



LIS Education And Data Science Integrated Network Group

LEADING Boot camp

Monday, June 13, 2022

1:00 PM-2:15 PM EDT/10:00 AM-12:15 PM PDT

Agenda:

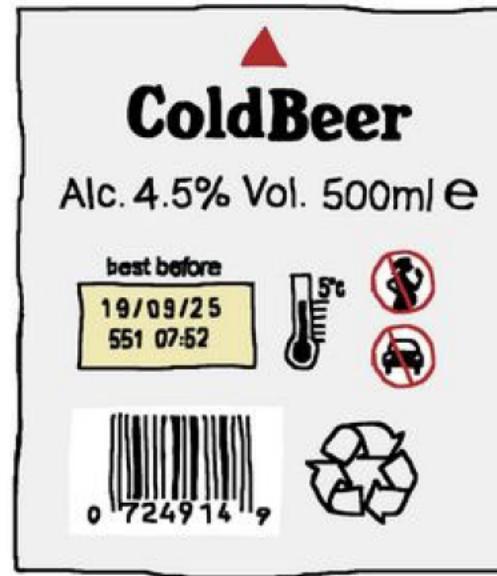
1. How's everyone doing
2. Big metadata (brief comments)
3. Ontology/briefly linked data
 - ~ Protégé
4. Other/open discussion



DATA



METADATA



Dataedo /cartoon

provenance tracking...)

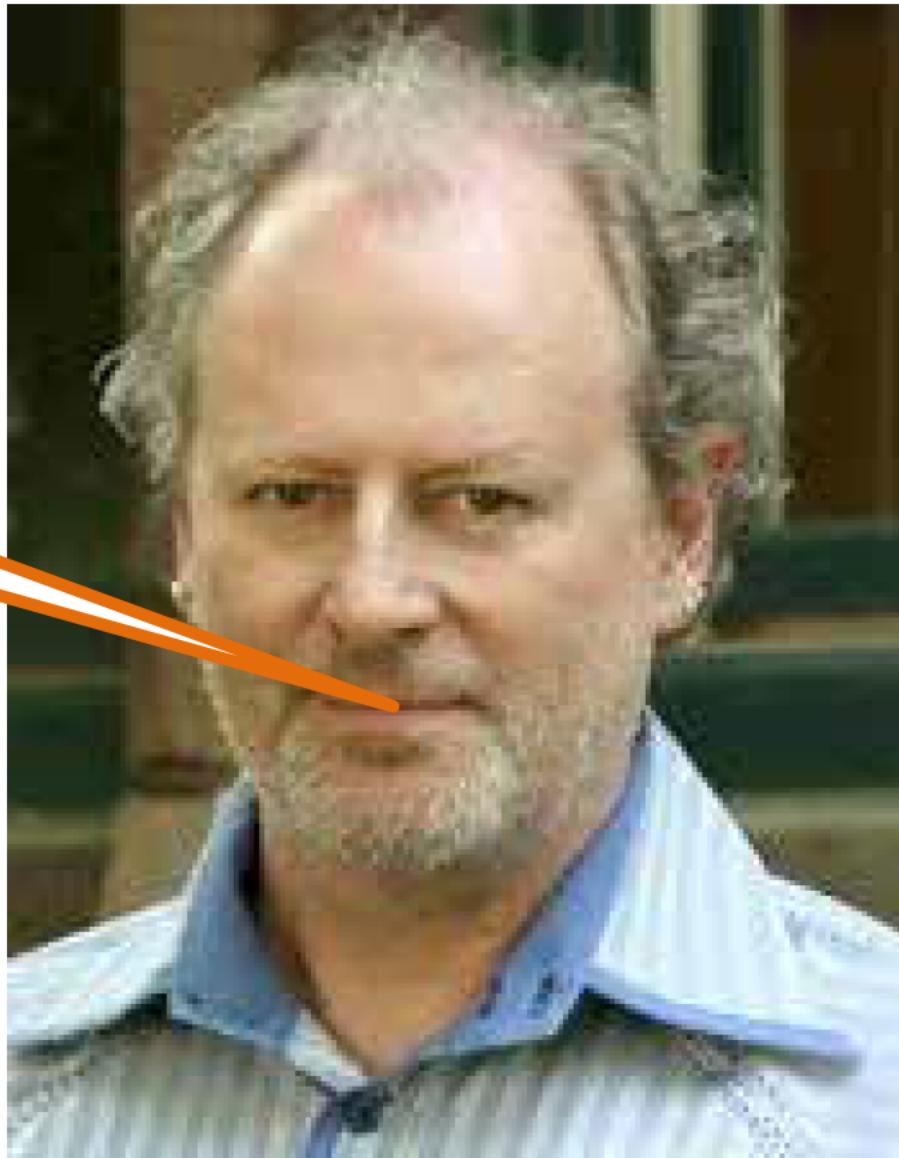
11.

Piotr@Dataedo

Research Data Alliance Conference

Peter Fox, RDA,
Sweden, 2014

Let's get rid of
the word
metadata



Components of Metadata Standards

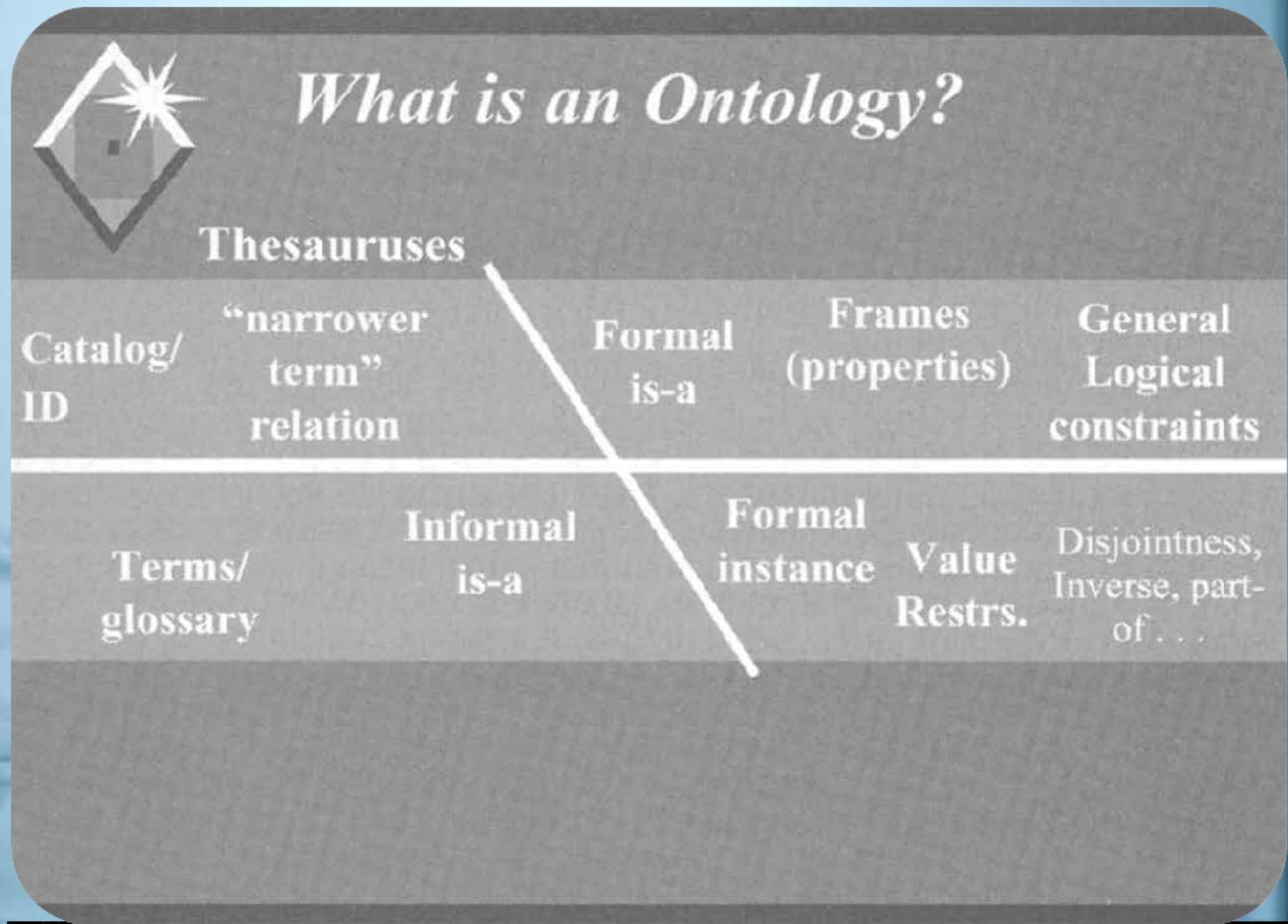
Data structure Standards <i>(container, label, semantics)</i>	Data communication standards <i>(encoding, markup, data exchange)</i>	Data value standards <i>(authority files, ontologies, taxonomies, etc.)</i>
Data syntax standards		
<i>element/property ordering</i>	<i>think grammar...</i>	<i>content syntax</i>

Table 1. The five Vs of big metadata.

Five Vs	Definition
Volume	The quantity and usefulness of metadata generated daily confirms the existence of big metadata. At times metadata is less than or equal to the extent of the data it describes in size (bytes). During other times the metadata exceeds the data being described or tracked, due to the complexity of the data lifecycle activity. Linked data offers an example, with metadata renderings that can be larger than the volume of data object(s) being represented. Like big data, not all big metadata is useful, and a challenge is to identify the big metadata that is useful for data science and analytic endeavors.
Velocity	Metadata is generated via automatic processes at immense speed correlating with rate of digital transactions. For example, searching Google, answering an email, purchasing an item online, and day-to-day office activities such as word processing of all log data, as well as associated metadata.
Variety	Metadata reflects the wide variety of data formats, types, and genres along with the extensive range of data and metadata lifecycles. In addition, the different types of metadata (e.g. discovery, technical, preservation, etc.) as well as unique domain specific metadata requirements intensify the variety.
Variability	There is an unmistakable unevenness of metadata across the digital ecosystem. Lack of uniformity is extensive for data descriptions across different domains, systems, and processes. This unevenness can even be profound within domains, given economic factors supporting metadata generation, competing standards, or, simply, differing adoption policies. For example, two organizations may use the same metadata standard, but have different implementation practices. Even when standardization is imposed, an organization, process, and human activity can contribute to inconsistencies.
Value	<i>If data is the new black gold*—akin to petroleum requiring purification, but also a money maker, then metadata is the <u>new platinum</u>—a malleable substance that keeps its toughness, and can serve as a catalyst, sparking a reaction.</i>
	Metadata, as the <i>new platinum</i> , can be modified, while remaining a strong, independent data type. Metadata stands as a durable data object that triggers various functions—the catalyst, and achieves results—a reaction. Metadata is vital to accurate data interpretation and use by both humans and machines, and the value of metadata for data science endeavors cannot be overstated or diminished.

(McGuinness, D. L. (2003). Ontologies Come of Age.
In Fensel, et al, *Spinning the Semantic Web*.
(Cambridge, MIT Press)

Ontology

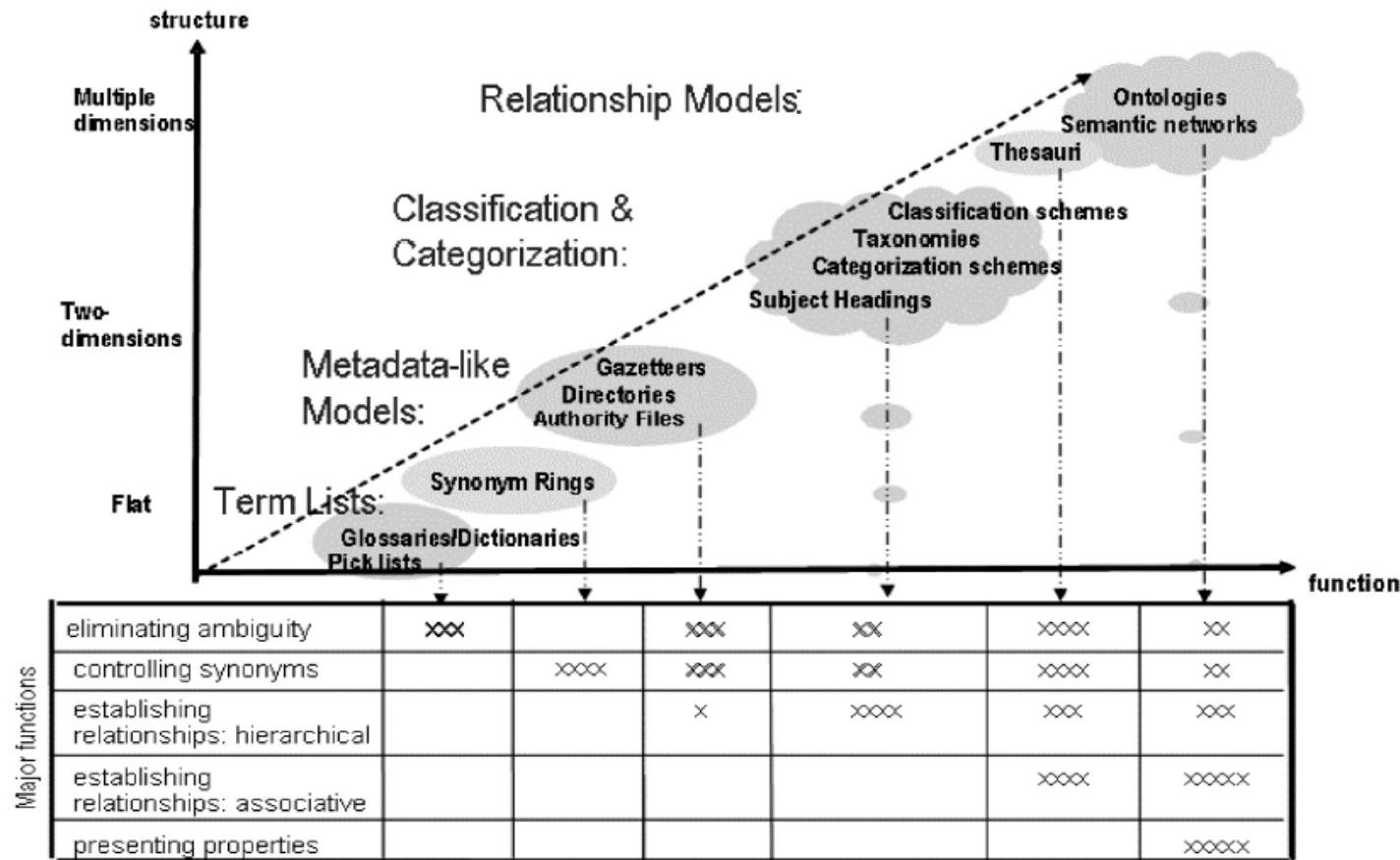


Ontologies and reality

ONE SIZE
DOESN'T FIT ALL



A Taxonomy of KOS



Zeng, Marcia Lei. 2008. "Knowledge Organization Systems". *Knowledge Organization* 35, nos. 2-3: 160-82.

Philosophy

Dates back to 5th Century B.C. when Empedocles divided the world into four elements – earth, fire, water and air.

Metaphysics:
Defined by philosophers as the *nature of being or existence.*

Aristotle · classification

Touches on “epistemology,” which” is about knowledge and knowing

Ontology defined

Webster's dictionary

1. A science or study of being specifically, a branch of metaphysics relating to the nature and relations of being.
2. A theory concerning the kinds of entities and specifically the kinds of abstract entities that are to be admitted to a language system.

Thomas Gruber

"An ontology is a specification of a conceptualization"

A must read!

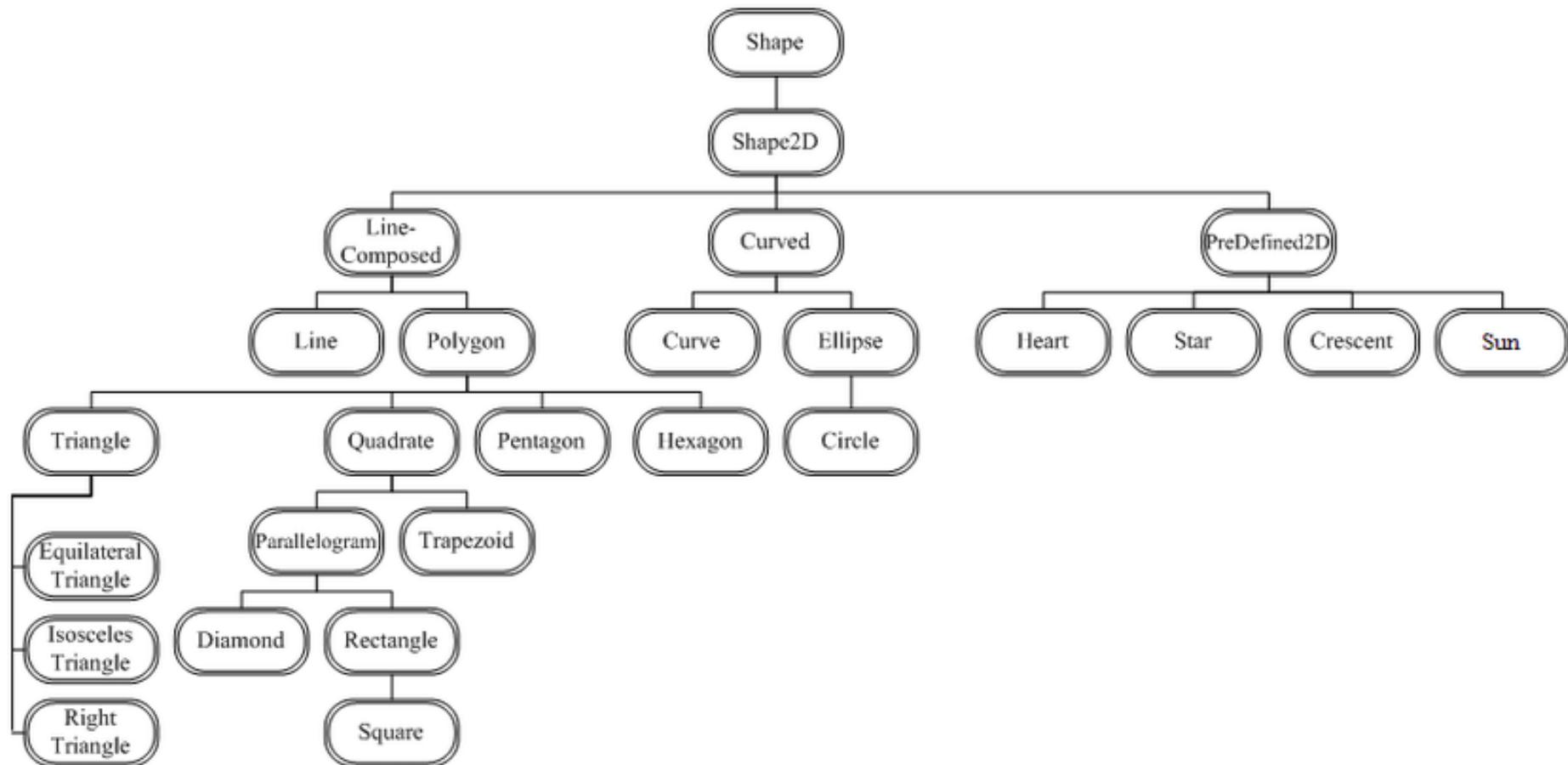
Other important stuff

(comfy space)

- No exact definition
- *Generally speaking - ontology is* a WAY to convey a theory on how to represent a class of things and their relationships
- Knowledge representation, Knowledge organization (ILS)
- Blurry lines – w/database structure and knowledge graphs



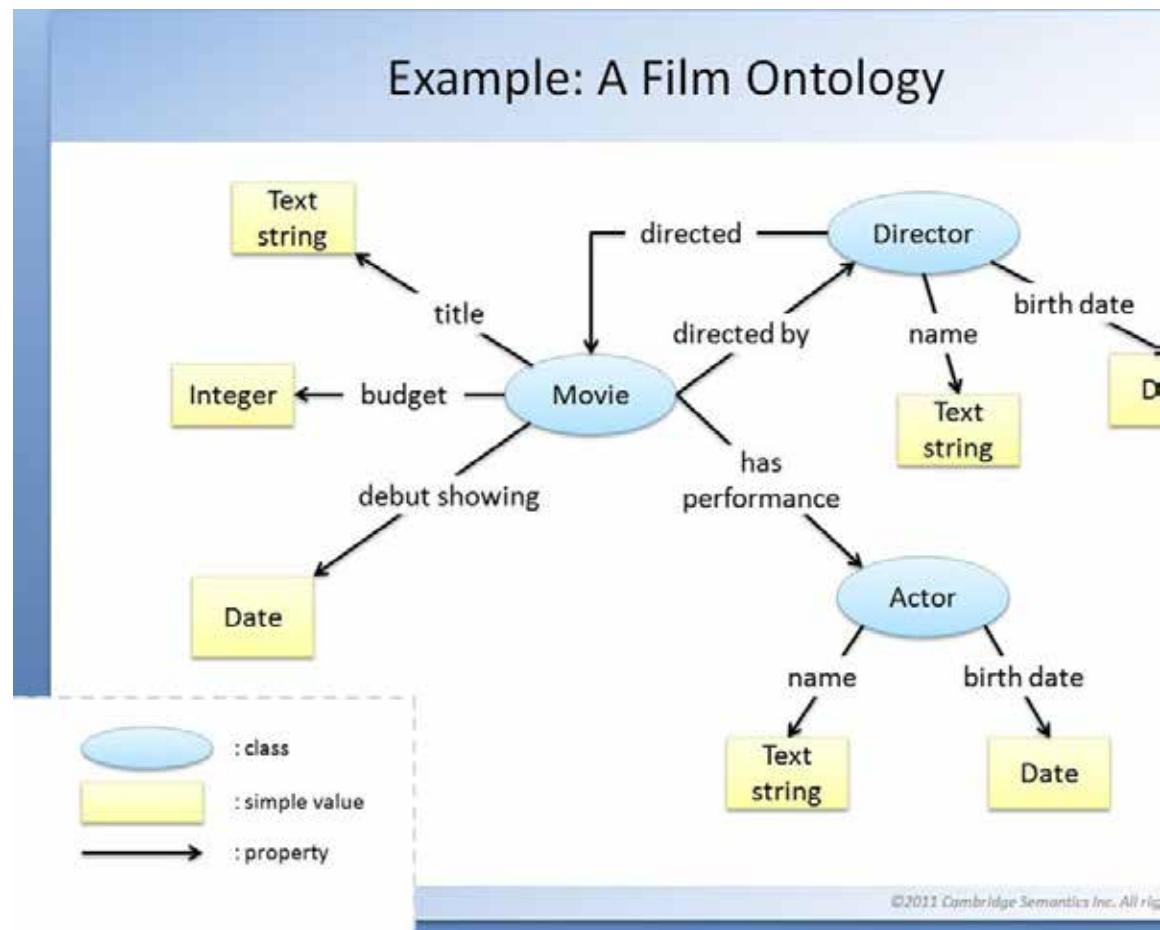
AN ONTOLOGY OF SHAPES



<http://ceur-ws.org/Vol-812/paper9.pdf>

Ontology

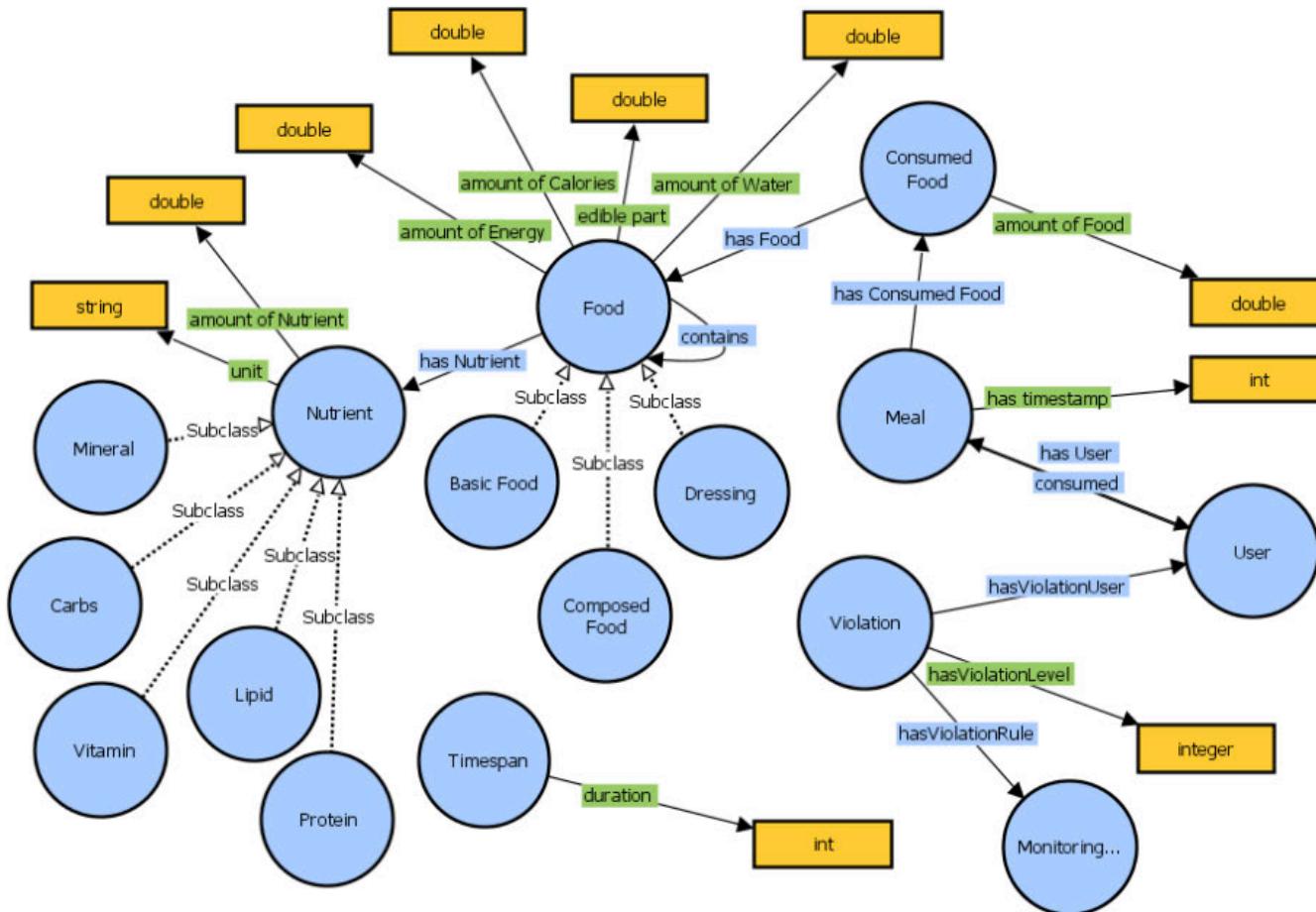
- National Center for Biological Ontologies:
<https://bioportal.bioontology.org/>
- Example of a Film ontologyà





- Most ontologies structured as taxonomies, hierarchies
- People talk about ontologies in many, many, many... ways
- Basic ontology has two classes of elements: the entities and the relationships between them
 - Class and sub-classes: Actor, Child actor
 - Relationships
 - Is a type of
 - Movie "is directed by" director
 - Movie "has performance" actor
- More involved / RDF triples
 - Predicate (has a property of X)
 - E.g., Book has title
 - predicates are linguistic entities
 - Subject-predicate-object (think reverse ABC)
- Organized according to axioms or rules that control how the world will be defined.

AN ONTOLOGY OF FOOD



https://www.w3.org/community/owled/files/2016/11/OWLED-ORE-2016_paper_3.pdf

Let's go to Protégé



We will work with OWL: Web Ontology Language: <https://www.w3.org/OWL/>

- Modeling language
- Enabling technology

SKOS – Simple Knowledge Organization System:
<https://www.w3.org/2004/02/skos/>

Important Facts



What exists is only what is represented in the ontology

Most ontologies focus on a specific area to conceptualize the world.

Must be updated to keep up with dynamic world

No set discipline or methodology!

Joined together with instances, get closer toward knowledge graph... but fuzzy line

The Role of Ontology

Application areas

- Indexing
- Knowledge Sharing & Reuse
- Artificial Intelligence (AI)
- Enterprise Modeling
- Software Design
- Molecular Biology
- eCommerce
- Semantic Web....

Indexing, browsing, finding, visualizing...

A language, reliable sharing of information

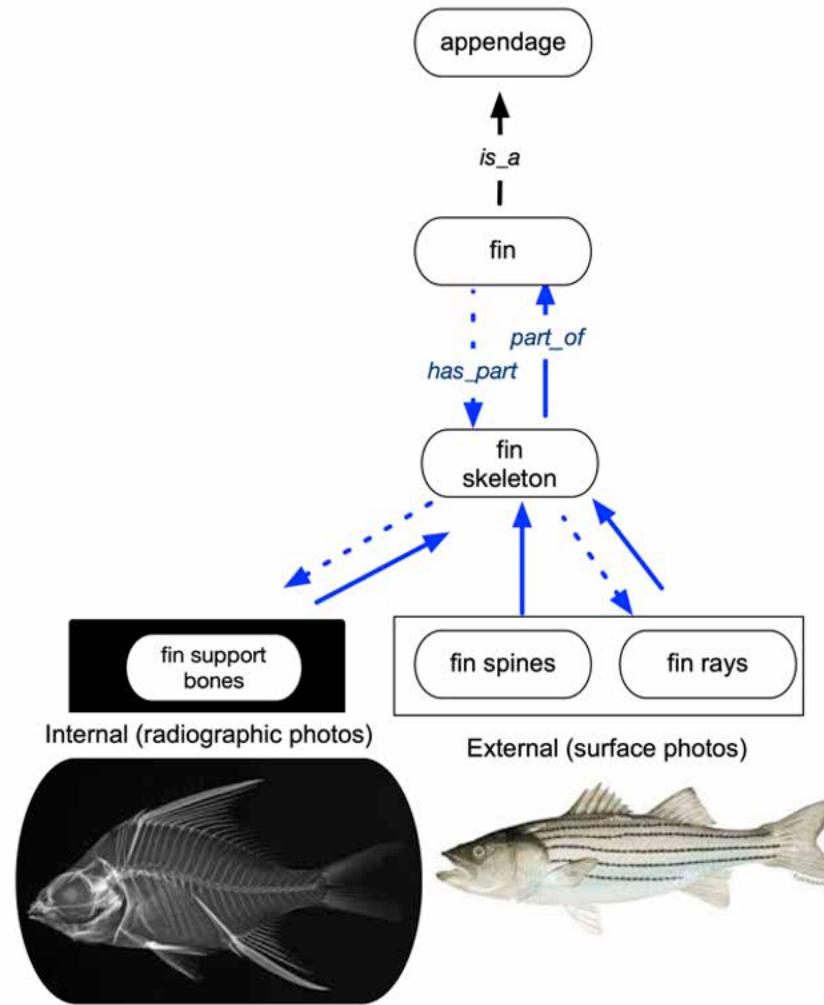
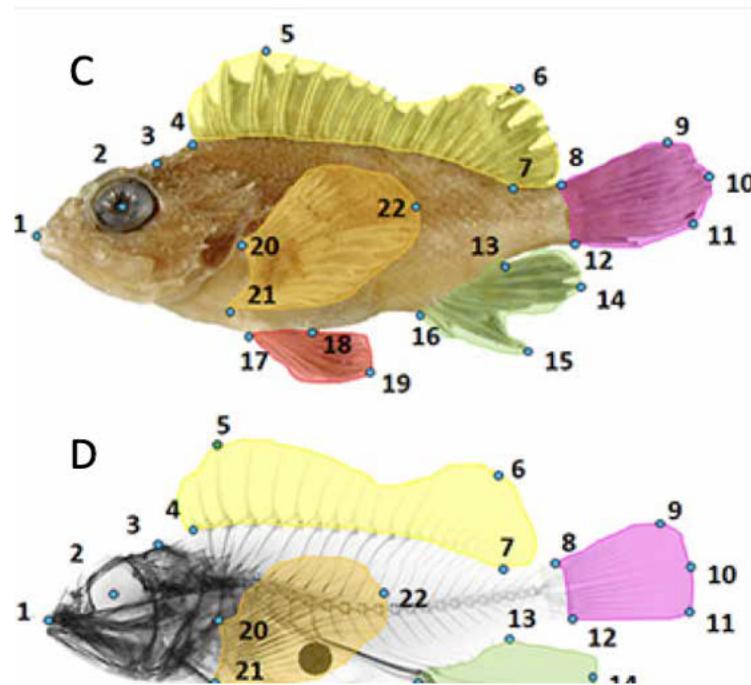
Enable reuse of domain knowledge

Makes assumptions more explicit

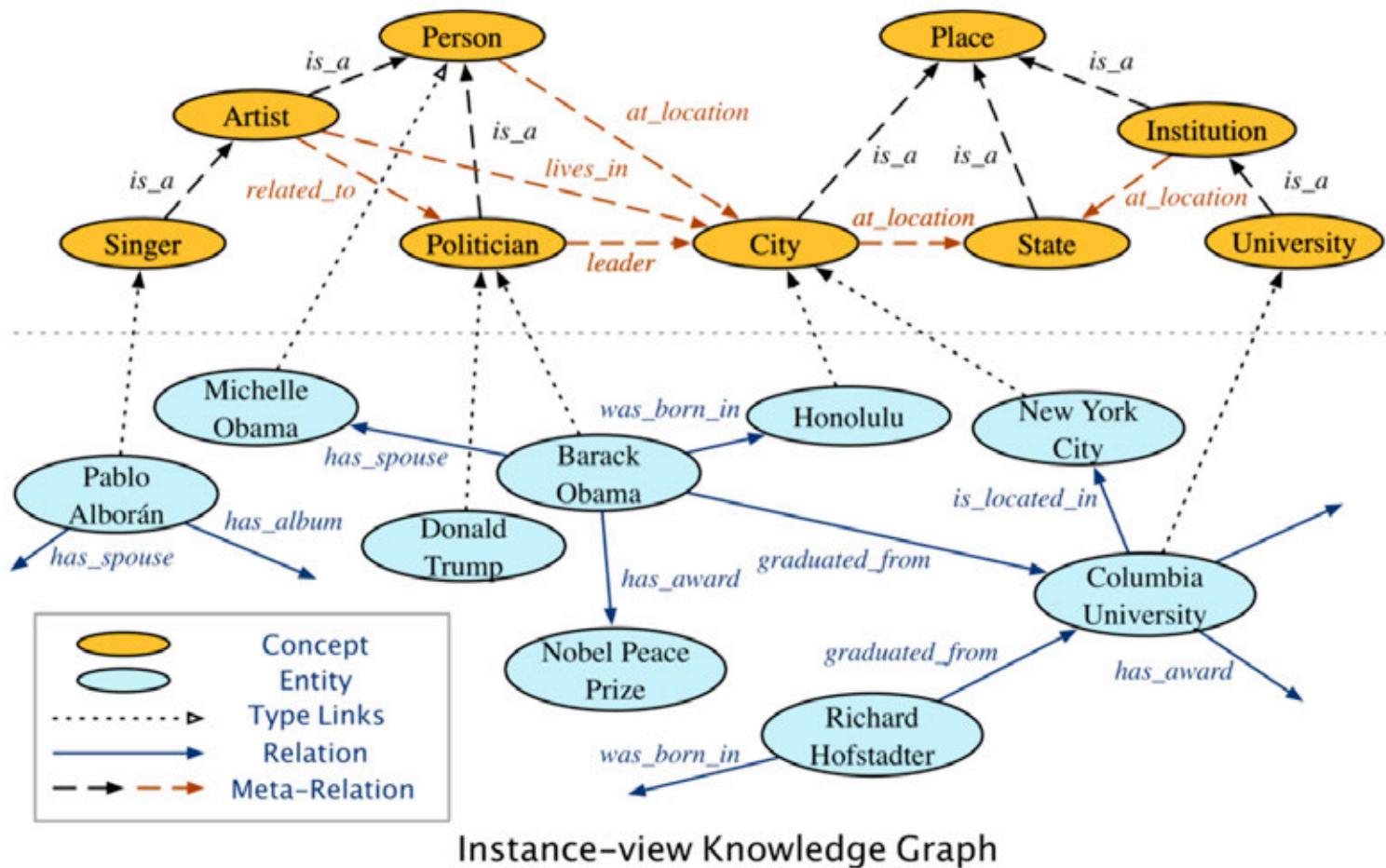
Analyze domain knowledge

Reasoning

Reasoning to identify species variation



Ontology-view Knowledge Graph



Publication: The 25th International ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2019)

<https://www.haojunheng.com/project/joie-kdd/>

Linked data

- Strings and things
- Semantic Web vision
- Everything should have a unique ID, in fact a Persistent ID (PID)



Evolution of the Web <http://www.w3.org>

The screenshot shows the W3C website's navigation bar with links for desktop, mobile, and print views, and tabs for STANDARDS, PARTICIPATE, MEMBERSHIP, and ABOUT W3C. A Google search bar is also present. The main content area displays the breadcrumb path W3C > Standards > Semantic Web, followed by the title "SEMANTIC WEB". Below the title, a sidebar lists various standards categories. The main text discusses the Semantic Web as a "Web of data" and mentions "W3C's vision of the Web of linked data". A red circle highlights the phrase "build vocabularies". The page also features three columns: "Linked Data", "Vocabularies", and "Query", each with a brief description.

Views: [desktop](#) [mobile](#) [print](#)

[STANDARDS](#) [PARTICIPATE](#) [MEMBERSHIP](#) [ABOUT W3C](#)

[Google™](#)

W3C » Standards » Semantic Web

SEMANTIC WEB

On this page → [technology topics](#) • [news](#) • [upcoming events and talks](#)

In addition to the classic "Web of documents" W3C is helping to build a technology stack to support a "Web of data," the sort of data you find in databases. The ultimate goal of the Web of data is to enable computers to do more useful work and to develop systems that can support trusted interactions over the network. The term "Semantic Web" refers to W3C's vision of the Web of linked data. Semantic Web technologies enable people to create data stores on the Web, **build vocabularies**, and write rules for handling data. Linked data are empowered by technologies such as RDF, SPARQL, OWL, and SKOS.

Linked Data

The Semantic Web is a Web of data — of dates and titles and part numbers and chemical properties and any other data one might conceive of. RDF provides the foundation for publishing and linking your data. Various technologies

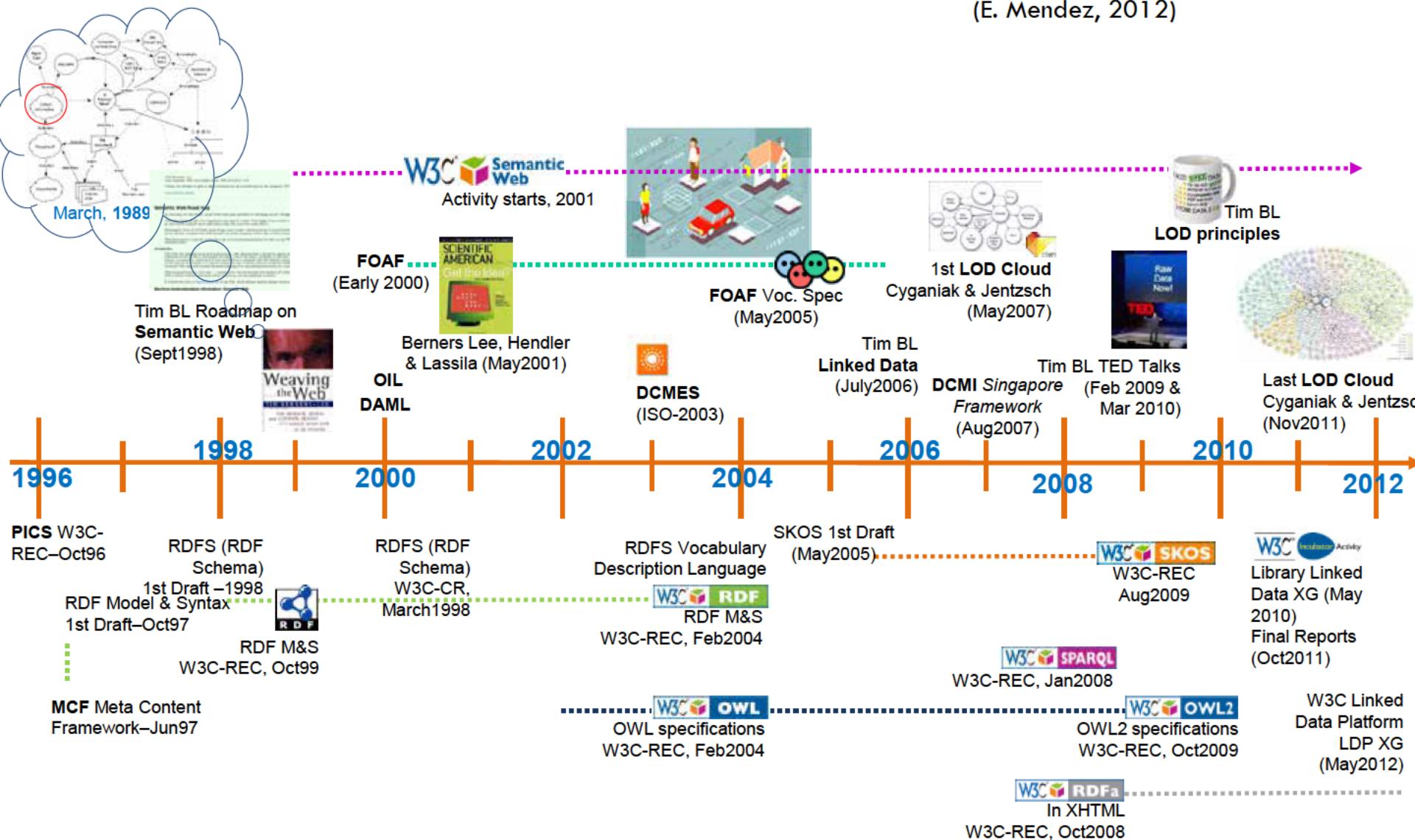
Vocabularies

At times it may be important or valuable to organize data. Using OWL (to build vocabularies, or "ontologies") and SKOS (for designing knowledge organization systems) it is possible to enrich data with additional meaning which

Query

Query languages go hand-in-hand with databases. If the Semantic Web is viewed as a global database, then it is easy to understand why one would need a query language for that data. SPARQL is the query language for the Semantic Web.

(E. Mendez, 2012)

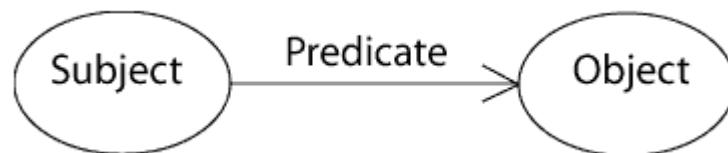


Linked Data based on RDF data model

Is similar to Entity-Relationship or Class diagrams, statements about resource in subject-predicate object expressions called “triples”.

subject= resource

predicate=traits or aspects of the resource and expresses a relationship between the subject and the object.



<http://www.w3.org/TR/rdf-concepts/>

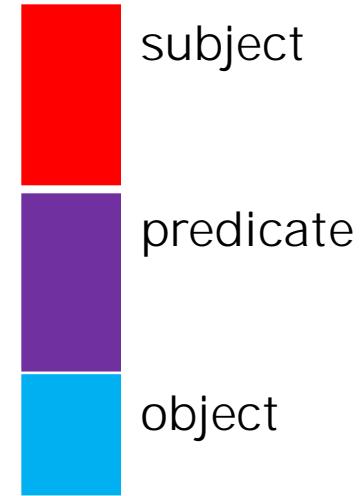
The sky has the color blue

RDF triple:

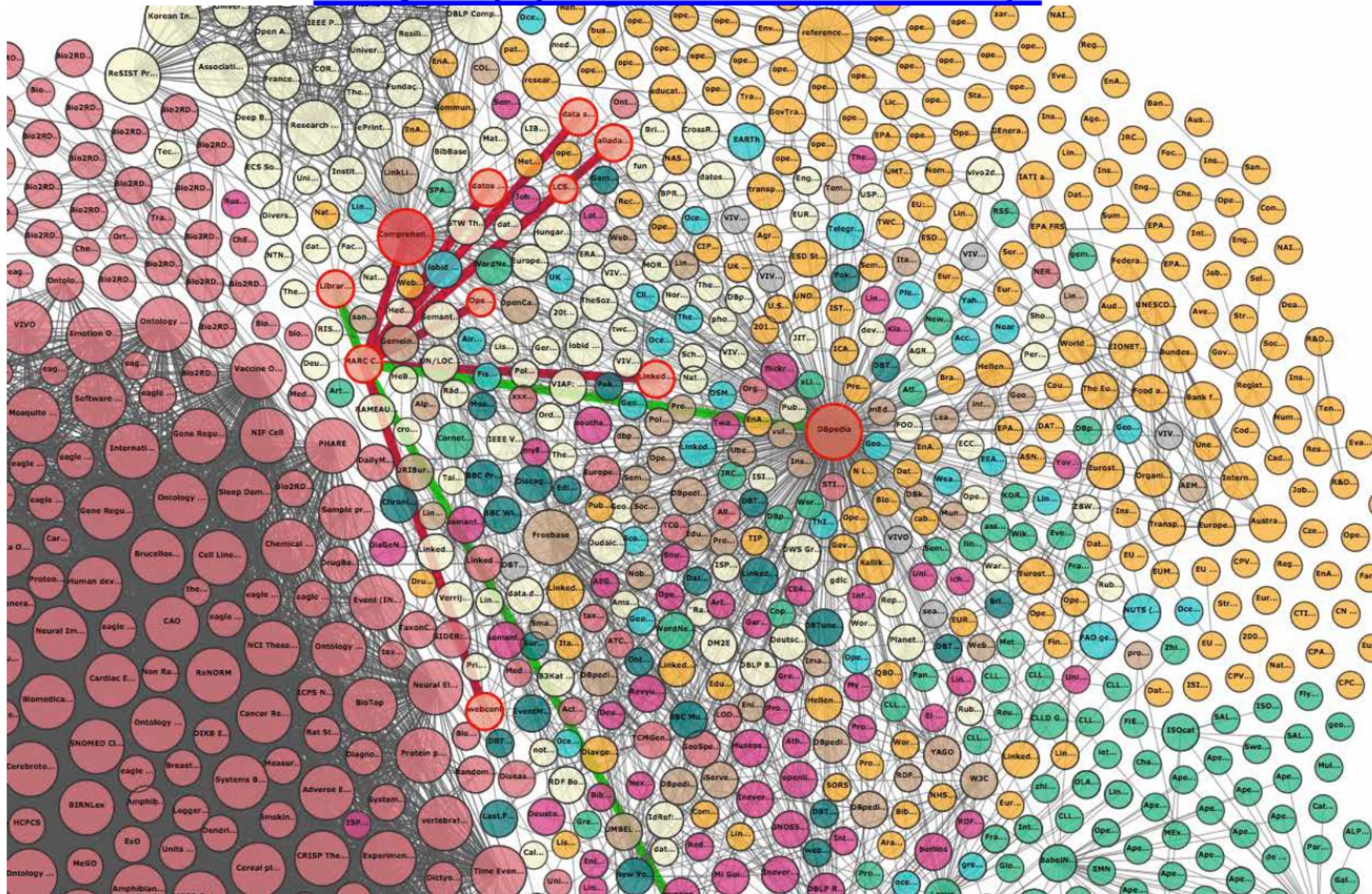
- A subject denoting “the sky”
- A predicate denoting “has the color”
- An object denoting “blue”

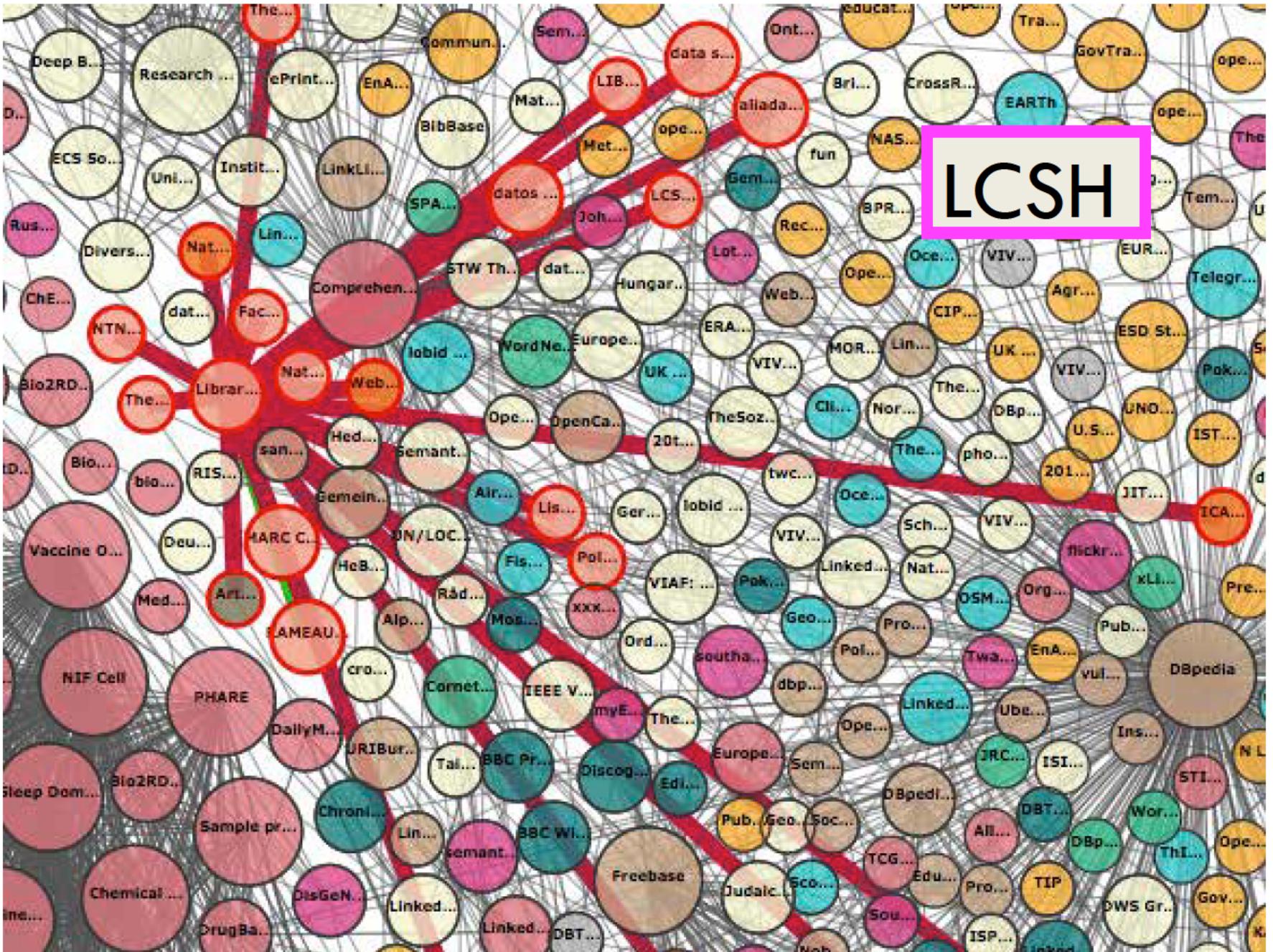
Example of RDF

```
<rdf:RDF  
    xmlns:rdf="http://www.w3.org/1999/02/2  
    2-rdf-syntax-ns#"  
    xmlns:dc="http://purl.org/dc/elements/1.1  
    /">  
  
<rdf:Description  
    rdf:about="https://www.sciencemag.org/ne  
    ws/2020/07/during-pandemic-students-do-  
    field-and-lab-work-without-leaving-home">  
    <dc:title>During the pandemic, students do  
    field...  
    </dc:title>  
    </rdf:Description>  
</rdf>  
  
https://www.sciencemag.org/news/2020/07/during-pandemic-students-do-field-and-lab-work-without-leaving-home
```



<https://lod-cloud.net/>





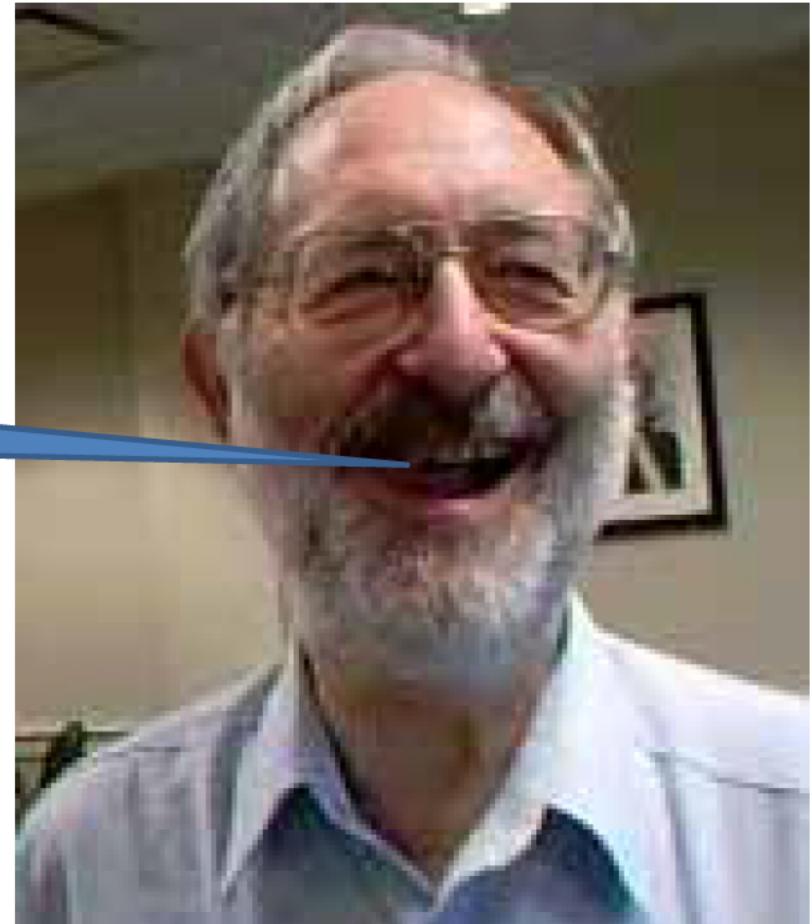
LCSH

Ontologies as “structured terminologies”

- An island “is a type of” land area
- A nation “is a type of” geopolitical area

Relationships

- *Is a type of*
- Causes
- Synonyms



I am an
ontologist

Dagobert Soergel, ACM/CL, 1997
<http://nkos.slis.kent.edu/busch/summary.pdf>