Building, Enhancing, and Integrating Taxonomies

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

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.

Outline

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 Types

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

Knowledge Organisation System Types

Common types of controlled vocabularies

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

“Taxonomy” sometimes means any controlled vocabulary.

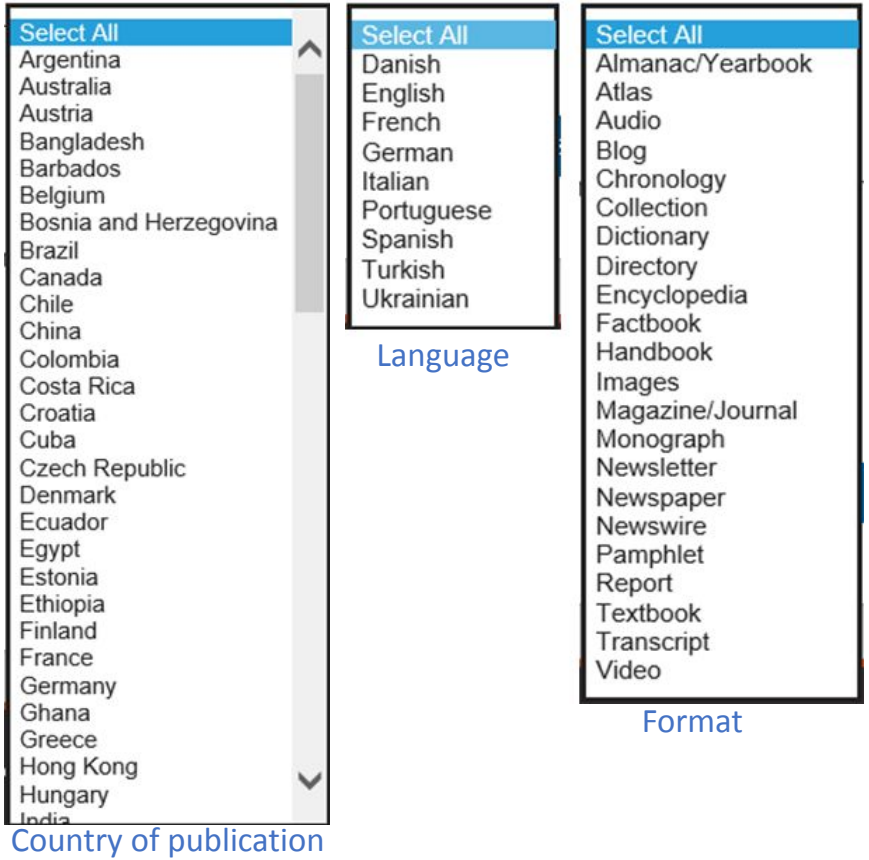
Other common kind of knowledge organisation system

- Ontology

Knowledge Organisation System Types

Term List

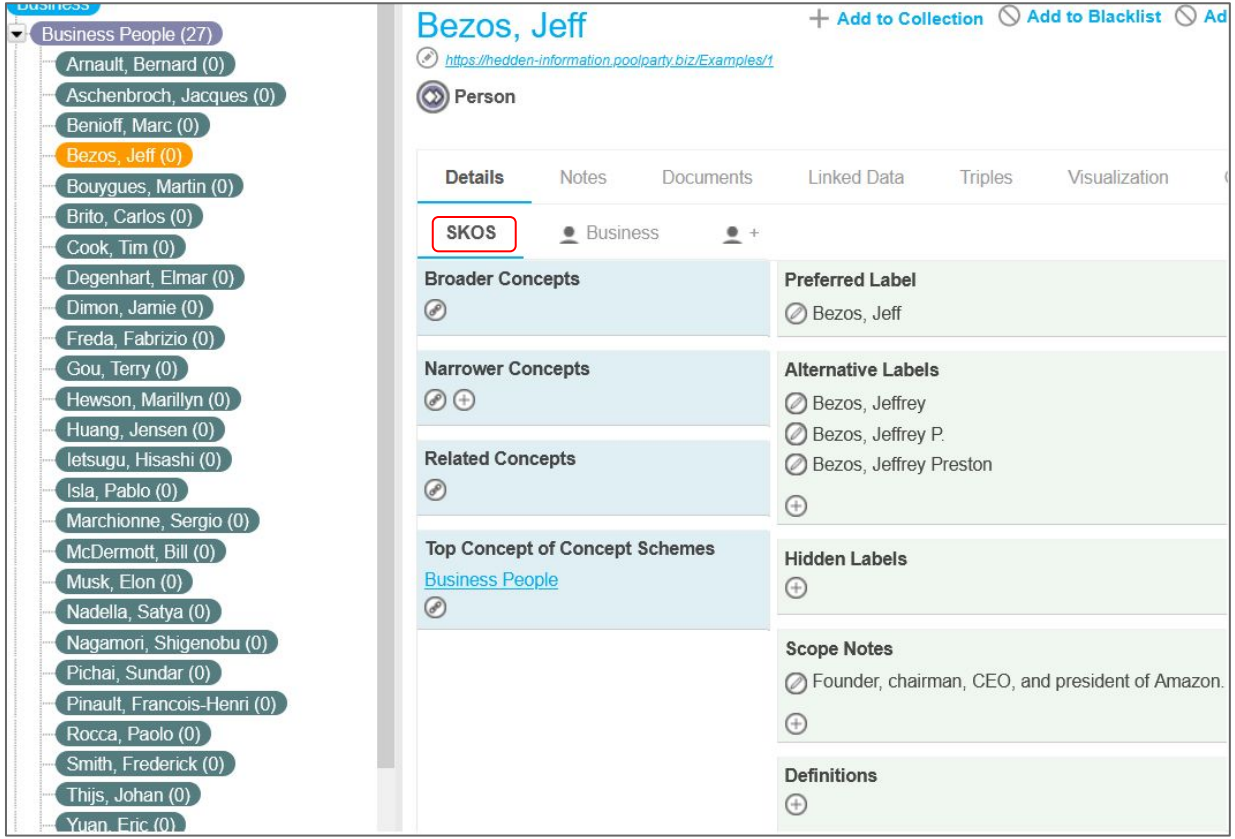
- A simple list of terms
- Usually alphabetical, but could be in other logical order
- Lacking synonyms, it is usually short enough for quick browsing
- Can appear in drop-down scroll boxes
- May be used for various metadata values, facets, concept schemes
- Part of a larger set of controlled vocabularies;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



Knowledge Organisation System Types

Name authority

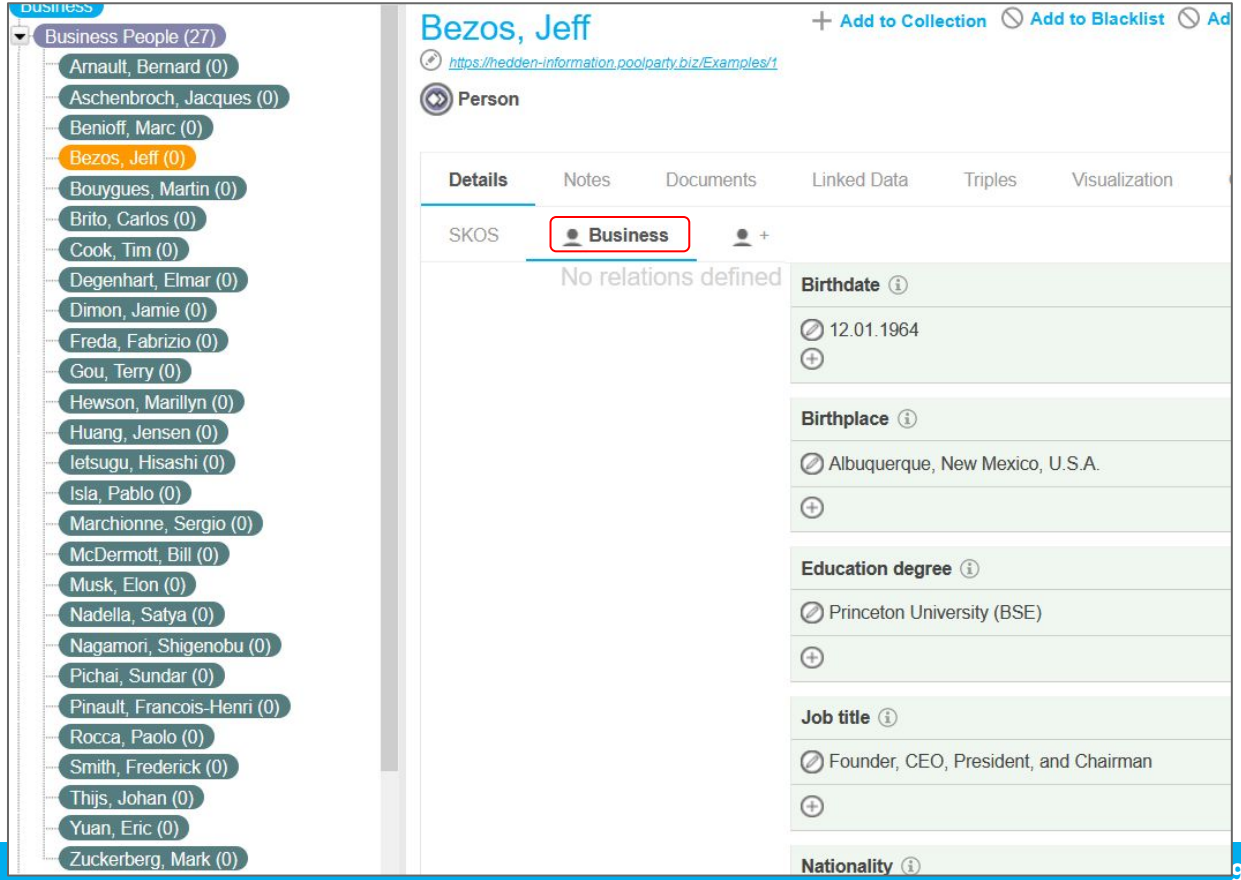
- For named entities, concrete entities, proper nouns
- A controlled vocabulary with preferred names and variant/alternative names.
- May or may not have hierarchical relationships between named entities.
- Usually has additional information/attributes (metadata) for each named entity.



Knowledge Organisation System Types

Name authority

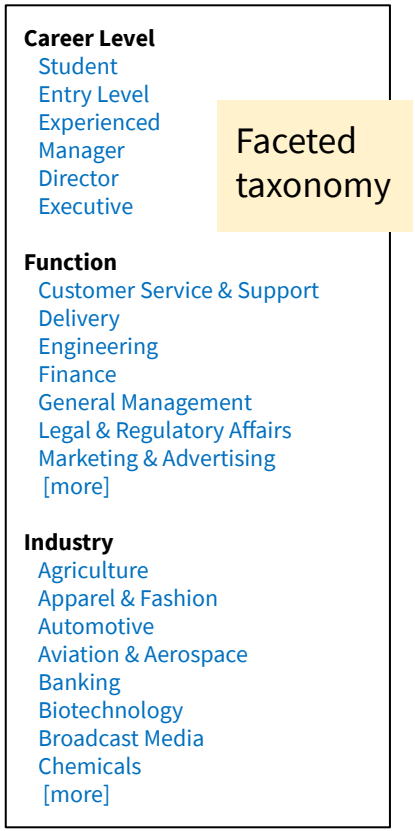
- The SKOS model can be extended by custom ontologies to support extended attributes, often desired for named entities.



Knowledge Organisation System Types

Taxonomy

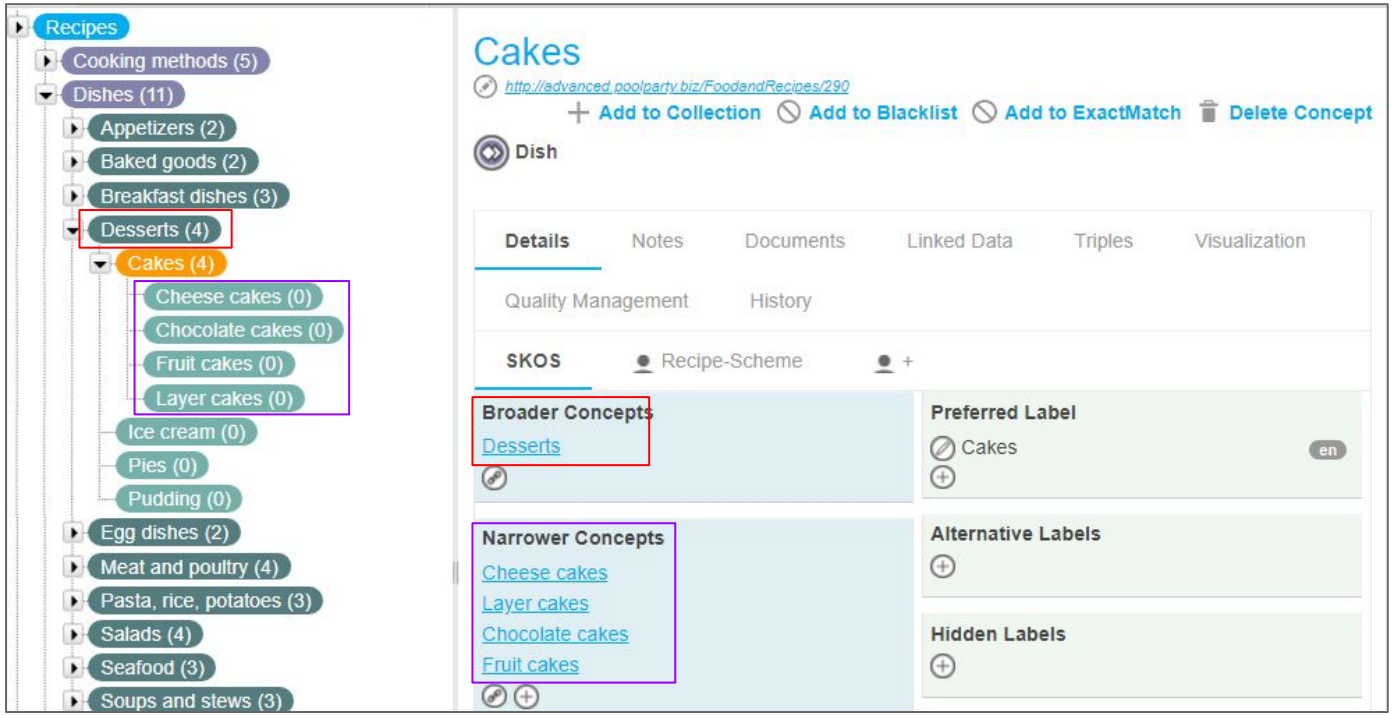
- A KOS with broader/narrower relationships that includes all concepts to create a hierarchical structure.
- Has a focus on categorising and organising concepts.
- May or may not have “synonyms” to point to the correct, preferred terms/labels.
- May comprise several hierarchies, concept schemes, or facets.
- (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.)



Knowledge Organisation System Types

Hierarchical taxonomy

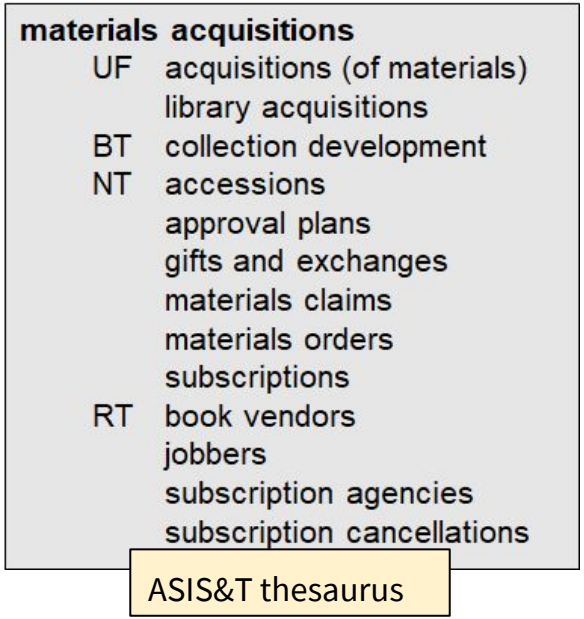
Concepts have broader-concept and/or narrower-concept relationships to other concepts.



Knowledge Organisation System Types

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
- “Thesaurus” is usually the kind of controlled vocabulary used in indexed articles databases, such as accessed through libraries.

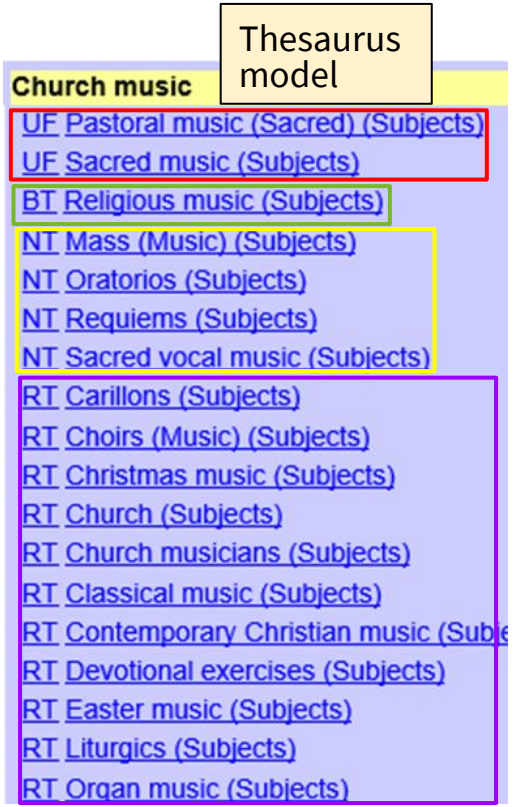


ASIS&T thesaurus

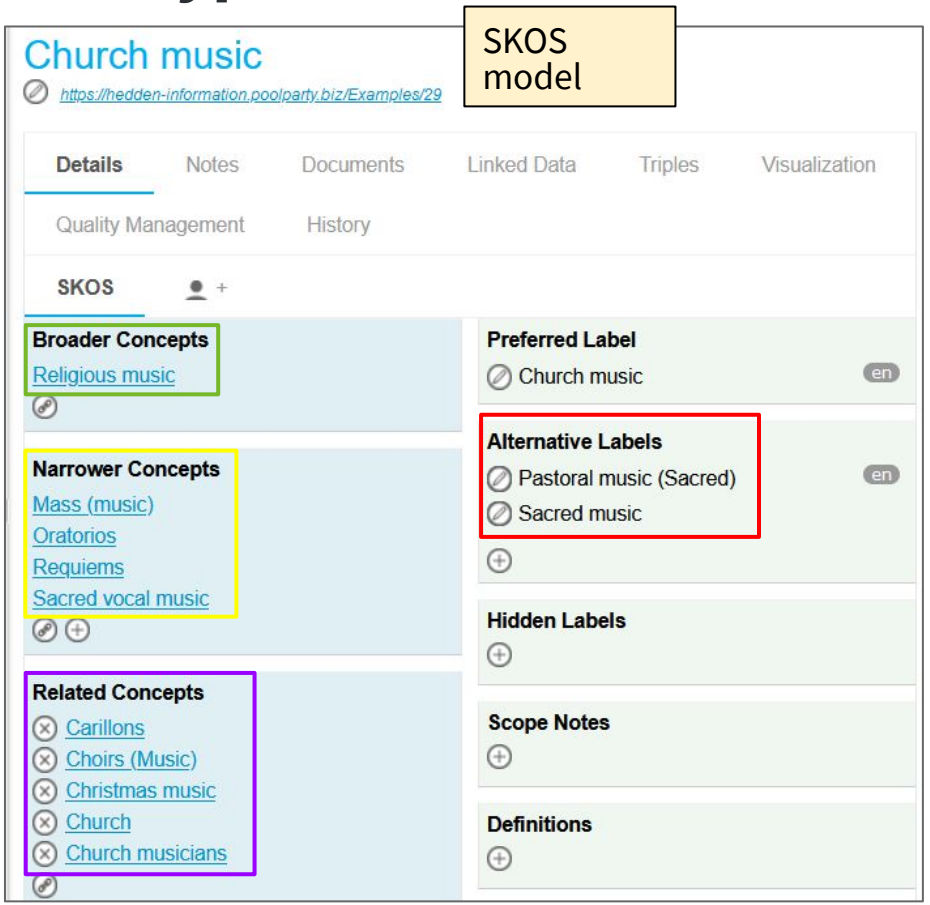
Knowledge Organisation System Types

Thesaurus

ANSI/NISO or ISO thesaurus model and SKOS model compared



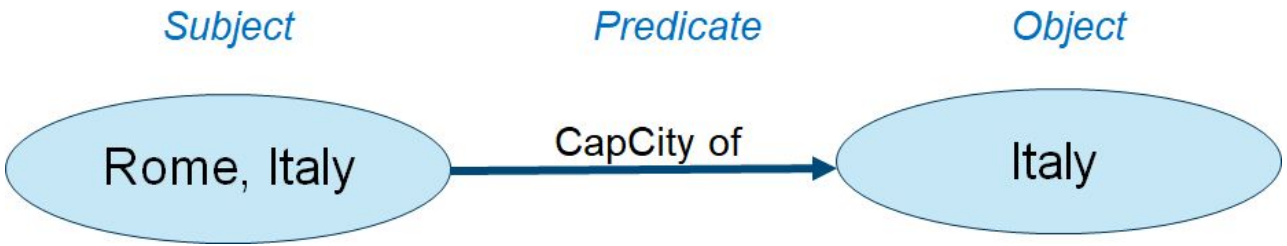
Gale Subject Thesaurus



Knowledge Organisation System Types

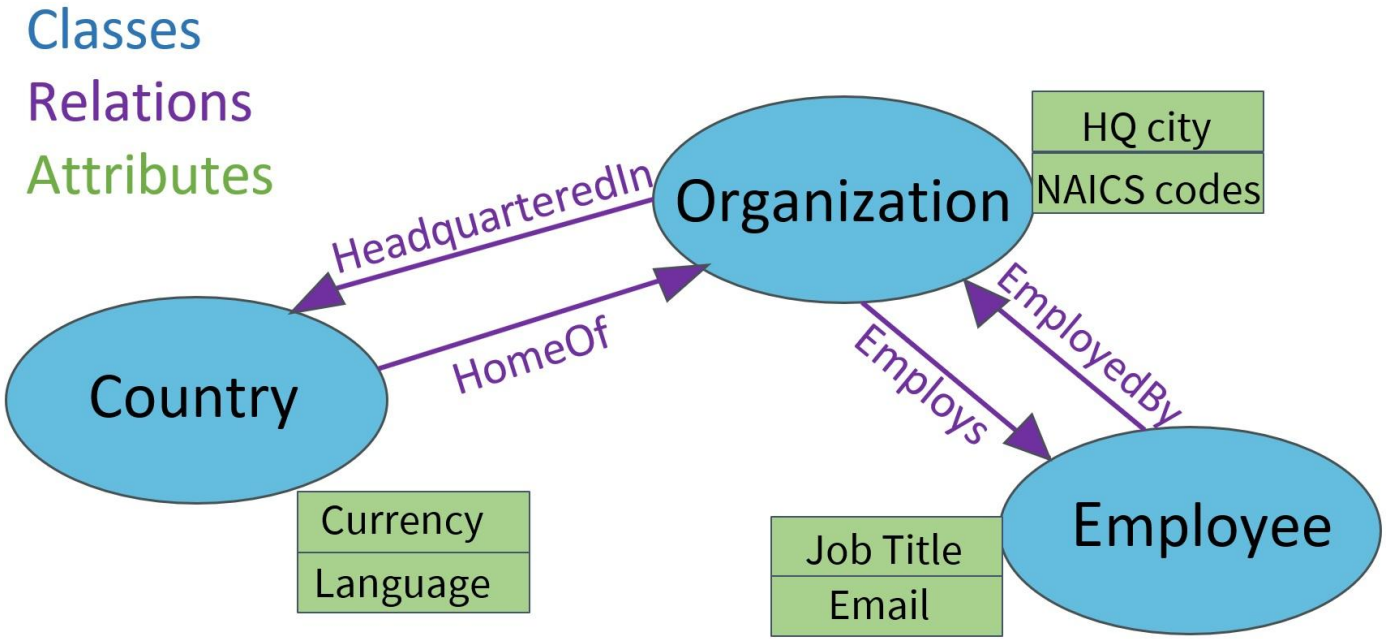
Ontology

- 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.

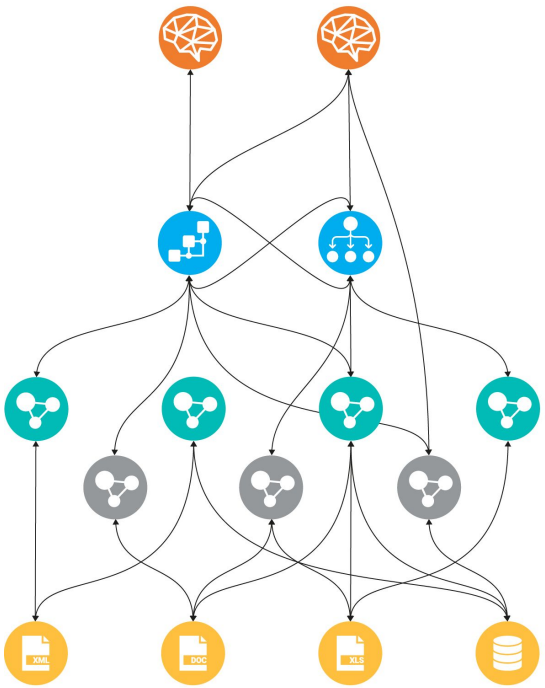


Knowledge Organisation System Types

Ontology



What is an Enterprise Knowledge Graph (EKG)?



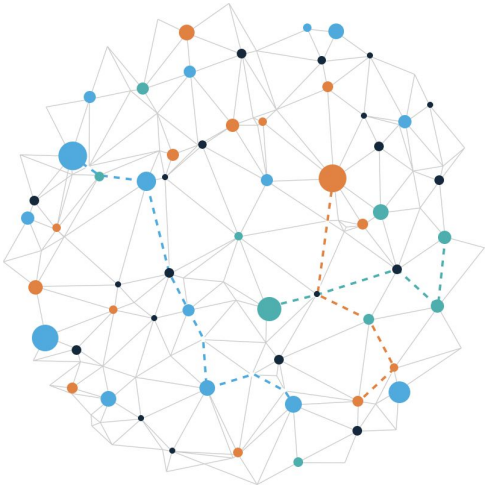
SEMANTIC AI APPLICATION

CONCEPTUAL AND LINGUISTIC MODEL

(VIRTUAL) DATA GRAPH

CONTENT & DATA LAYER

ENTERPRISE KNOWLEDGE GRAPH



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

Standards & Recommendations

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
2. Standards for specifications and interoperability
   Dublin Core, MARC, ZThes, DD 8723-5, SKOS, RDF, RDFS, and OWL

Standards & Recommendations

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.

Standards & Recommendations

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).

Standards

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.

Standards

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		

Other W3C Recommendations

RDF (Resource Description Framework)

- A World Wide Web (W3C) recommendation 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.

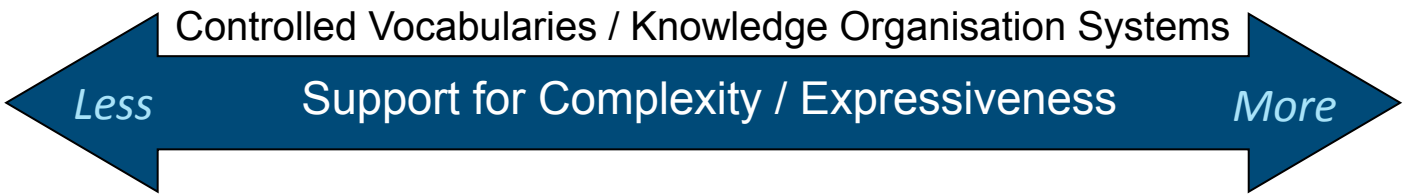
RDFS (RDF-Schema)

- A W3C recommendation www.w3.org/2001/sw/wiki/RDFS
- 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
- “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

Comparison of Knowledge Organisation Systems

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.

Comparison of Knowledge Organisation Systems

poolparty

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

Names of Knowledge Organisation Systems

“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

Uses of Knowledge Organisation Systems

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

Uses of Knowledge Organisation Systems

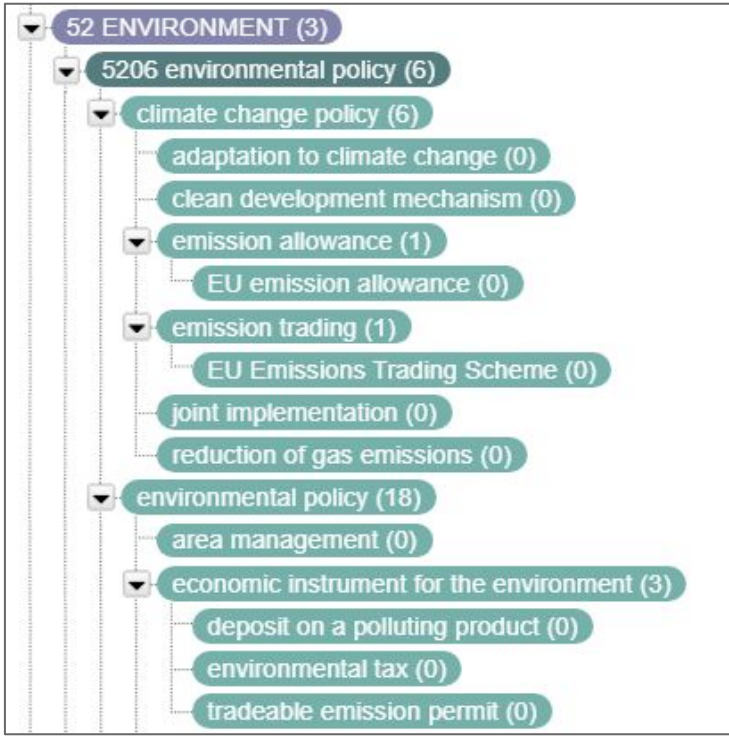
Multiple purposes and uses of controlled vocabularies

- 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

Uses of Knowledge Organisation 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

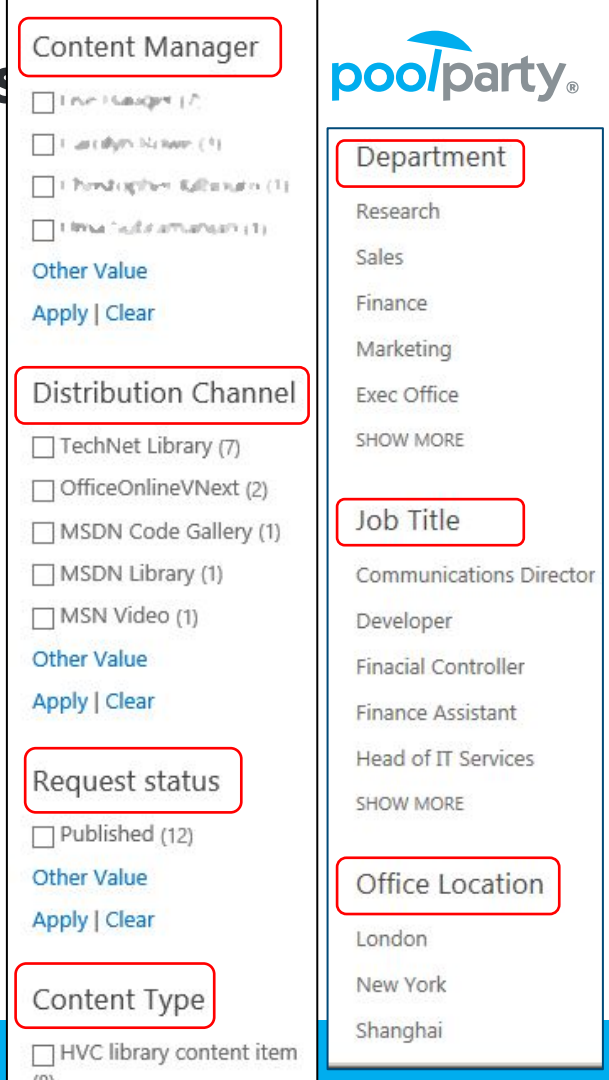
Uses of Knowledge Organisation Systems

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



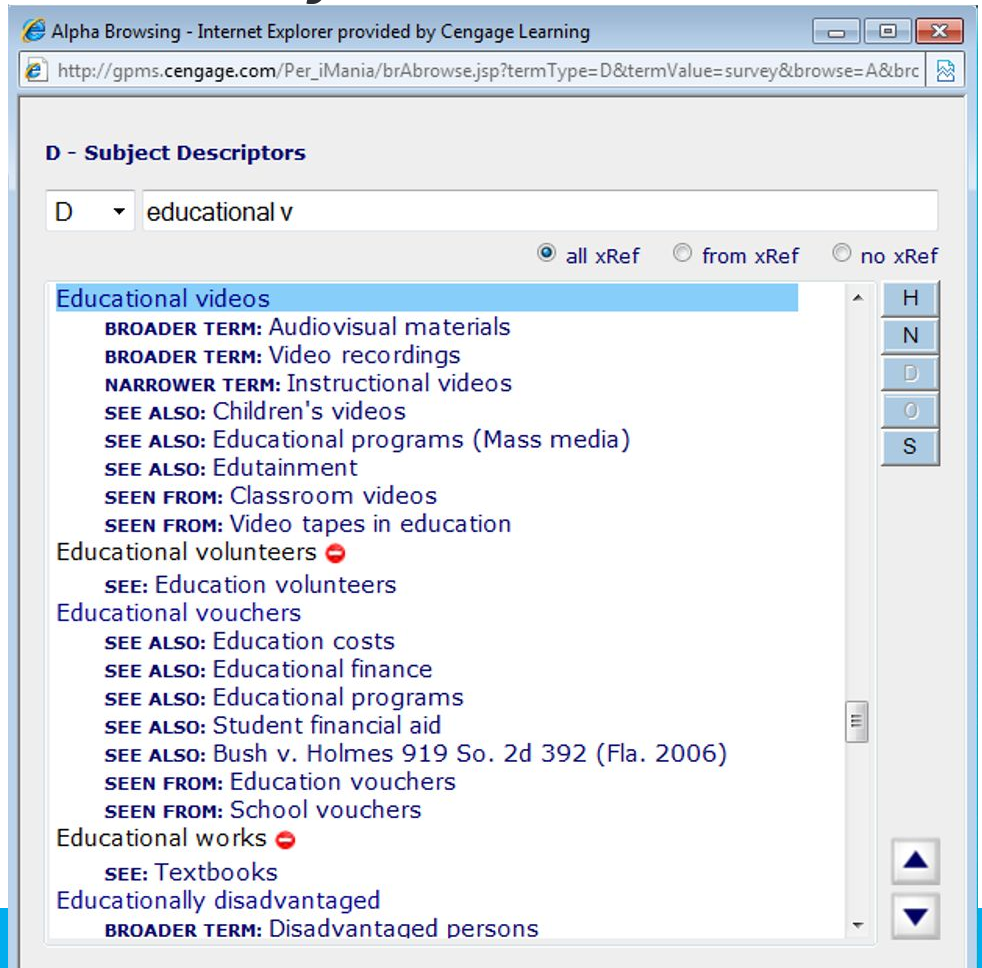
Uses of Knowledge Organisation Systems

Thesaurus purpose

- Support for manual indexing

Manual indexing
user interface example

Cengage/Gale Subject Thesaurus
Internal indexer alphabetical browse view



Uses of Knowledge Organisation Systems

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

Combinations of KOS Types

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.

Combinations of KOS Types

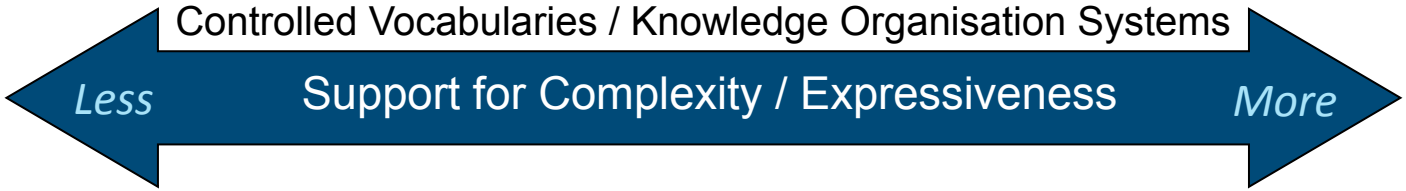
An organisation may use multiple KOS types for a single domain of content.

- 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.

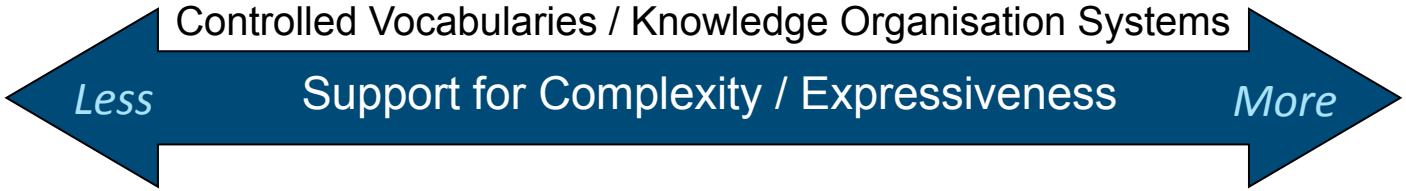
KOS or Scheme (from)	Relation	KOS or scheme (to)
Law Legislative body	issuedBy issues	Legislative body Law
Law Thesaurus subject	dealsWith isSubjectOf	Thesaurus subject Law

Combinations of KOS Types



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

Combinations of KOS Types



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

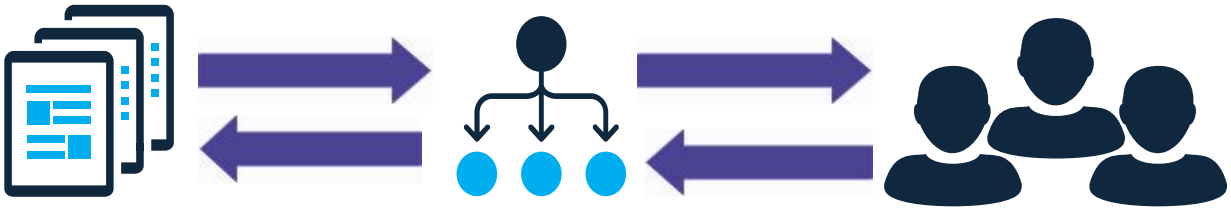
Part 2:
Starting to build or enhance a knowledge organisation system

Gathering terms
From users
From content

Building and Enhancing a KOS

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.

User Input in Taxonomy Development

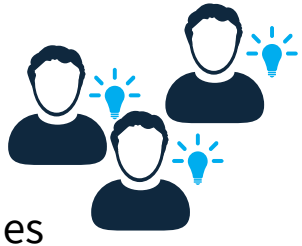
Methods of obtaining user input

- 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

User Input in Taxonomy Development

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

User Input in Taxonomy Development

Card sorting

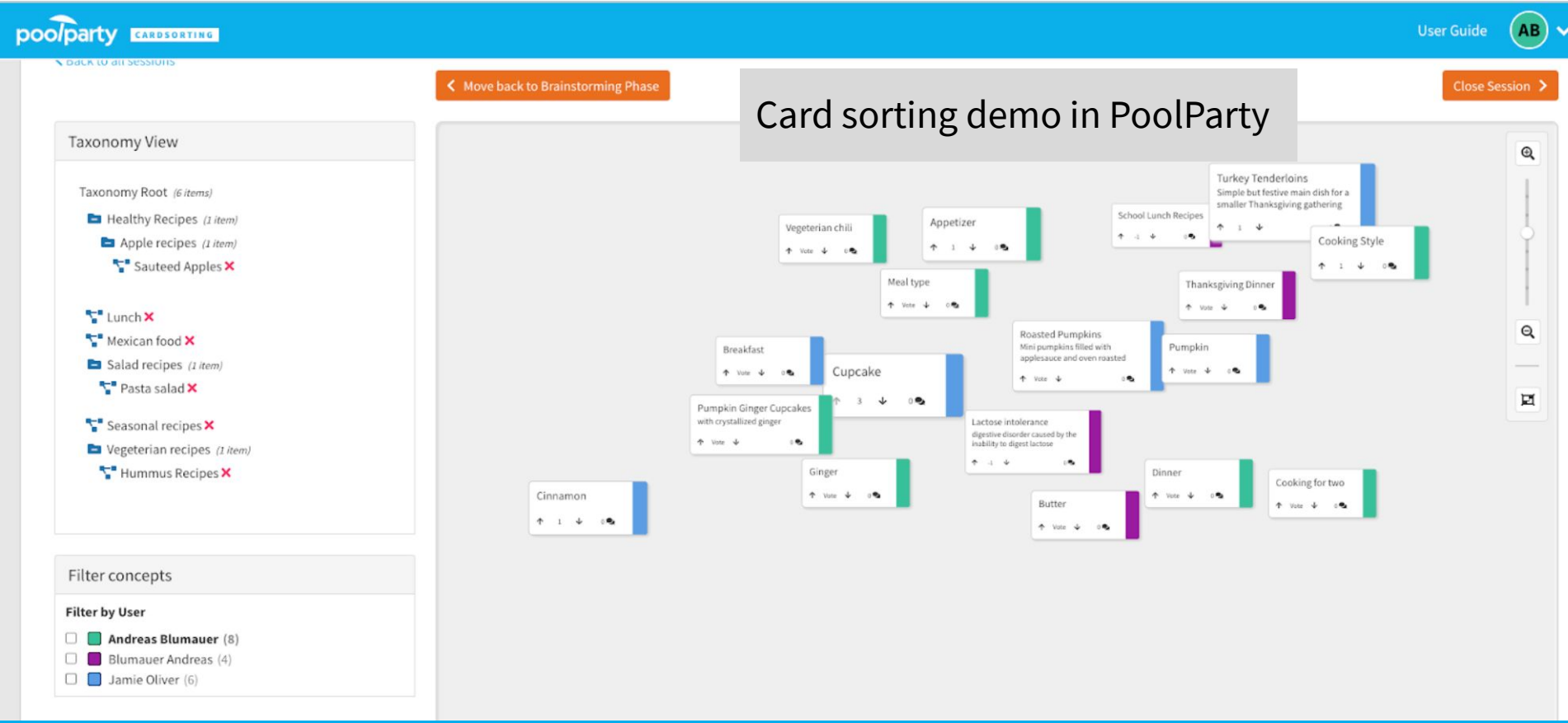
- Method common in information architecture for website menu label organisation
- Term names/label/topics are written down each one to a card, and the cards can be sorted into groups.
- Traditionally done with actual index cards. Now usually done through software, usually drag-and-drop and online to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

User Input in Taxonomy Development

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



Content Sources in Taxonomy Development

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

Content Sources in Taxonomy Development

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.

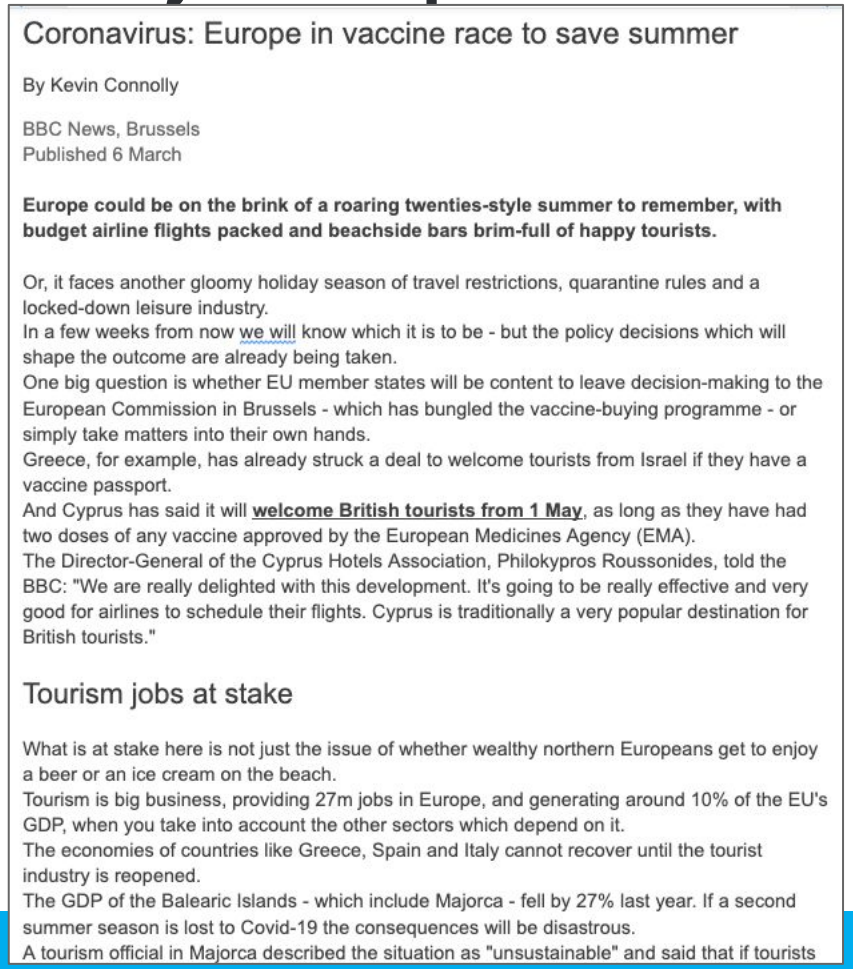
- Look for concepts especially within:
  - Document titles and section headings
  - Website navigation menu labels, site maps, Web page titles
  - Existing metadata (keywords, titles, short description)
- Look for main idea concepts, as if indexing.
- Consider desired search strings to retrieve the content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

Content Sources in Taxonomy Development

Manually identifying main concepts

Concept identification, Similar to tagging without a controlled vocabulary.

- Tourism industry
- Summer holidays
- Europe
- EU
- Coronavirus pandemic
- Coronavirus vaccines
- Vaccine distribtuion
- 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

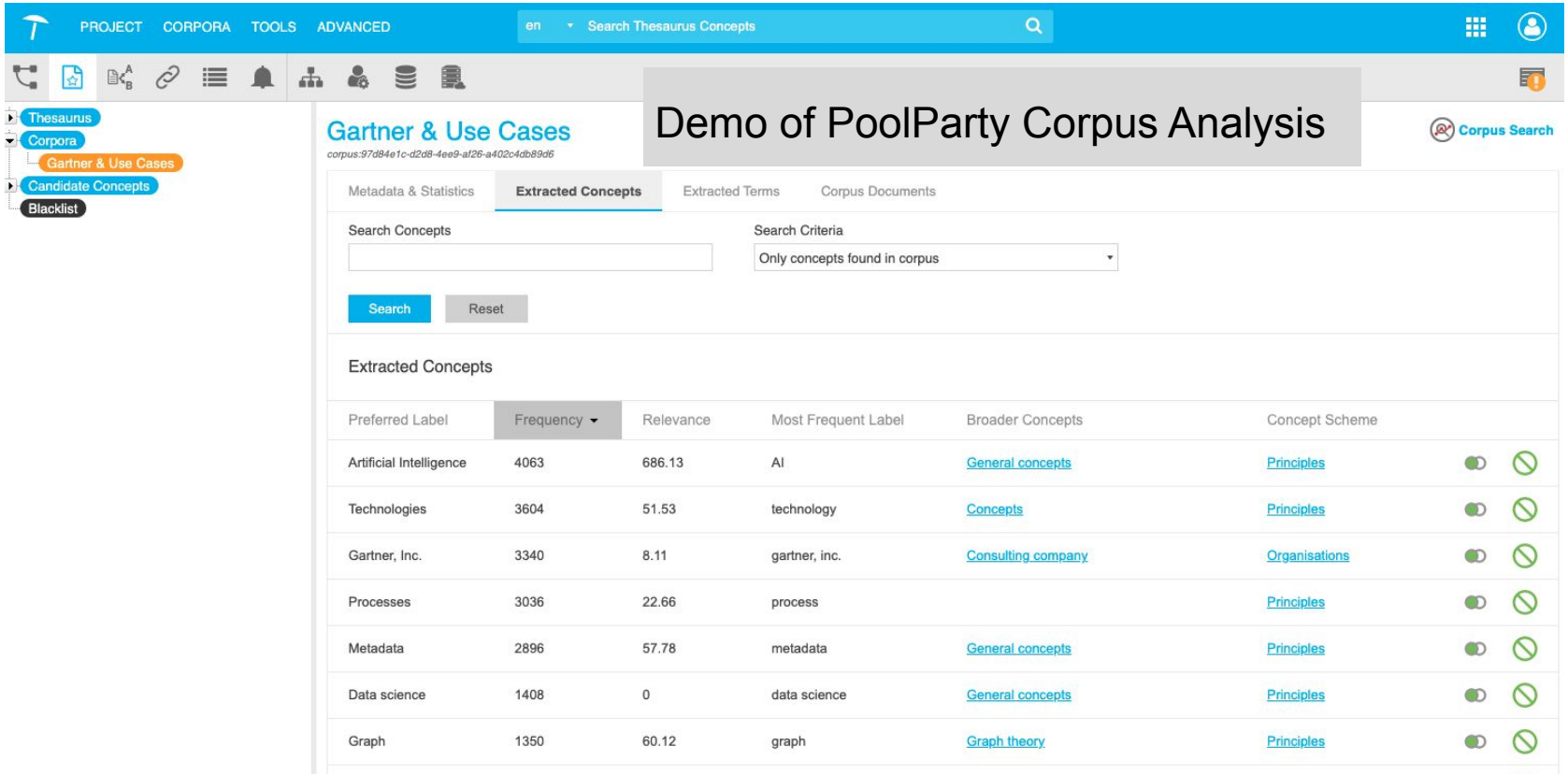
What is at stake here is not just the issue of whether wealthy northern Europeans get to enjoy a beer or an ice cream on the beach.
Tourism is big business, providing 27m jobs in Europe, and generating around 10% of the EU's GDP, when you take into account the other sectors which depend on it.
The economies 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

Content Sources in Taxonomy Development

Empirical Approach

- Concepts are extracted form 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



Part 3:
Integrating existing taxonomies

Reuse & Extending
Linking & Mapping
Merging

Integrating Existing Taxonomies

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).

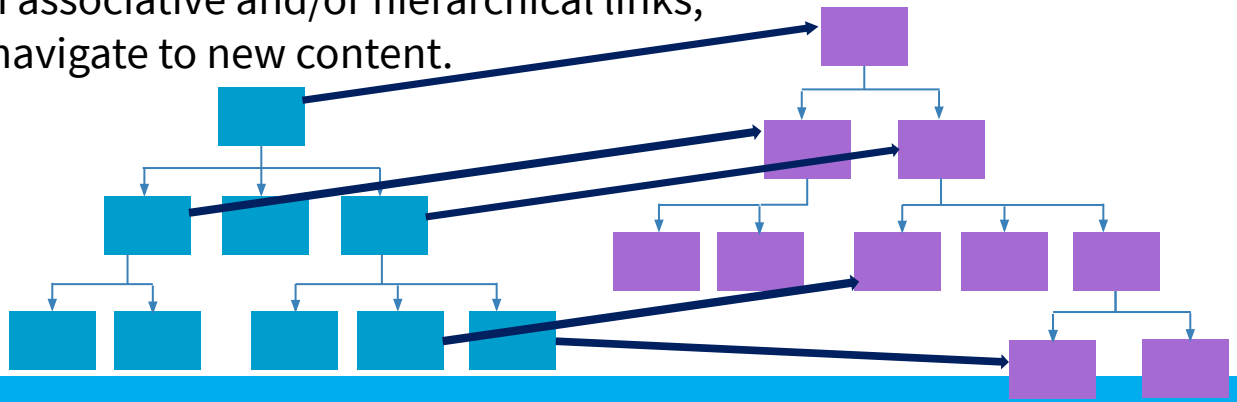
Merging

- 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 Taxonomies

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.



Linking Taxonomies

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

Linking Taxonomies

Standards for linking

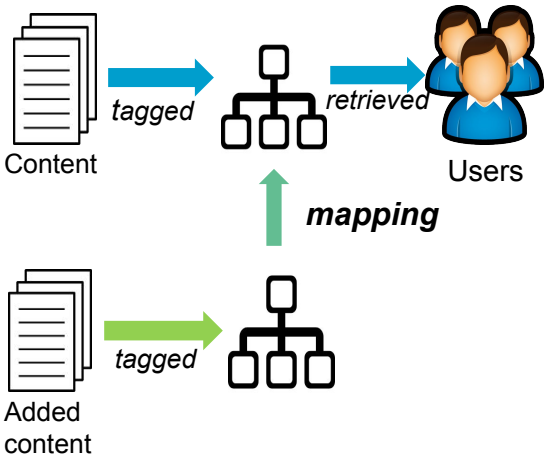
ISO 25964-2 *Thesauri and interoperability with other Vocabularies*
Part 2: Interoperability with other vocabularies (2013)

Inter-vocabulary mapping is the principal focus.

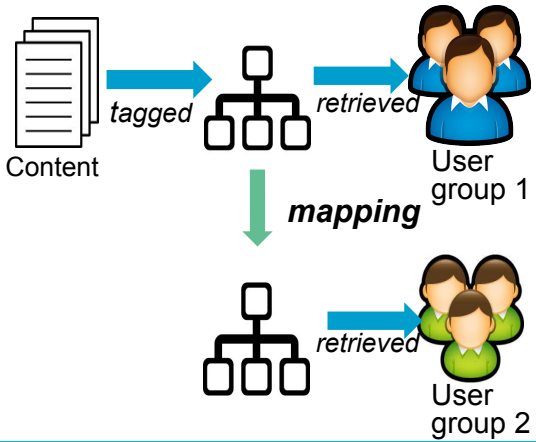
- Addresses the theory and method of various kinds of mappings.
- Addresses both one-way directional mapping, and multi-directional.
- Considers also mapping between thesauri and other kinds of vocabularies: synonym rings, classification schemes, subject heading schemes, taxonomies, terminologies, name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

Mapping Taxonomies: Situations

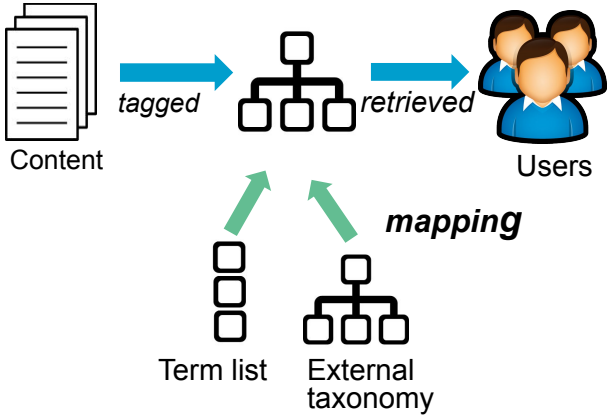
An expanded set of content, tagged with a different taxonomy, will be retrieved by users with their existing taxonomy.



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: Methods

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

Mapping Taxonomies: Methods

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.

Merging Existing Taxonomies

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.

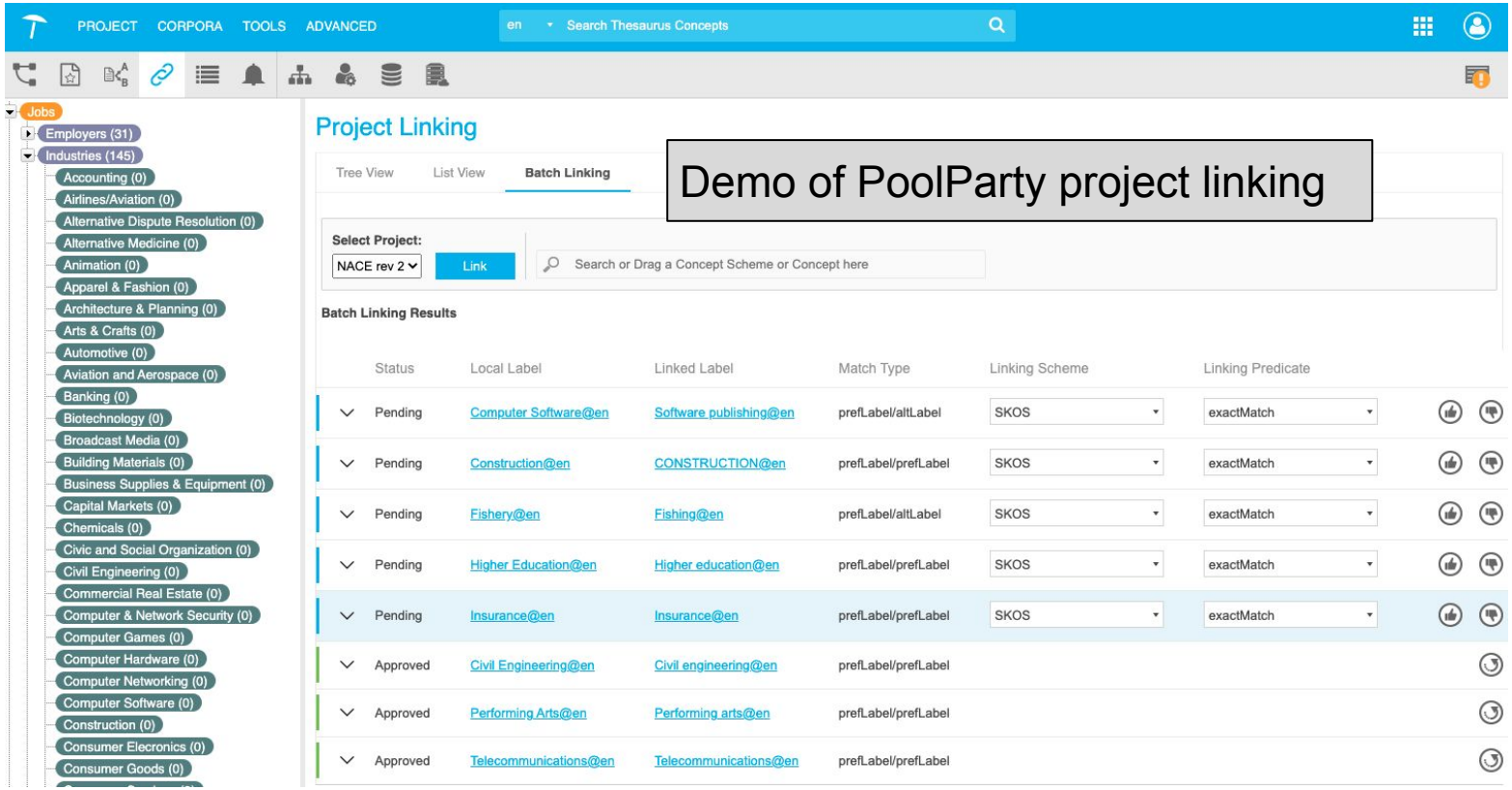
Merging Taxonomies

Compare vocabularies - automatically and with human review

- 

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

Linking/Mapping Taxonomies



Questions/Contact

Helmut Nagy
COO
Semantic Web Company GmbH
Neubaugasse 1, Top 8
1070 Vienna
Austria
+43 1-4021235
helmut.nagy@semantic-web.com
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
www.linkedin.com/in/hedden

Semantic Web Company www.semantic-web.com

PoolParty software www.poolparty.biz