ▶ Consistent metadata for identification, comparison, analysis
- ▶ Visualisation of topics (importance and/or relations)
- ▶ Curated content in feeds or info boxes
- ▶ Automatic linking of relevant topics for personalisation or recommendation systems

## Hierarchical taxonomy purposes

1. Serving users who are browsing, exploring, discovering, not searching, to whom the hierarchy is displayed.
2. Instructing users on appropriate classification
3. Providing context to terms for manual indexers/taggers so that they apply the correct term.
4. Providing the context of a broader concept and thus meaning to aid in auto-classification.
5. Enabling “recursive”/“rolled up” retrieval results (A term retrieves what is indexed to it and what is indexed to each on of its narrower terms, all together.)



Eurovoc Thesaurus excerpt

# Uses of Knowledge Organisation Systems

## Faceted taxonomy purposes

- ▶ Ensures comprehensive tagging and comprehensive search/retrieval by multiple different aspects/vocabulary types.
- ▶ Supports filtering search results by different aspects/vocabulary types.
- ▶ Provides guided Boolean “AND” searching upon a combination of terms in different facets.
- ▶ Allows users to control the search refinement, narrowing or broadening in any manner or order.

Suitable for content of a similar type that shares the same facets.

Examples: all research literature, all internal policies & procedures, all person profiles, all media (image/video) files

### Content Manager

- TechNet Library (7)
- OfficeOnlineVNext (2)
- MSDN Code Gallery (1)
- MSDN Library (1)
- MSN Video (1)

Other Value

Apply | Clear

### Distribution Channel

- Published (12)

Other Value

Apply | Clear

### Request status

- HVC library content item (0)

Other Value

Apply | Clear

### Content Type

- HVC library content item (0)



### Department

- Research
- Sales
- Finance
- Marketing
- Exec Office
- SHOW MORE

### Job Title

- Communications Director
- Developer
- Financial Controller
- Finance Assistant
- Head of IT Services
- SHOW MORE

### Office Location

- London
- New York
- Shanghai

# Uses of Knowledge Organisation Systems

## Thesaurus purpose

- ▶ Support for manual indexing

Manual indexing  
user interface example

Cengage/Gale Subject Thesaurus  
Internal indexer alphabetical browse view

Alpha Browsing - Internet Explorer provided by Cengage Learning

http://gpms.cengage.com/Per\_iMania/brAbrowse.jsp?termType=D&termValue=survey&browse=A&brc

D - Subject Descriptors

D educational v

all xRef  from xRef  no xRef

**Educational videos**

- BROADER TERM:** Audiovisual materials
- BROADER TERM:** Video recordings
- NARROWER TERM:** Instructional videos
- SEE ALSO:** Children's videos
- SEE ALSO:** Educational programs (Mass media)
- SEE ALSO:** Edutainment
- SEEN FROM:** Classroom videos
- SEEN FROM:** Video tapes in education

**Educational volunteers** -

- SEE:** Education volunteers

**Educational vouchers**

- SEE ALSO:** Education costs
- SEE ALSO:** Educational finance
- SEE ALSO:** Educational programs
- SEE ALSO:** Student financial aid
- SEE ALSO:** Bush v. Holmes 919 So. 2d 392 (Fla. 2006)
- SEEN FROM:** Education vouchers
- SEEN FROM:** School vouchers

**Educational works** -

- SEE:** Textbooks

**Educationally disadvantaged**

- BROADER TERM:** Disadvantaged persons

Which kind of KOS is more suitable in each case?

- ▶ Enterprise/intranet search → faceted taxonomy
- ▶ A government agency public website → hierarchical taxonomy
- ▶ A digital asset management system → faceted taxonomy
- ▶ A repository of published research articles → thesaurus + name authority
- ▶ An database of researchers and projects → ontology + thesaurus + name authority



An organisation may use multiple KOSs, of different types, for a single domain of content.

A single document or content item may be tagged from multiple KOSs.

- ▶ **Term Lists** – for a short list of values
  - ▷ For example: Document type, Source, Audience, Language, Rights, Phase
- ▶ **Name authorities** – for named entities described
  - ▷ For example: Names of organizations, agencies, departments, companies, people, places, laws/statutes/treaties, events, products/services
- ▶ **Thesaurus or taxonomy** – for detailed subjects or topics
  - ▷ For example: a single large subject thesaurus, or taxonomy with hierarchies for subject disciplines

In SKOS, each KOS may be a separate Concept Scheme of the same project.

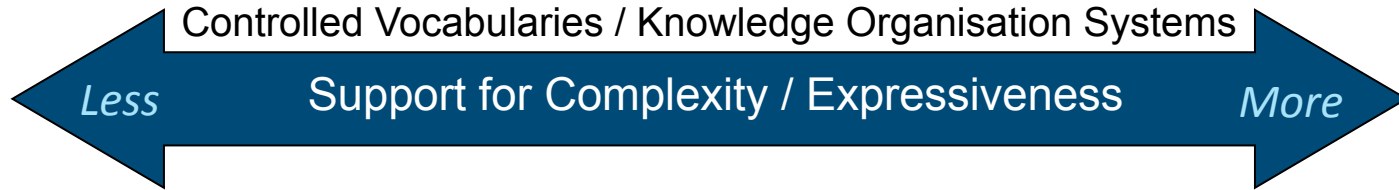
An organisation may use multiple KOS types for a single domain of content.

- ▶ **Ontologies** – to model classes, shared attributes, and semantic relationships across different classes.
  - ▷ Ontologies, by their nature, link to other specific vocabularies, such as term lists, named entity files, taxonomies and thesauri.
  - ▷ Creating customised semantic relationships between different controlled vocabulary types (different concepts schemes) is a simple form of (first step toward) an ontology.

In SKOS, these do not need to be separate knowledge organisation systems, but just separate Concept Schemes.

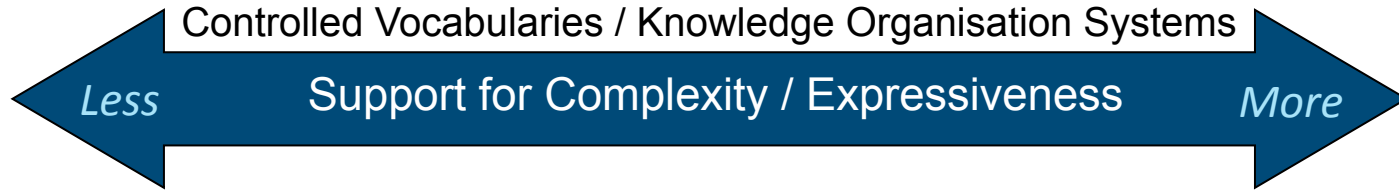
KOS or Scheme (from)	Relation	KOS or scheme (to)
Law Legislative body	issuedBy issues	Legislative body Law
Law Thesaurus subject	dealsWith isSubjectOf	Thesaurus subject Law

# Combinations of KOS Types



<b>Term List</b>	<b>Name Authority</b>	<b>Taxonomy</b>	<b>Thesaurus</b>	<b>Ontology</b>
Ambiguity control	Ambiguity control	Ambiguity control	Ambiguity control	Semantic relationships
	Synonym control	(Synonym control)	Synonym control	Classes
	(Attributes)	Hierarchical relationships	Hierarchical relationships	Attributes
			Associative relationships	

# Combinations of KOS Types



Ontology			
Term List	Name Authority	Taxonomy	Thesaurus
Ambiguity control	Ambiguity control	Ambiguity control	Ambiguity control
	Synonym control	(Synonym control)	Synonym control
	(Attributes)	Hierarchical relationships	Hierarchical relationship
			Associative relationships

An ontology does not exist in isolation, but in combination with other controlled vocabularies.

# Combinations of KOS Types

Which KOS type is most suitable for each vocabulary / set of concepts?

Subjects – [thesaurus](#) or [taxonomy](#)

Activities – [term list](#) or [taxonomy](#)

Languages – [term list](#)

Document types – [term list](#) or [taxonomy](#)

Countries – [term list](#) or [name authority](#)

Laws – [name authority](#)

Organisations – [name authority](#)

Events – [name authority](#)

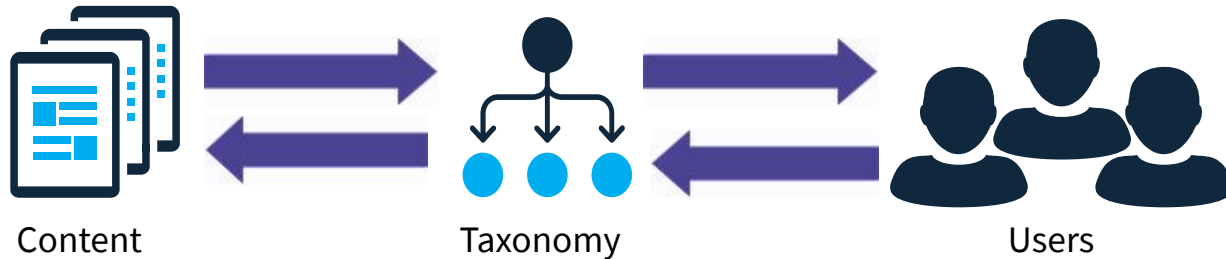
Access right	Environmental impact	Organization type
Accessibility	Event	Other place service
Accreditation	File status	Permission
Address type	File type	Place
Administrative territorial unit	Form type	Position grade
Administrative territorial unit type	Formation of the Court	Position status
Applicability	Framework agreement	Position type
Assessment	Frequency	Procedure nature
Asset classification	Grammatical alternation	Procedure phase
Award criterion type	Grammatical consciousness	Procurement procedure type
Browser	Grammatical gender	Product form
Buyer legal type	Grammatical number	Public event type
COM internal consultation type	Honorific	Publication theme
COM internal event	Human sex	Received submission type
COM internal procedure	Innovative acquisition	Remedy type
Capital classification	Interinstitutional procedure	Requirement stage
Case report	Internal procedure	Reserved procurement
Case status	Irregularity type	Resource type
Change corrig justification	K4p list	Review body type
Communication channel	Label type	Review decision type
Communication channel use	Language	Role
Communication justification	Learning activity	Role nature
Concept status	Learning and verification	Role qualifier
Continent	Learning assessment	Scoring
Contract nature	Learning opportunity	Script
Corporate body	Learning schedule	Selection criterion
Corporate body classification	Learning setting	Site
Correction status	Legal basis	Social objective
Country	Legal proceeding	Strategic priority
Court type	Legal proceeding result	Strategic procurement
Crawler	Legal proceeding type	Subcontracting indication
Credential	Licence	Subcontracting obligation
Currency	Licence domain	Subdivision
DPS usage	Main activity	Subdivision content
Data theme	Measurement unit	Subdivision position
Dataset status	Membership classification	Subject matter
Dataset type	Missing info submission	Summaries of EU legislation classification
Direct award justification	Modification justification	Target audience
Directory of EU legal acts	Modification type	Target group
Distribution type	Multilingual	Time period
Document collection	Non award justification	Treaty
Documentation type	Non publication justification	Treaty classification
EU budget amount status	Notation type	Usage
EU budget stage	Notice type	Use context
EU budget status	Number	Verification
EU programme	Number fixed	Verification status
Economic operator size	Number threshold	Website identifier
Education credit	Number type	Winner selection status
Encoding	Number weight	Writing system
Entitlement	Organisation role	Entitlement status
Organisation subrole		

**Part 2:**  
**Starting to build or  
enhance a knowledge  
organisation system**

**Gathering terms**  
**From users**  
**From content**

## What is a knowledge organisation system for?

- ▶ Concepts are used to tag/index/categorise pages or content to make them easier to be found and retrieved
  - ▷ supporting better findability than search alone
- ▶ The KOS is an intermediary that links the user to the desired content.



- ▶ Consider the users' needs and input.
- ▶ Consider the terms in the content.

## Methods of obtaining user input

- ▶ **Brainstorming workshop**

For designating vocabularies and facets and gathering top concepts

- ▶ **Interviews of sample users and stakeholders**

For building or enhancing KOSs

- ▶ **Card sorting**

For designing and high-level building of hierarchical taxonomies

- ▶ **Search log reports**

For enhancing (not building) a taxonomy or thesaurus, especially for alternative labels





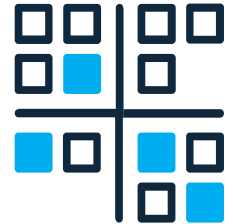
## Interviews of sample users and stakeholders

- ▶ From different functions that deal with the content
- ▶ In person or by phone
- ▶ 1-2 people at once (if from the same function)
- ▶ Have prepared sets of questions sent to stakeholders in advance
- ▶ Different sets of questions for information users and for information curators (uploading/tagging)
- ▶ For information users, different question about how they:
  - ▷ find/discover information
  - ▷ find desired content items
- ▶ For information curators, questions about decisions and issues for tagging and categorising content



## Card sorting

- ▶ Method common in information architecture for website menu label organisation
- ▶ Term names/label/topics are written down each one to a card, and the cards can be sorted into groups.
- ▶ Traditionally done with actual index cards. Now usually done through software, usually drag-and-drop and online to allow remote access.
- ▶ Involves participation of multiple stakeholders or test-user subjects



## Card sorting types

### 1. Open sort

- ▷ Participants group terms and assign the groups category names of their own choosing
- ▷ At the beginning in the taxonomy development process
- ▷ For taxonomy creation, not enhancing

### 2. Closed sort

- ▷ Categories are pre-defined, and participants place terms in the appropriate categories
- ▷ Not at the beginning of the taxonomy development process
- ▷ Could be for taxonomy creation or enhancing



# User Input in Taxonomy Development

[BACK TO ALL SESSIONS](#)[Move back to Brainstorming Phase](#)[Close Session](#)

### Taxonomy View

Taxonomy Root (6 items)

- Healthy Recipes (1 item)
  - Apple recipes (1 item)
    - Sauteed Apples ✕
- Lunch ✕
- Mexican food ✕
- Salad recipes (1 item)
  - Pasta salad ✕
- Seasonal recipes ✕
- Vegeterian recipes (1 item)
  - Hummus Recipes ✕

### Filter concepts

Filter by User

- Andreas Blumauer (8)
- Blumauer Andreas (4)
- Jamie Oliver (6)

## Card sorting demo in PoolParty

## Content/material as sources for candidate concepts

- ▶ **Manually identify main concepts**
  - ▷ From the content to be tagged with the taxonomy
  - ▷ From a representative sample of content
- ▶ **Automatically extract concepts**
  - ▷ From the full set (corpus) of content to be tagged
  - ▷ From additional, very similar sources of content



## Manually identifying main concepts

Identify sample content items of all kinds:

- ▶ Web pages
- ▶ Intranet pages
- ▶ Images
- ▶ Videos
- ▶ Word documents
- ▶ Spreadsheet documents
- ▶ Presentation files
- ▶ PDF documents



Perform a content “audit” or “inventory,” and “extract” a log of candidate terms.

- ▶ Look for concepts especially within:
  - ▶ Document titles and section headings
  - ▶ Website navigation menu labels, site maps, Web page titles
  - ▶ Existing metadata (keywords, titles, short description)
- ▶ Look for main idea concepts, as if indexing.
- ▶ Consider desired search strings to retrieve the content item or page.

## Manually identifying main concepts

Concept identification,  
Similar to tagging without a  
controlled vocabulary.

Tourism industry  
Summer holidays  
Europe  
EU  
Coronavirus pandemic  
Coronavirus vaccines  
Vaccine distribution  
Vaccine passports

### Coronavirus: Europe in vaccine race to save summer

By Kevin Connolly

BBC News, Brussels

Published 6 March

**Europe could be on the brink of a roaring twenties-style summer to remember, with budget airline flights packed and beachside bars brim-full of happy tourists.**

Or, it faces another gloomy holiday season of travel restrictions, quarantine rules and a locked-down leisure industry.

In a few weeks from now we will know which it is to be - but the policy decisions which will shape the outcome are already being taken.

One big question is whether EU member states will be content to leave decision-making to the European Commission in Brussels - which has bungled the vaccine-buying programme - or simply take matters into their own hands.

Greece, for example, has already struck a deal to welcome tourists from Israel if they have a vaccine passport.

And Cyprus has said it will **welcome British tourists from 1 May**, as long as they have had two doses of any vaccine approved by the European Medicines Agency (EMA).

The Director-General of the Cyprus Hotels Association, Philokypros Roussonides, told the BBC: "We are really delighted with this development. It's going to be really effective and very good for airlines to schedule their flights. Cyprus is traditionally a very popular destination for British tourists."

### Tourism jobs at stake

What is at stake here is not just the issue of whether wealthy northern Europeans get to enjoy a beer or an ice cream on the beach.

Tourism is big business, providing 27m jobs in Europe, and generating around 10% of the EU's GDP, when you take into account the other sectors which depend on it.

The economies of countries like Greece, Spain and Italy cannot recover until the tourist industry is reopened.

The GDP of the Balearic Islands - which include Majorca - fell by 27% last year. If a second summer season is lost to Covid-19 the consequences will be disastrous.

A tourism official in Majorca described the situation as "unsustainable" and said that if tourists

## Empirical Approach

- ▷ Concepts are extracted from content objects (e.g. documents) automatically per text extraction.
- ▷ Principle of term extraction and named entity recognition.
- ▷ Finding new concepts, synonyms
- ▷ Evaluating the existing taxonomy (what concepts are found in documents)
- ▷ Calculating Co-occurrences
- ▷ Suggesting relations
- ▷ Statistical model of the language in your content





# Content Sources in Taxonomy Development

The screenshot displays the PoolParty Corpus Search interface. The main title is "Demo of PoolParty Corpus Analysis". The interface is divided into several sections:

- Navigation:** PROJECT, CORPORA, TOOLS, ADVANCED. Search bar: "en Search Thesaurus Concepts".
- Left Sidebar:** Thesaurus, Corpora (selected), Gartner & Use Cases (selected), Candidate Concepts, Blacklist.
- Main Content:**
  - Search Concepts:** Search Criteria dropdown: "Only concepts found in corpus". Buttons: Search, Reset.
  - Extracted Concepts Table:**

Preferred Label	Frequency	Relevance	Most Frequent Label	Broader Concepts	Concept Scheme		
Artificial Intelligence	4063	686.13	AI	<a href="#">General concepts</a>	<a href="#">Principles</a>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Technologies	3604	51.53	technology	<a href="#">Concepts</a>	<a href="#">Principles</a>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Gartner, Inc.	3340	8.11	gartner, inc.	<a href="#">Consulting company</a>	<a href="#">Organisations</a>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Processes	3036	22.66	process		<a href="#">Principles</a>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Metadata	2896	57.78	metadata	<a href="#">General concepts</a>	<a href="#">Principles</a>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Data science	1408	0	data science	<a href="#">General concepts</a>	<a href="#">Principles</a>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Graph	1350	60.12	graph	<a href="#">Graph theory</a>	<a href="#">Principles</a>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

# **Part 3: Integrating existing taxonomies**

**Reuse & Extending  
Linking & Mapping  
Merging**

## Reusing & Extending

- ▶ Simplest way to integrate existing taxonomies is reusing them and extending them based on need.

## Linking & Mapping

- ▶ Taxonomies are linked at individual concepts, and the taxonomies are retained as distinct, but can be used in combination, extending each other.
  - ▶ **Mapping** is a form of linking for exact or close matches, so that one taxonomy can be used for another, and the taxonomies are retained as distinct.
  - ▶ They are used in combination but one is the backend, and one is the frontend (not alongside each other).

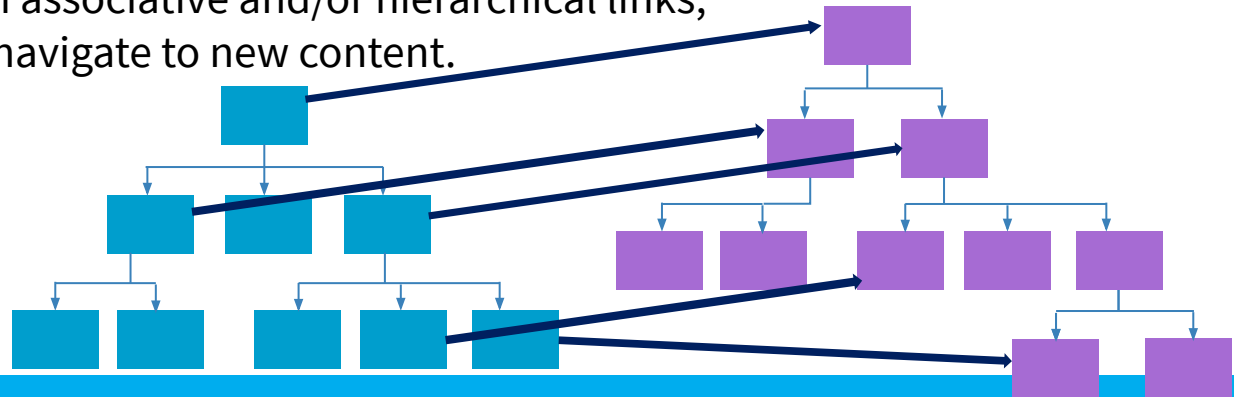


## Merging

- ▶ Taxonomies are combined permanently, removing duplicates, without any longer retaining them as distinct.
- ▶ First step is to link the taxonomies, then incorporate the unlinked concepts.

## Linking scenarios and directions

- ▶ Directional from one KOS to another with sufficiently equivalent links, so that one KOS may be used for another. > [Mapping](#)
- ▶ Directional from a term set to a KOS with equivalent and hierarchical links, so that a KOS can be enriched with added concepts. > [Merging](#)
- ↔ ▶ Bidirectional, with equivalent links, so that content can be shared.
- ↔ ▶ Bidirectional, with associative and/or hierarchical links, so that users can navigate to new content.



## Standards for linking

SKOS supports links across different concept schemes

SKOS names these cross-scheme links “mapping properties”



<https://www.w3.org/TR/skos-reference/#mapping>

"These properties are used to state mapping (alignment) links between SKOS concepts in different concept schemes, where the links are inherent in the meaning of the linked concepts."

- ▶ **exactMatch** – exact match, bidirectional, in all circumstances
- ▶ **closeMatch** – close match, bidirectional, in some (sufficient) circumstances or in a certain context
- ▶ **broadMatch** – has broader concept in the other KOS; inverse of narrowMatch
- ▶ **narrowMatch** – has narrower concept in the other KOS; inverse of broadMatch
- ▶ **relatedMatch** – has related concept in the other KOS; bidirectional

## Standards for linking

ISO 25964-2 *Thesauri and interoperability with other Vocabularies*  
*Part 2: Interoperability with other vocabularies* (2013)

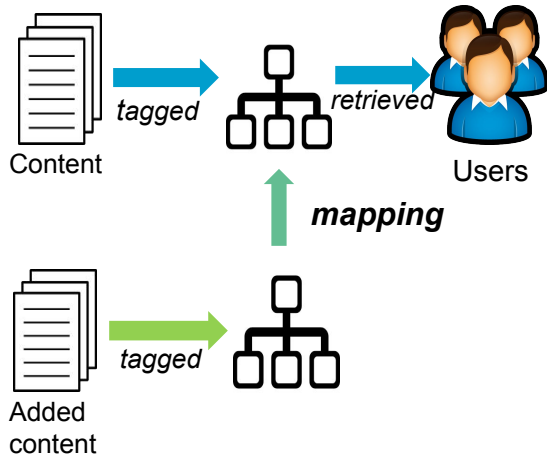


Inter-vocabulary mapping is the principal focus.

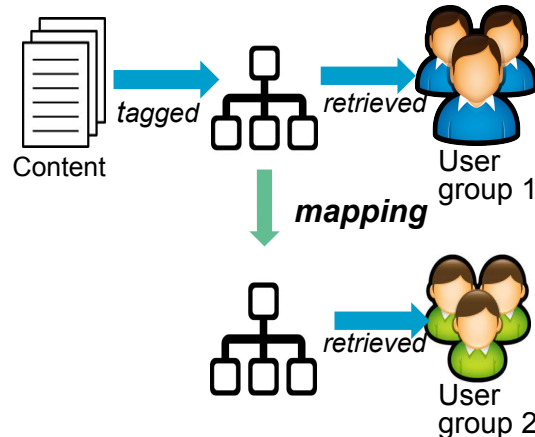
- ▶ Addresses the theory and method of various kinds of mappings.
- ▶ Addresses both one-way directional mapping, and multi-directional.
- ▶ Considers also mapping between thesauri and other kinds of vocabularies: synonym rings, classification schemes, subject heading schemes, taxonomies, terminologies, name authority lists, and ontologies.

# Mapping Taxonomies: Situations

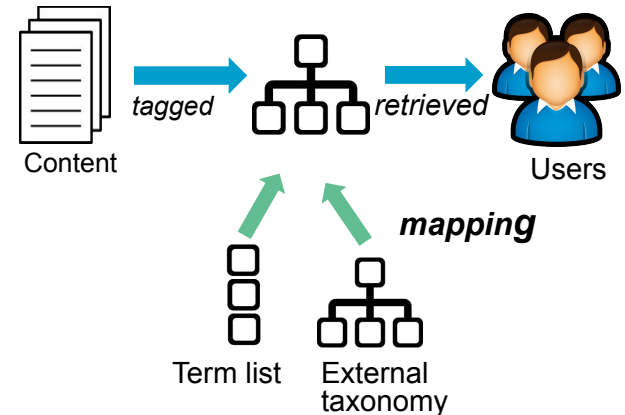
An expanded set of content, tagged with a different taxonomy, will be retrieved by users with their existing taxonomy.



A set of content will be retrieved by different audiences, each accessing their own taxonomy.



A front-end taxonomy will be used to retrieve various content sets, each tagged with its own taxonomy.



## Mapping taxonomies: linking equivalent concepts across taxonomies

Designate a dominant/primary taxonomy into which to merge the other into:

- ▶ The larger taxonomy
- ▶ The taxonomy with greater breadth
- ▶ The taxonomy with greater depth
- ▶ The more structured taxonomy
- ▶ The higher quality taxonomy
- ▶ The taxonomy of the leading, acquiring organisation





## Directional mapping



Directional mapping is easier when:

- ▶ The scope of both is identical.
- ▶ The retrieval taxonomy has fewer terms than the tagged taxonomy.
- ▶ The tagged taxonomy is more specific/granular than the retrieval taxonomy.

Directional mapping is more complex when:

- ▶ Mapping from a hierarchical taxonomy to a faceted taxonomy.
- ▶ There is inconsistency, and one taxonomy is more detailed (with more specific/granular concepts) in some areas, and the other KOS is more detailed in other areas.

Directional mapping does not work when:

- ▶ From a *faceted* taxonomy to a hierarchical taxonomy, thesaurus, or ontology.

## Combining two taxonomies in the same subject area into one

- ▶ Enhancing a taxonomy with concepts from another that won't be used any more
- ▶ Taxonomies are combined permanently, removing duplicates.
  - ▷ Concepts are added, where there is none equivalent match.
  - ▷ Equivalent concepts may gain additional alternative labels.
  - ▷ Legacy content can be retrieved through added alternative labels.
- ▶ Situations for merging:
  - ▷ An enterprise taxonomy replaces multiple taxonomies of separate administrative departments.
  - ▷ An organisation acquires or merges with another organisation, and their redundant vocabularies are merged.
  - ▷ A folksonomy or uncontrolled keywords are incorporated into a taxonomy.
  - ▷ An internally created taxonomy is combined with an external, licensed taxonomy.

# Merging Taxonomies

Compare vocabularies - automatically and with human review

Merging taxonomy (will go away)	Primary taxonomy ( <i>Keep and</i>	Taxonomist Reviews
<b>Exact matches of:</b>		
<i>Preferred label: Cars</i>	<i>Preferred label: Cars</i>	no need
<i>Preferred label: Automobiles</i>	<i>Alternative label: Automobiles</i> <i>For preferred: Cars</i>	no need
<i>Alternative label: Cars</i> <i>For preferred: Automobiles</i>	<i>Preferred label: Cars</i>	yes
<i>Alternative label: Cars</i> <i>For preferred: Automobiles</i>	<i>Alternative label: Cars</i> <i>For preferred: Autos</i>	yes
<b>Inexact matches of:</b>		
<i>Preferred label: Automobile</i>	<i>Preferred label: Automobiles</i>	yes

# Linking/Mapping Taxonomies

**Project Linking**

Tree View | List View | **Batch Linking**

Select Project: NACE rev 2

**Batch Linking Results**

Status	Local Label	Linked Label	Match Type	Linking Scheme	Linking Predicate	
▼ Pending	<a href="#">Computer Software@en</a>	<a href="#">Software publishing@en</a>	prefLabel/altLabel	SKOS	exactMatch	<input type="button" value="like"/> <input type="button" value="comment"/>
▼ Pending	<a href="#">Construction@en</a>	<a href="#">CONSTRUCTION@en</a>	prefLabel/prefLabel	SKOS	exactMatch	<input type="button" value="like"/> <input type="button" value="comment"/>
▼ Pending	<a href="#">Fishery@en</a>	<a href="#">Fishing@en</a>	prefLabel/altLabel	SKOS	exactMatch	<input type="button" value="like"/> <input type="button" value="comment"/>
▼ Pending	<a href="#">Higher Education@en</a>	<a href="#">Higher education@en</a>	prefLabel/prefLabel	SKOS	exactMatch	<input type="button" value="like"/> <input type="button" value="comment"/>
▼ Pending	<a href="#">Insurance@en</a>	<a href="#">Insurance@en</a>	prefLabel/prefLabel	SKOS	exactMatch	<input type="button" value="like"/> <input type="button" value="comment"/>
▼ Approved	<a href="#">Civil Engineering@en</a>	<a href="#">Civil engineering@en</a>	prefLabel/prefLabel			<input type="button" value="undo"/>
▼ Approved	<a href="#">Performing Arts@en</a>	<a href="#">Performing arts@en</a>	prefLabel/prefLabel			<input type="button" value="undo"/>
▼ Approved	<a href="#">Telecommunications@en</a>	<a href="#">Telecommunications@en</a>	prefLabel/prefLabel			<input type="button" value="undo"/>

# Questions/Contact

---

## Helmut Nagy

COO

Semantic Web Company GmbH

Neubaugasse 1, Top 8

1070 Vienna

Austria

+43 1-4021235

[helmut.nagy@semantic-web.com](mailto:helmut.nagy@semantic-web.com)

[www.linkedin.com/in/alexisdimitriadis](https://www.linkedin.com/in/alexisdimitriadis)

## Heather Hedden

Data and Knowledge Engineer

Semantic Web Company Inc.

One Boston Place, Suite 2600

Boston, MA 02108

USA

+1 857-400-0183

[heather.hedden@semantic-web.com](mailto:heather.hedden@semantic-web.com)

[www.linkedin.com/in/hedden](https://www.linkedin.com/in/hedden)

Semantic Web Company [www.semantic-web.com](http://www.semantic-web.com)

PoolParty software [www.poolparty.biz](http://www.poolparty.biz)

