5. Ontologies and languages in the biomedical semantic web

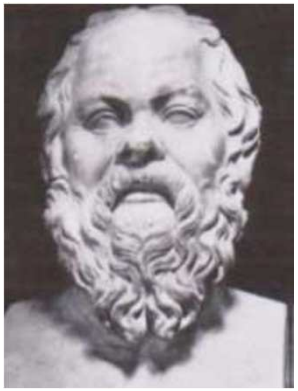
Dr. Isabella Hinterleitner
hinterleitner.ih@gmail.com

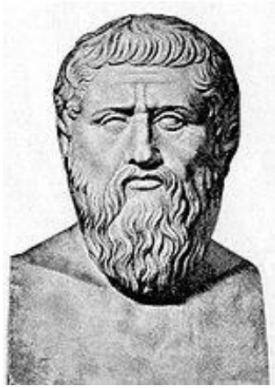
Agenda

- Ontologies
  - Definition
  - Examples
  - Ontology Types
  - Ontology Management & Design
  - Connecting Ontologies
- Semantic Web
  - Standards
  - XML Examples
  - OWL and RDF Triples

What is an Ontology? (1)

- **Philosophy**: The word „ontology“ comes from the philosophical discipline which deals with being (dem Sein), the being as a person and fundamental types of entitites.
- Separation between
  - Concept / class (= idea)
  - Instance (object of the real world, “shadow” of ideas)
  - Hierarchy of concepts
- (Sokrates, Platon)



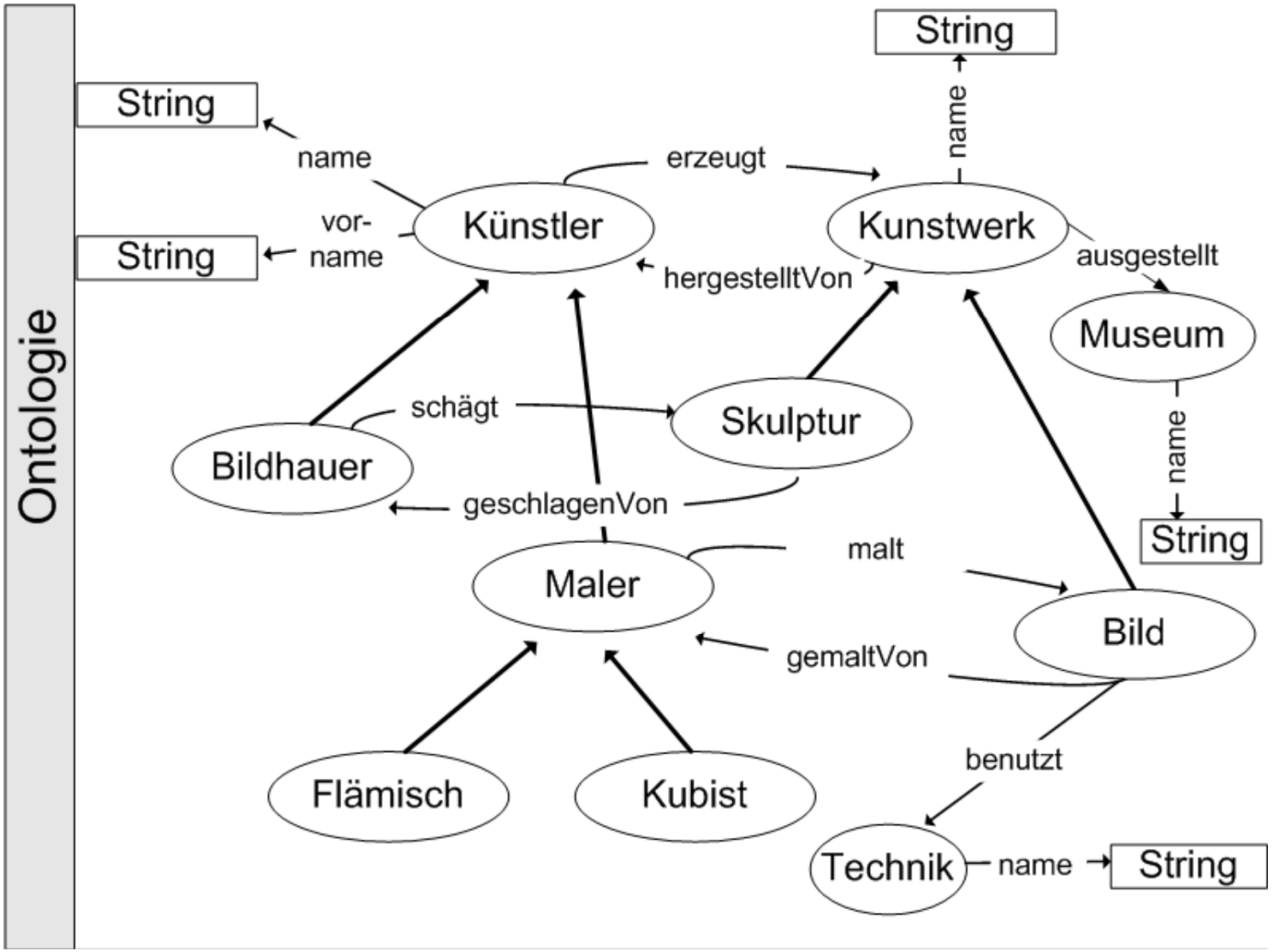


What is an Ontology? (2)

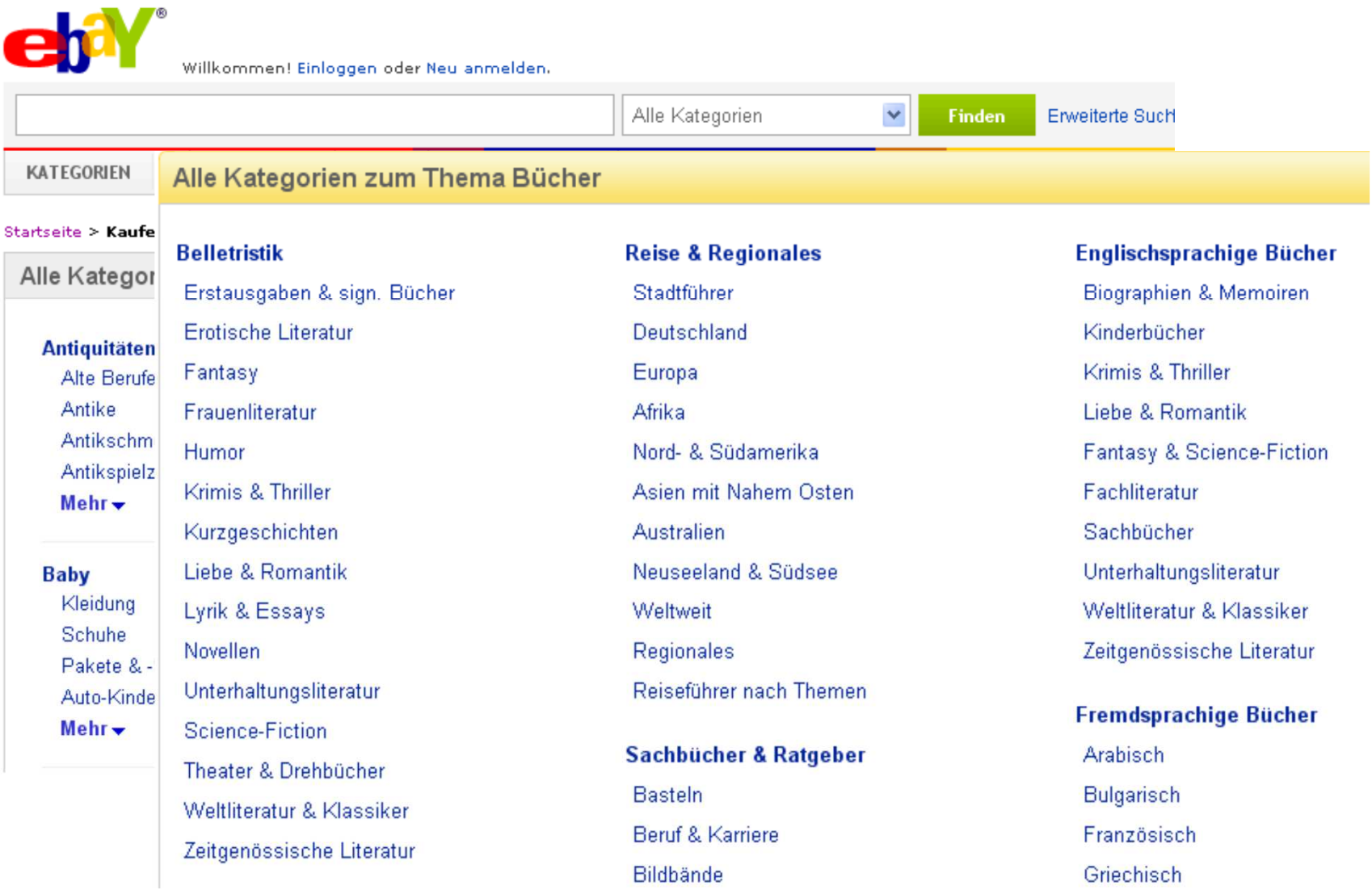
An ontology is an explicit, formal specification of a shared conceptualization. The term is borrowed from philosophy, where an ontology is a systematic account of Existence. For knowledge-based systems, what “exists” is exactly that which can be” represented. *(Thomas R. Gruber, 1993)*

- **Conceptualization** abstract modell (domain, identified relevant expressions, relationships)
- **Explicit** meaning of all expressions is defined
- **Formal** machine-readable
- **Shared** consense regarding ontology

Simple Ontology



Ontology in e-business



Existing Ontologies in Science

- NCI Metatheasaurus

  Cancer Biology

- Gene Ontology

  Gene Products & Biological Processes

- BioPAX

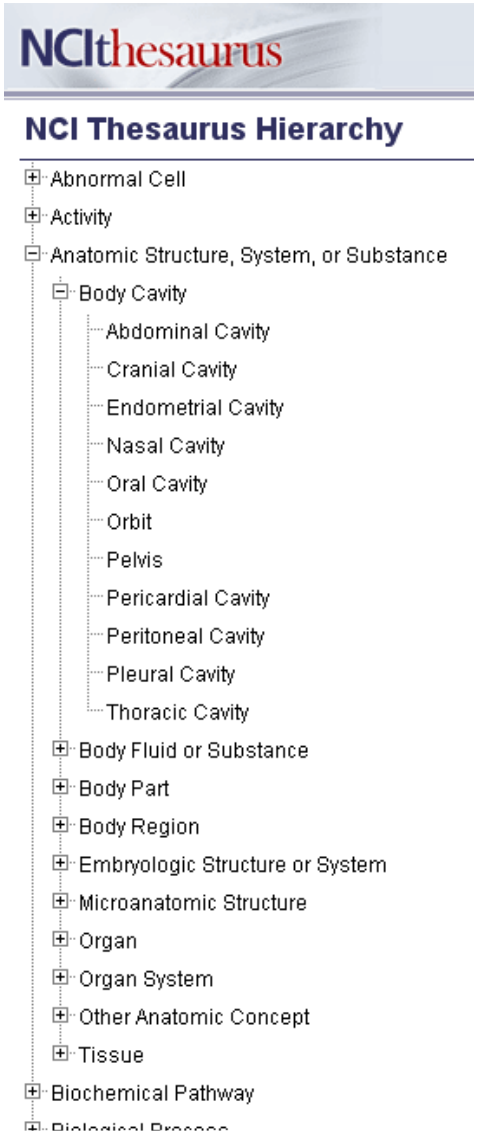
  Biological Pathways

- TAMBIS

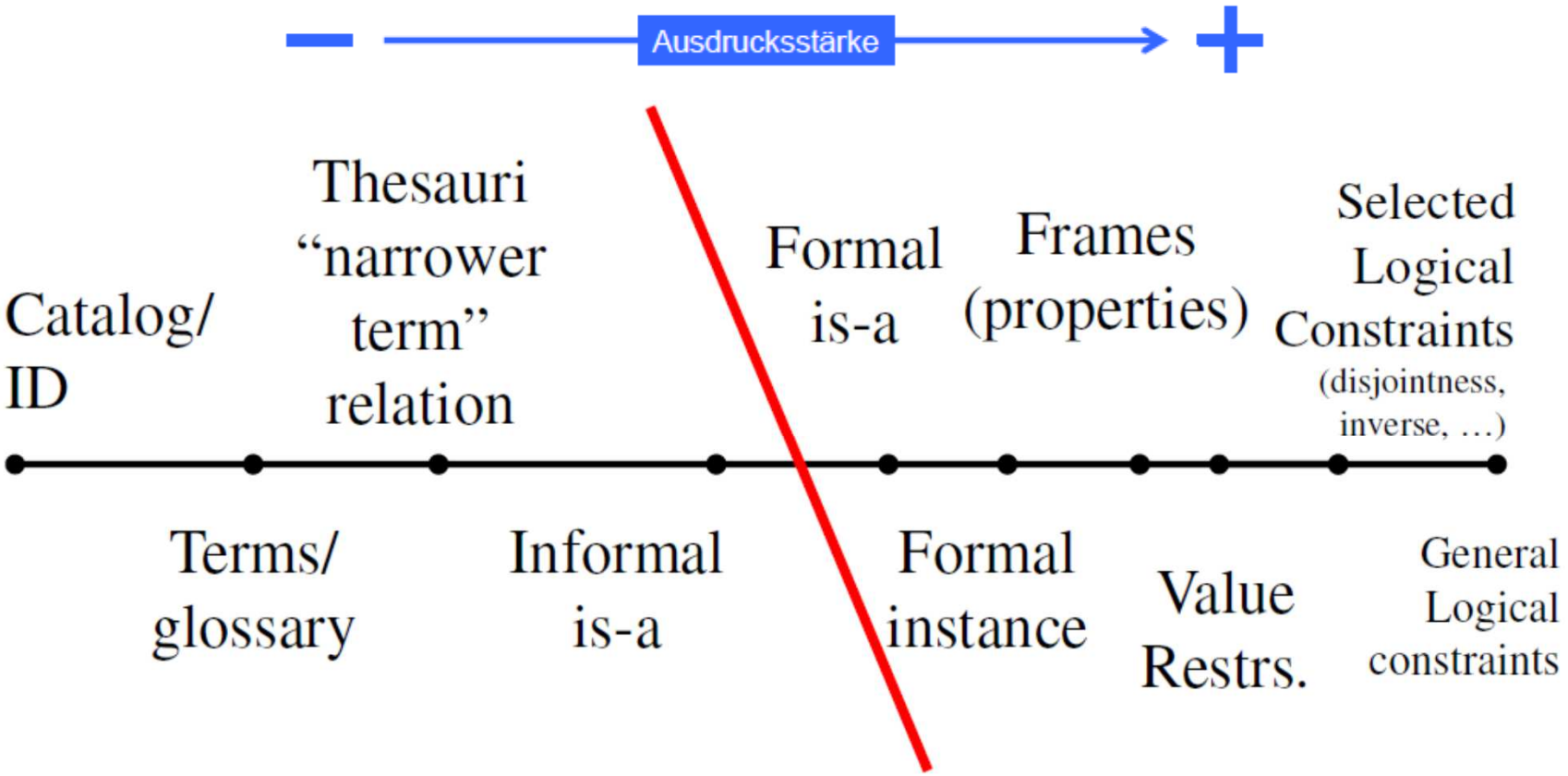
  Molecular Biology

- MGED

  Microarray Data

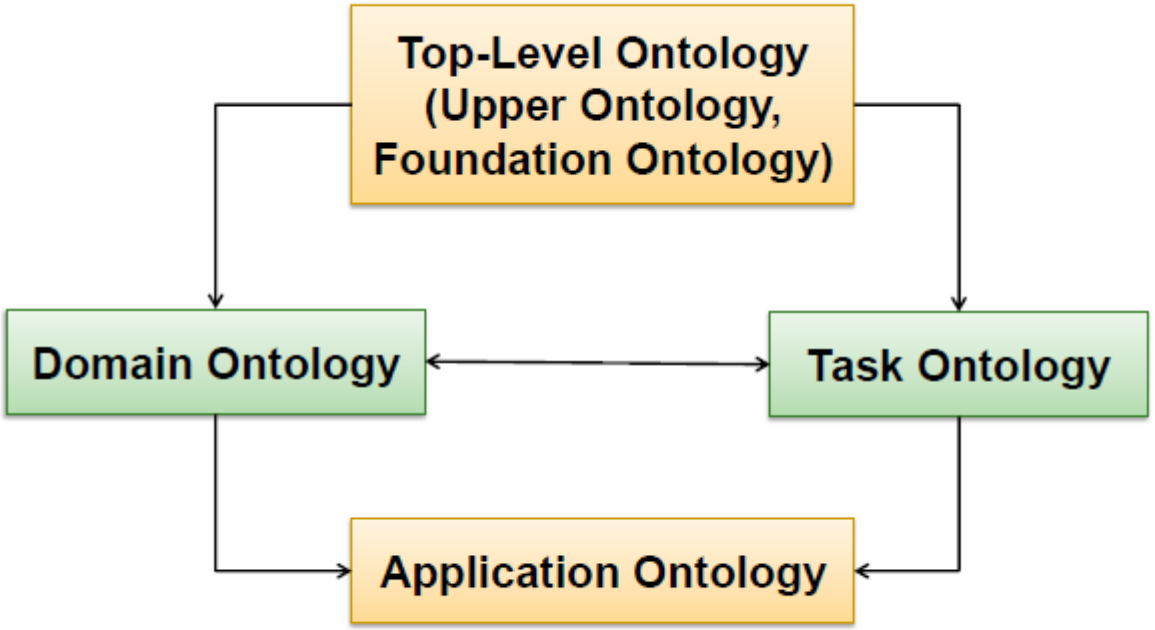


Expressiveness of Ontologies



Ontologie-Spektrum nach McGuinness, 2001

Type of Ontologies



Klassifikation nach Guarino, 1998

Ontology Types

Top-Level Ontology

- General, interdisciplinary ontology
- Describes general concepts (Zeit, Raum, Vorgang) independen of
- A certain domain or problem statement

Domain Ontology

- General concepts regarding a generic domain

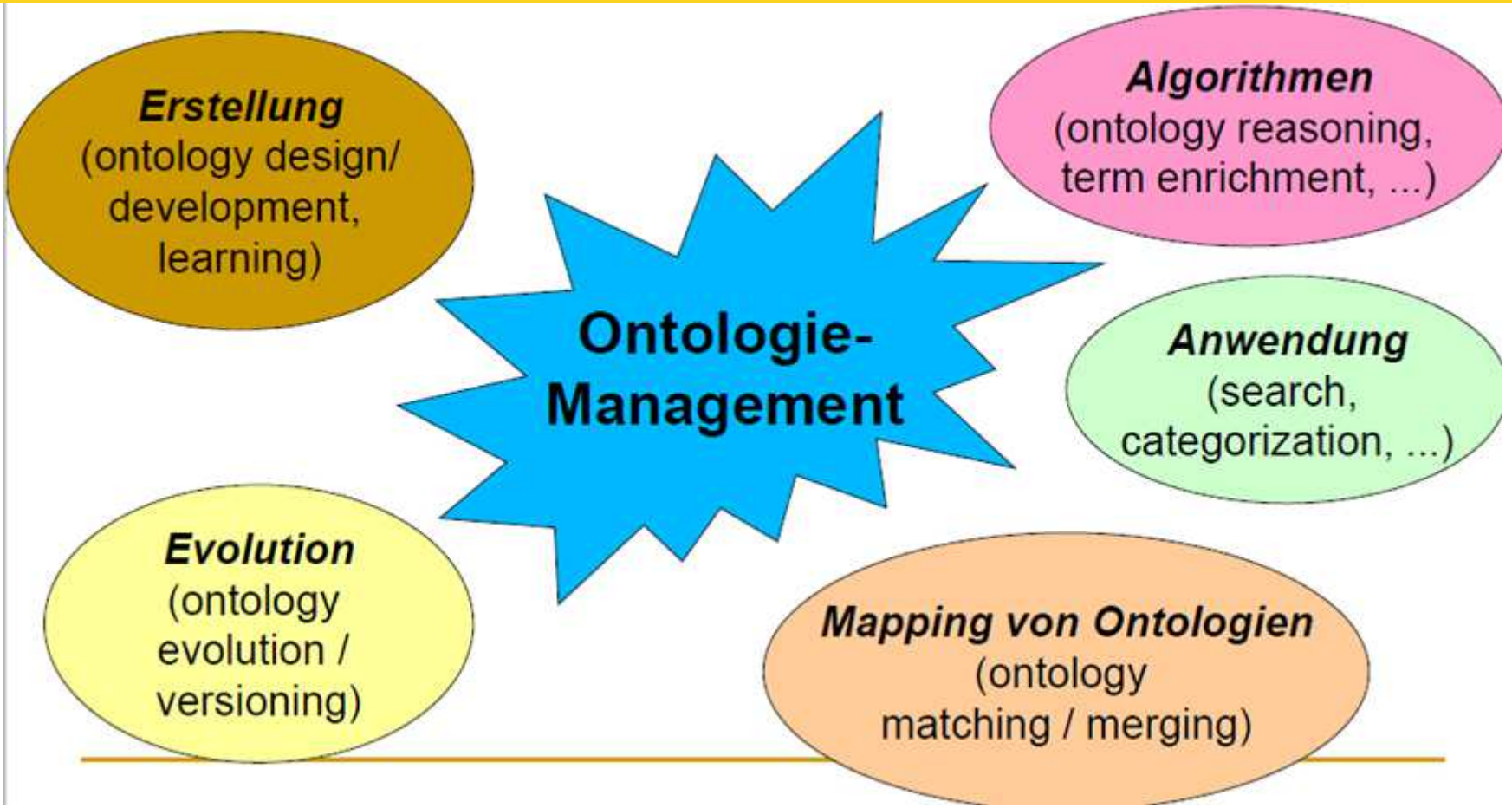
Task Ontology

- Basic concepts regarding tasks

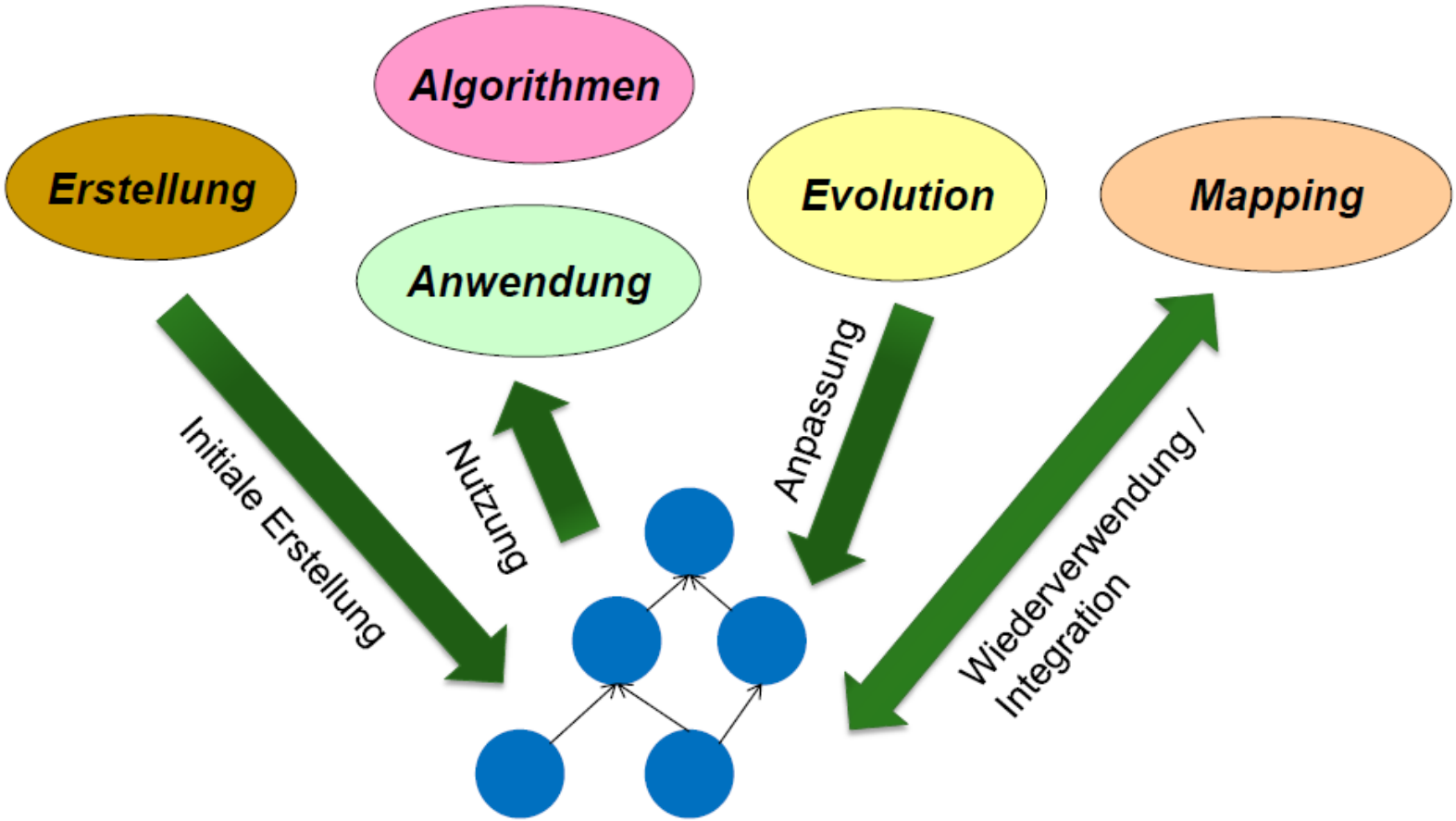
Application Ontology

- Especially regarding a concrete focussed domain or tasks
- is focussed in general on a domain or task ontology

Ontology Management



Ontology Management



Design and Development

Methods of Design

- Includes all activities that are needed for the design of ontologies

Why?

- Development of consistent ontologies
- Efficient development of complex ontologies
- Distributed development of ontologies

Support by means of Tools

- Partly Semi(automatic) solutions
- Ontology Learning

Ontology Design

1) There is no one correct way to model a domain — there are always viable alternatives. The best solution almost always depends on the application that you have in mind and the extensions that you anticipate.

2) Ontology development is necessarily an iterative process.

3) Concepts in the ontology should be close to objects (physical or logical) and relationships in your domain of interest. These are most likely to be nouns (objects) or verbs (relationships) in sentences that describe your domain.

N.F. Noy, D. McGuinness: Ontology Development 101: A Guide to Creating Your First Ontology. Stanford Knowledge Systems Laboratory, 2001.

Tools

protégé

KAON

OntoWiki

LinKFactory®

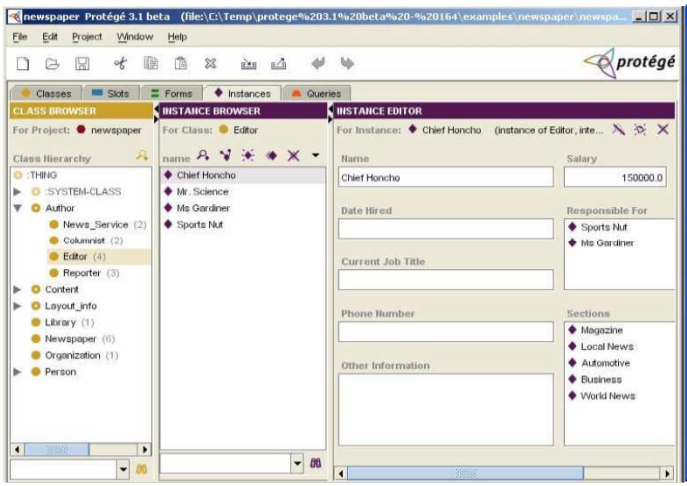
OILed

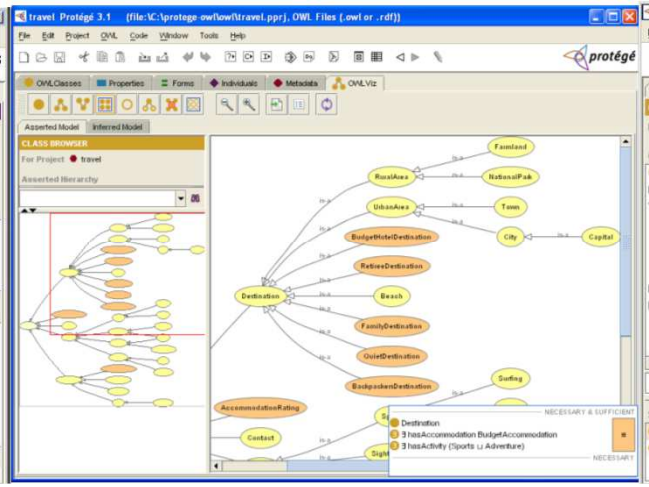
www.ontoknowledge.org/oil

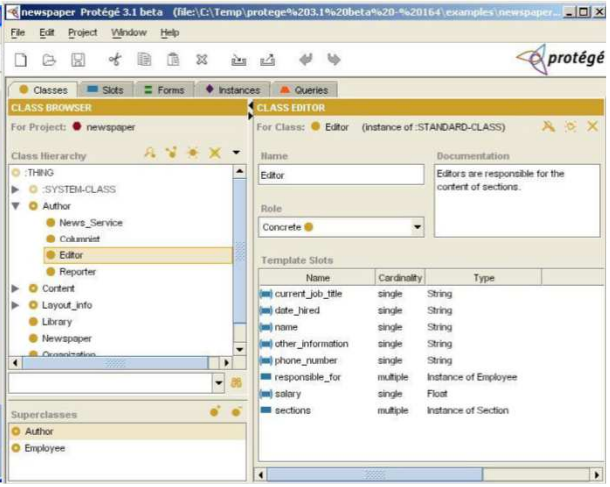
TopBraid Composer™

KSL Ontolingua

Joe Java Ontology Editor







Dynamics of Ontologies

Consistent Evolution

- Implementation of a new requirement/adaption of the ontology
- consistency: After the modification the ontology should be consistent again
- Example: Deleting/Adding a concept
- Versioning
- Accessing different versions of ontology
- How is versioning realized?
- Differences between two ontology versions
- How did version 2 develop from version 1

Connecting Ontologies

Ontologie-Mapping

- Mapping between two ontologies **A** and **B**
- Which Ontologieconcepts from **A** resemble the concepts in **B**?
- *Ontology Matching* → (semi) automatic comparison of ontologies
- Different techniques can be applied
- Integration (merging) of different ontologies
- Support new analysis methods for ontologies
- Reuse of knowledge!

The Semantic Web Vision

"The Semantic Web is a vision: the idea of having data on the Web defined and linked in such a way that it can be used by machines not just for display purposes, but for automation, integration and reuse of data across various applications."

Tim Berners - Lee

Semantic Web Standards

- XML – E**x**tensible **M**arkup **L**anguage
- RDF – **R**esource **D**escription **F**ramework
- OWL – **W**eb **O**ntology **L**anguage

W3C WORLD WIDE WEB consortium

The metadata revolution

- E**x**tensible **M**arkup **L**anguage
- Formatting documents and data by assigning

Metadata

(Data about data)

- Syntax resembles HTML

Examples for XML Formats

- SBML *(Systems Biology Markup Language)*

- CellML

http://sbml.org http://cellml.org

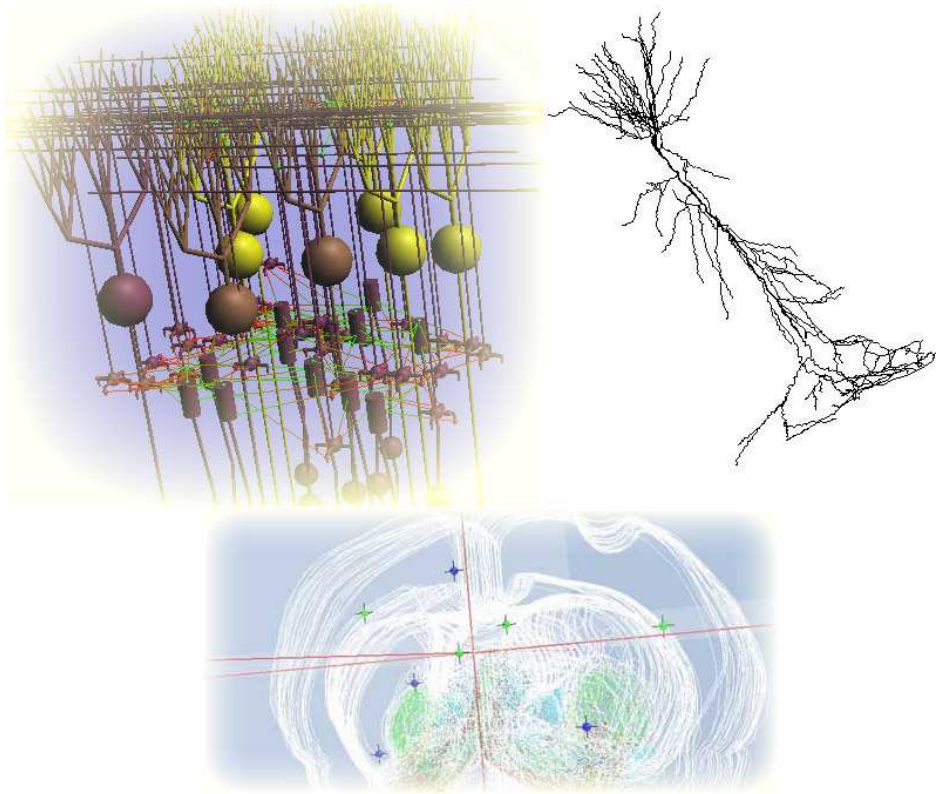
Examples for XML Formats

```
<variable name="i_NaK" public_interface="in" units="picoA" />
<variable name="i_SR_U" public_interface="in" units="picoA" />

<math xmlns="http://www.w3.org/1998/Math/MathML">
 <apply id="ATP_contraction_diff_eq"><eq />
   <apply><diff />
     <bvar><ci> time </ci></bvar>
     <ci> ATP_contraction </ci>
   </apply>
   <apply><times />
     <cn cellml:units="dimensionless"> 0.4 </cn>
     <ci> TCa_ </ci>
     <ci> troponin </ci>
   </apply>
 </apply>

 <apply id="ATP_synthesis_eq"><eq />
   <apply><diff />
     <bvar><ci> time </ci></bvar>
     <ci> ATPi </ci>
   </apply>
   <apply><minus />
     <apply><times />
       <cn cellml:units="dimensionless"> 0.003 </cn>
       <ci> ADP </ci>
     </apply>
     <apply><plus />
       <apply><times />
         <cn cellml:units="dimensionless"> 0.4 </cn>
         <ci> TCa_ </ci>
```

CellML Model: Current Systems in Ventricular Cells

Examples for XML Formats

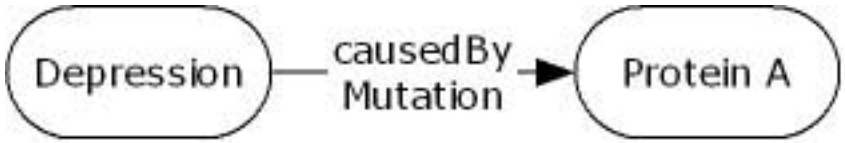
- Many other domain – specific formats
- e.g. BrainML, NeuroML, MorphML…



RDF/OWL

- XML gives structure to data and documents
- RDF and OWL add logic, meaning and connectivity

RDF/OWL is based on triples



Subject - Predicate - Object

RDF is based on triples

<Depression> <causedByMutation> <Protein A>

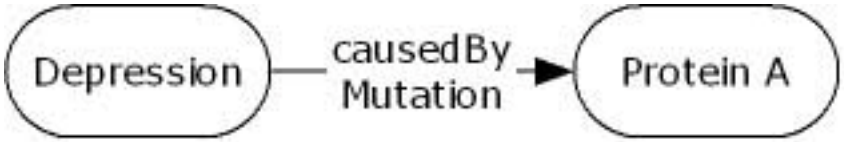
Subject - Predicate - Object

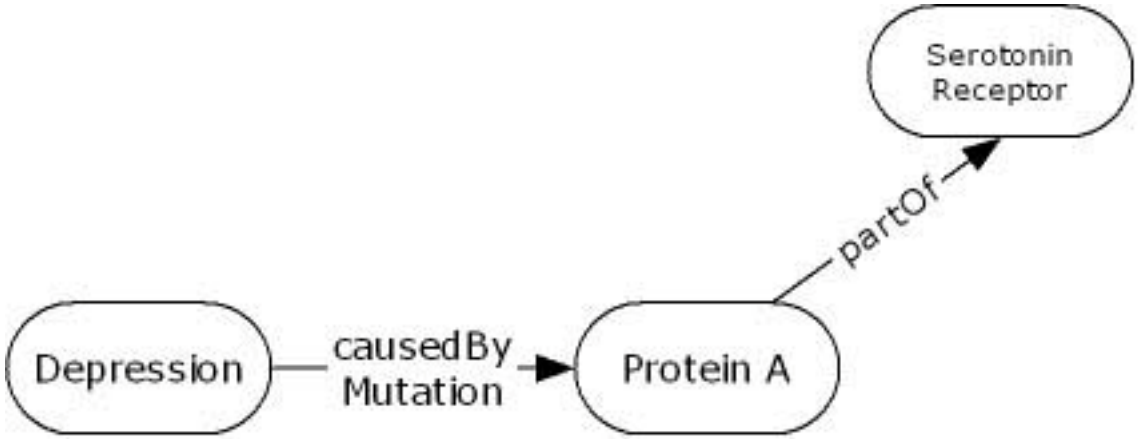
RDF is based on triples

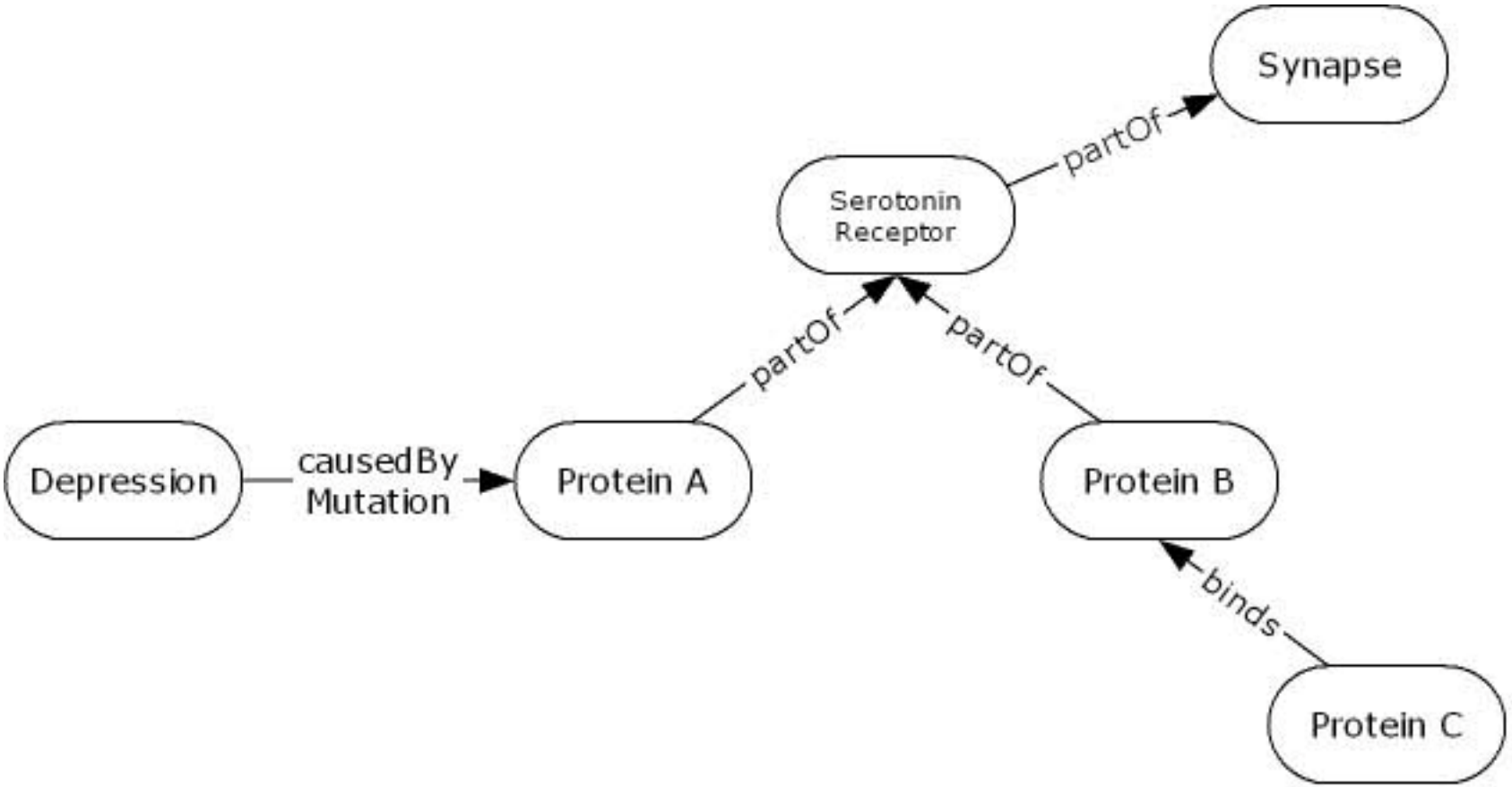
URIs (Unified Resource Identifieres) are used to identify resources

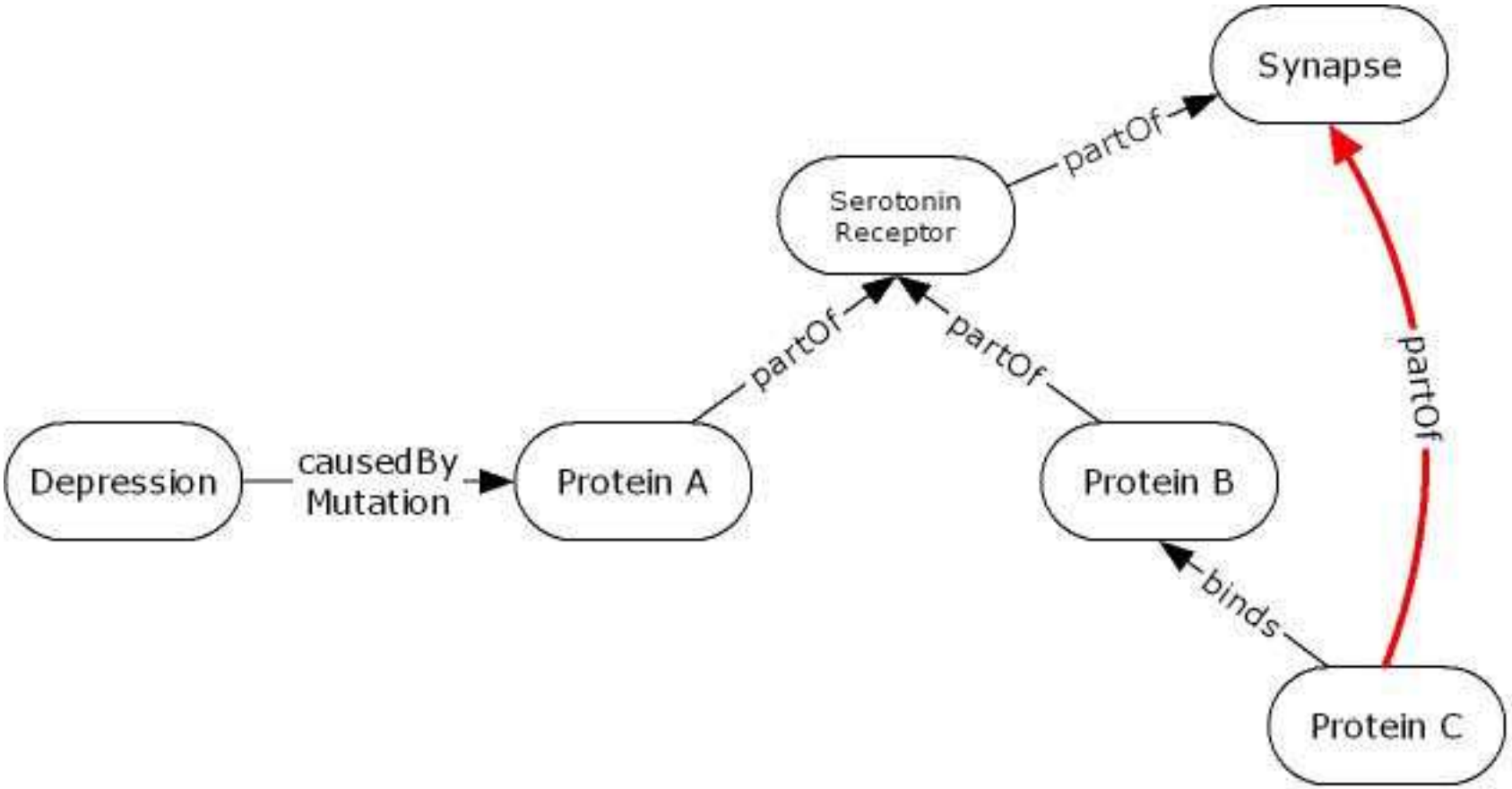
<http://psychiatry.org/disease#**depression**>
<http://genomics.org/ontology#**causedByMutation**>
<http://proteindatabank.org/data#**ProteinA**>

Subject - Predicate - Object

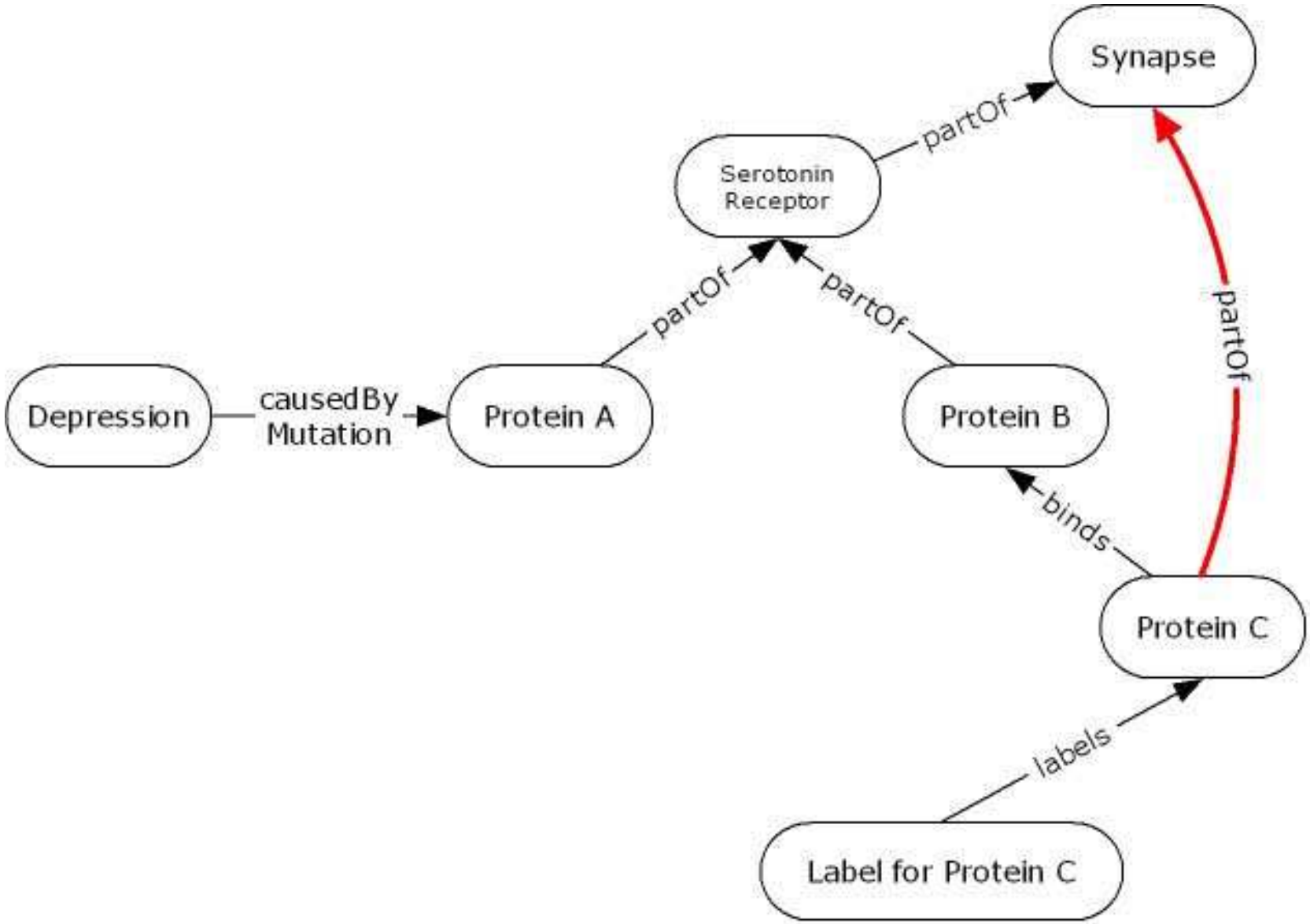

Depression
causedBy
Mutation
Protein A

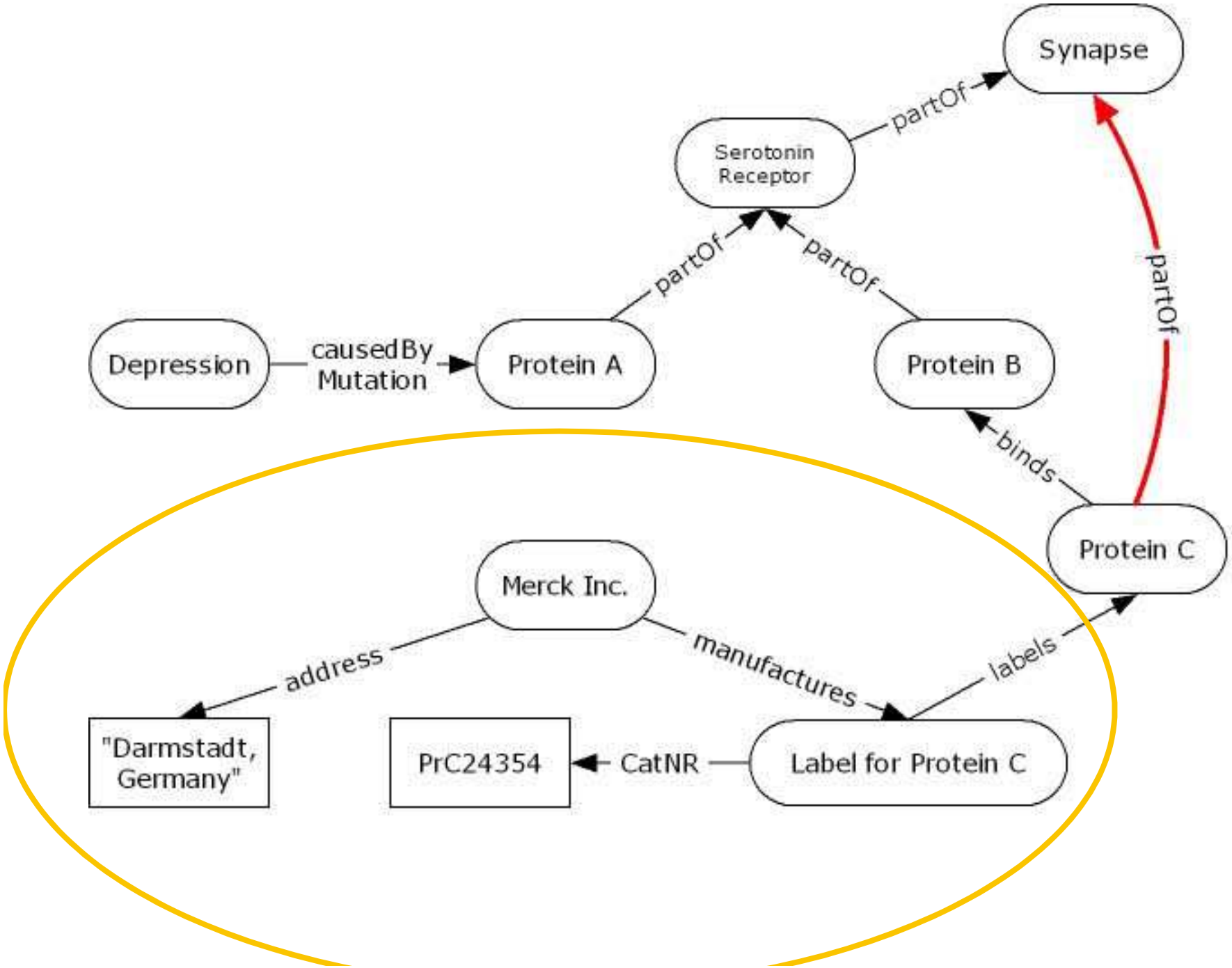

Serotonin
Receptor
partOf
Depression
causedBy
Mutation
Protein A


Synapse
partOf
Serotonin Receptor
partOf
partOf
Depression
causedBy Mutation
Protein A
Protein B
binds
Protein C


Synapse
partOf
Serotonin Receptor
partOf
partOf
partOf
Depression
causedBy Mutation
Protein A
Protein B
binds
Protein C

Inference


Synapse
partOf
Serotonin Receptor
partOf
partOf
partOf
Depression
causedBy Mutation
Protein A
Protein B
binds
Protein C
labels
Label for Protein C


Synapse
partOf
Serotonin Receptor
partOf
partOf
partOf
Depression
causedBy Mutation
Protein A
Protein B
binds
Protein C
Merck Inc.
address
manufactures
labels
"Darmstadt, Germany"
PrC24354
CatNR
Label for Protein C

Thanks for your attention