



Core Concepts of Semantic Intelligence

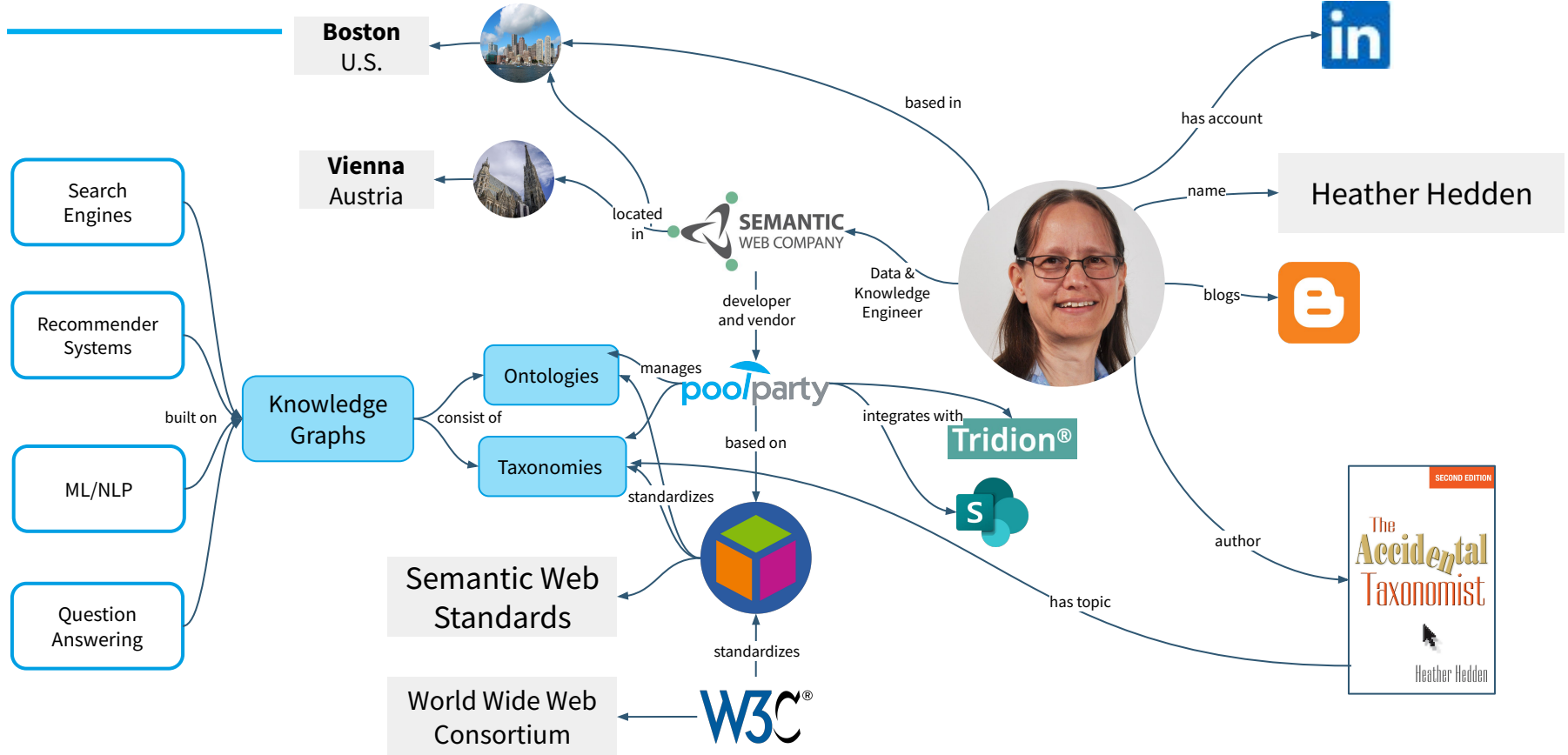
Semantic Content Graph Guild

June 22, 2022

The background of the slide is a light blue-grey color. It features a top-down view of a person's hands and arms working at a desk. One hand is holding a pen over a tablet displaying a video conference grid. Another hand is holding a pen over a tablet displaying a colorful Venn diagram. A network diagram with various icons (globe, gear, bar chart, code) is overlaid on the right side of the image.

Heather Hedden
Data and Knowledge Engineer,
Semantic Web Company

Click the Graph—get in contact with us!



- ▶ Introduction to knowledge organization systems
- ▶ Semantic Web standards for knowledge organization systems
- ▶ Why taxonomies
- ▶ Why ontologies and knowledge graphs

What is semantics?

- ▶ Semantics = meaning
- ▶ Thoughts, ideas, concepts, values for things and kinds of relations between things
- ▶ Not mere words, text strings, and unqualified links
- ▶ Semantics is about “things, not strings” (of text)

Why semantics?

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

Semantics is not new; it has existed in:

- ▶ Dictionaries, glossaries, terminologies, indexes, classification systems, taxonomies, thesauri, mind maps/semantic networks
- ▶ In print or in stand-alone files

Semantics that is machine-readable, interoperable, and supports data linking, is newer:

- ▶ Guidelines, standards, and methods (languages) from the World Wide Web Consortium (W3C)
- ▶ “for representing interconnected data on the web”
to enable the Semantic Web (Web 2.0) for semantically linked data
- ▶ Supporting knowledge organization systems (taxonomies, thesauri, ontologies, etc.) in new standard machine-readable forms

A blue rectangular box with a white vertical line on its left side. Inside the box, the text 'Knowledge organization systems' is written in a bold, black, sans-serif font, arranged in two lines.

**Knowledge organization
systems**

Knowledge Organization system (KOS)

- ▶ Any system of terms, terminology, classification, etc.
- ▶ to organize, define, manage, and/or retrieve information.
- ▶ Not any 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 information
retrieval

Knowledge Organization System Types

Controlled vocabularies

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

Defined vocabularies

- ▶ Dictionaries
- ▶ Glossaries
- ▶ Gazetteers
- ▶ Terminologies

Classification systems

- ▶ Cataloging systems
- ▶ Categorization schemes
- ▶ Classification schemes

Semantic models

- ▶ Mind maps
- ▶ Topic maps
- ▶ Semantic networks
- ▶ Ontologies

Common types of controlled vocabularies

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

“Taxonomy” sometimes means any controlled vocabulary.

Knowledge Organization Systems: Term List

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
- ▶ Often used for various metadata values or search refinements/facets
- ▶ Part of a larger set of controlled vocabularies, part of a KOS



Country of publication



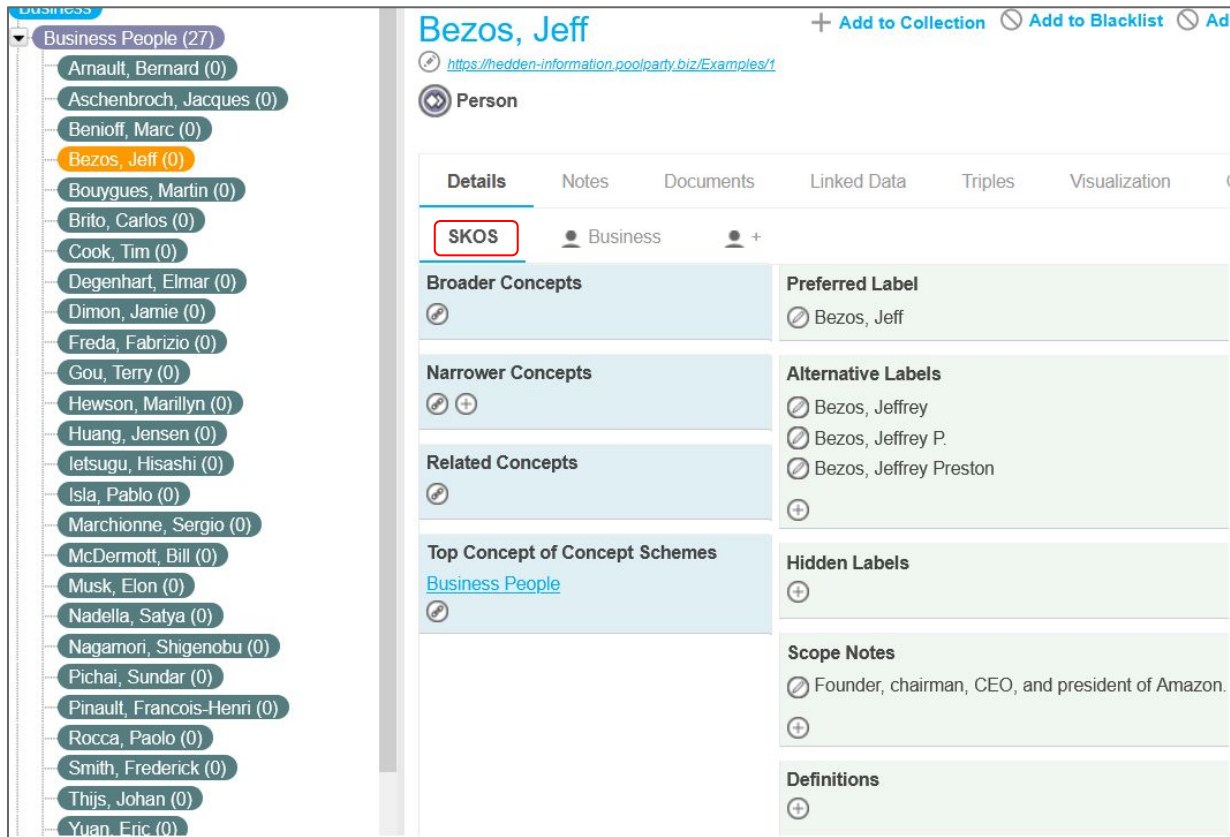
Language



Format

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.

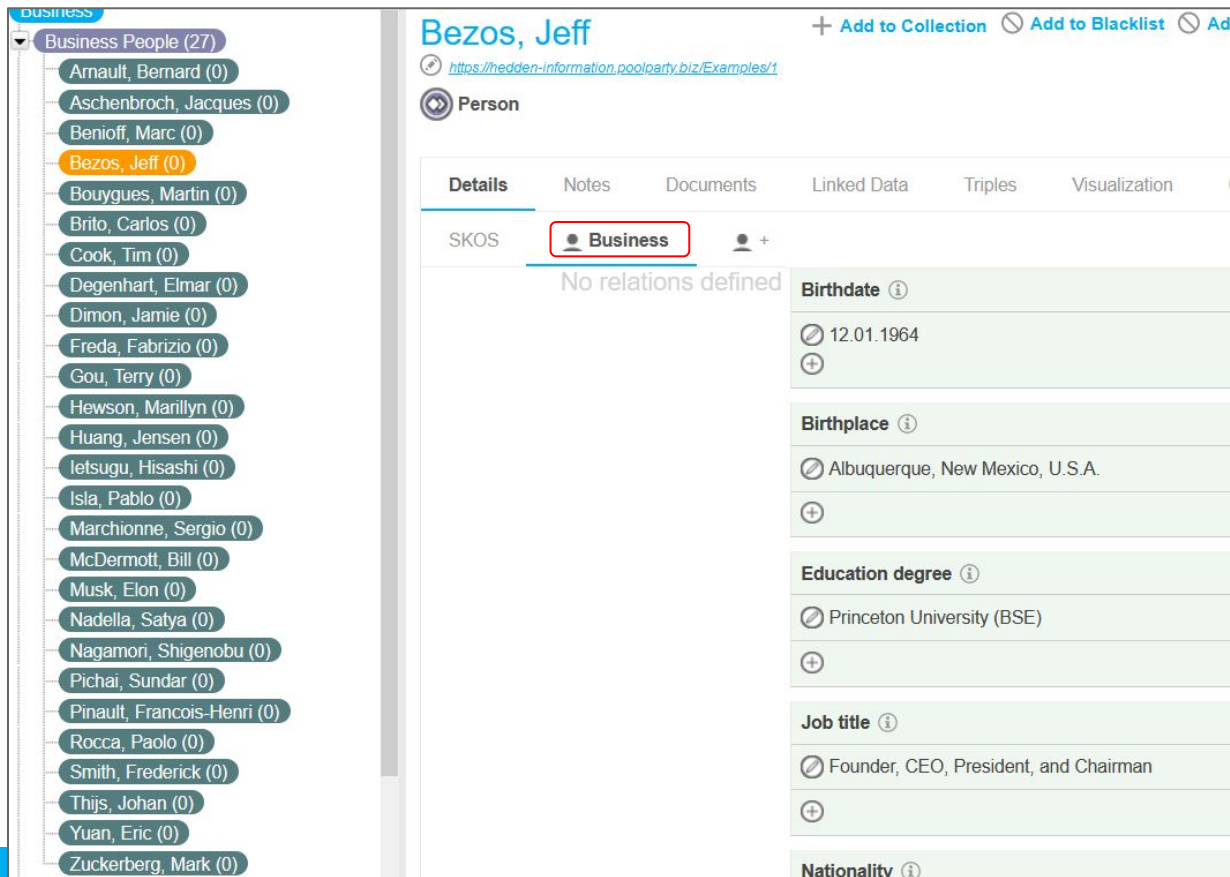


The screenshot displays the poolparty interface. On the left, a list titled "Business People (27)" contains names in blue pills, with "Bezos, Jeff (0)" highlighted in orange. On the right, the "Bezos, Jeff" entity page is shown. It includes a URL, a "Person" type, and tabs for "Details", "Notes", "Documents", "Linked Data", "Triples", and "Visualization". The "Details" tab is active, showing a red-bordered "SKOS" label and a "Business" category. The page is divided into several sections: "Broader Concepts" (with a search icon), "Narrower Concepts" (with search and add icons), "Related Concepts" (with a search icon), and "Top Concept of Concept Schemes" (with a search icon and a link to "Business People"). On the right side of the details view, there are sections for "Preferred Label" (Bezos, Jeff), "Alternative Labels" (Bezos, Jeffrey; Bezos, Jeffrey P.; Bezos, Jeffrey Preston), "Hidden Labels" (with an add icon), and "Scope Notes" (Founder, chairman, CEO, and president of Amazon. with an add icon). A "Definitions" section at the bottom also has an add icon.

Knowledge Organization Systems: Name Authority

Name authority

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



The screenshot displays the poolparty interface. On the left, a sidebar shows a list of 'Business People (27)' with names in rounded rectangles. 'Bezos, Jeff (0)' is highlighted in orange. The main area shows the profile for 'Bezos, Jeff' with a URL and a 'Person' icon. Below this are tabs for 'Details', 'Notes', 'Documents', 'Linked Data', 'Triples', and 'Visualization'. The 'SKOS' tab is active, showing a 'Business' ontology. The profile details include:

- Birthdate**: 12.01.1964
- Birthplace**: Albuquerque, New Mexico, U.S.A.
- Education degree**: Princeton University (BSE)
- Job title**: Founder, CEO, President, and Chairman
- Nationality**: (empty)

Taxonomy

- ▶ A KOS with broader/narrower relationships that includes all concepts to create a hierarchical structure.
- ▶ Has a focus on categorising 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 can be considered as a hierarchy.)
- ▶ “Taxonomy” sometimes refers to any kind of controlled vocabulary (term list, authority file, classification scheme, thesaurus, etc.)



Hierarchical taxonomy

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

The screenshot displays a hierarchical taxonomy on the left and a detailed view of the 'Cakes' concept on the right.

Left Panel: Hierarchical Taxonomy

- Recipes
 - Cooking methods (5)
 - Dishes (11)
 - Appetizers (2)
 - Baked goods (2)
 - Breakfast dishes (3)
 - Desserts (4)
 - Cakes (4)
 - Cheese cakes (0)
 - Chocolate cakes (0)
 - Fruit cakes (0)
 - Layer cakes (0)
 - Ice cream (0)
 - Pies (0)
 - Pudding (0)
 - Egg dishes (2)
 - Meat and poultry (4)
 - Pasta, rice, potatoes (3)
 - Salads (4)
 - Seafood (3)
 - Soups and stews (3)

Right Panel: Cakes Concept Details

Cakes
<http://advanced.poolparty.biz/FoodandRecipes/290>
+ Add to Collection - Add to Blacklist - Add to ExactMatch Delete Concept

Dish

Details Notes Documents Linked Data Triples Visualization

Quality Management History

SKOS Recipe-Scheme

Broader Concepts
[Desserts](#)

Narrower Concepts
[Cheese cakes](#)
[Layer cakes](#)
[Chocolate cakes](#)
[Fruit cakes](#)

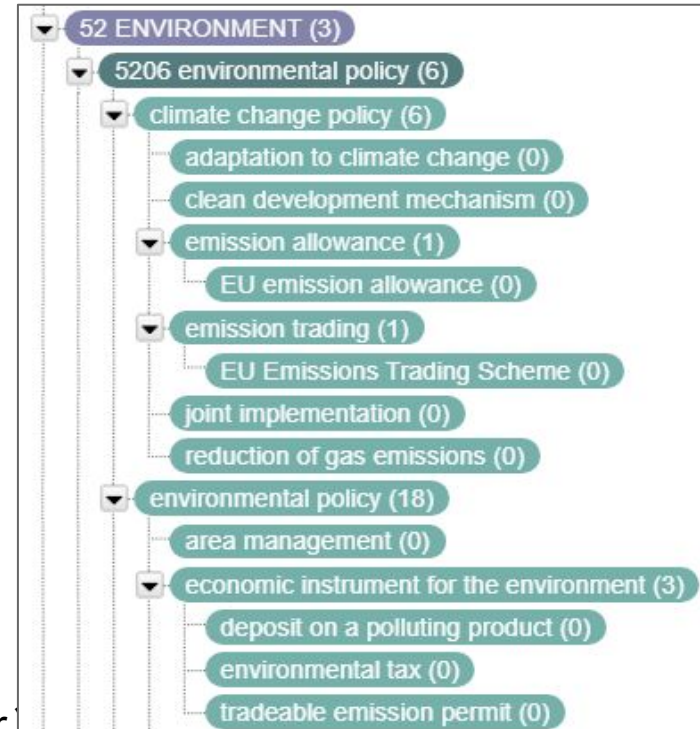
Preferred Label
Cakes (en)

Alternative Labels

Hidden Labels

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

Knowledge Organization System

Faceted taxonomies

- ▶ Taxonomies organized into aspects, each for a “dimension” of a query
- ▶ Intended for searching with multiple terms, each from different facets, in combination
- ▶ Each as a refinement, filter, limit by, narrow by
- ▶ Also called “faceted classification” to classify content objects by multiple classification types

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

Examples: all research literature, all internal policies & procedures, all person profiles, all digital media files

The screenshot displays a search interface with various facets for filtering content. The facets are organized into columns and sections:

- Format:** CD, DVD, E-Mail
- Platform:** Mac, Universal, Windows
- Language:** Arabic, English, Multilingual, Russian, Spanish
- Brand:** Adobe Systems, Corel, Intuit, McAfee, Microsoft, Nero, Norton, Unbranded
- Operating Systems:** Microsoft Windows 10, Microsoft Windows 7
- FILTER BY CATEGORY:** Evaluations, Event resources, Multimedia, News and stories, Reports and publications
- TOPICS:** Air, Biosafety, Chemicals & waste, Cities and lifestyles, Climate change, Disasters & conflicts, Ecosystems and biodiversity, Education & environment, Energy, Environmental rights and governance, Environment under review, Extractives, Forests, Gender, Green economy, Oceans & seas, Resource efficiency, Sustainable development goals, Technology, Transport, Water
- REGION:** Choose (dropdown)
- COUNTRY:** Select country (dropdown), Select country (highlighted), Afghanistan, Albania, Algeria, Angola
- 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

The poolparty logo is visible in the top right corner.

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

Higher Education

Scope Note: All education beyond the secondary level leading to a form appears in the list of mandatory educational level Descriptors)

Category: [Educational Levels, Degrees, and Organizations](#)

 [Search collection using this descriptor](#)

Broader Terms

[Postsecondary Education](#)

Narrower Terms

[Graduate Study](#)

[Postdoctoral Education](#)

[Undergraduate Study](#)

Use this term instead of

[Advanced Education](#)

[College Deans \(1968 1980\)](#)

[Private Higher Education](#)

[Public Higher Education](#)

Related Terms

[Academic Advising](#)

[Academic Degrees](#)

[College Admission](#)

[College Attendance](#)

[College Bound Students](#)

[College Curriculum](#)

[College Faculty](#)

[College Instruction](#)

[College Preparation](#)

[College Programs](#)

[Colleges](#)

[Developing Institutions](#)

[Doctoral Programs](#)

[Educational Benefits](#)

[Extension Education](#)

[Graduate Students](#)

[Masters Programs](#)

[Postsecondary Education as](#)

[Undergraduate Students](#)

[Universities](#)

ERIC - Education thesaurus

Thesaurus

As expressed in the SKOS model:

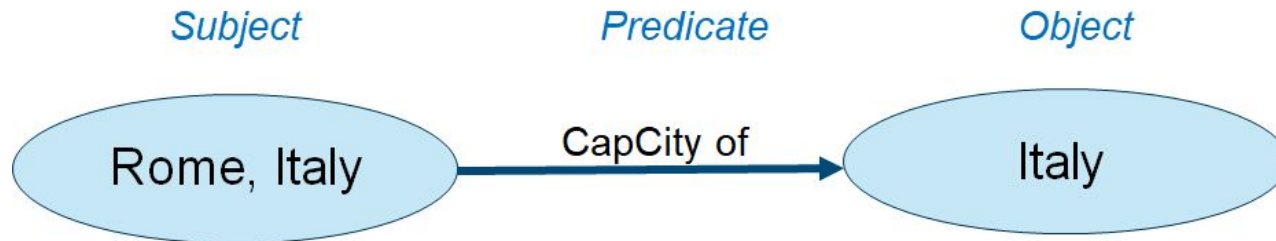
- Broader concepts
- Narrower concepts
- Related concepts
- Preferred label
- Alternative labels

The screenshot displays a SKOS interface. On the left is a hierarchical tree of concepts, with 'herbicides (5)' selected. The tree includes categories like pesticides (11), biopesticides (3), insecticides (7), and others. The right pane shows the details for 'herbicides', including a URL, action buttons (Add to Collection, Add to Blacklist, Add to ExactMatch, Delete Concept), and tabs for Details, Notes, Documents, Linked Data, and Triples. Under the SKOS tab, there are sections for Broader Concepts (pesticides), Narrower Concepts (algicides, arboricides, herbicide selectivity, oxyfluorfen, selective herbicides), Related Concepts (defoliants), Preferred Label (herbicides), Alternative Labels (brush killers, weed killers), Hidden Labels, Notation, and Scope Notes.

Ontology



- ▶ A model of a knowledge domain
- ▶ Similar to (most of) a knowledge graph, but excludes the data
- ▶ 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.

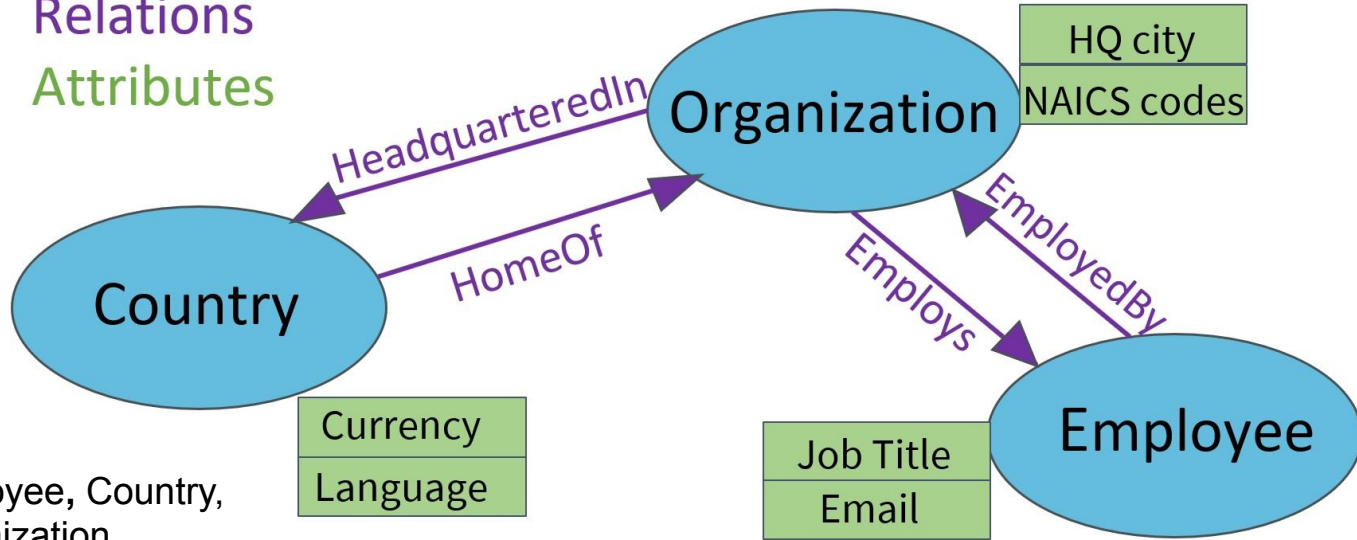


Knowledge Organization Systems: Ontology

Classes

Relations

Attributes



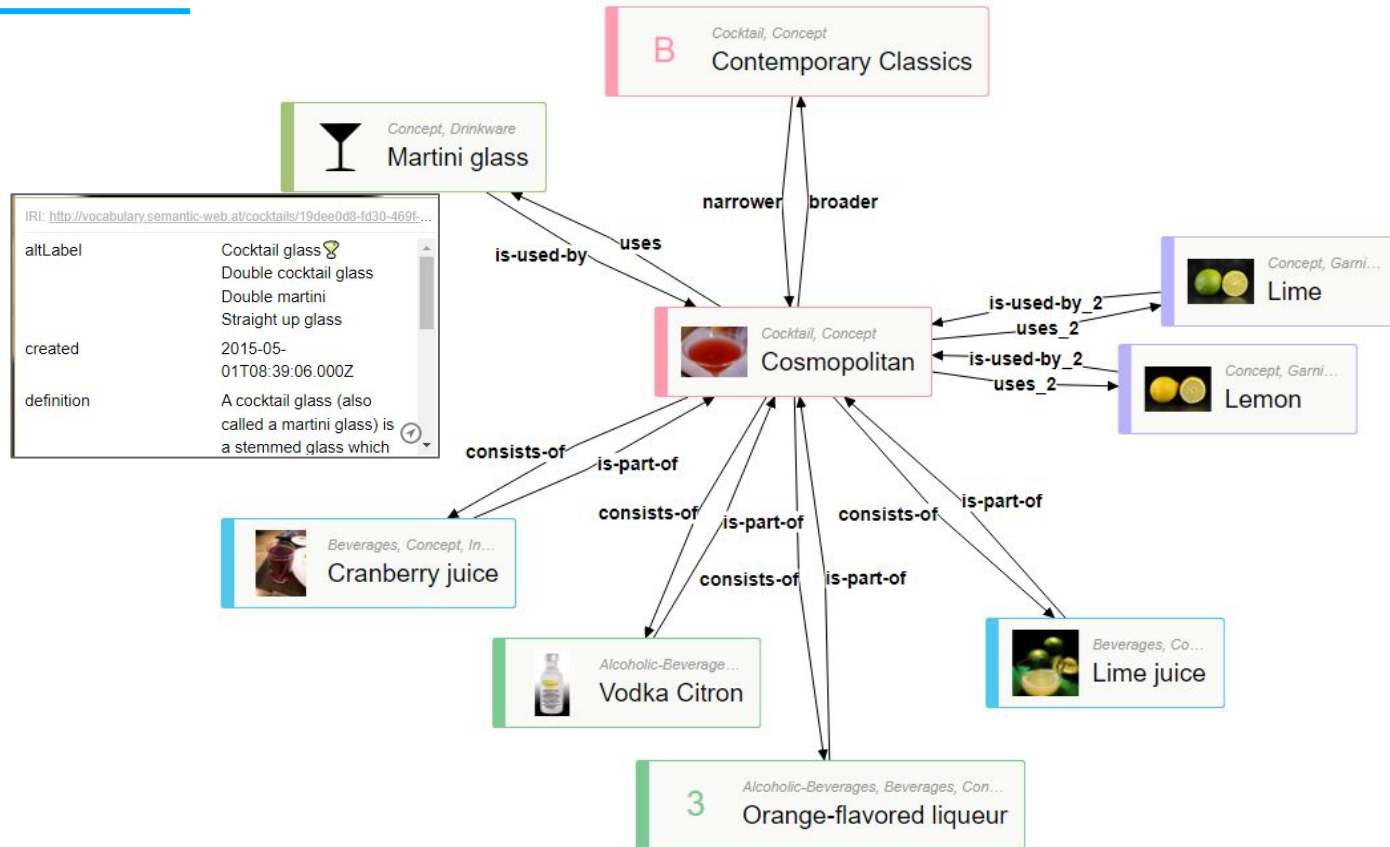
Classes

Employee, Country,
Organization

Relations: HeadquarteredIn < > HomeOf
EmployedBy < > Employs

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

Knowledge Organization Systems: Ontology



Comparison of Knowledge Organization 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 relationship Associative relationships	Semantic relationships Classes Attributes

Comparison of Knowledge Organization Systems



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

A blue rectangular box on the left side of the slide contains the text 'Knowledge organization system standards' in a bold, black, sans-serif font. A thin white vertical line is positioned to the left of the text. The background of the slide is a light grey color with a network diagram of white circles and lines.

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 organization systems of each type:

1. **Standards for design**
 - ANSI/NISO Z39.19-2005 *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*
 - ISO 25964 (2011 and 2013) *Thesauri and Interoperability with Other Vocabularies*
2. **Standards for specifications and interoperability**
 - Dublin Core, MARC, ZThes, DD 8723-5, SKOS, RDF, RDFS, and OWL

ANSI/NISO Z39.19-2005 (R2010)

Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies

www.niso.org/publications/ansiniso-z3919-2005-r2010 (167 pages)

- ▶ Despite its title, focus is on **thesauri**, but much is applicable to taxonomies, in general.
- ▶ Comparable international standard is ISO 25964 .
(which also includes multilingual thesauri and more on interoperability)

Guidelines/best practices for:

- ▶ Definitions, types, structure of various controlled vocabularies
- ▶ Forming relationships
- ▶ Term choice, scope, form, and style
- ▶ Display formats
- ▶ Vocabulary construction approaches
- ▶ Vocabulary management system considerations



ANSI/NISO Z39.19-2005 (R2010)

Examples of quality standards

- ▶ Concepts/terms are nouns or noun phrases.
- ▶ Use plural for countable nouns.
- ▶ 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 extend:
 - ▷ 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 Organization System)

- ▶ A data model to represent knowledge organization systems.
- ▶ A World Wide Web (W3C) recommendation (initial version 2004 - revised 2009)
- ▶ “A common data model for sharing and linking knowledge organization 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		



Standards & Models

Basic SKOS view
for taxonomy editing











Budgeting

<https://hidden-information.poolparty.biz/Examples/26>

[+ Add to Collection](#) [⊘ Add to Blacklist](#) [⊘ Add to ExactMatch](#)

Details | Notes | Documents | Linked Data | Triples | Visualization | Quality Management

SKOS | [Business](#) | [+](#)

Broader Concepts Finance 	Preferred Label <input type="radio"/> Budgeting
Narrower Concepts Budget analysis  	Alternative Labels <input type="radio"/> Budget management <input type="radio"/> Budget planning <input type="radio"/> Budgets 
Related Concepts  Accounting 	Hidden Labels 
Top Concept of Concept Schemes 	Scope Notes 
	Definitions 

Standards: 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.



Standards: OWL (Web Ontology Language)

OWL-Defined Ontology Components

Entities – subjects or objects of properties (domains and ranges)

- ▶ **Classes**
 - ▷ Named sets of concepts that share characteristics and relations
 - ▷ May group subclasses or individuals (instances of the class)
- ▶ **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*.



<https://www.w3.org/TR/2012/REC-owl2-primer-20121211/>

A large blue rectangular box is positioned on the left side of the slide. A thin white vertical line is located on the left edge of this box, extending from the top to the bottom.

Why taxonomies

Taxonomy Approaches

“Taxonomy” may refer to

1. A hierarchy of terms/topics/categories arranged with narrower topics/subcategories displayed under their broader/parent categories.



- ▶ To guide users to find the desired topic (and its linked content of pages or documents)
- ▶ Similar to navigation and site maps, but more topical and not just based on page titles

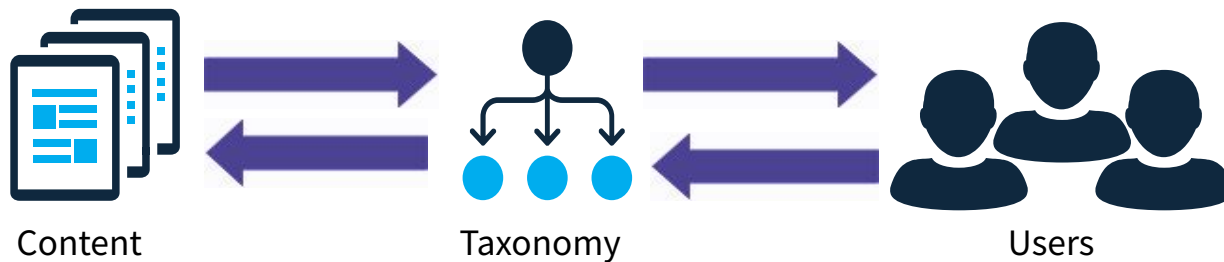
2. A controlled vocabulary of metadata tags/labels to apply to pages, posts, or documents, so that they can be more precisely and comprehensively retrieved (than by search algorithms alone on keywords in text).



- ▶ Implemented as search suggestion terms, search refinement filters, or related topics and searches

Why taxonomies? Traditional use

- ▶ Concepts/terms are used to tag/index/categorize content to make it easier to be found and retrieved
 - ▷ supporting better findability than search alone
- ▶ The taxonomy is an intermediary that links the user to the desired content.



- ▶ Taxonomies usually link to content; ontologies can also link to data.

Taxonomy Uses and Benefits

The screenshot shows the Office Depot website's navigation menu. The 'Products' dropdown is open, listing various office supply categories. A search bar is visible at the top right. A blue box at the bottom left contains the URL www.officedepot.com.

Products	Services	Deals	Search
Office Supplies >			
Furniture >			
Cleaning >			
Breakroom >			
Paper >			
School Supplies >			
Technology >			
Ink & Toner >			
Tax Return Solutions			
Cold & Flu			
	Pens, Pencils & Markers		
	Pens		
	Markers & Highlighters		
	Pencils		
	Correction Fluid & Tape		
	Basic Supplies		
	Staplers & Staples		
	Paper & Binder Clips		
	Scissors		
	Paper Punches & Cutters		
	Mailing & Shipping		
	Envelopes		
	Shipping & Moving Boxes		
	Mailers		
	Packing Tape		
	Packing Materials		
	Filing & Folders		
	File Folders		
	Hanging File Folders		
	Expanding File Folders		
	Classification Folders		
	Binders & Accessories		
	Binders		
	Dividers & Tabs		
	Report Covers & Portfolios		
	Sheet Protectors		
	Labels & Label Makers		
	Standard Labels		
	Label Maker Tapes		
	Address Labels		
	Label Makers		
	Name Badges & Holders		
	Desk Accessories		
	Desk Organizers		
	Desk Trays		
	File Organizers		
	Drawer Organizers		
	Calendars & Planners		
	Calendars		
	Planners		
	Organizers		
	More Ways to Shop		
	Tax Return Solutions		
	Post-It & Sticky Notes		
	Tape & Adhesives		
	Storage Boxes & Bins		
	Presentation Boards		

World Bank Open Data

Free and open access to global development data

- Population, male
- Population, total
- Population, female
- Rural population
- Urban population
- Population, male (% of total population)
- Population, female (% of total population)
- Population in largest city

World Bank

<https://data.worldbank.org>

Benefits of taxonomies and other 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

Organizes information into a logical structure

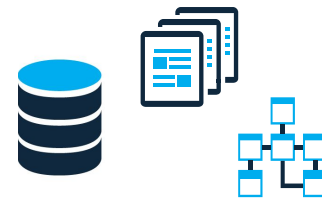
- ▶ Helps people browse or navigate for information
- ▶ Provides context and meaning for concepts for indexing and retrieval



Scope issues: “Taxonomy” sometimes refers to any kind of controlled vocabulary

Relationships between data management/data science and taxonomies

- ▶ **Graph databases** - used with taxonomies and ontologies to create knowledge graphs.
- ▶ **Metadata** - is often populated with controlled vocabularies/taxonomy values.
- ▶ **Reference data and master data** - may be structured into hierarchies as taxonomies.
- ▶ **Data catalogs** - can be improved with taxonomies or taxonomy principles.
- ▶ **Linked data** - is based on W3C Semantic Web Standards and so are taxonomies.
- ▶ **Machine learning and AI** - may be used for text analytics on unstructured data for:
 - ▷ term extraction to automatically generate candidate taxonomy terms
 - ▷ auto-tagging content items with concepts from a taxonomy



Traditional uses of taxonomies:

- ▶ Topic/category browsing
- ▶ Search (matching search strings to concepts)
- ▶ Consistent tagging/indexing

Additional uses:

- ▶ Discovery (related concept links, or content sharing the same concepts)
- ▶ Filtering or sorting results
- ▶ Content management workflow (structured authoring, rights management, retention)
- ▶ Consistent metadata for identification, comparison, analysis, and reporting
- ▶ Visualization of topics (importance and/or relations)
- ▶ Curated content in feeds or info boxes
- ▶ Knowledge-based personalization or recommendation systems
- ▶ When integrated with ontologies, support for knowledge graphs and AI-driven applications

A solid blue rectangular box is positioned on the left side of the slide, containing the title text. A thin white vertical line is located on the left edge of this box.

Why ontologies and knowledge graphs

What you cannot do with a taxonomy alone, but can with an added ontology

- ▶ Modeling complex interrelationships (e.g. in product approval or supply chain processes) and also connect to content
- ▶ Complex multi-part searches: e.g. find contacts in a specific location, who are employed by companies which belong to certain industries
- ▶ Exploring explicit relationships between concepts (not just broader, narrower, related)
- ▶ Search across datasets, not just search for content
- ▶ Search on more specific criteria that vary based on category (class)
- ▶ Visualization of concepts and semantic relationships
- ▶ Reasoning and inferencing
- ▶ Build a knowledge graph...



What is a knowledge graph?

- ▶ A model of a knowledge domain combined with instance data
- ▶ Represents unified information across a domain or an organization, enriched with context and semantics that are meaningful across information silos.
- ▶ Contains business objects and topics that are closely linked, classified, semantically enriched, and connected to existing data and documents.
- ▶ A layer between the actual content/data and the querying layer.



Why Knowledge Graphs

Problems

- ▶ Data silos
- ▶ Heterogeneous data sources
- ▶ Mix of unstructured and structured data
- ▶ Same things with different names
- ▶ Localized meanings
- ▶ Change

Causing

- ▶ Inefficiencies
- ▶ Missed opportunities
- ▶ Poor decisions

Provided by

- ▶ Data-centric architecture
- ▶ Ontologies
- ▶ Knowledge graphs

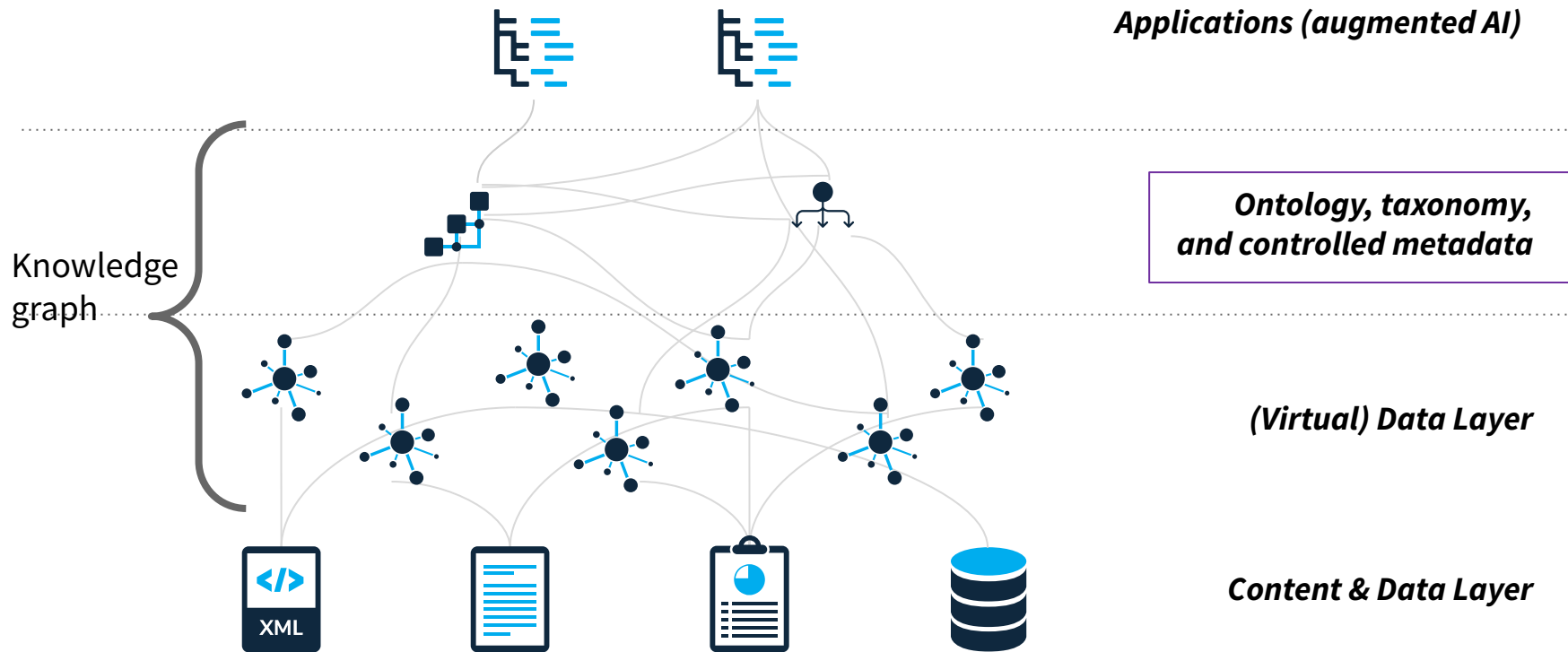
Solutions

- ▶ Sharing data
- ▶ Reusable data sets
- ▶ Semantic links
- ▶ Semantic data fabric
- ▶ Unified views

Results in

- ▶ Knowledge discovery
- ▶ Better decisions
- ▶ Customer satisfaction

Knowledge Graphs



Knowledge graph components

- ▶ Extracted **data**, stored in either
 - ▷ A graph database, of either
 - RDF-based triple store
 - Labeled property graph
 - ▷ A search index
 - RDF-based triple store
- ▶ which are tagged/linked with metadata or concepts in **controlled vocabularies** (taxonomies and other types) to label and organize the data
- ▶ which in turn are semantically linked to represent conceptual relationships and enriched with additional metadata attributes in an **ontology**

Scope issue/question: What kind of data is included or not? Data vs. content

Benefits of combining a high-level ontology as a semantic layer with a taxonomy

- ▶ Makes use of existing taxonomies, even multiple taxonomies
- ▶ Easier to model the ontology
 - ▷ Existing taxonomies and controlled vocabularies provide a basis for knowledge modeling.
 - ▷ No need to distinguish between sub-classes and individuals.
- ▶ Supports expert specialization
 - ▷ Taxonomists develop and maintain taxonomies.
 - ▷ Ontologists develop and maintain the ontology.
- ▶ More flexible and adaptable
 - ▷ The taxonomy changes more frequently than does the ontology.
 - ▷ Taxonomies can easily be added.
- ▶ Different purposes served
 - ▷ The ontology is for modeling, reasoning, and analysis.
 - ▷ The taxonomy is for tagging and information/data retrieval.

The word 'Resources' is written in a bold, black, sans-serif font. It is positioned on the left side of a large, solid blue rectangular area that occupies the middle-left portion of the slide. A thin white vertical line is located to the left of the text.

- ▶ ANSI/NISO Z39.19-2005 (2010) *Guidelines for Construction, Format, and Management of Monolingual Controlled Vocabularies*
www.niso.org/publications/ansiniso-z3919-2005-r2010
- ▶ Accidental Taxonomist book websites
<http://www.hedden-information.com/accidental-taxonomist/websites> (updating in July)
- ▶ The Accidental Taxonomist Blog
<http://accidental-taxonomist.blogspot.com>
- ▶ Hedden Information Management past presentations
<http://www.hedden-information.com/presentations>
- ▶ Taxonomy Talk, taxonomists community on Discord
<https://discord.com/invite/3qyMVYCAsw>

Courses, Workshops, Conferences

- ▶ “[Taxonomies and Controlled Vocabularies](#)” self-paced online course any time
Hedden Information Management
- ▶ “[From Taxonomy to Ontology](#)” online tutorial, July 14, 2:00 - 4:30 pm EDT
Special Libraries Association (SLA)
- ▶ “Knowledge Engineering of Taxonomies, Thesauri, and Ontologies” tutorial, September 14 or 15,
[SEMANTiCS conference](#), Vienna (hybrid)
- ▶ [Taxonomy Boot Camp](#) conference
November 7-8, 2022, Washington DC
- ▶ [Taxonomy Boot Camp London](#) Virtual half-day conference
October 12, 2022
- ▶ “[Taxonomy and Metadata Design](#),” 8-hour 2-day online seminar, November 14-15,
Technology Transfer S.r.l.

Questions/Contact

Heather Hedden

Data and Knowledge Engineer

Semantic Web Company Inc.

One Boston Place, Suite 2600

Boston, MA 02108

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

