What are ontologies?

Eva Blomqvist

eva.blomqvist@liu.se

Some slides by Patrick Lambrix

LINKÖPING UNIVERSITY

Ontologies

- Definition(s)
- Use
- Components
- Knowledge representation

Ontologies

"*Ontologies define the basic terms and relations comprising the vocabulary of a topic area, as well as the rules for combining terms and relations to define extensions to the vocabulary.*"

(Neches, Fikes, Finin, Gruber, Senator, Swartout, 1991)

Definitions

- An ontology is an explicit specification of a conceptualization (Gruber)
- An ontology is a hierarchically structured set of terms for describing a domain that can be used as a skeletal foundation for a knowledge base. (Swartout, Patil, Knight, Russ)
- An ontology provides the means for describing explicitly the conceptualization behind the knowledge represented in a knowledge base. (Bernaras, Lasergoiti, Correra)
- An ontology is a formal, explicit specification of a shared conceptualization (Studer, Benjamins, Fensel)
- ...

Ontologies on the Semantic Web

- Symbolic models expressed in some formal (logical?) language, preferably a web-enabled language such as OWL, to allow for describing online datasets and automated reasoning

Ontologies used …

- for communication between people and organizations
- for enabling knowledge reuse and sharing
- as basis for interoperability between systems
- as repository of information
- as query model for information sources
- as vocabularies/schemas for Linked Data

Key technology for the Semantic Web

Example

immune response
- **i-** acute-phase response
- **i-** anaphylaxis
- **i-** antigen presentation
- **i-** antigen processing
- **i-** cellular defense response
- **i-** cytokine metabolism
  - **i-** cytokine biosynthesis synonym cytokine production
    - …
  - **p-** regulation of cytokine biosynthesis
    - …
- …
- **i-** B-cell activation
  - **i-** B-cell differentiation
  - **i-** B-cell proliferation
- **i-** cellular defense response
- …
- **i-** T-cell activation
  - **i-** activation of natural killer cell activity
    - …

Biomedical Ontologies - efforts

- International Health Terminology Standards Development Organisation

  http://www.ihtsdo.org

SNOMED CT (**S**ystematized **No**menclature of **Med**icine-**C**linical **T**erms)

Ontologies in biomedical research

- many biomedical ontologies

  e.g. GO, OBO, SNOMED-CT

- practical use of biomedical ontologies

  e.g. databases annotated with GO

GENE ONTOLOGY (GO)

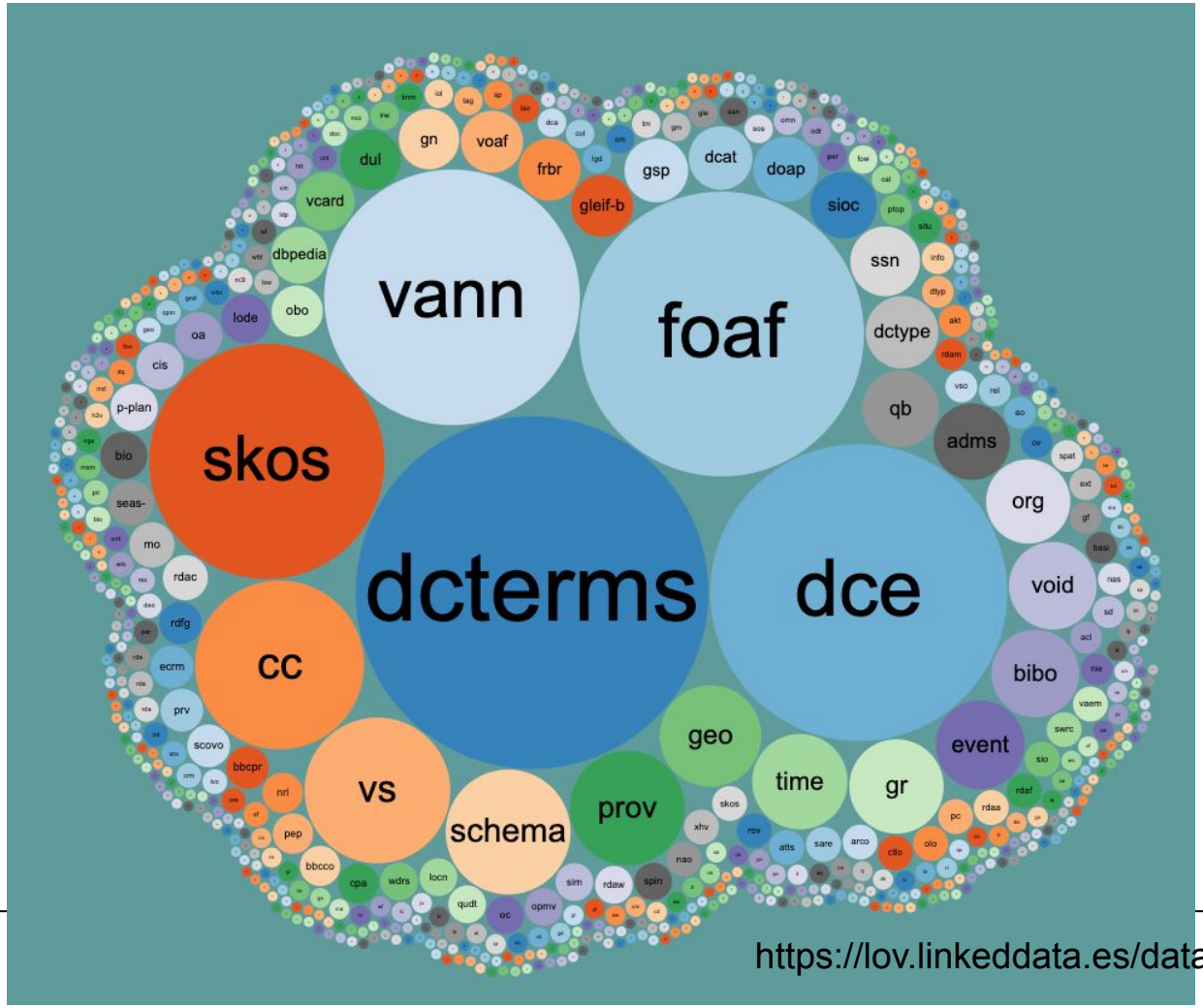
immune response
- i- acute-phase response
- i- anaphylaxis
- i- antigen presentation
- i- antigen processing
- i- cellular defense response
- i- cytokine metabolism
  - i- cytokine biosynthesis
    - synonym cytokine production
    - …
  - p- regulation of cytokine biosynthesis
  - …
- …
- i- B-cell activation
  - i- B-cell differentiation
  - i- B-cell proliferation
- i- cellular defense response
- …
- i- T-cell activation
  - i- activation of natural killer cell activity
  - …

Catalogues:

BioPortal - https://bioportal.bioontology.org/

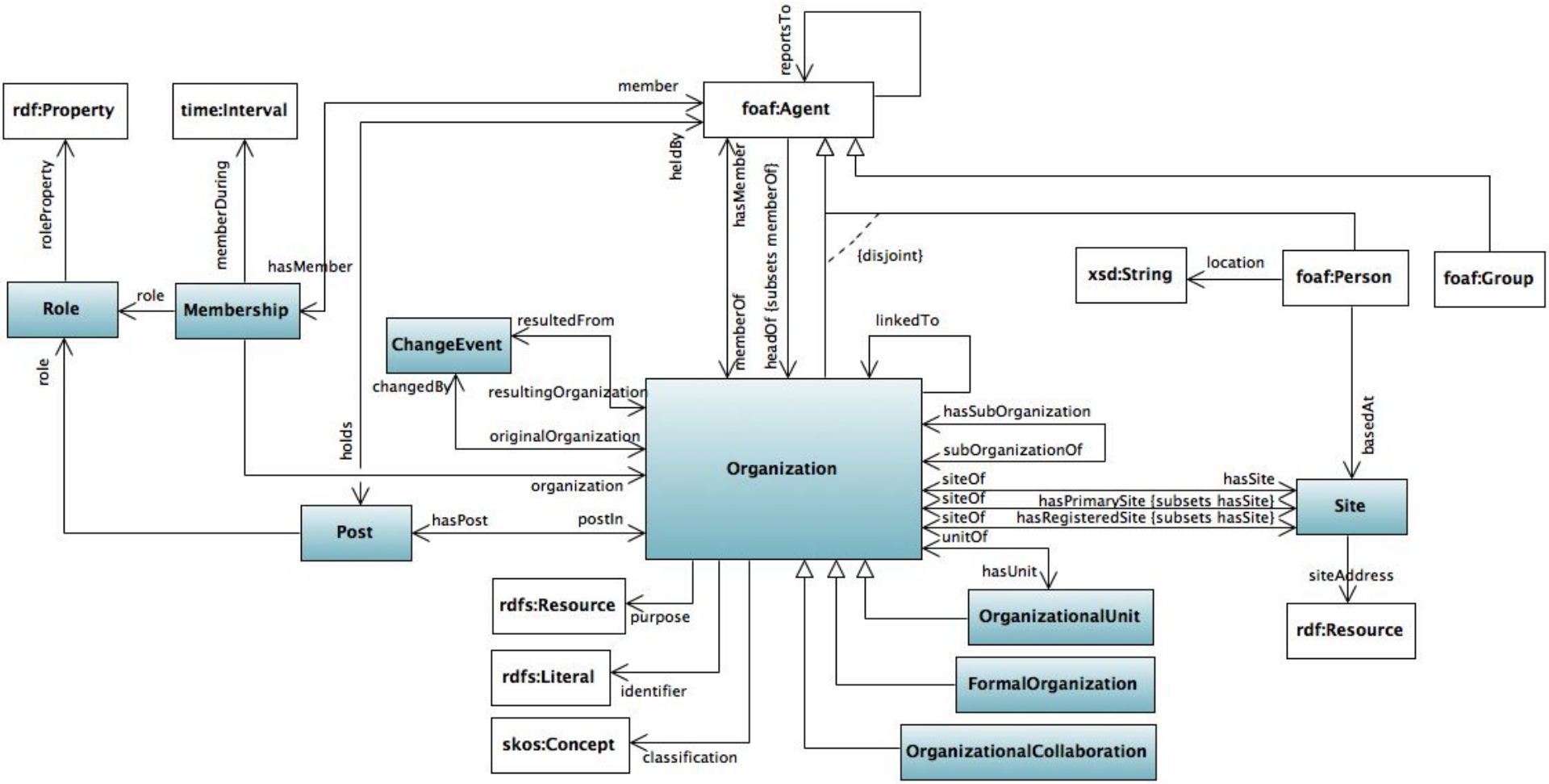
OBO Foundry - http://www.obofoundry.org/

Common ontologies on the Web



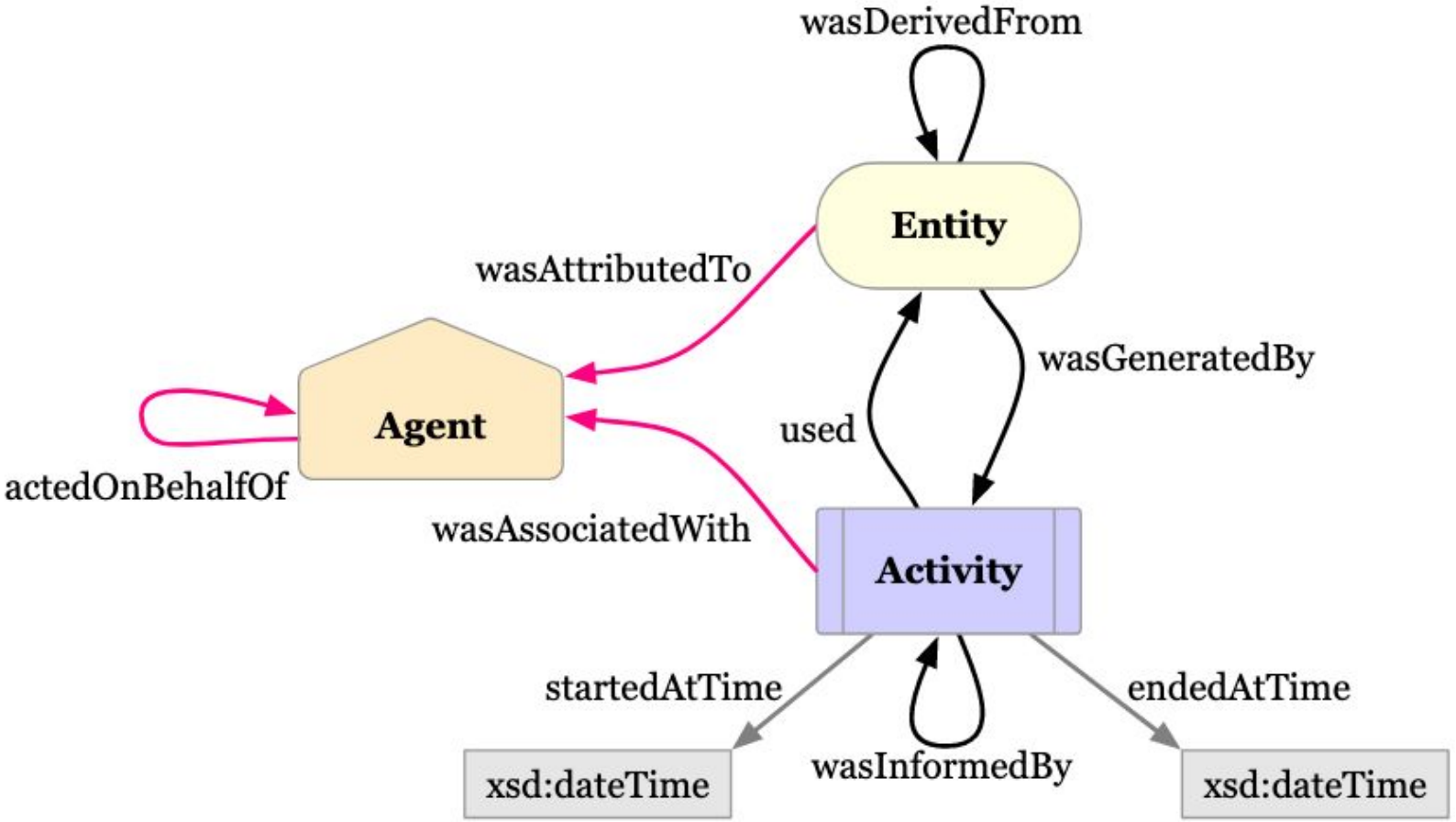
https://lov.linkeddata.es/dataset/lov/

W3C Organisation ontology



https://www.w3.org/TR/vocab-org/

W3C PROV-O



https://www.w3.org/TR/prov-o/

Components

- concepts
  - represent a set or class of entities in a domain
    immune response
  - organized in taxonomies
    (hierarchies based on e.g. *is-a* or *is-part-of*)
    immune response is-a *defense response*

- instances
  - often not represented in an ontology but in the related dataset/knowledge base/knowledge graph

Components

- relations

  R: C1 x C2 x … x Cn

Protein hasName ProteinName

Chromosone hasSubcellularLocation Nucleus

Components

- axioms

  ‘facts that are always true’

The origin of a protein is always of the type ‘gene coding origin type’

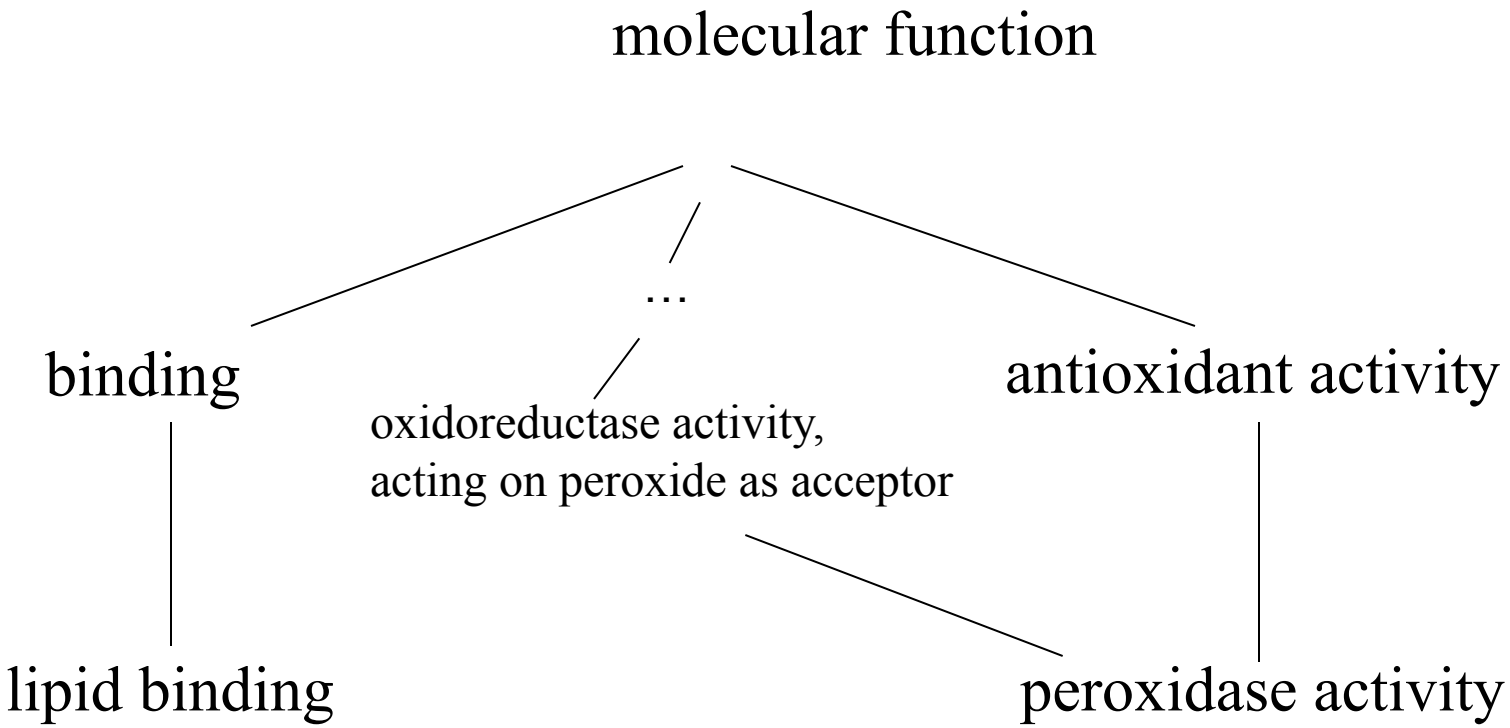
Each protein has at least one source.

A helix can never be a sheet and vice versa.

Different kinds of "ontologies"

- Controlled vocabularies
  Concepts
- Taxonomies
  Concepts, is-a
- Thesauri
  Concepts, predefined relations
- Data models (e.g. EER, UML)
  Concepts, relations, axioms
- Logic-based ontologies
  Concepts, relations, axioms

Taxonomy - GeneOntology



Thesaurus

- graph
- fixed set of relations

(synonym, narrower term, broader term, similar)

Thesaurus - WordNet

thesaurus, synonym finder
=> wordbook
=> reference book, reference, reference work, book of facts
=> book
=> publication
=> print media
=> medium
=> means
=> instrumentality, instrumentation
=> artifact, artefact
=> object, inanimate object, physical object
=> entity
=> work, piece of work
=> product, production
=> creation
=> artifact, artefact
=> object, inanimate object, physical object
=> entity

Tables vs. Knowledge Graph Example

We have a lot of data in tables and want to be able to ask for all research articles

Employment	Person
Press contact	Hans P.
Information off.	Ann M.
Lecturer	Eva B.

Person	Article
Eva B.	Ontology design patterns
Ann M.	Research at LiU 2016-17
Hans P.	Universitetsnytt 22/4 2017

Example (cont.)

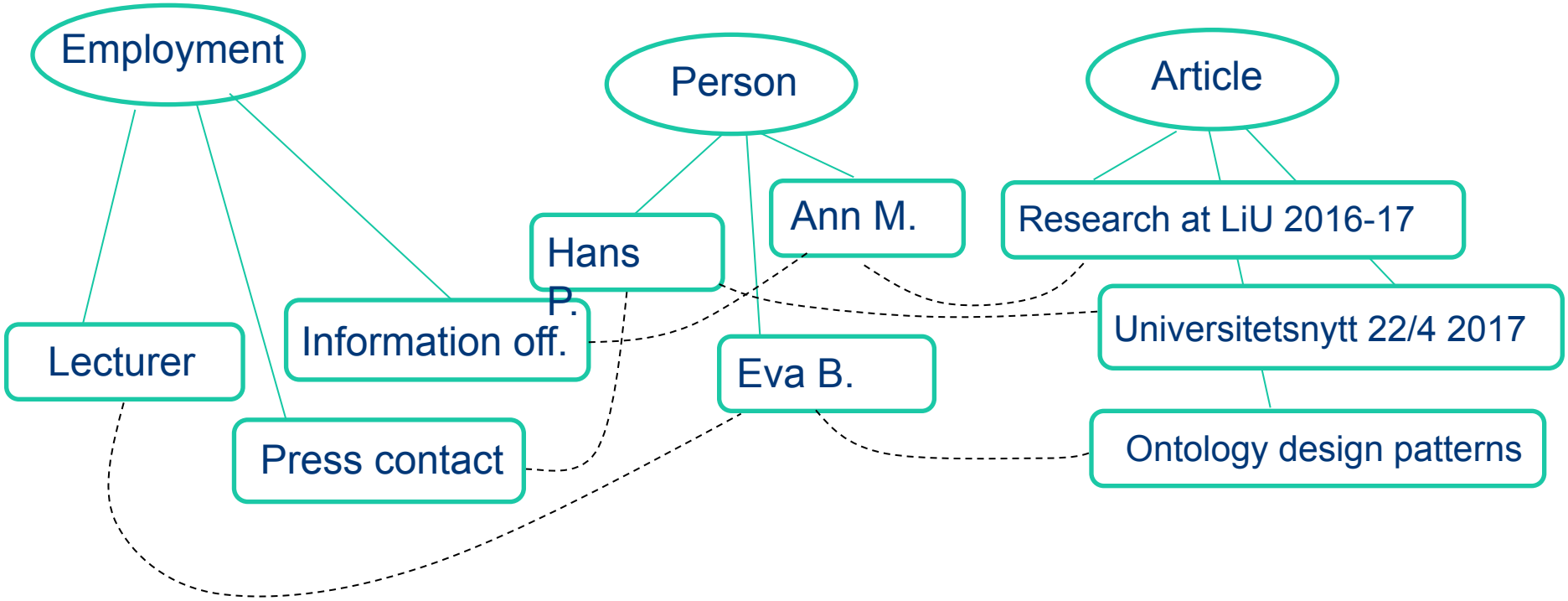
We have a lot of data in tables and want to be able to ask for all research articles

Employment	Person
Press contact	Hans P.
Information off.	Ann M.
Lecturer	Eva B.

Person	Article	Res. art.
Eva B.	Ontology design patterns	yes
Ann M.	Research at LiU 2016-17	no
Hans P.	Universitetsnytt 22/4 2017	no

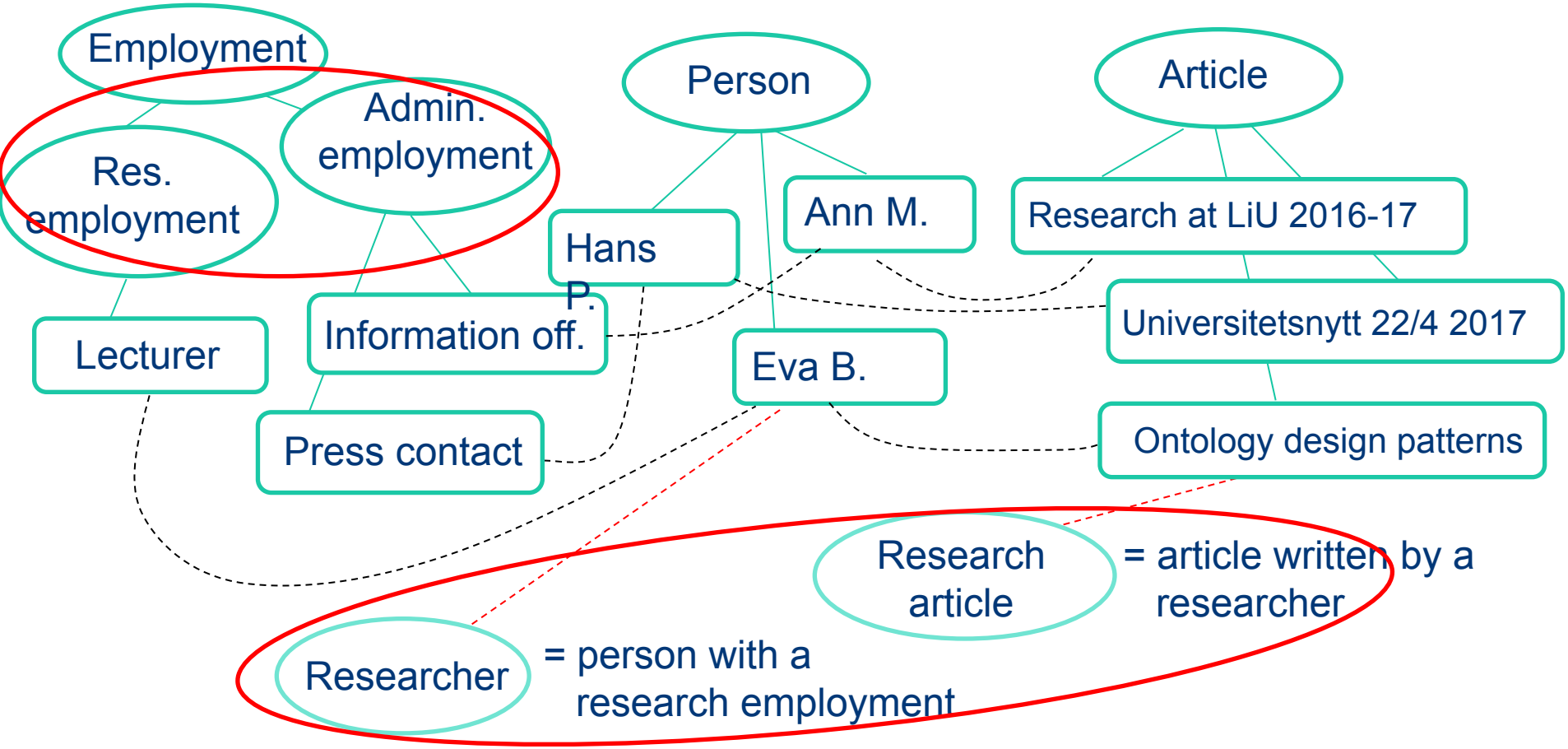
Example (cont.)

We have a lot of data, put it in a graph, and want to be able to ask for all research articles



Example (cont.)

We have a lot of data, put it in a graph, and want to be able to ask for all research articles



RDF(S): RDF Schema

- RDF gives a data representation format and ways to serialize, but it does not give any special meaning to vocabulary such as “subClassOf” or “range”

- Triple interpretation is an arbitrary binary relation

- RDF Schema extends RDF with elements to express a vocabulary for the data
  - Classes as types for individuals: rdfs:Class, rdfs:Literal, rdfs:Datatype, rdf:type and rdfs:subClassOf, etc.
  - Property relations: rdf:Property, rdfs:subPropertyOf, rdfs:range, rdfs:domain, etc.
  - Annotations: rdfs:label, rdfs:comment, etc.

RDF/RDF(S) “Liberality”

- No distinction between classes and instances (individuals)
- Properties can themselves have properties
- No distinction between language constructors and ontology vocabulary, so constructors can be applied to themselves/each other

What does RDF(S) give us?

- Ability to use simple schema/vocabularies when describing our resources
- Consistent vocabulary use and sharing
- Simple inference, e.g. inheritance in a taxonomy
- But…
  - In some cases too weak (inexpressive) to describe resources in sufficient detail
  - Not formally based on any logic

Logics

- Formal languages
- Syntax, semantics, inference mechanisms
- Description Logics is the foundation of the Web Ontology Language (OWL)
  - W3C standard
  - Builds on top of RDF(s) but also restricts it

OWL

- DL-based language
- OWL2 is the latest version of the standard (https://www.w3.org/TR/owl2-primer/)
- Different language profiles
  - OWL EL, OWL QL, OWL RL
- Add language constructs such as intersection, union of concepts, but also owl:sameAs for connecting data

Defining ontologies is not so easy ...

The Celestial Emporium of Benevolent Knowledge, Borges (1942)

"On those remote pages it is written that animals are divided into:

a. those that belong to the Emperor

b. embalmed ones

c. those that are trained

d. suckling pigs

e. mermaids

f. fabulous ones

g. stray dogs

h. those that are included in this classification

i. those that tremble as if they were mad

j. innumerable ones

k. those drawn with a very fine camel's hair brush

l. others

m. those that have just broken a flower vase

n. those that resemble flies from a distance"

Slide from talk by C. Goble

Defining ontologies is not so easy ...

Dyirbal classification of objects in the universe

- Bayi: men, kangaroos, possums, bats, most snakes, most fishes, some birds, most insects, the moon, storms, rainbows, boomerangs, some spears, etc.
- Balan: women, anything connected with water or fire, bandicoots, dogs, platypus, echidna, some snakes, some fishes, most birds, fireflies, scorpions, crickets, the stars, shields, some spears, some trees, etc.
- Balam: all edible fruit and the plants that bear them, tubers, ferns, honey, cigarettes, wine, cake.
- Bala: parts of the body, meat, bees, wind, yamsticks, some spears, most trees, grass, mud, stones, noises, language, etc.

Slide from talk by C. Goble

Ontology tools

- Ontology development tools
- Ontology merge and alignment tools
- Ontology evaluation tools
- Ontology-based annotation tools
- Ontology storage and querying tools
- Ontology learning tools
- Ontology debugging and completion tools

Further reading ontologies

Books about ontologies and ontology engineering:

- Semantic Web for the Working Ontologist, by Allemang, Hendler & Gandon - 3rd edition (https://workingontologist.org/)
- An introduction to ontology engineering, by Keet (https://open.umn.edu/opentextbooks/textbooks/590)

Some older articles:

- KnowledgeWeb (http://knowledgeweb.semanticweb.org/) and its predecessor OntoWeb (http://ontoweb.aifb.uni-karlsruhe.de/)
- Lambrix, Tan, Jakoniene, Strömbäck, Biological Ontologies, chapter 4 in Baker, Cheung, (eds), Semantic Web: Revolutionizing Knowledge Discovery in the Life Sciences, 85-99, Springer, 2007. ISBN: 978-0-387-48436-5.

(general about ontologies)

- Lambrix, Towards a Semantic Web for Bioinformatics using Ontology-based Annotation, Proceedings of the 14th IEEE International Workshops on Enabling Technologies: Infrastructures for Collaborative Enterprises, 3-7, 2005. Invited talk.

The OWL language:

- OWL, https://www.w3.org/OWL/

www.liu.se

LINKÖPING UNIVERSITY