Semantic Standards and Methods for Information Linking

Heather Hedden

International Workshop on Open-Access Publishing

Indian Institute of Technology, Delhi

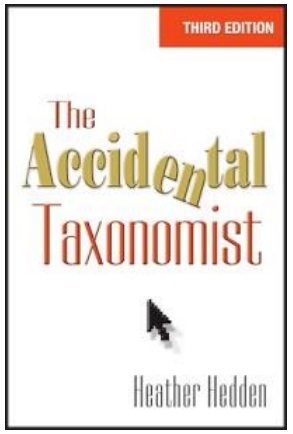
16 January 2025

About Heather Hedden

- Independent Consultant, Hedden Information Management
- Instructor of online and corporate taxonomy training courses and workshops
- Former taxonomy-related roles at Enterprise Knowledge, Semantic Web Company, Gale/Cengage, Viziant, Project Performance Corp., First Wind
- Author of *The Accidental Taxonomist, 3rd ed.* (2022, Information Today, Inc.)

Hedden Information Management

Making information findable



Outline

- Introduction: Web Publishing, Findability, Discovery, and Semantics
- Semantic Web Principles and Standards
- Knowledge Organization Systems
- Standards for Linking Knowledge Organization Systems
- Linked Knowledge Organization Systems on the Web

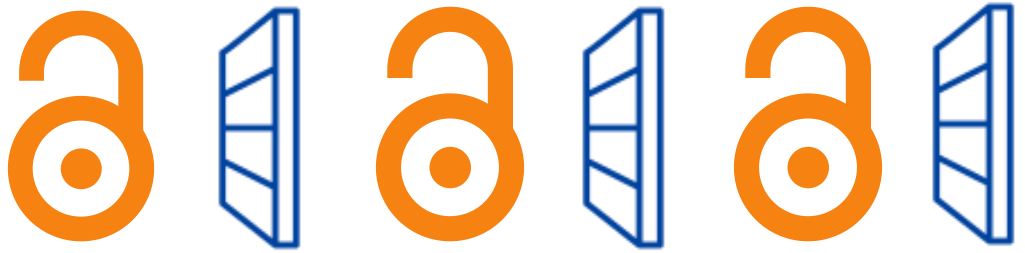
Web Publishing, Findability, Discovery, and Semantics

Open Access Publishing to the Web

- Open access means publishing for free access on the Web.
- Publishing to the Web, however, should ensure that articles and information are not just freely available, but also *findable* and *discoverable*.
- Open access publisher platforms usually have metadata and taxonomies to help users search and retrieve desired articles.

Problems:

- Different, separate (siloed) publisher platforms
- Cannot easily explore related information on different platforms (even if free)
- Does not support discovery well



Findability vs. Discoverability

- Researchers seek to *find* what they are looking for when they *search*
- Research also depends on ***discovery***.
  - Finding something one was not looking for, but also of value

Discovery

- For students:
  - Learning broader connections
  - Identifying related areas of interest
  - Expanding/refining the scope of a research topic area
- For research scholars
  - Identifying connections, correlations, patterns, sources, etc.
  - Finding out who else is researching and publishing on the subject
- Taxonomies and thesauri support discovery through their relationships between the terms: broader, narrower, and related.
  - But this is typically *within* a single repository or publisher's platform.

Linking Data on the Web

- The Web supports linking published information.
- The *Semantic Web* is an extension of the World Wide Web through standards set by the World Wide Web Consortium (W3C).
  - The goal of the Semantic Web is to make Internet data machine-readable.
- The Semantic Web is "a web of data that can be processed directly and indirectly by machines." - Tim Berners Lee
- HTML only presents and links pages but does not relate data or metadata on or between web pages.
  - Encoding with Semantic Web standards can further describe things.
- Semantic technologies and standards are used to formally represent metadata and also taxonomies, thesauri, and ontologies.
- By following Semantic Web technologies, open data can be meaningfully (with semantics) linked across different repositories and publisher platforms.
  - Enabling external discovery

Semantics and Semantic Technology

What is semantics?

- Meaning in language: words, phrases, sentences
- Thoughts, ideas, concepts, values for things and kinds of relations between things
- Not mere words, text strings, or unqualified links
- Semantics is about “things, not strings” (of text)

Why semantics?

- To find information and answers, not just matching text strings.
- To explore specific kinds of linkages, not just anything “related.”
- To formulate complex, multi-part queries, and not just information “about.”

Semantic technology

- Standards and frameworks for data models to encode meanings to help machines interpret data.
- The encoding is separate from the data, content, and application code.

Semantically enriched data

- Providing meaning to data, such as standardized descriptions, categories, types, purposes, relationships to other data.

Semantic Web
Principles and Standards

Semantic Web Overview

The Semantic Web is a **vision** for linking data across webpages, applications and files.

Semantic Web Principles (2001)

1. Everything can be identified by URI's.
2. Web resources and links can have types.
3. Partial information is tolerated.
4. There is no need for absolute truth.
5. Evolution is supported.
6. Minimalist design.

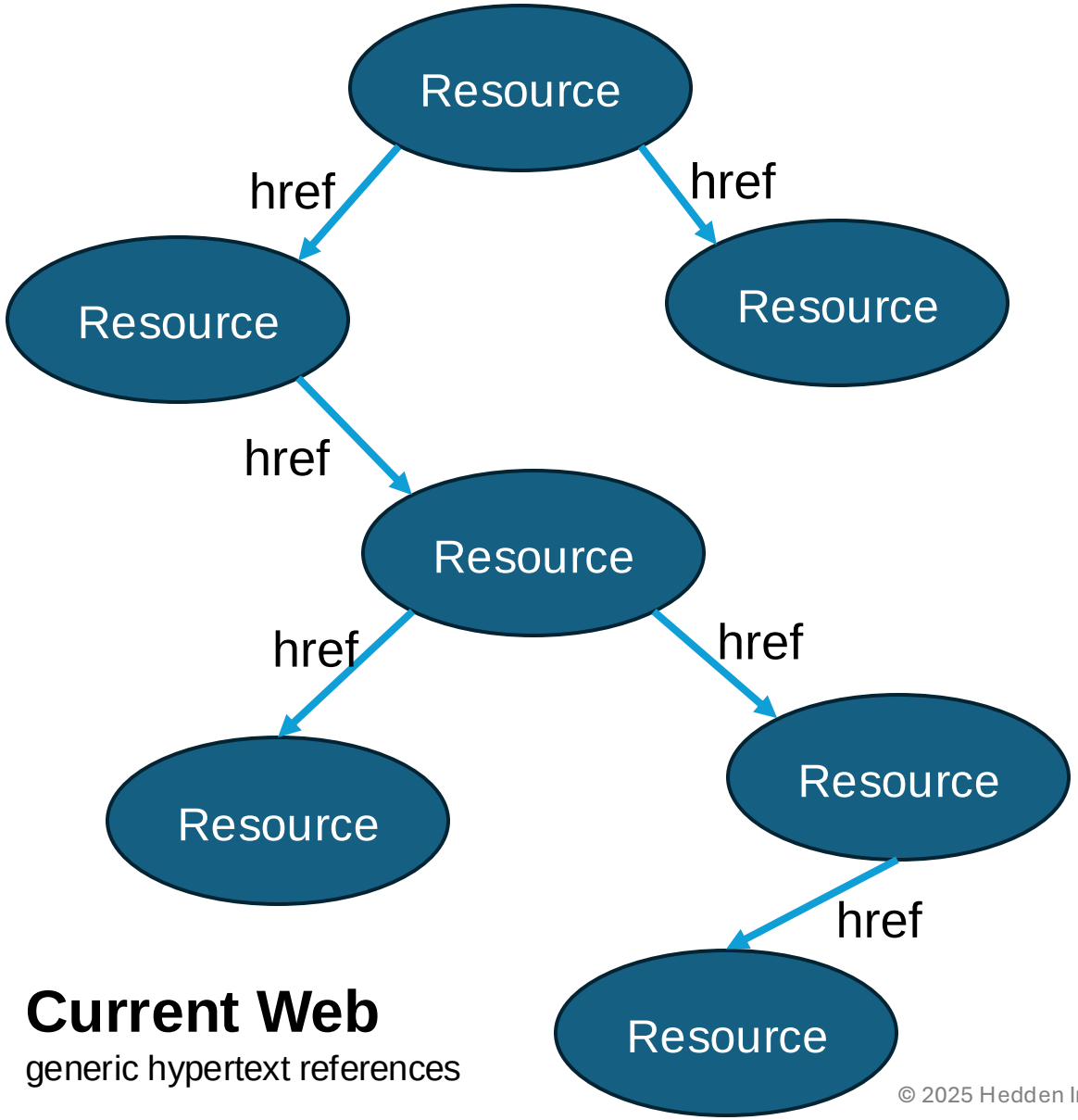
www.w3.org/2001/12/semweb-fin/w3csw

Semantic Ethical Principals (2024)

- There is one web
- The web does not cause harm to society
- The web supports healthy community and debate
- The web is for all people
- The web is secure and respects people's privacy
- The web enables freedom of expression
- The web makes it possible to verify information
- The web enhances individuals' control and power
- The web is an environmentally sustainable platform
- The web is transparent
- The web is multi-browser, multi-OS, and multi-device
- The web can be consumed in any way that people choose

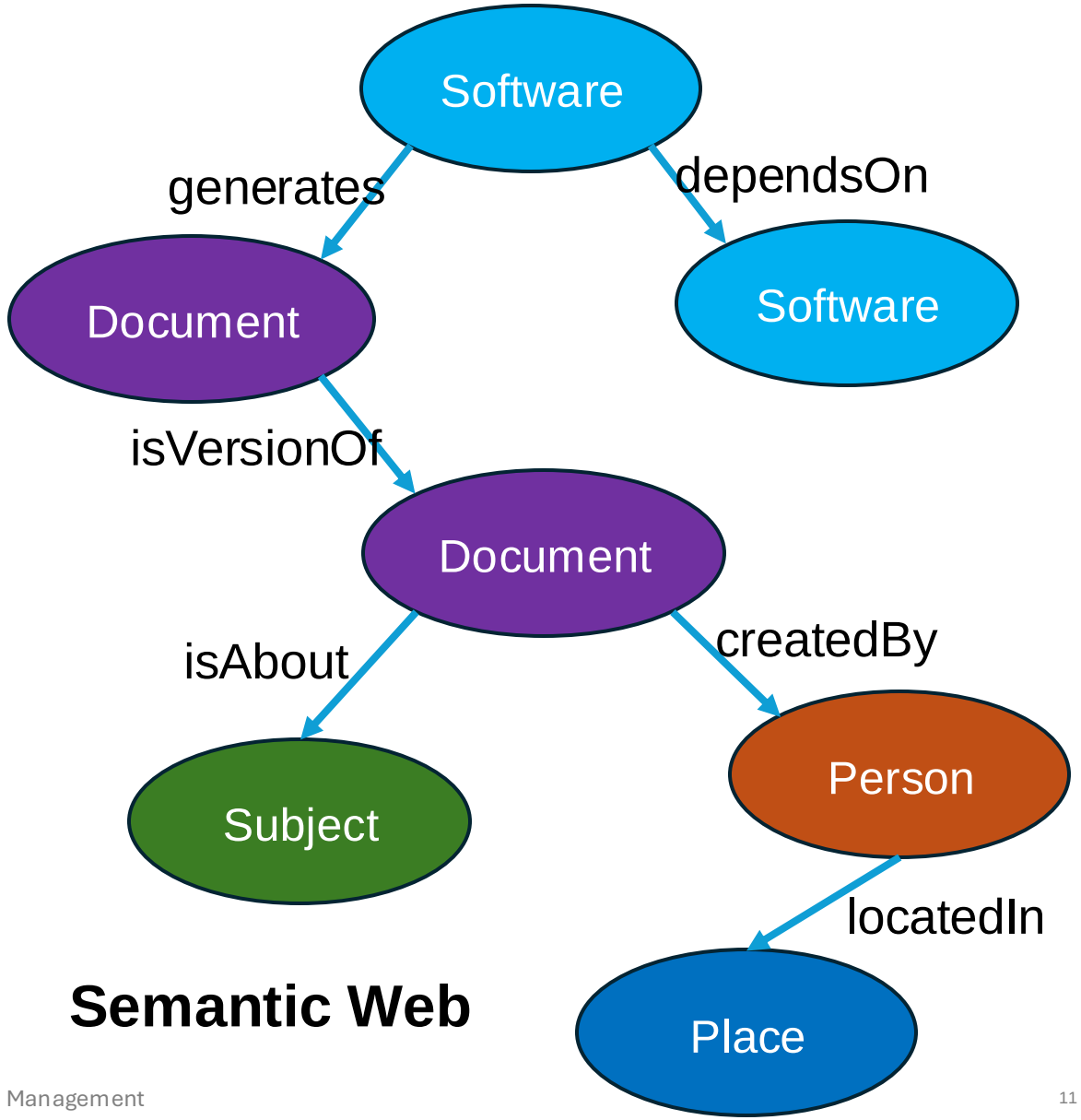
W3C Statement www.w3.org/TR/2024/STMT-ethical-web-principles-20241212/

Semantic Web Overview: Resources and links can have types



Current Web

generic hypertext references



Semantic Web

Semantic Web Standards Overview

Semantic Web guidelines published by the W3C (World Wide Web Consortium) relevant to knowledge organization systems:

- **RDF** (Resource Description Framework)
  1999, revised in 2014 as RDF 1.1
  www.w3.org/2001/sw/wiki/RDF
  Basis for any semantic data
- **RDFS** (RDF-Schema)
  2004, revised in 2014 as RDFS 1.1
  www.w3.org/2001/sw/wiki/RDFS
  Basics for ontologies
- **OWL** (Web Ontology Language),
  2004, OWL 2 in 2009 and revised in 2012
  www.w3.org/2001/sw/wiki/OWL
  More details for ontologies
- **SKOS** (Simple Knowledge Organization System), 2009
  www.w3.org/2001/sw/wiki/SKOS
  For knowledge organization systems, including thesauri and taxonomies
- **SPARQL** (SPARQL Protocol and RDF Query Language)
  2008, revised in 2013 as SPARQL 1.1
  www.w3.org/2001/sw/wiki/SPARQL
  For querying and editing an RDF-based data (including RDFS, OWL, and SKOS)

Semantic Web Standards: RDF

RDF (Resource Description Framework)

- Published by the W3C in 1999, and as RDF 1.1 in 2014
- “A standard model for data interchange on the Web”
- Facilitates data merging even if the underlying schemas differ.
- Requires the use of URIs (Uniform Resource Identifiers) to specify **things** and to specify **relationships.** May also be IRIs (Internationalized Resource Identifiers)
- Models information as **subject – predicate – object** triples.
  - Example: *Concept A* – *relates to* – *Concept B*
  - Example: *Concept A* – *has label* – *Label name*
- Models information on a graph-based model.
- More fundamental, basic, and generic than other, subsequent standards based on it.
- Detailed specification: www.w3.org/TR/rdf11-concepts

Semantic Web Standards: RDF

RDF is an abstract framework.

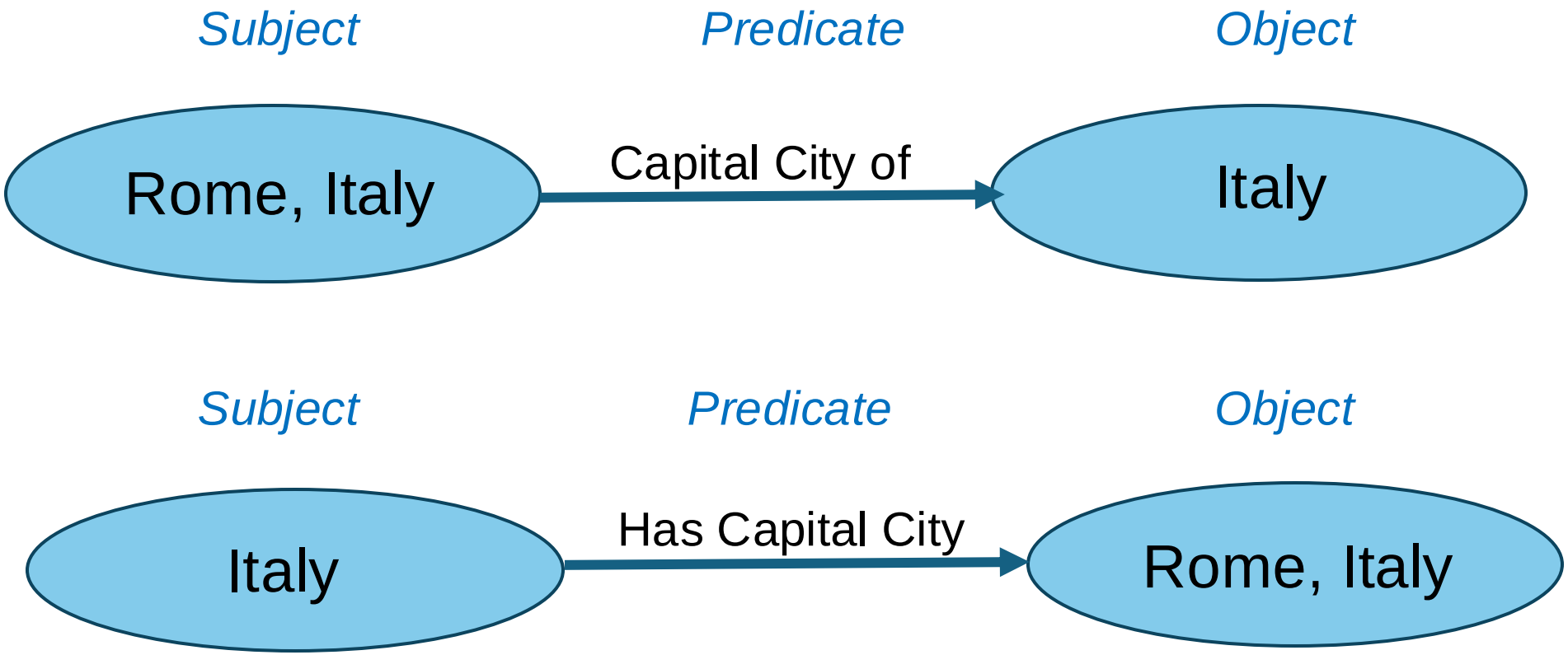
As a standard format for exchange/interoperability of data, there are various *serialization formats*:

- RDF/XML – XML-based syntax, the first standard format for serializing RDF
- Turtle – compact, human-friendly format
- N-Triples – very simple, easy-to-parse, line-based format, not as compact as Turtle
- N-Quads – superset of N-Triples, for serializing multiple RDF graphs
- JSON-LD – JSON-based serialization
- RDF/JSON – alternative syntax for expressing RDF triples using a simple JSON notation
- N3 (Notation3) – non-standard serialization similar to Turtle, but has additional features

Semantic Web Standards: RDF

RDF triple: (1) Subject – (2) Predicate – (3) Object

Example



Semantic Web Standards: RDFS

RDF Schema – RDFS or RDF/S or RDF(S)

- Also called: RDF Vocabulary Description Language 1.0
- Published as part of the RDF Specification Suite Recommendations in 2004
- “A general-purpose language for representing simple RDF vocabularies on the Web”
- A flexible data model adaptable to specific needs
- Goes beyond RDF to provide a vocabulary for designating **classes** and **properties** of RDF resources.
- RDFS serves as the basic standard for ontologies
- Detailed specification: https://www.w3.org/2001/sw/wiki/RDFS

Class: A type or category of resources or things.

- RDFS also describes subclasses and instances.

Property: Used to describe characteristics of things.

Semantic Web Standards: OWL

OWL – Web Ontology Language

- First published in 2004,
  OWL 2 (with extended features), published in 2009
- “A Semantic Web language designed to represent rich and complex knowledge about things, groups of things, and relations between things”
- Provides a common way to process the content of web information
- A computer-readable language, usually written in XML,
  a declarative language (not a programming or schema language)
- Enables knowledge linking on the web/Semantic Web
- Based on RDF and RDFS. OWL is an extension of RDFS.
- Detailed specification: www.w3.org/TR/owl2-overview

Semantic Web Standards: SKOS

SKOS – Simple Knowledge Organization System

- A data model to represent knowledge organization systems.
- A KOS built on SKOS is machine-readable and interchangeable.
- Published by the W3C in 2009, following the ontology standards.
- Based on RDF and encoded in XML or other various RDF serializations
- Different knowledge organization system types (taxonomies, thesauri, classification systems, etc.) can all be built on the SKOS standard.
- Enables easy publication and use of such vocabularies as linked data.
- Detailed specification: www.w3.org/TR/skos-reference

Semantic Web Standards: SKOS

SKOS Principles

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

Semantic Web Standards: SKOS

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	broadMatch
orderedCollection		notation	changeNote		narrowMatch
member			editorialNote		relatedMatch
memberList			historyNote		

In computer-readable form, for example: skos:Concept

SKOS

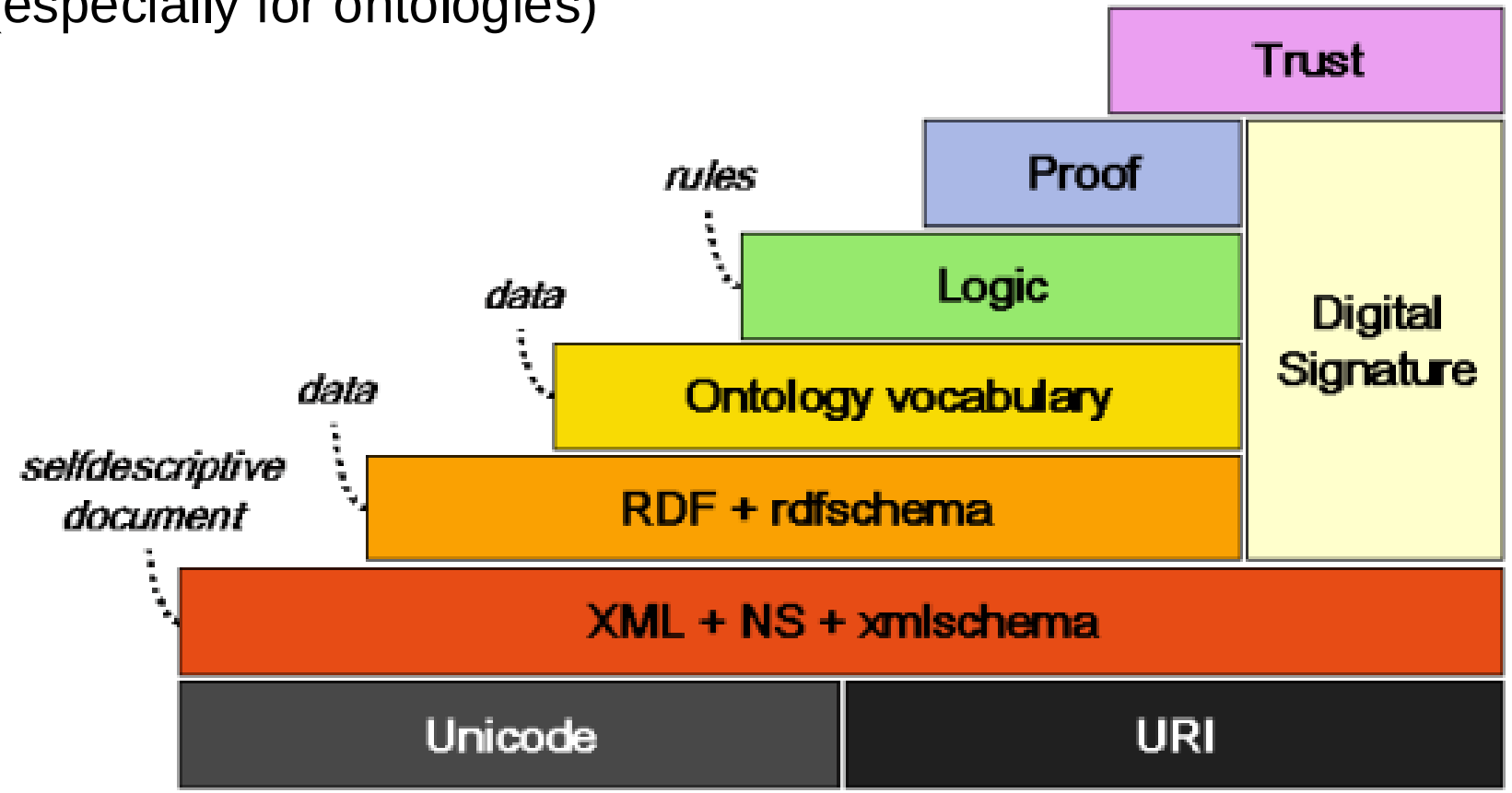
SKOS elements in a human-readable form in the user interface of a taxonomy management system.

Users can add relationships, labels, and documentation notes.

Semantic Web Technology Summary

Semantic Web Layers

(especially for ontologies)



Knowledge Organization Systems

Knowledge Organization System Types

Knowledge Organization system (KOS)

- Any system of concepts, terminology, classification, etc.

  to organize, define, manage, and/or retrieve information.

- Not a method to organize knowledge directly,

  but rather a *scheme* to organize concepts for organizing, classifying, defining, tagging, or retrieving information.

- Broader, includes more than just “controlled vocabularies”

KOS types:

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

Controlled Vocabularies for tagging and information retrieval

Knowledge Organization System Types

Controlled vocabularies

- Term lists/pick lists
- Authority files
  - Name authorities
- Taxonomies
- Subject heading schemes
- Thesauri

Defined vocabularies

- Glossaries
- Gazetteers
- Terminologies

Classification systems

- Categorization schemes
- Classification schemes

Semantic models

- Mind maps
- Topic maps
- Ontologies

Supported by SKOS, fully

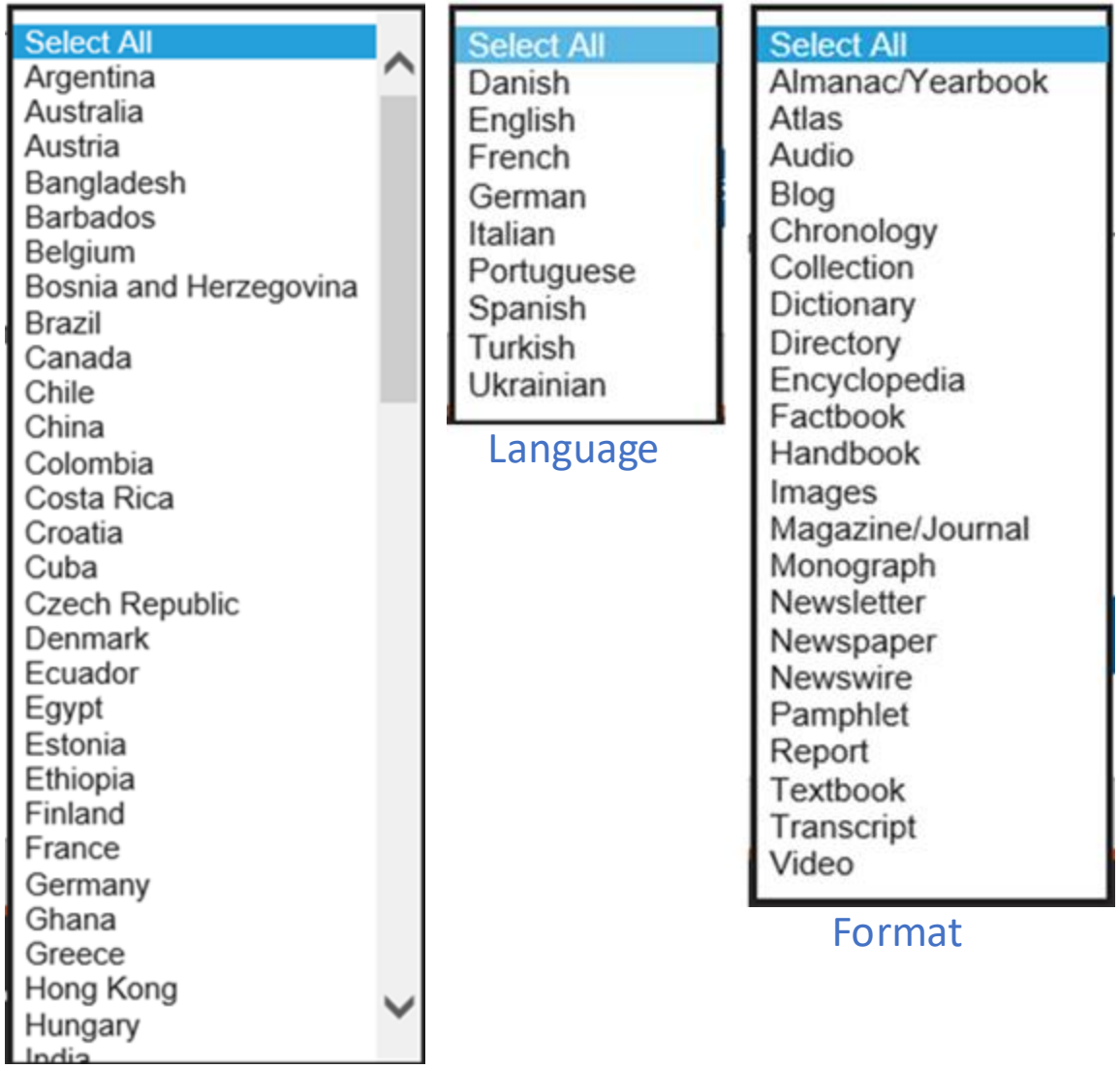
Supported by SKOS, partially

Supported by RDFS and OWL

Controlled Vocabulary Types

Term list

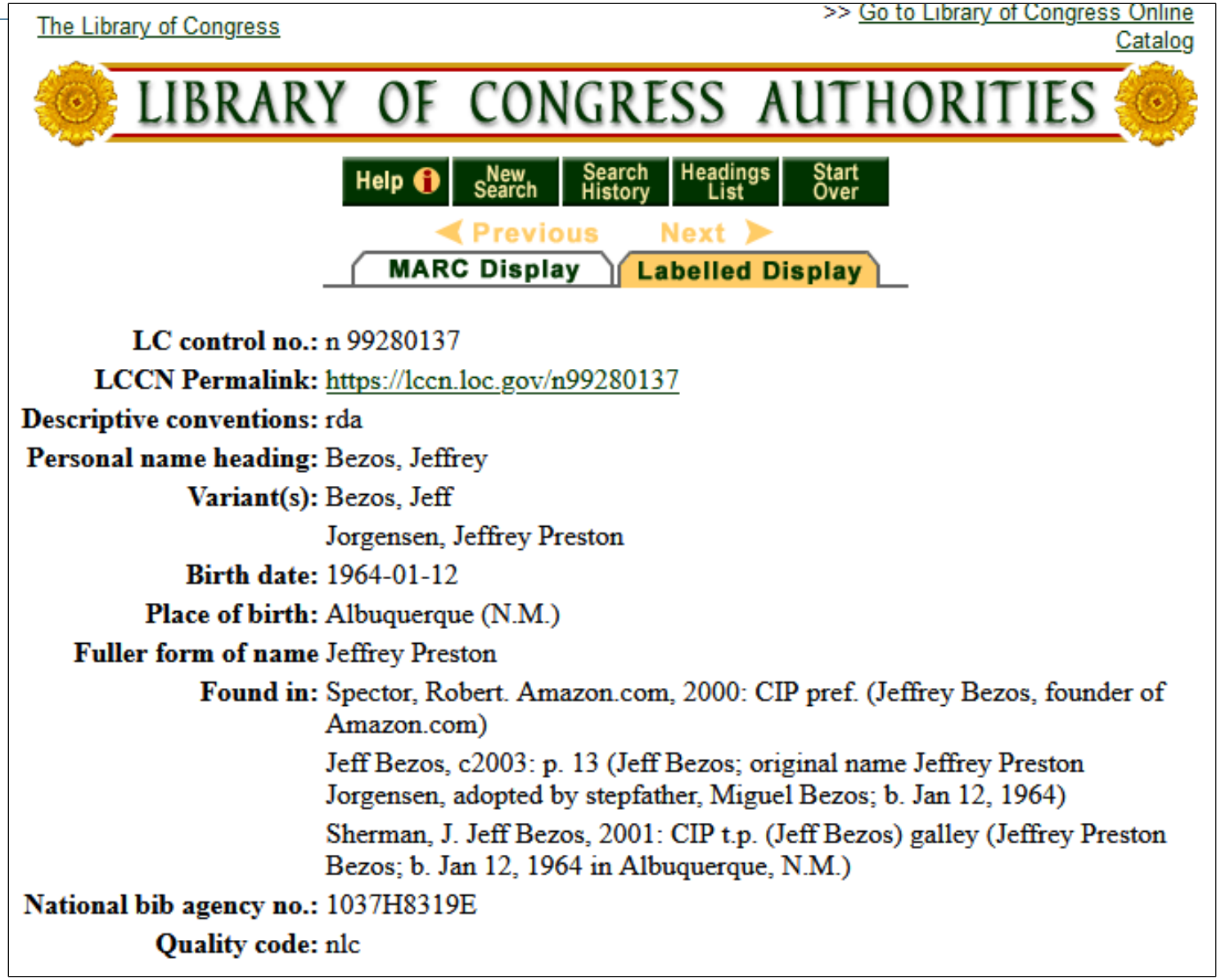
- A simple list of terms
- Usually alphabetical, but could be in other logical order
- Lacking synonyms, it is usually short enough for quick browsing
- Can display to users in drop-down scroll boxes
- May be used for various metadata values, facets, concept schemes
- Part of a larger set of controlled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



Controlled Vocabulary Types

Name authority

- For named entities, concrete entities, proper nouns
- Has preferred names and variant/alternative names.
- Typically has no hierarchical relationships.
- Usually has additional information/attributes (metadata) for each entity.


The Library of Congress

>> Go to Library of Congress Online Catalog

LIBRARY OF CONGRESS AUTHORITIES

Help | New Search | Search History | Headings List | Start Over

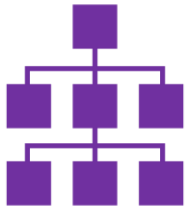
◀ Previous Next ▶

MARC Display | Labelled Display

LC control no.: n 99280137
LCCN Permalink: https://lccn.loc.gov/n99280137
Descriptive conventions: rda
Personal name heading: Bezos, Jeffrey
Variant(s): Bezos, Jeff
Jorgensen, Jeffrey Preston
Birth date: 1964-01-12
Place of birth: Albuquerque (N.M.)
Fuller form of name Jeffrey Preston
Found in: Spector, Robert. Amazon.com, 2000: CIP pref. (Jeffrey Bezos, founder of Amazon.com)
Jeff Bezos, c2003: p. 13 (Jeff Bezos; original name Jeffrey Preston Jorgensen, adopted by stepfather, Miguel Bezos; b. Jan 12, 1964)
Sherman, J. Jeff Bezos, 2001: CIP t.p. (Jeff Bezos) galley (Jeffrey Preston Bezos; b. Jan 12, 1964 in Albuquerque, N.M.)
National bib agency no.: 1037H8319E
Quality code: nlc

Controlled Vocabulary Types

Taxonomy



- A KOS with broader/narrower relationships that includes all concepts to create a hierarchical structure.
- Has a focus on categorizing and organizing 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 is usually a concept scheme and can be considered as a hierarchy.)
- May or may not include named entities.

Hierarchical taxonomy

- 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

Faceted taxonomy

Career Level
- Student
- Entry Level
- Experienced
- Manager
- Director
- Executive

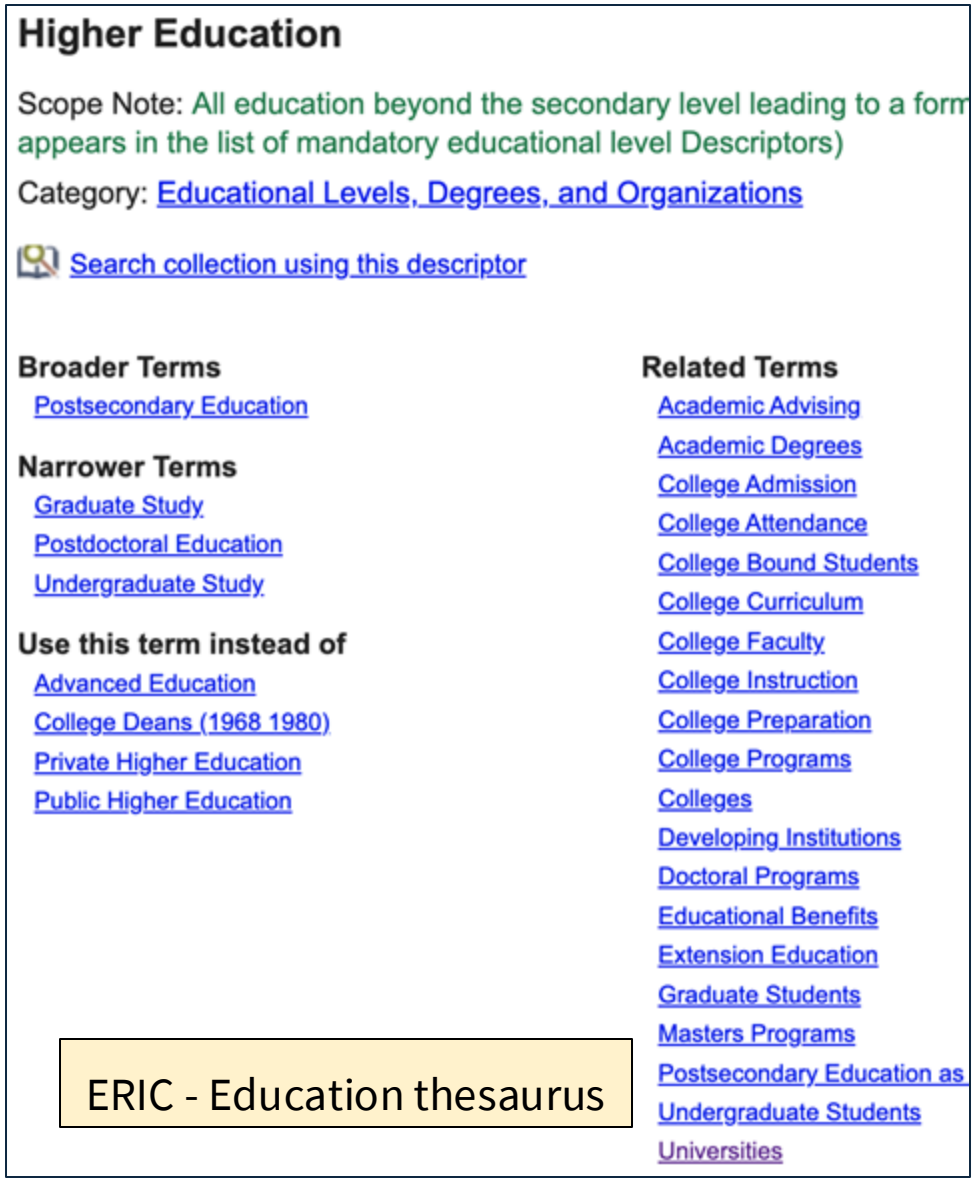
Function
- Customer Service & Support
- Delivery
- Engineering
- Finance
- General Management
- Legal & Regulatory Affairs
- Marketing & Advertising
- [more]

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

Controlled Vocabulary Types

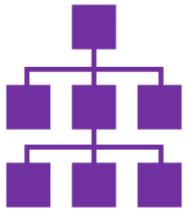
Thesaurus

- 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 best-practice 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

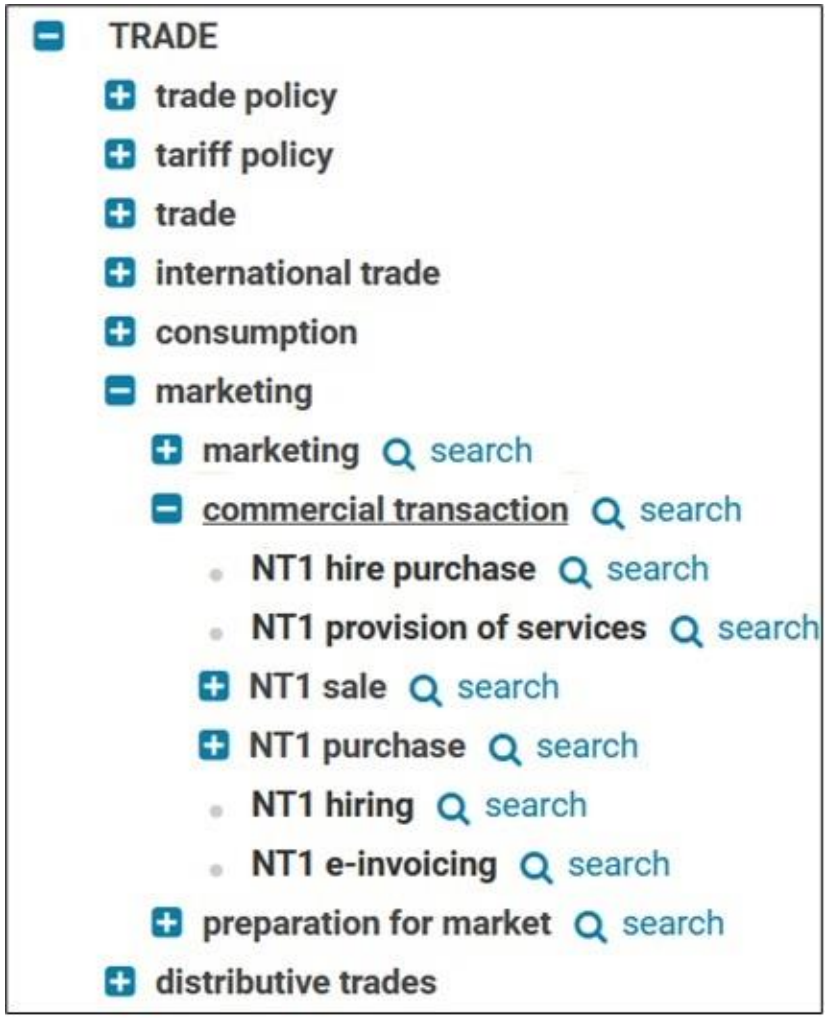


Controlled Vocabulary Types

Taxonomy/thesaurus hierarchy purposes



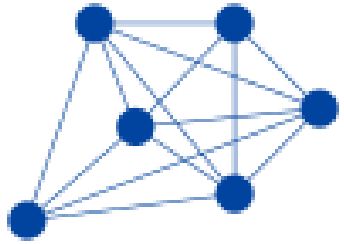
1. Serving users who are browsing, exploring, discovering, not searching.
2. Instructing users on appropriate classification
3. Providing context to terms for manual indexers/taggers so that they apply the correct term.
4. Providing the context of a broader concept and thus meaning to aid in auto-classification/auto-tagging.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



Eurovoc Thesaurus excerpt

Ontologies

Ontology definition



- A model of a knowledge domain
- A form of knowledge representation; not just knowledge organization
- Comprises classes, relations, and attributes, which are linked in triples.
  - Relations contain meaning, are “semantic.”
- A set of precise descriptive statements about a particular domain.
  - Statements as subject-predicate-object are expressed as triples.
- A more abstract layer in describing a knowledge organization system
  - Overlays and connects to a taxonomy or other controlled vocabulary to add semantics
- Common standards provided by W3C:
  Web Ontology Language (OWL) and RDF-Schema (RDFS)

Ontology Components

OWL-Defined Ontology Components

Entities – subjects or objects of properties, of triples

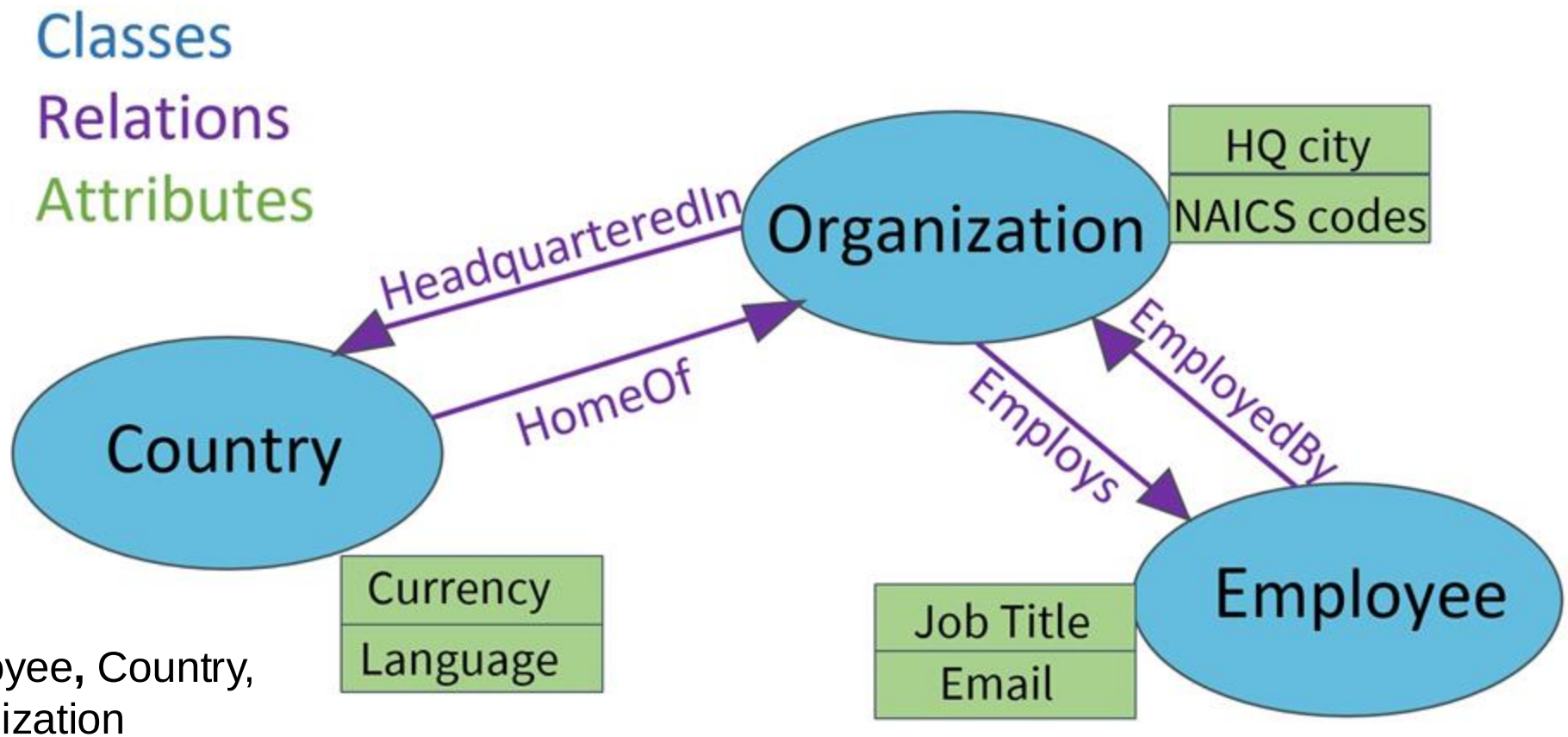
- **Classes**
  - Named sets of things that share characteristics and relations
  - May group subclasses or individuals
- **Individuals**
  - Members or instances of a class

Properties – predicates about individuals (instances)

- **Object properties**
  - Relations between individuals
  - May be directed (single direction), symmetric, or with an inverse (different in each direction)
- **Datatype properties**
  - Attributes or characteristics of individuals
  - The object of a datatype property is a *value*

Literals – values of attributes, with just a *lexical form* and a *datatype*

Ontology Components



Classes

Employee, Country, Organization

Relations: HeadquarteredIn < > HomeOf
EmployedBy < > Employs

Attributes: Email address, Job title, HQ city, NAICS codes, Currency, Language

Knowledge Organization with Increasing Semantics

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

Knowledge Organization with Increasing Semantics

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

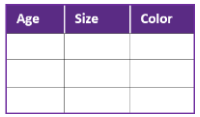
Ontologies, thus link other knowledge organization systems together and enable linking their information

Knowledge Organization with Increasing Semantics

Metadata

provides:

standardized structured information about content items consistently across a set of content.



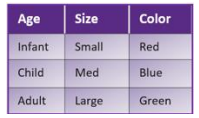
Enables:

- Machine-readability
- Workflow management
- Content reuse
- Basic search and findability
- Compliance/rights management

Controlled vocabularies

provide:

standardized *values* for many metadata properties.



Enables:

- Accuracy
- Consistency
- Completeness
- Better search and findability

Taxonomies & Thesauri

provide:

a user-friendly way to *browse*, *navigate, and filter* with controlled vocabularies

Enables:

- Efficiency
- Better user experience
- Better tagging
- Discovery

Ontologies

provide:

a data model and a semantic method of linking taxonomies and other controlled vocabularies

Enables:

- Search across datasets, not just content
- Complex multi-part searches
- Reasoning and inferencing
- Semantic access to linked repositories

Standards for Linking Knowledge Organization Systems

Linking Vocabularies with SKOS

SKOS supports links across different concept schemes: different thesauri and taxonomies

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."

- Exact Match exact match in meaning, bidirectional, in all circumstances/contexts
- Close Match match in meaning, bidirectional, in a specific circumstance/context
- Broad Match has broader concept in the other vocabulary; inverse of Narrow Match
- Narrow Match has narrower concept in the other vocabulary; inverse of Broad Match
- Related Match has related concept in the other vocabulary; bidirectional

- Exact Match and Close Match may link taxonomies in use to expand the number of documents that can be retrieved within the same, existing subject area scope.
- Broad Match, Narrow Match, and Related Match expand the subject scope of vocabularies by linking them together, thus expanding the scope of content retrieval.

Linking Vocabularies with SKOS

Possible reasons to link vocabularies

- Link to a standard, published vocabulary/classification scheme for alignment.
  - Involves **Exact Match** only
- Use one taxonomy in the user interface to retrieve additional content already tagged with a different taxonomy (also called “mapping”).
  - Involves **Exact Match**, possibly **Close Match**, and **Narrow Match** in one direction
- Enrich a taxonomy with concepts from another controlled vocabulary (“mapping”).
  - Involves **Exact Match**, possibly **Close Match**, and **Narrow Match** in one direction
- Combine two or more taxonomies to extend them, but each still remains intact.
  - May involve all match types
- Compare and align taxonomies prior to fully merging them (with one absorbed into the other taxonomy).
  - May involve all match types

Linking Vocabularies with SKOS

Vocabulary linking type directions

Directional - "Mapping"
from one taxonomy to another with sufficiently equivalent or narrower-to-broader hierarchical links

- One taxonomy may be used for another in the front end.
- A taxonomy can be enriched with added concepts.
- May serve as first step for merging controlled vocabularies.

Bidirectional

- With equivalent (exact/close match) links, so that taxonomies and content can be shared.
- With associative and/or hierarchical links, so that users can navigate to new content.

Linking Ontologies

OWL supports links across different ontologies, also referred to as “mapping”

https://www.w3.org/TR/2004/REC-owl-guide-20040210/#OntologyMapping

- Equivalent class – owl:equivalentClass
  States that two *classes* with different URIs have identical meaning and thus may have the same set of members.
- Equivalent property – owl:equivalentClass
  States that two *relations* or *attribute types* with different URIs have identical meaning.
- SameAs – owl:sameAs
  States that two *individuals* with different URIs, and often also different names, have the same identity. Everything stated about one entity hold for the corresponding entity (i.e., they share all of their properties).
  - In Linked Data, the use of *owl:sameAs* is the most common method of interlinking data-sets, but it has been misused for links more generic than its strict meaning.

RDFS also has a linking relationship: See Also - rdfs:seeAlso

seeAlso does not suggest full identity between the linked entities, but indicates that a related entity provides additional information.

ISO Standards for Vocabulary Linking

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

Standards for best practices, rather than for machine-readable data exchange

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: classification schemes, subject heading schemes, taxonomies, terminologies, name authority lists, and ontologies.

Linked Knowledge Organization Systems on the Web

Linked Open Data

Linked Data

- Structured data, interlinked with other data, so is more useful through semantic queries.
- Built upon standard Web technologies such as HTTP, RDF and URIs.
- A set of design principles for sharing machine-readable interlinked data on the Web.

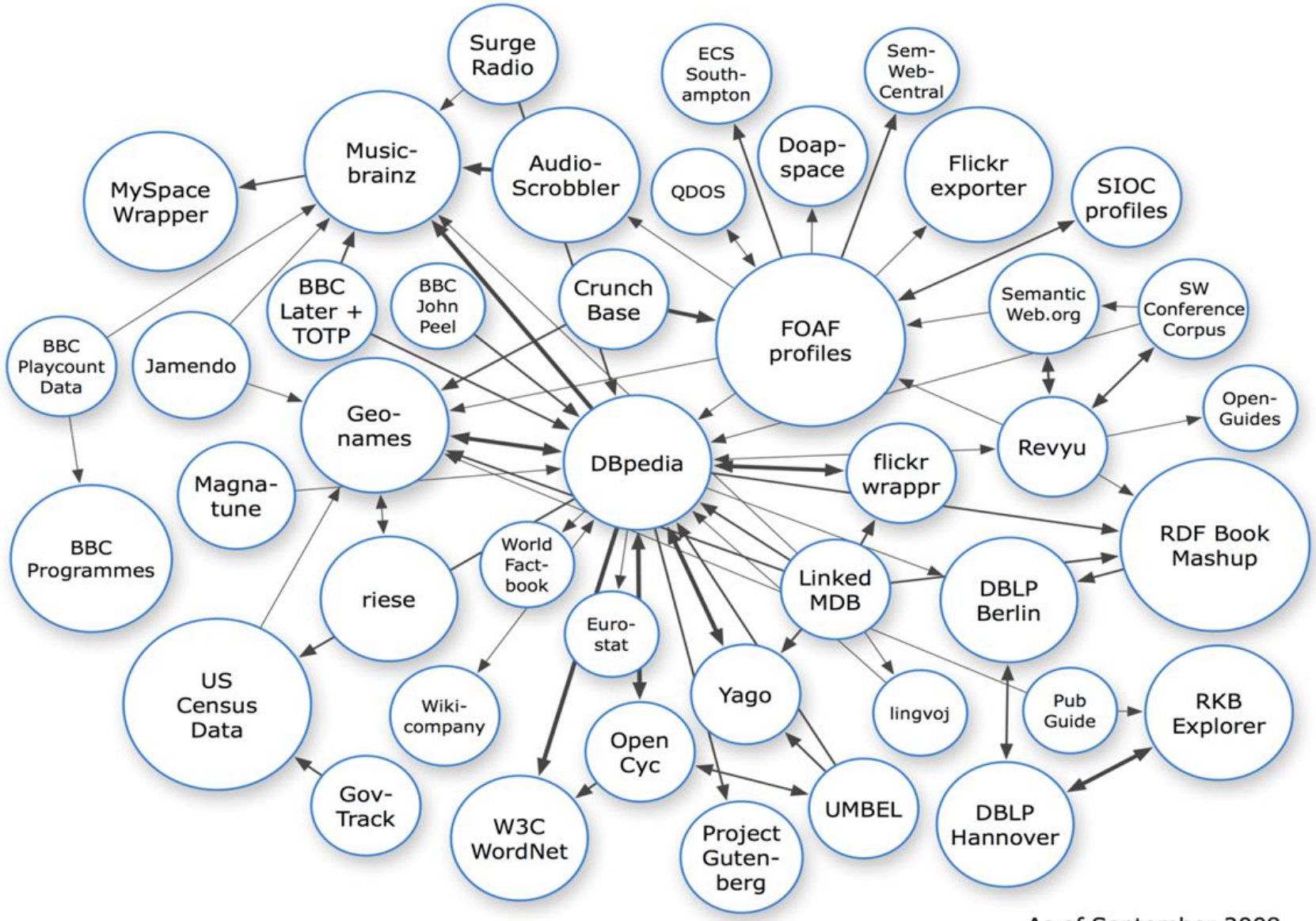
Linked Open Data (LOD)

- Linked Data which is released under an open license, which does not impede its reuse for free.
- A set of best practices for sharing data on the Web that can be accessed and reused by both humans and machines.
- A core part of the Semantic Web.

LOD Principles and Features

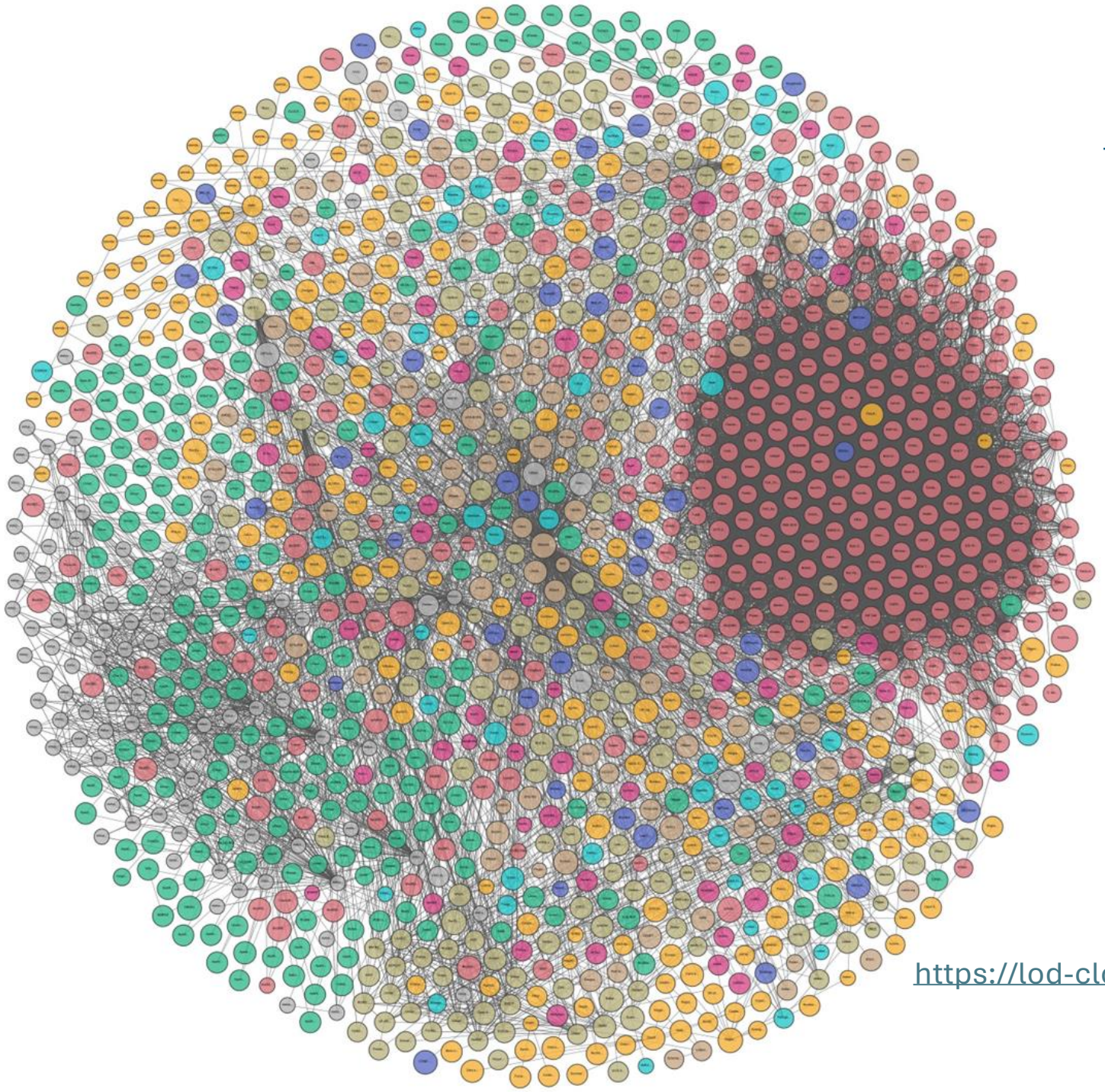
- The data must be accessible under an open license.
- Data must be linked using common and predefined standards.
- Each element is assigned a unique identifier, such as a uniform resource identifier (URI).
- Relationships between elements are expressed using triples.

Linked Open Datasets



https://lod-cloud.net/

Linked Open Datasets



https://lod-cloud.net/

Linked Open Datasets: DBpedia

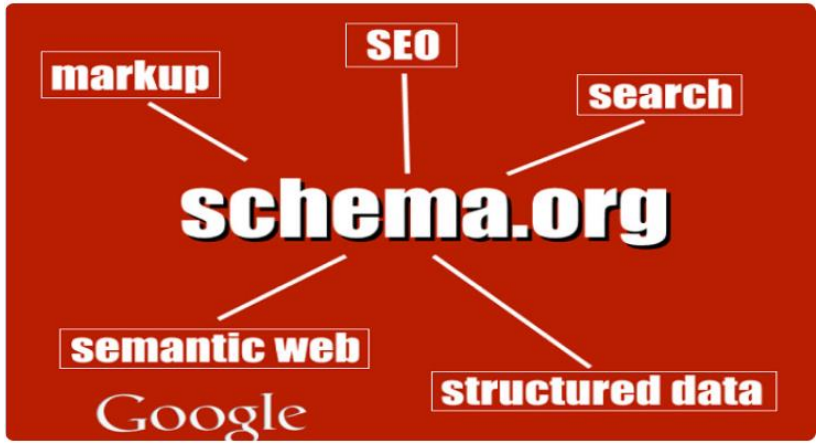
- DBpedia is the most interlinked LOD dataset and a central point of the Linked Open Data Cloud since 2008.
- An RDF-based dataset containing extracted structured content and data from Wikipedia, first published in 2007.
- Includes an ontology of 768 classes, 3000 different properties, 4,233,000 instances.
- Comprises over 1 billion triples, with data in 11 different languages.
- Structured information is made available on the World Wide Web using OpenLink Virtuoso.
- Allows users to semantically query relationships and properties of Wikipedia resources, including links to other related datasets.
- Project started at Free University of Berlin and Leipzig University, now maintained by people at the University of Mannheim and Leipzig University.

https://www.dbpedia.org/

Supporting Search via Linked Open Data: Schema.org

- An initiative launched on June 2, 2011, by major web search engine companies (Bing, Google, Yahoo!, and Yandex) to create and support a common set of schemas for structured data markup on web pages.
- Considered part of the Semantic Web initiative, by making document mark-up codes more readable and meaningful to both humans and machines.
- A collection of vocabularies and schemas to enrich HTML pages with additional information.
- The vocabulary of Schema.org includes a set of classes and their properties, including Organization, Person, Event, and Place. Resembles an ontology.
- There are mappings from Schema.org vocabularies and microdata to RDFS.

http://schema.org



Linked Open Vocabularies (LOV)

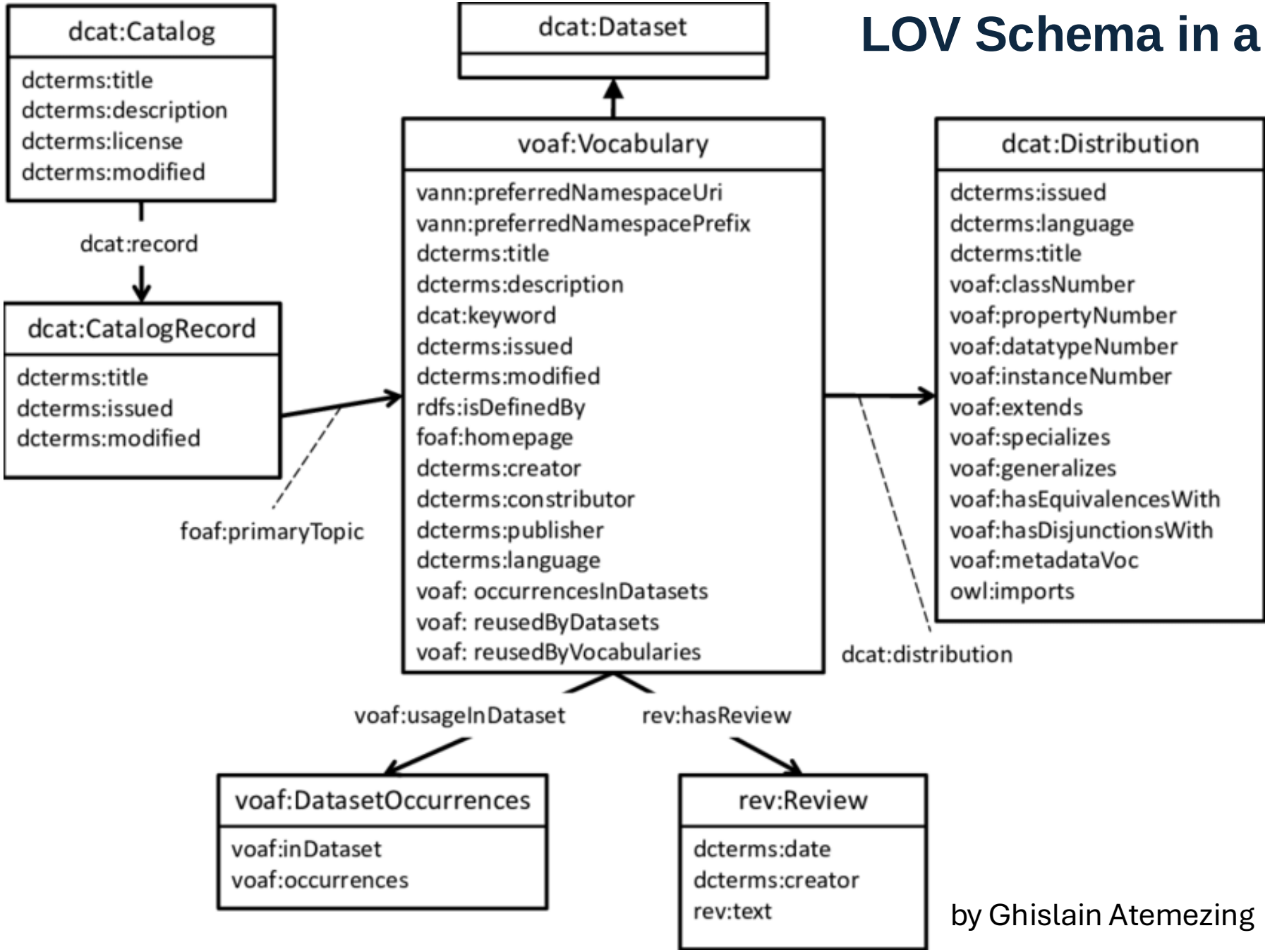
Linked Open Vocabularies (LOV) initiative

- A curated set of vocabularies (metadata element sets or ontologies) in OWL / RDFS used to describe linked data. (lov.linkeddata.es/dataset/lov) Now contains 835 vocabularies.
- Hosted by the Ontology Engineering Group at Universidad Politécnica de Madrid.
- Provides a single point of access to multiple vocabularies.
- Helps improve vocabularies' understanding, visibility, usability, synergy, sustainability and quality.
- Promotes a technically and socially sustainable management of the vocabularies
- A community and open project, started in 2011

Activities:

- Collecting new vocabularies from the LOV Community
- Tracking and analysis of the LOV vocabulary catalogue
- Giving access to the data, using various methods: search engine, metadata search, ontology search, data dumps, SPARQL endpoint, and APIs
- Gathering indicators of linkages between vocabularies: versioning history, maintenance policy, past and current referents of individuals and organization.

LOV Schema in a UML Class Diagram



by Ghislain Atemezing

Prefixes - namespaces:

dcat – Data Catalog Vocabulary – describes metadata about datasets

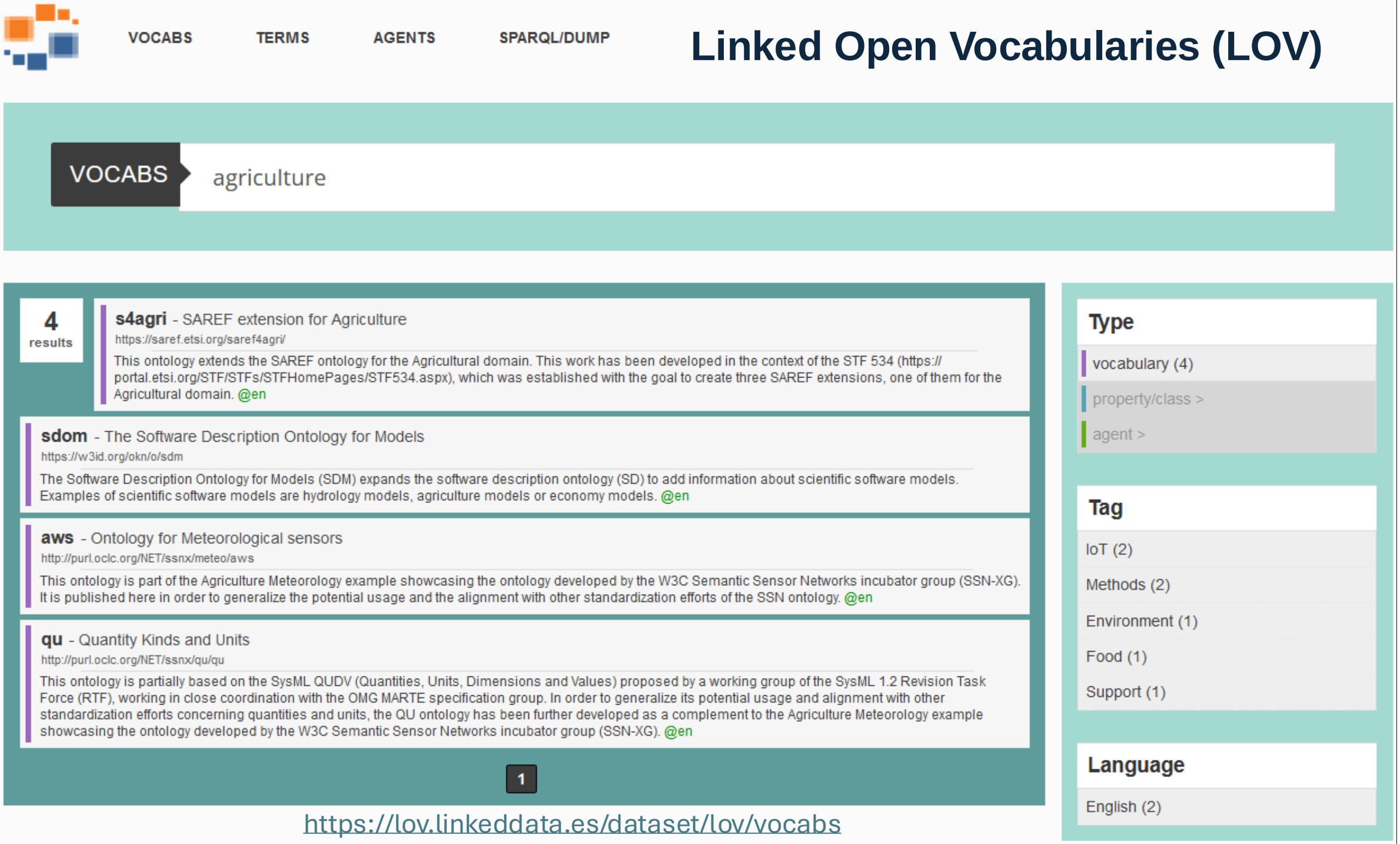
dcterms - Dublin Core Metadata Terms - set of common properties for describing digital resources: creator, title, subject, description, etc.

vann - Vocabulary of Annotation for Describing Namespaces

voaf - Vocabulary of a Friend - describes and relates vocabularies to each other

foaf - Friend of a friend - ontology in RDF and OWL describing people, their activities, and their relationships.

Linked Open Vocabularies (LOV)



https://lov.linkeddata.es/dataset/lov/vocabs

TERMS price

179 results

cerif:price (cerif) 0.556
n/a (use in LOD)
http://www.eurocris.org/ontologies/cerif/1.3#price
rdfs:label **Price**
localName **price**

geop:GDPTotalInCurrentPrices (geop) 0.527
192 occurrences in 1 LOD datasets
http://aims.fao.org/aos/geopolitical.owl#GDPTotalInCurrentPrices
localName GDPTotalInCurrentPrices

cbo:price (cbo) 0.470
n/a (use in LOD)
http://comicmeta.org/cbo/price
rdfs:label **price** @en
rdfs:comment A general **price**. @en
localName **price**

dcndl:price (dcndl) 0.455
n/a (use in LOD)
http://ndl.go.jp/dcndl/terms/price
rdfs:label **Price**
localName **price**

schema:price (schema) 0.455
n/a (use in LOD)
http://schema.org/price
rdfs:comment The offer **price** of a product, or of a **price**
rdfs:label **price**
localName **price**

Type

vocabulary >
property/class
property (121)
class (58)
agent >

Tag

General & Upper (31)
Environment (28)
eBusiness (23)
Methods (18)
Contracts (17)
Industry (11)
Metadata (10)
Multimedia (7)

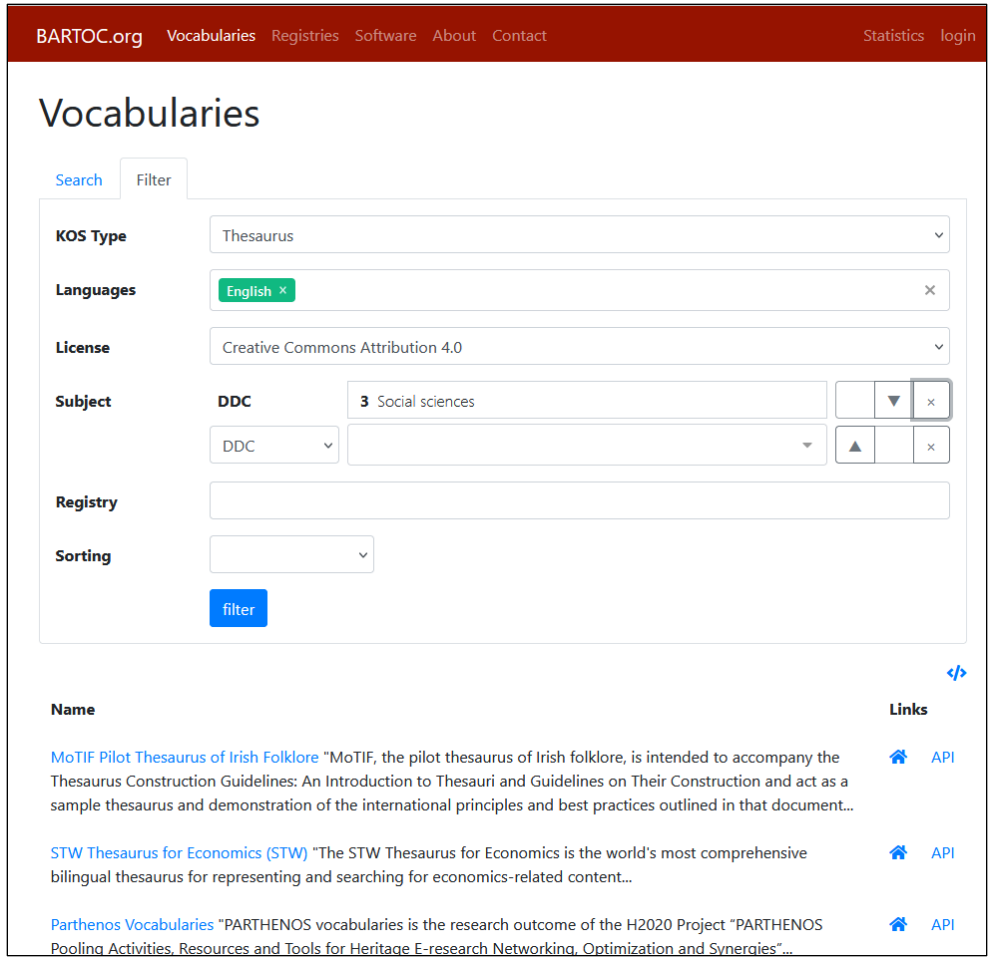
https://lov.linkeddata.es/dataset/lov/terms

Re-Using Knowledge Organization Systems

Using a shared open vocabulary to link information

The same KOS could be reused for different content repositories by different organizations for information sharing.

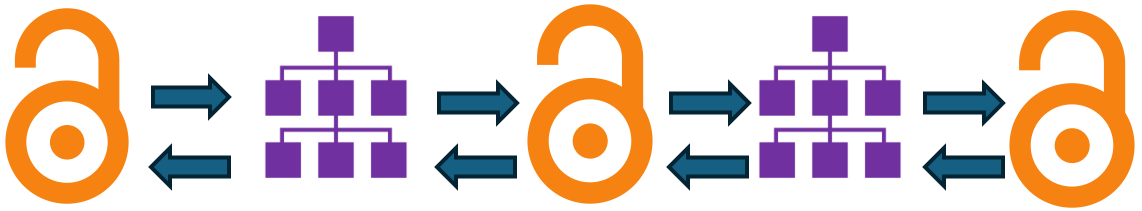
- BARTOC (Basic Register of Thesauri, Ontologies & Classifications), the largest registry of knowledge organization systems
  - Many of which are available by open license for sharing and reuse (Creative Commons Attributions, Open Data Commons, Open Government License, etc.), especially if not for commercial use and not for creating derivatives.
- NISO published a technical report in 2017: TR-06-2017 *Issues in Vocabulary Management,* which addresses taxonomy reuse. www.niso.org/publications/tr-06-2017-issues-vocabulary-management



https://bartoc.org/vocabularies

Conclusions

- Open Access Publishing principles alone do not ensure findability and discoverability of information.
- Information needs to be assigned metadata and tagged/indexed with controlled vocabularies, and these vocabularies can and should be linked together, if not shared.
- Semantic Web standards and guidelines provide a way to link metadata, controlled vocabularies, and knowledge organization systems on the Web so that information is linked and shared.
- Ontologies, also based on Semantic Web standards, by their nature, also link knowledge organization systems together, and with greater, added semantics
- Initiatives making use of these semantic standards on methods to link information, such as Linked Open Data (LOD), Linked Open Vocabularies (LOV) and DBpedia, have grown, but could be implemented much further.
- Open Access Publishing should take greater advantage of the Semantic Web.



Further Reading

- Vandenbussche, Pierre-Yves & Atemezing, Ghislain & Poveda-Villalón, María & Vatant, Bernard. (2017). “Linked Open Vocabularies (LOV): A gateway to reusable semantic vocabularies on the Web.” *Semantic Web Journal*. 8. 437-452. 10.3233/SW-160213. https://www.semantic-web-journal.net/system/files/swj1178.pdf
- Abdel-Qader, M., Vagliano, I., Scherp, A. (2019). “Analyzing the Evolution of Linked Vocabularies.” In: Bakaev, M., Frasincar, F., Ko, IY. (eds) *Web Engineering*. ICWE 2019. Lecture Notes in Computer Science, vol 11496. Springer, Cham. https://doi.org/10.1007/978-3-030-19274-7_29
- Bauer, Florian, and Martin Kaltenböck. (2011). “Linked Open Data: The Essentials.” *Edition mono/monochrom, Vienna* 710.21 https://cdn.semantic-web.com/wp-content/uploads/2017/05/LOD-the-Essentials_0.pdf
- Yu, Liyang, and Liyang Yu. (2011) “Linked open data.” In *A Developer’s Guide to the Semantic Web* : 409-466. https://books.google.com/books?id=kf8sBAAAQBAJ&lpg=PR5&ots=BGIyTC7Pbb&dq=%22Linked%20Open%20data%22
- Lehmann, Jens, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N. Mendes, Sebastian Hellmann et al. (2015) "Dbpedia – A large-scale, Multilingual Knowledge Base Extracted from Wikipedia." *Semantic web* 6, no. 2 : 167-195. https://www.researchgate.net/publication/259828897_DBpedia_-_A_Large-scale_Multilingual_Knowledge_Base_Extracted_from_Wikipedia
- Godby, Carol Jean. (2016.) “A Division of Labor: The Role of Schema.org in a Semantic Web Model of Library Resources.” In *Linked Data for Cultural Heritage* (Alcts Monograph), Ed Jones (ed.). Chicago: Amer Library Assn Editions. https://www.alastore.ala.org/content/linked-data-cultural-heritage-alcts-monograph
- Vandenbussche, Pierre-Yves & Vatant, Bernard. (2011). “Metadata Recommendations For Linked Open Data Vocabularies.” https://lov.linkeddata.es/Recommendations_Vocabulary_Design.pdf

Questions/Contact

Heather Hedden
Taxonomy Instructor/Author
Hedden Information Management
Carlisle, MA USA
+1 978-467-5195
heather@hedden.net
www.hedden-information.com
accidental-taxonomist.blogspot.com
www.linkedin.com/in/hedden