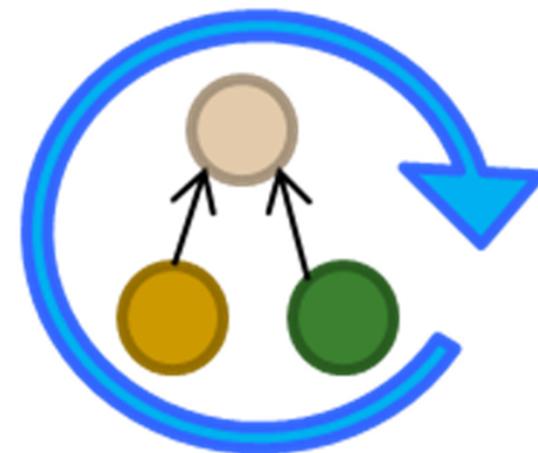

Ontologie-Management

Kapitel 3: Anwendungen

Dr. Michael Hartung

Wintersemester 2012/13

Universität Leipzig
Institut für Informatik
<http://dbs.uni-leipzig.de>

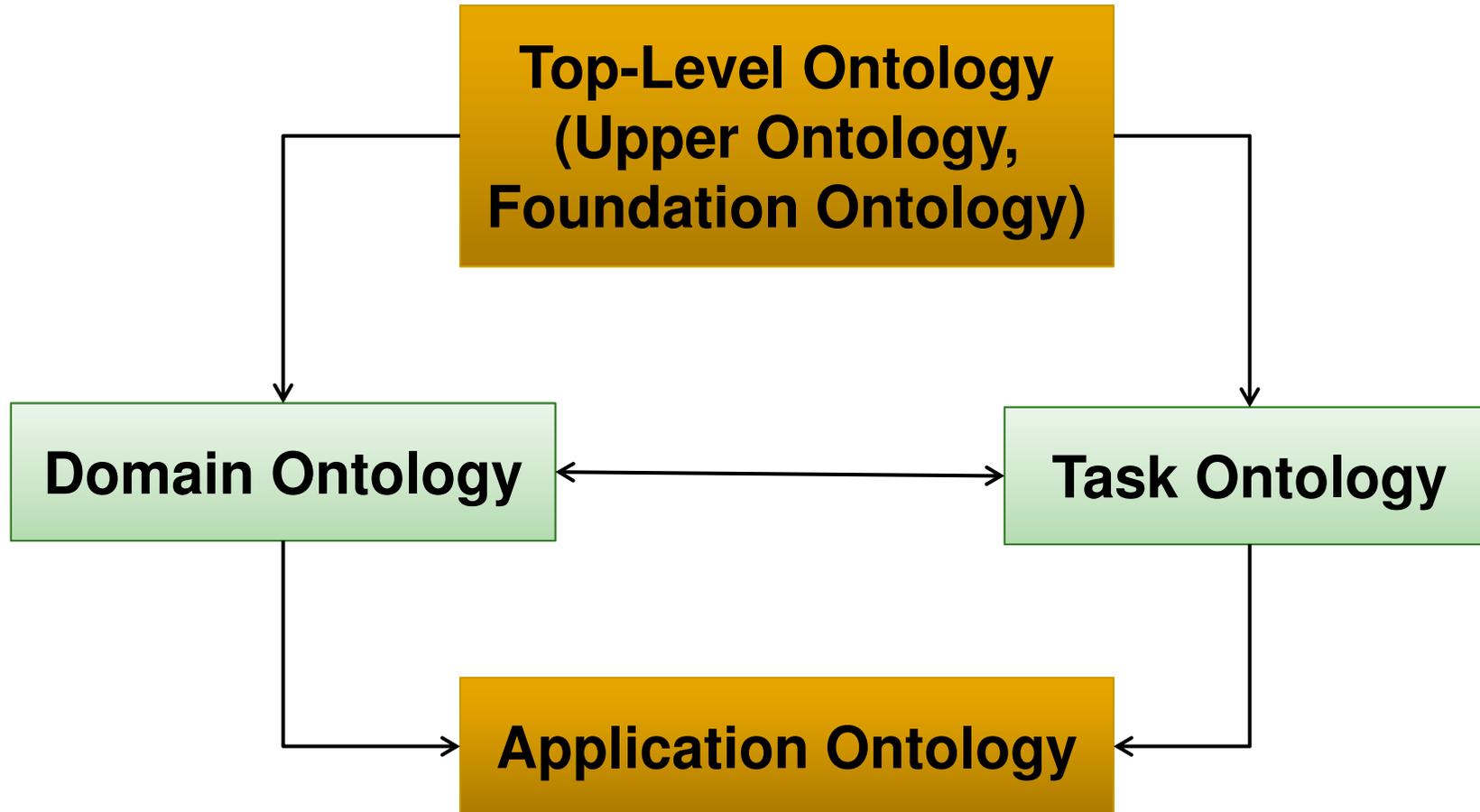


Inhalt

- Beispiele von Realwelt-Ontologien
- Annotation
- Semantische Suche und Navigation
- Linked Data



Klassifikation von Ontologien



*Klassifikation nach
Guarino, 1998*



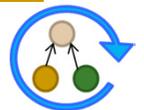
Ontologiebeispiele im Überblick

■ Top-Level Ontologien

- ❑ Basic Formal Ontology (BFO), General Formal Ontology (GFO)
- ❑ Standard Upper Merged Ontology (SUMO)
- ❑ OpenCyc
- ❑ WordNet
- ❑ ...

■ Domänenontologien (Lebenswissenschaften)

- ❑ Gene Ontology (GO)
- ❑ National Cancer Institute Thesaurus (NCIT)
- ❑ ...



BFO / GFO

- Basic Formal Ontology (BFO)
 - Unterstützung der Informationsintegration bei wissenschaftlichen Arbeiten
 - Hilfreich zur Entwicklung von Domänenontologien → Sicherung von Konsistenz sowie Vermeidung von Redundanzen
 - 39 Klassen: <http://purl.bioontology.org/ontology/BFO>
- General Formal Ontology (GFO)
 - Integration von Objekten und Prozessen
 - Informationen zu Zeit, Raum, Relationen, Rollen, ...
 - 45 Klassen: <http://purl.bioontology.org/ontology/GFO>

BFO: <http://www.ifomis.org/bfo>

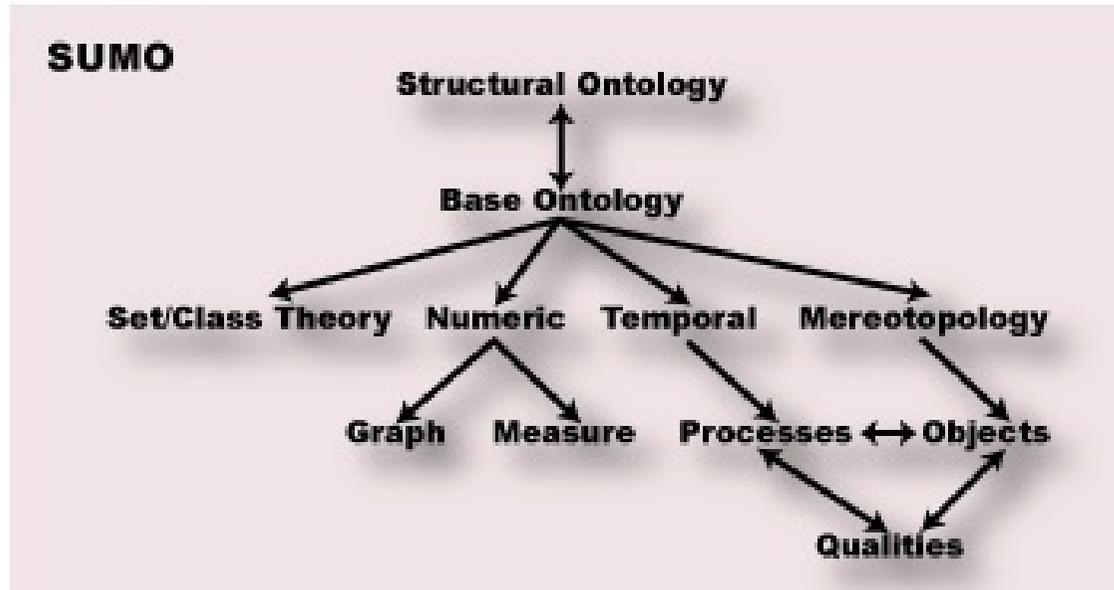
GFO: <http://www.onto-med.de/ontologies/gfo/>



SUMO

- Standard Upper Merged Ontology
 - Drei Teile
 - Upper Structural Ontology
 - Mid-level Base Ontology
 - Domain Ontologies: Wirtschaft, Transport, Geographie, ...
 - Ursprünglich nur generelle Konzepte → Erweiterung im Laufe der Zeit

SUMO:
<http://www.ontologyportal.org/>



OpenCyc / Cyc

- Cyc
 - Von „encyclopedia“
 - Aufbau einer umfassenden Ontologie und Wissensbasis für Allgemeinwissen (seit 1984)
 - Teile frei verfügbar → OpenCyc, ResearchCyc
 - CycL sentences zur Wissensrepräsentation
 - Unterscheidung von Individuen und Kollektionen
 - (`#$isa` `#$BillClinton` `#$UnitedStatesPresident`)
 - (`#$genls` `#$Tree-ThePlant` `#$Plant`)
 - (`#$capitalCity` `#$France` `#$Paris`)

Cyc: <http://www.cyc.com/cyc>

OpenCyc: <http://opencyc.org/>



WordNet

- Gruppierung englischsprachiger Wörter in Synonymmengen (synsets)
 - Kurze, allgemeine Definitionen
 - Semantische Relationen zwischen Synsets
 - Unterscheidung von Substantiven (N), Verben (V), Adjektiven (Adj) und Adverbien (Av)
- 117.659 Synsets in Version 3.0
- Relationen
 - Hypernym/hyponym (Ober- / Unterbegriffe)
 - Meronym/holonym (Teil-Ganzes)
 - Antonym (Gegensatzwort)
 - ...

WordNet: <http://wordnet.princeton.edu/>



WordNet

WordNet Search - 3.1

WordNet Search - 3.1

- [WordNet home page](#) - [Glossary](#) - [WordNet home page](#) - [Glossary](#) - [Help](#)

Word to search for: Word to search for:

Display Options: Display Options:

Key: "S:" = Show Synset (semantic) rel Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (gloss) "an e Display options for sense: (gloss) "an example sentence"

Noun

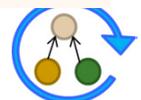
- [S:](#) (n) **mouse** (any of numerous s having pointed snouts and small e tails)
- [S:](#) (n) [shiner](#), [black eye](#), **mouse** (
- [S:](#) (n) **mouse** (person who is quie
- [S:](#) (n) **mouse**, [computer mouse](#) (coordinates of a cursor on your c the bottom of the device is a ball t *much more room than a trackba*

Verb

- [S:](#) (v) [sneak](#), **mouse**, [creep](#), [puss](#) *around spying on the neighbor's*
- [S:](#) (v) **mouse** (manipulate the mo

Noun

- [S:](#) (n) **mouse** (any of numerous small rodents typically resembling diminutive rats having pointed snouts and small ears on elongated bodies with slender usually hairless tails)
 - [direct hyponym](#) / [full hyponym](#)
 - [S:](#) (n) [house mouse](#), [Mus musculus](#) (brownish-grey Old World mouse now a common household pest worldwide)
 - [S:](#) (n) [harvest mouse](#), [Micromyx minutus](#) (small reddish-brown Eurasian mouse inhabiting e.g. cornfields)
 - [S:](#) (n) [field mouse](#), [fieldmouse](#) (any nocturnal Old World mouse of the genus Apodemus inhabiting woods and fields and gardens)
 - [S:](#) (n) [nude mouse](#) (a mouse with a genetic defect that prevents them from growing hair and also prevents them from immunologically rejecting human cells and tissues; widely used in preclinical trials)
 - [S:](#) (n) [wood mouse](#) (any of various New World woodland mice)
 - [direct hypernym](#) / [inherited hypernym](#) / [sister term](#)
 - [derivationally related form](#)



Gene Ontology (GO)

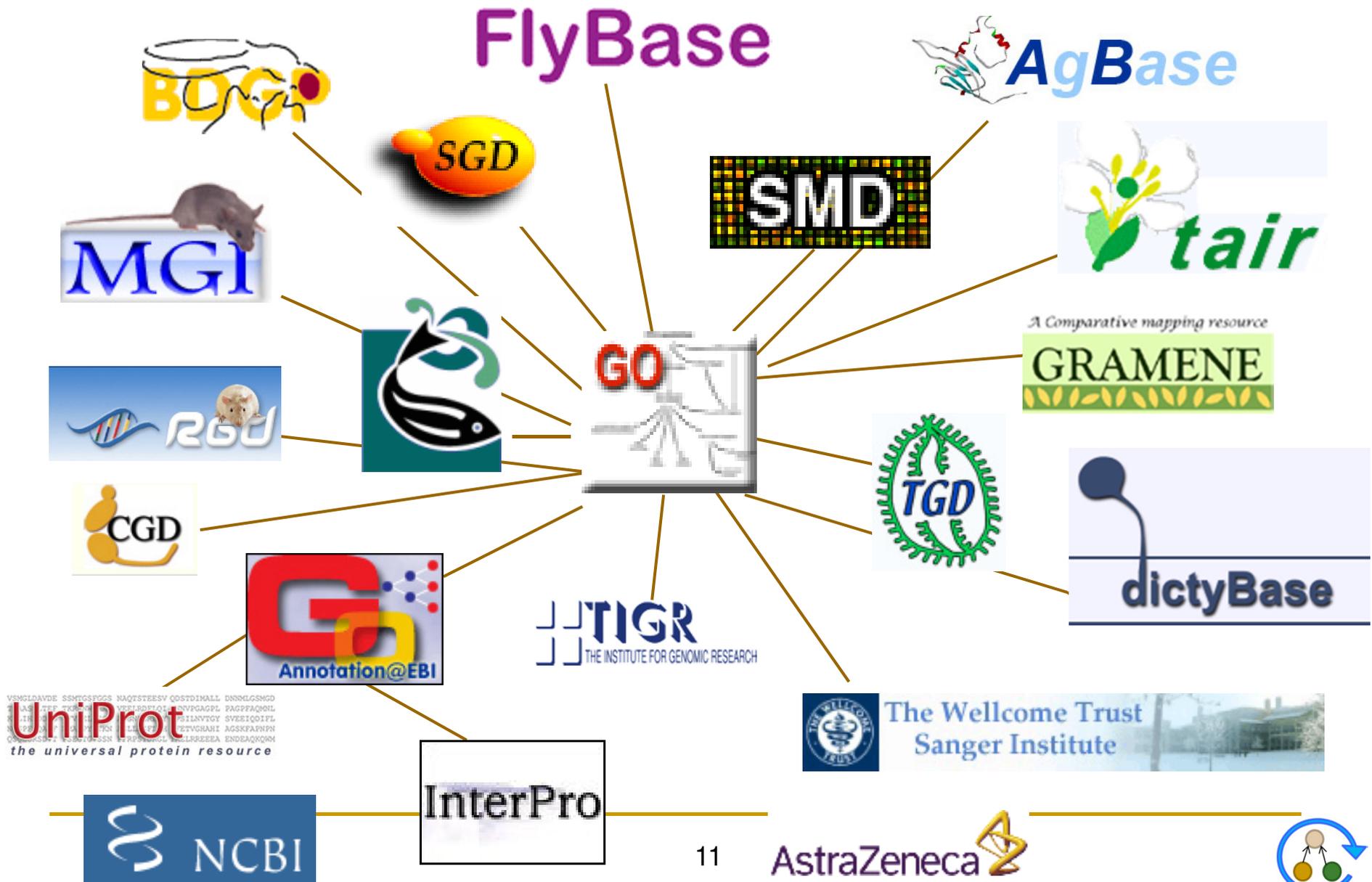


- Gene Ontology Consortium
 - Entwicklung einer gemeinsamen Sprache zur Annotation von molekularen Charakteristiken in Organismen → Wissensaustausch
 - Teilnehmer stellen GO-basierte Annotationen zur Verfügung → Möglichkeit neuartiger Datenanalysen und Zusammenarbeit
- Projektziele:
 - Management / Entwicklung der eigentlichen Ontologie
 - Annotation von Genen / Genprodukten
 - Tools / Werkzeuge für den Umgang mit GO

Gene Ontology: <http://www.geneontology.org/>



Verbreitung von GO

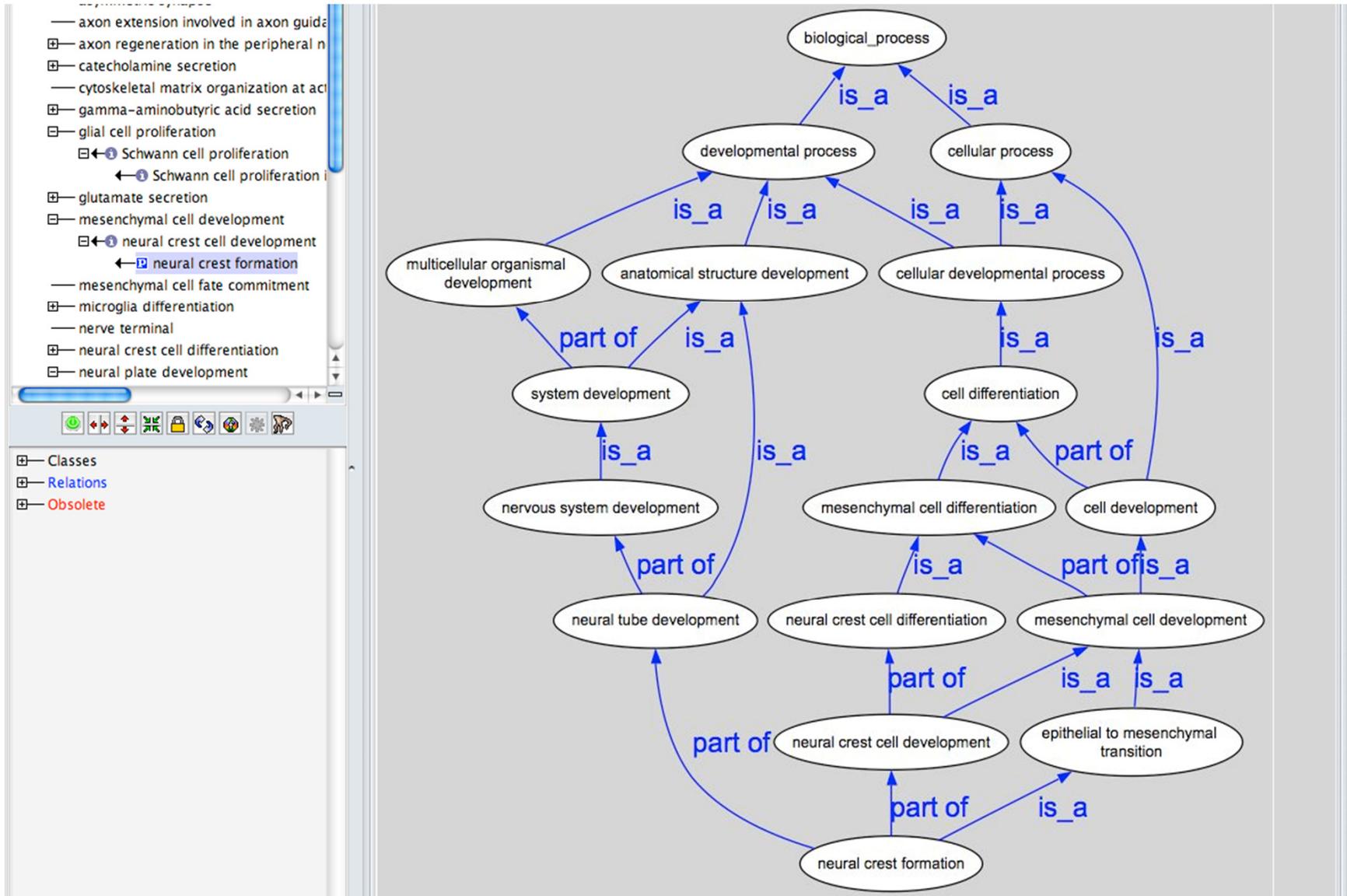


Gene Ontology (GO)

- Einteilung in drei Subdomänen
 - *Molecular Function* (ca. 10.000 Konzepte)
 - Elementare Aktivitäten / Aufgaben, durchgeführt von individuellen Genprodukten
 - Bsp.: carbohydrate binding, ATPase activity
 - *Biological Process* (ca. 20.000 Konzepte)
 - Biologisches Ziel realisiert durch eine Menge molekularer Funktionen
 - Bsp.: mitosis, purine metabolism
 - *Cellular Component* (ca. 3.000 Konzepte)
 - Ort oder Komplex (subzellulare Strukturen)
 - Bsp.: nucleus, telomere, RNA polymerase II holoenzyme



Konzepte in GO



National Cancer Institute Thesaurus (NCIT)

- NCI Thesaurus
 - Referenzterminologie des NCI
 - Krebsbezogene Themen für kollaborative Forschung
 - Ca. 90.000 Konzepte organisiert in 20 Hauptkategorien
 - Anatomie
 - Genprodukte
 - Krankheiten
 - Medikamente
 - ...
 - Konzeptinformationen
 - Bevorzugter Name, Synonyme, Definitionen, ...



NCI Thesaurus: <http://ncit.nci.nih.gov/ncitbrowser/>



Annotation – Motivation

Informationsflut in den Lebenswissenschaften

- Zehntausende Gene und gen-regulatorische Elemente (in versch. Spezies)
- Zahlreiche Interaktionen in komplexen molekularen Netzwerken
- Millionen von Publikationen
- ...

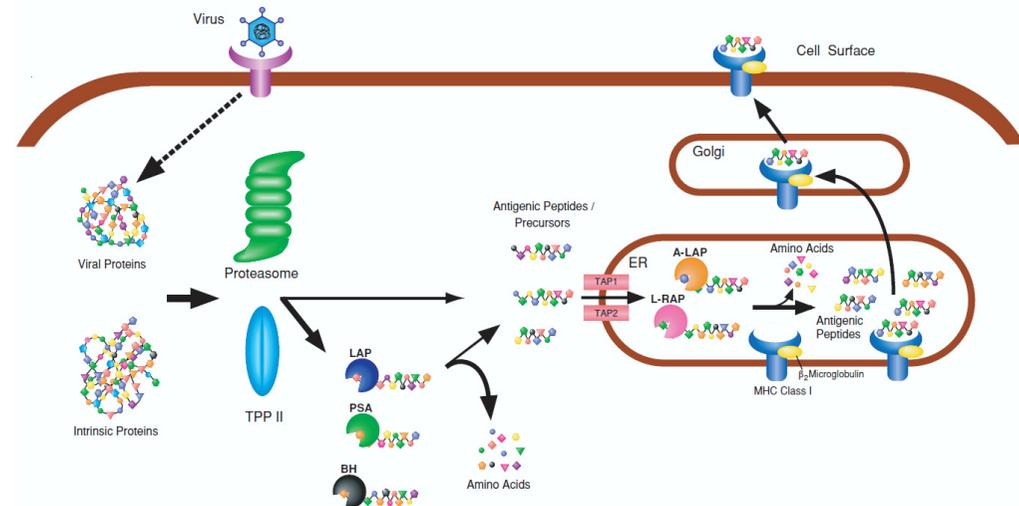


Fig. 1. Processing Pathway of Antigenic Peptides Presented to MHC Class I Molecules

Hattori A, Tsujimoto M.: Processing of antigenic peptides by aminopeptidases. *Biol Pharm Bull.*, 2004

- Zuordnung
 - der Funktionen, Prozesse, Interaktionen zu biologischen Objekten
 - der enthaltenen Themen zu Publikationen
 - ...

→ Nutzen maschinenverstehbarer Annotationen



Begriff Annotation

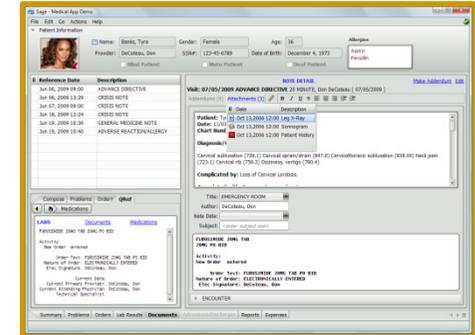
- Annotationen dienen der semantischen Beschreibung der Eigenschaften von Objekten
 - Erfassung von Metadaten
 - Möglichst einheitlicher Gebrauch der domänenspezifischen Begriffe
 - Assoziationen eines zu beschreibenden Objekts zu den Begriffen eines Vokabulars bzw. den Konzepten einer Ontologie
 - Genprodukte, Gene in Lebenswissenschaften mit zugeordneten Funktionen
 - Produkte in Online-Shops annotiert mit ihren Eigenschaften



Arten von Annotationen in den Lebenswissenschaften



- Biologische Objekte (Gene, Proteine, ...)
- Pathways, Netzwerke
- ...



Experiment-Annotation
MIAME,
MGED...

Annotation biologischer Objekte / Pathways
GO, KEGG, ...

Annotation von Publikationen (indexing)
MeSH, GO, ...

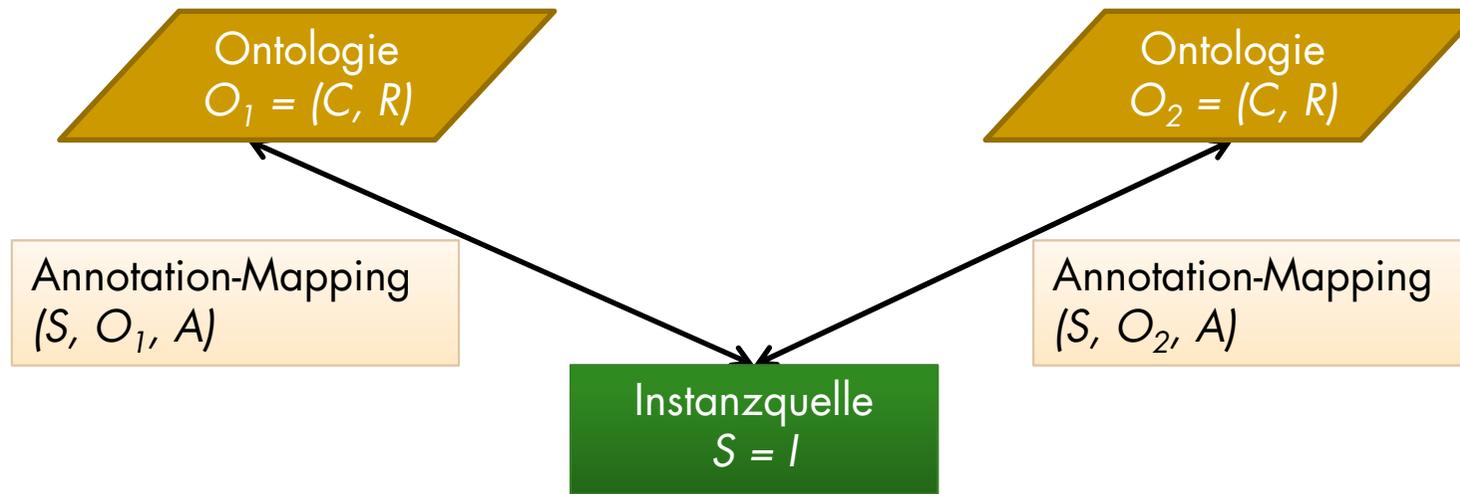
Annotation von klinischen Dokumenten
ICD, SNOMED, LOINC, ...

MIAME Minimum Information About a Microarray Experiment
MGED Microarray Gene Expression Data
KEGG Kyoto Encyclopedia of Genes and Genomes
MeSH Medical Subject Headings

ICD International Classification of Diseases
SNOMED Systematized Nomenclature of Medicine
LOINC Logical Observation Identifiers Names and Codes
NCIT NCI (National Cancer Institute) Thesaurus



Allgemeines Annotationsmodell



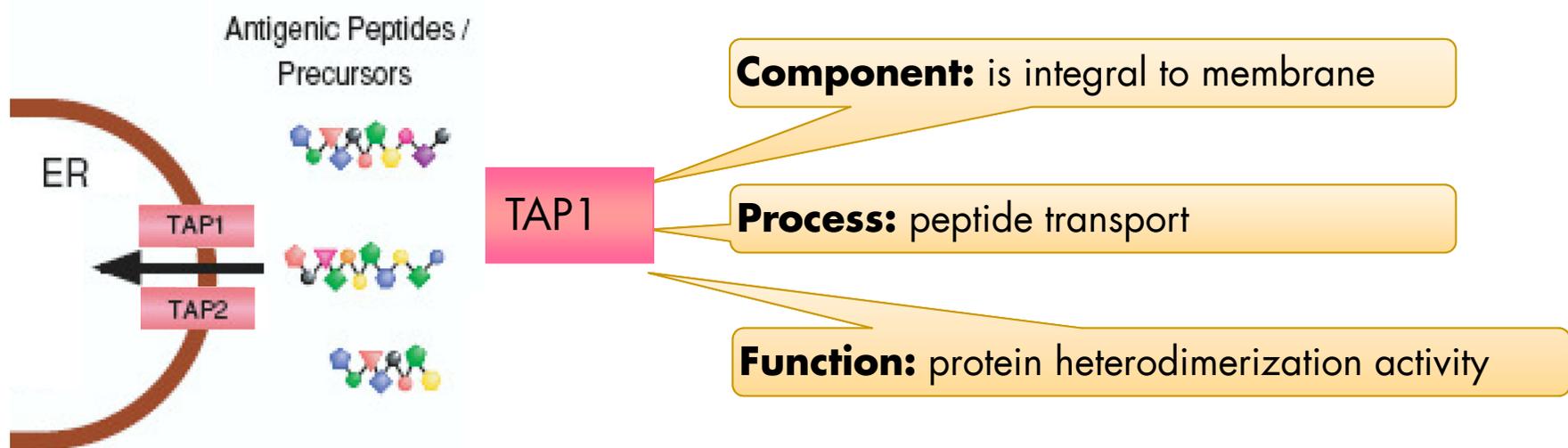
- C** Menge von Konzepten
- R** Menge von Relationen
- I** Menge von Instanzen/Objekten
- A** Menge von Annotationen / Korrespondenzen

Annotation $a=(i,c)$, $a \in A$, $i \in I$, $c \in C$
object_id - concept_id



Beispiel: GO Annotationen

- Assoziation eines Proteins zu einem/r bestimmten biologischen Prozess / zellulären Komponente / molekularen Funktion



Beispiel: GO Annotationen

UniProtKB/Swiss-Prot entry Q03518

Note: most headings are highlighted in yellow in the original image.

Ensembl Home > Human
Location: 6:32,920,965-32,929

Transcript-based displays

- Transcript summary
- Exons (11)
- Supporting evidence (28)
- Sequence
 - cDNA
 - Protein
- External References
 - General identifiers (88)
 - Oligo probes (16)
 - Gene ontology (24)**
- Genetic Variation
 - Population comparison
 - Comparison image
- Protein Information
 - Protein summary
 - Domains & features (36)
 - Variations (18)
- External Data
 - ID History
 - Transcript history
 - Protein history
 - Export transcript data

Name and origin

- Protein name
- Synonyms

Ontologies

- GO
- Quick

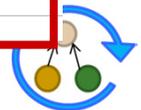
TAP1 Homo sapiens Q03518

Accession	Q03518
Gene	TAP1
Taxonomy	Homo sapiens
Description	Antigen peptide transporter 1

Annotation

Help: filtering, analyzing and downloading annotation

Columns:	DB	ID	Alt	Symbol	Taxon	Ev	GO ID	GO Term name
1 Filter:	Any	Q03518			Any	Any	Any	
2 Statistics:		1			1	7	33	
3 View	1-25 of bookmark this annotation set							
								Process
	UniProtKB/Swiss-Prot	Q03518		TAP1	9606	IEA	GO:0055085	transmembrane transport
	UniProtKB/Swiss-Prot	Q03518		TAP1	9606	IEA	GO:0055085	transmembrane transport
	UniProtKB/Swiss-Prot	Q03518		TAP1	9606	IEA	GO:0006810	transport
	UniProtKB/Swiss-Prot	Q03518		TAP1	9606	IMP	GO:0046967	cytosol to ER transport
	UniProtKB/Swiss-Prot	Q03518		TAP1	9606	IMP	GO:0015833	peptide transport
	UniProtKB/Swiss-Prot	Q03518		TAP1	9606	IEA	GO:0006810	transport
								Function
	UniProtKB/Swiss-Prot	Q03518		TAP1	9606	NAS	GO:0042605	peptide antigen binding
	UniProtKB/Swiss-Prot	Q03518		TAP1	9606	IEA	GO:0005524	ATP binding
	UniProtKB/Swiss-Prot	Q03518		TAP1	9606	IEA	GO:0042626	ATPase activity, coupled to transmembrane movement of substances
	UniProtKB/Swiss-Prot	Q03518		TAP1	9606	IEA	GO:0005524	ATP binding
	UniProtKB/Swiss-Prot	Q03518		TAP1	9606	IEA	GO:0005524	ATP binding
	UniProtKB/Swiss-Prot	Q03518		TAP1	9606	IEA	GO:0016887	ATPase activity
	UniProtKB/Swiss-Prot	Q03518		TAP1	9606	IEA	GO:0046978	TAP1 binding
	UniProtKB/Swiss-Prot	Q03518		TAP1	9606	IEA	GO:0016887	ATPase activity
	UniProtKB/Swiss-Prot	Q03518		TAP1	9606	IPI	GO:0046979	TAP2 binding
	UniProtKB/Swiss-Prot	Q03518		TAP1	9606	IPI	GO:0046979	TAP2 binding
	UniProtKB/Swiss-Prot	Q03518		TAP1	9606	IEA	GO:0046982	protein heterodimerization activity
	UniProtKB/Swiss-Prot	Q03518		TAP1	9606	IPI	GO:0005515	protein binding
	UniProtKB/Swiss-Prot	Q03518		TAP1	9606	IEA	GO:0005524	ATP binding



Beispiel: MeSH Annotationen

NCBI Resources How To

PubMed.gov
U.S. National Library of Medicine
National Institutes of Health

Search: PubMed Limits Advanced search Help

Search Clear

Display Settings: Abstract Send to:

BMC Bioinformatics. 2009 Aug 13;10:250.

OnEX: Exploring changes in life science ontologies.

Hartung M, Kirsten T, Gross A, Rahm E.
Interdisciplinary Centre for Bioinformatics, University of Leipzig, Härtelstrasse 16-18, 04107 Leipzig, Germany. hartung@izbi.uni-leipzig.de

Abstract
BACKGROUND: Numerous ontologies have recently been developed in life sciences to support a consistent annotation of biological objects, such as genes or proteins. These ontologies underlie continuous changes which can impact existing annotations. Therefore, it is valuable for users of ontologies to study the stability of ontologies and to see how many and what kind of ontology changes occurred. RESULTS: We present OnEX (Ontology Evolution Explorer) a system for exploring ontology changes. Currently, OnEX provides access to about 560 versions of 16 well-known life science ontologies. The system is based on a three-tier architecture including an ontology version repository, a middleware component and the OnEX web application. Interactive workflows allow a systematic and explorative change analysis of ontologies and their concepts as well as the semi-automatic migration of out-dated annotations to the current version of an ontology. CONCLUSION: OnEX provides a user-friendly web interface to explore information about changes in current life science ontologies. It is available at <http://www.izbi.de/onex>.

PMID: 19678926 [PubMed - indexed for MEDLINE] PMID: PMC2746816 **Free PMC Article**

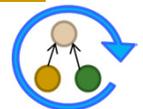
Publication Types, MeSH Terms

Publication Types:
[Research Support, Non-U.S. Gov't](#)

MeSH Terms:
[Computational Biology/methods*](#)
[Databases, Factual](#)
[Gene Expression Profiling](#)
[Internet](#)
[Software*](#)
[User-Computer Interface](#)
[Vocabulary, Controlled](#)

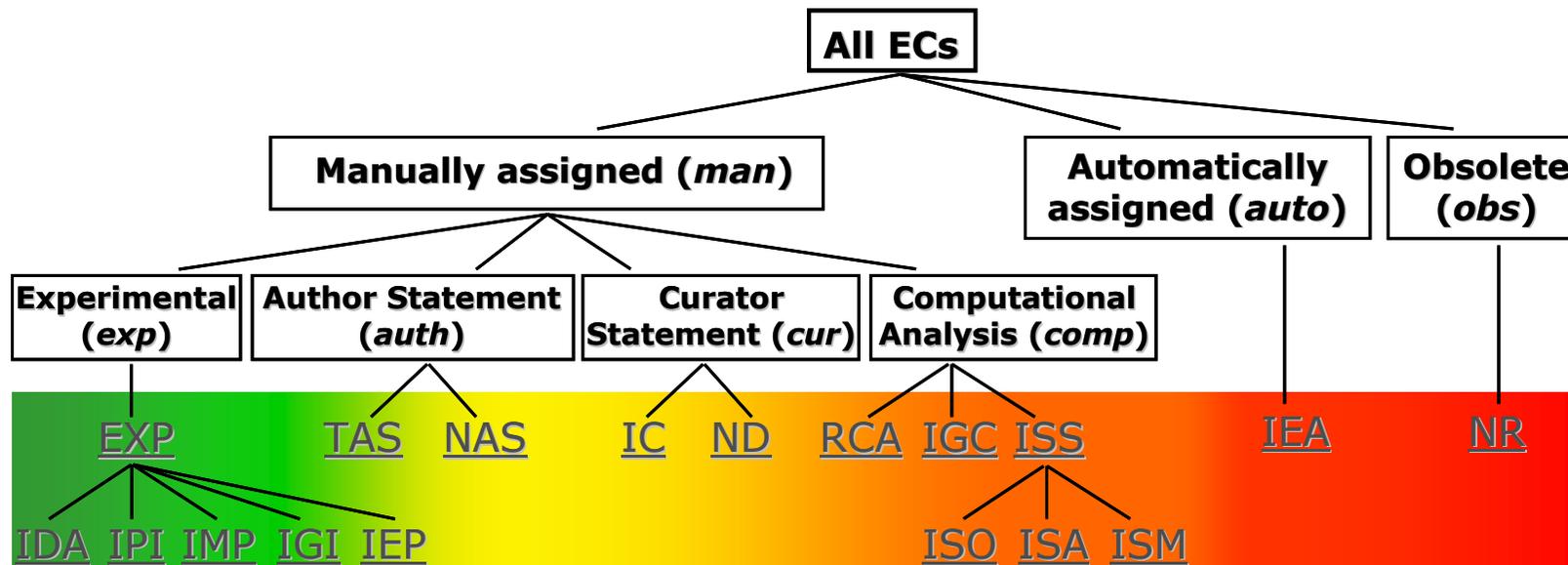
LinkOut - more resources

MeSH = Medical Subject Headings



Qualität von Annotationen

- Annotationen können auf unterschiedlichste Art und Weise entstehen
 - Erfassung weiterer Metadaten zu Annotationen
 - Herkunft (provenance) → Evidence Codes
 - Kurator, Datum, Methode, Quelle, Referenzen, ...
 - GO Evidence Code (EC) Taxonomy



* <http://www.geneontology.org/GO.evidence>



Klassische Informationssuche

The screenshot shows a Google search for "golf von 2000". The search bar contains the text "golf von 2000" and a magnifying glass icon. Below the search bar, the text "Search" is followed by "About 43,300,000 results (0.22 seconds)".

On the left side, there is a sidebar with the following options: "Everything", "Images", "Maps", "Videos", "News", "Shopping", "More", "Leipzig", "Change location", and "Show search tools".

The main search results are as follows:

- GOLFENTWICKLUNG | Golfer, Clubs und Plätze von 2000 bis 2009 | golf.de**
www.golf.de/dgw/statistiken.cfm?objectid... - Translate this page
27. Jan. 2010 – GOLFENTWICKLUNG: Golfentwicklung, Golfer, Clubs und Plätze von 2000 bis 2009.
- VW Golf IV – Wikipedia**
de.wikipedia.org/wiki/VW_Golf_IV - Translate this page
2000: Einführung 1.6 16V mit 105 PS, 2.3 V5 20V mit 170 PS und 1.9 TDI PD mit Der erste Golf R32 entstand als geplante limitierte Auflage von 5000 Autos ...
- Golf von Morbihan – Wikipedia**
de.wikipedia.org/wiki/Golf_von_Morbihan - Translate this page
- Block all de.wikipedia.org results
Der Golf von Morbihan (französisch Golf de Morbihan) ist ein Golfplatz in Morbihan, Frankreich.
- Golf 2000 Angebote bei mobile.de**
suchen.mobile.de/auto/volkswagen.golf.2000.html - Translate this page
mobile.de: Golf 2000 kaufen. Finden Sie eine Vielzahl von günstigen Angeboten bei mobile.de - Deutschlands größter Fahrzeugmarkt.
- Volkswagen Zubehör Shop - Kleiderbügel, Chrom 00V061127**
www.volkswagen-zubehoer-shop.de/.../homestage... - Translate this page
Bora (A4) von 1999 bis 2005. Bora Var. (A4) von 2000 bis 2004. Eos ab 2006. Eos GP ab 2011. Fox ab 2005. Golf (A3) von 1992 bis 1997. Golf Var. (A3) von ...
- Segeln im Golf von Korinth**
www.fr-yachting.de/greecede.htm - Translate this page
Am Ende des Kanals bei der Durchfahrt von Ost nach West erreicht man die Stadt Korinth. Wir sind mit westlichem Kurs auf den Golf von Korinth hinausgesegelt ...

There are three images: a golf ball on a green, a dark Volkswagen Golf IV car, and a wide view of a golf course with water features.

■ Keyword-Suche oftmals mit **nicht relevanten** Ergebnissen

- ❑ Bedeutungsunterschiede
- ❑ Homonyme
- ❑ Unterschiedlicher Kontext

Keyword-Suche findet **nicht alle relevanten** Ergebnisse

- ❑ Synonyme
- ❑ Fehlende Kontextpräzisierung



Semantische Suche

- Suchmaschinen – Document Retrieval

- Eingabe: „Golf von 2000“



Nachfrage zu Golf:

(1) PKW, (2) Sportart, (3) Geographie, ...
→ Verfeinerung: PKW



Semantische Suche

■ Suchmaschinen – Document Retrieval

□ Eingabe: „Golf von 2000“



- Name: Volkswagen **Golf**
- Kategorie: **Auto**
 - ...
 - Baujahr **<Integer>**
 - Tech. Parameter
 - Hubraum **<Integer>**
 - Leistung **<Integer>**
 - ...
 - ...

Nachfrage zu 2000:
Baujahr oder technischer Parameter
→ Verfeinerung: Baujahr

Wissenrepräsentation
(Ontologie, Linked Data)



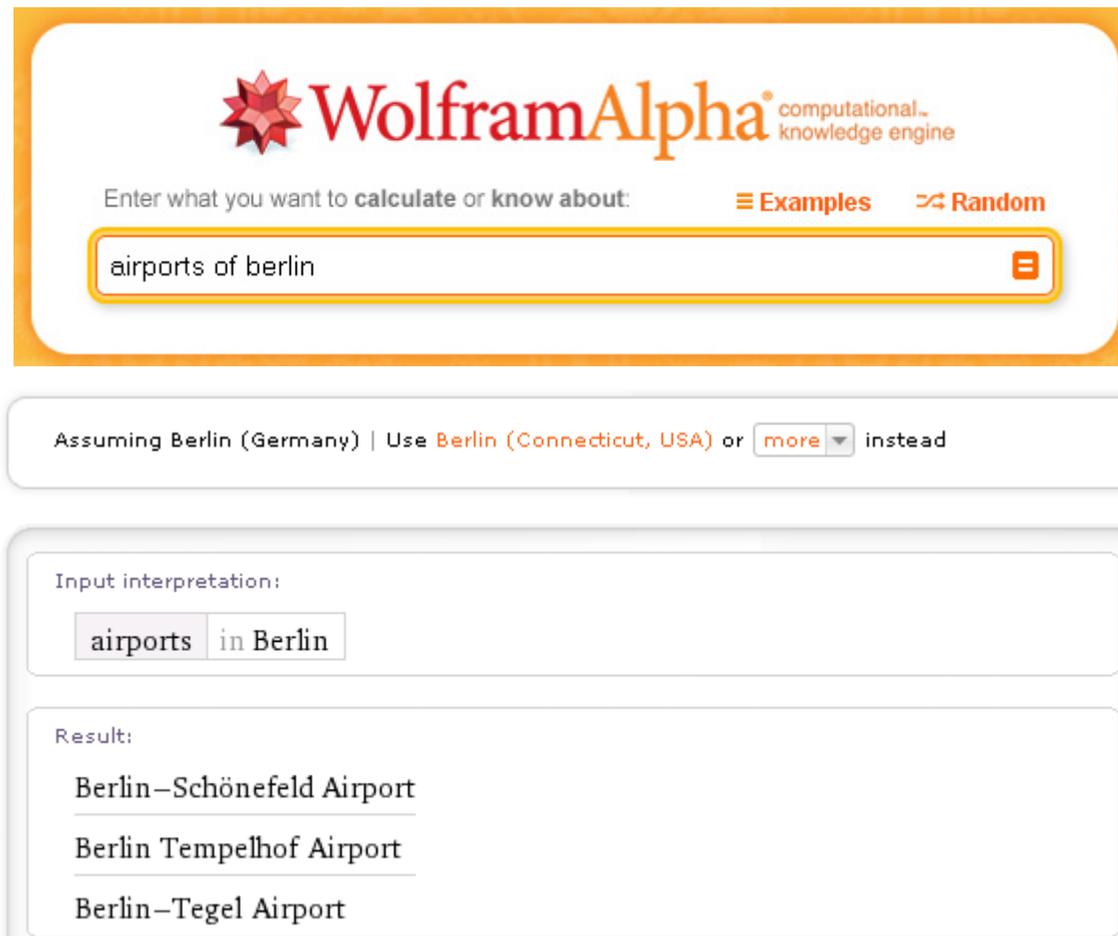
Semantische Suche

- Suchmaschinen – Document Retrieval
 - Generelle Probleme
 - Korrekte Interpretation der Suchphrase (Keywords)
 - Korrekte Identifikation von Entitäten
 - Automatische Disambiguierung
 - Usability
 - Personalisierung



Semantische Suche

- Suchmaschinen – Fact Retrieval
 - Eingabe in Suchmaschine: „Welche Flughäfen gibt es in Berlin?“ (airports of Berlin)



The image shows a screenshot of the WolframAlpha search engine interface. At the top, the WolframAlpha logo is displayed with the tagline "computational knowledge engine". Below the logo, there is a search bar containing the text "airports of Berlin". To the right of the search bar, there are links for "Examples" and "Random". Below the search bar, there is a section for "Assuming Berlin (Germany) | Use Berlin (Connecticut, USA) or more instead". The "Input interpretation" section shows the query broken down into "airports" and "in Berlin". The "Result" section lists three airports: Berlin-Schönefeld Airport, Berlin Tempelhof Airport, and Berlin-Tegel Airport.



GOPubMed

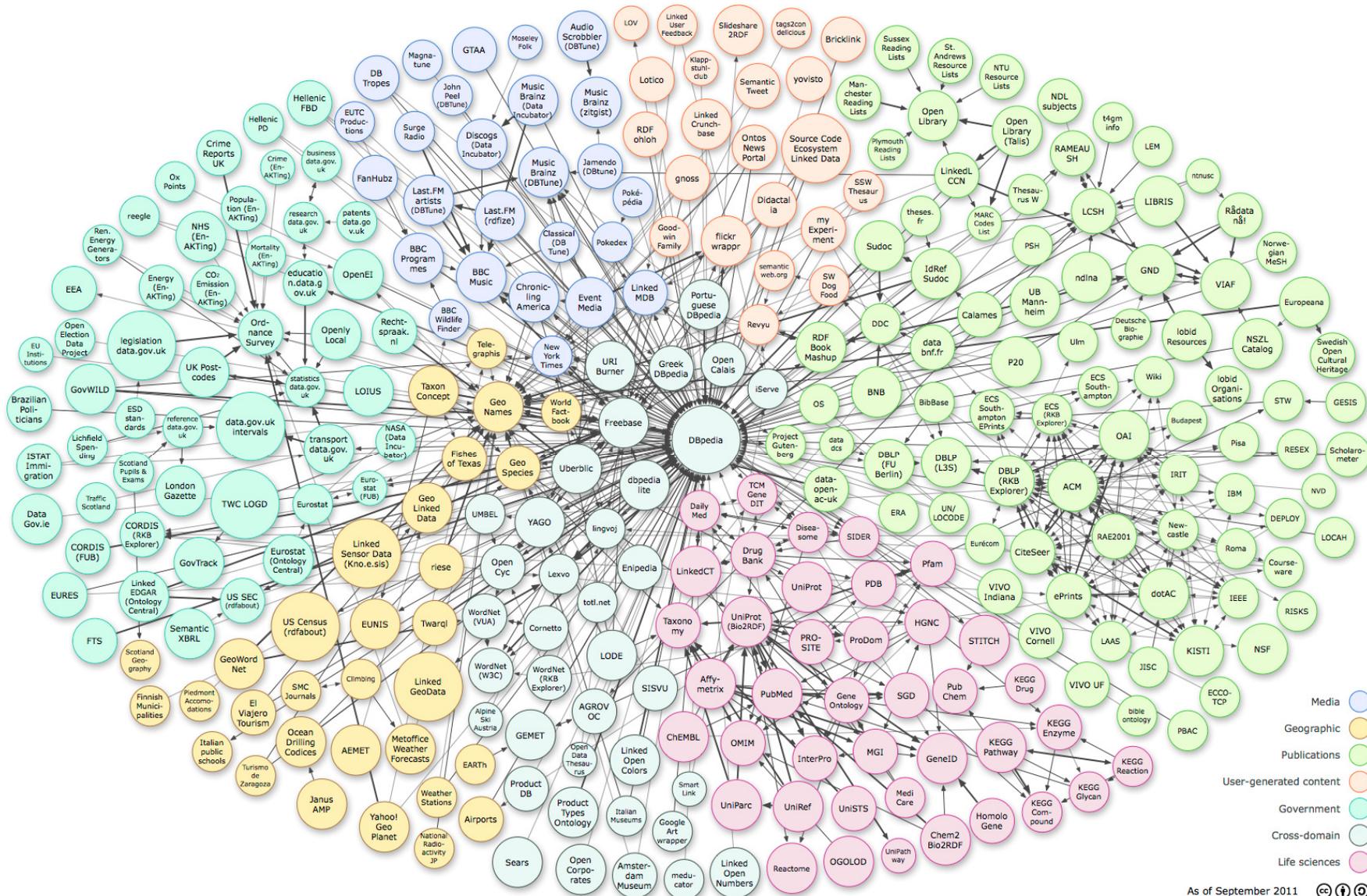
- Klassifikation von PubMed-Artikeln unter Verwendung von GO und MeSH

The screenshot displays the GOPubMed interface. On the left, a sidebar contains navigation options: 'my search', 'what' (with 'Top Terms' and 'Knowledge Base' lists), 'who' (with 'All Authors' and 'Find specific author...'), 'where' (with 'Earth' and 'All Journals'), and 'when'. The main content area shows a search for 'TAP1 binding' resulting in 944 documents. Below this, there are sections for 'top author', 'statistics', and 'documents'. Three articles are listed:

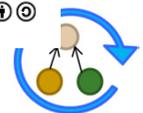
- Aflatoxin G1 reduces the molecular expression of HLA-I, TAP-1 and LMP-2 of adult esophageal epithelial cells in vitro.**
Author: Li, Z. et al.
Journal: *Toxicology Letters*, Vol. 195 (2-3): 169-73, 2010
Affiliation: Department of nutrition, The First Attached Hospital, Hebei Medical University, Shijiazhuang, China. Zengningli@hebmh.edu.cn
- The CRAL/TRIO and GOLD domain protein TAP-1 regulates RAF-1 activation.**
Author: Johnson, K G, et al.
Journal: *Developmental biology*, Vol. 341 (2): 464-71, 2010
Affiliation: Department of Developmental Biology, Washington University School of Medicine, St. Louis, MO 63110, USA.
Order related Antibodies online: antiEphB2 antibody (ABIN317801), SerineThreonine Protein Kinase RAF RAF1 Human antibody (ABIN209203)
- Down-regulation of HLA class I antigen in human papillomavirus type 16 E7 expressing HaCaT cells: correlate with TAP-1 expression.**
Author: Li, W, et al.
Journal: *International journal of gynecological cancer: official journal of the International Gynecological Cancer Society*, Vol. 20 (2): 227-32, 2010
OBJECTIVES: High-risk human papillomaviruses (HPVs) are the major causative agents of cervical cancer, and the E6 and E7 genes encode the major HPV oncoproteins.
Affiliation: Department of Clinical Laboratory, Chinese Ministry of Education and Chinese Ministry of Health, Qilu Hospital, Shandong University, Jinan, China.
- Structure and function of the GINS complex, a key component of the eukaryotic replisome.**
Author: MacNeill, S A
Journal: *The Biochemical journal*, Vol. 425 (3): 489-500, 2010
High-fidelity chromosomal DNA replication is fundamental to all forms of cellular life and requires the complex interplay of a wide variety of essential and non-essential protein factors in a spatially and temporally co-ordinated manner.



LOD (September 2011)



As of September 2011



Beispiel BBC Music

a-ha

Formed 1983. Disbanded 2010.



David Redfern/Redferns

Biography

A-ha (spelled "a-ha" in lower-case on their releases) were a Norwegian pop band formed in Oslo in 1982. The band was founded by Morten Harket (vocals), Magne Furuholmen (keyboard), and Pål Waaktaar (guitars). The group initially rose to fame during the mid 1980s after being discovered by musician and producer John Ratcliff and had continued global success in the 1990s and 2000s.

[Read more at Wikipedia...](#)

WIKIPEDIA This entry is from [Wikipedia](#), the user-contributed encyclopedia. It may not have been reviewed by professional editors and is licensed under the [GNU Free Documentation License](#). If you find the biography content factually incorrect, defamatory or highly offensive you can [edit this article at Wikipedia](#). [Find out more about our use of this data.](#)

PLAYED MOST ON **BBC RADIO 2**



Share This Page

3 so far

[Share](#) [facebook](#) [twitter](#)

BBC Music Showcase

BBC MUSIC SHOWCASE
Watch and listen to exclusive music clips

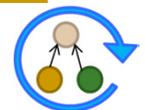
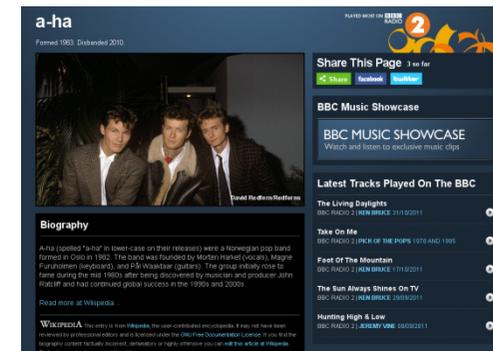
Latest Tracks Played On The BBC

- The Living Daylights**
BBC RADIO 2 | [KEN BRUCE](#) 31/10/2011 
- Take On Me**
BBC RADIO 2 | [PICK OF THE POPS](#) 1978 AND 1985 
- Foot Of The Mountain**
BBC RADIO 2 | [KEN BRUCE](#) 17/10/2011 
- The Sun Always Shines On TV**
BBC RADIO 2 | [KEN BRUCE](#) 29/09/2011 
- Hunting High & Low**
BBC RADIO 2 | [JEREMY VINE](#) 08/09/2011 



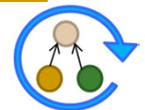
BBC Music

- Was ist das Besondere?
 - ❑ Informationen werden aus externen, öffentlich verfügbaren Daten automatisch aktuell zusammengestellt (Wikipedia, MusicBrainz, ...)
 - ❑ Kein Screen Scraping der Website
 - ❑ Kein spezielles API
 - ❑ Daten liegen im (Semantic) Web als Linked Open Data vor
 - ❑ Zugriff über gewöhnlichen HTTP Request
 - ❑ Daten immer hoch aktuell, kein manueller Eingriff nötig



BBC Music als RDF

```
- <rdf:RDF>
- <rdf:Description rdf:about="/music/artists/7364dea6-ca9a-48e3-be01-b44ad0d19897.rdf">
  <rdfs:label>Description of the artist a-ha</rdfs:label>
  <foaf:primaryTopic rdf:resource="/music/artists/7364dea6-ca9a-48e3-be01-b44ad0d19897#artist"/>
</rdf:Description>
- <mo:MusicArtist rdf:about="/music/artists/7364dea6-ca9a-48e3-be01-b44ad0d19897#artist">
  <rdf:type rdf:resource="http://purl.org/ontology/mo/MusicGroup"/>
  <foaf:name>a-ha</foaf:name>
  <ov:sortLabel>a-ha</ov:sortLabel>
- <bio:event>
  - <bio:Birth>
    <bio:date rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">1983</bio:date>
  </bio:Birth>
</bio:event>
<foaf:page rdf:resource="/music/artists/7364dea6-ca9a-48e3-be01-b44ad0d19897.html"/>
<owl:sameAs rdf:resource="http://dbpedia.org/resource/A-ha"/>
<mo:musicbrainz rdf:resource="http://musicbrainz.org/artist/7364dea6-ca9a-48e3-be01-b44ad0d19897.html"/>
<mo:image rdf:resource="http://static.bbc.co.uk/music/images/artists/234x132/7364dea6-ca9a-48e3-be01-b44ad0d19897.jpg"/>
<foaf:homepage rdf:resource="http://www.a-ha.com"/>
<mo:wikipedia rdf:resource="http://en.wikipedia.org/wiki/A-ha"/>
<mo:imdb rdf:resource="http://www.imdb.com/name/nm2165250"/>
<mo:myspace rdf:resource="http://www.myspace.com/aha"/>
<mo:member rdf:resource="/music/artists/1ac73985-2cfa-433a-a6f8-379480e87179#artist"/>
<mo:member rdf:resource="/music/artists/2093207d-cb69-4789-866b-5ac772c9ebd3#artist"/>
<mo:member rdf:resource="/music/artists/0ef9b4c4-4a41-4ad8-9c64-98c0419c0150#artist"/>
- <foaf:made>
  - <mo:Record>
    <dc:title>25 (disc 1)</dc:title>
    <mo:musicbrainz rdf:resource="http://musicbrainz.org/release/27ae8b3f-a4f8-408c-bbfe-1b4e3d56b3c7.html"/>
    <rev:hasReview rdf:resource="/music/reviews/qrd3#review"/>
  </mo:Record>
</foaf:made>
- <foaf:made>
  - <mo:Record>
    <dc:title>Foot of the Mountain</dc:title>
```



Linked Data Grundkonzept

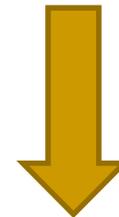
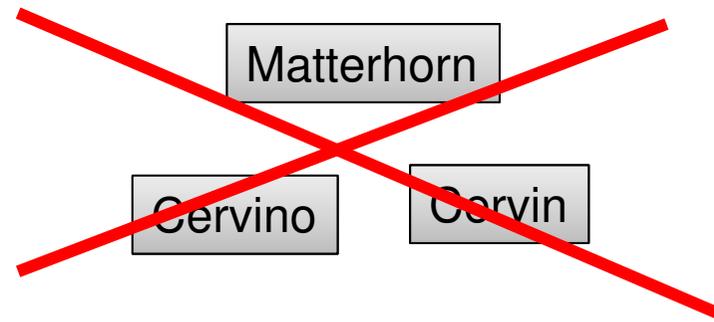
- Aus dem Netz von Webseiten soll ein Netz aus Daten entstehen
- „4 Regeln“ von Berners Lee
 1. Use URIs as names for things
 2. Use HTTP URIs so that people can look up those names
 3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)
 4. Include links to other URIs, so that they can discover more things.

Linked Data Design Issues: <http://www.w3.org/DesignIssues/LinkedData.html>



Regel 1

- Eindeutige URIs als Bezeichnung für Objekte



<http://dbpedia.org/resource/Matterhorn>



Regel 3

■ Nutzvolle Informationen bei Lookup

rdf:type

- owl:Thing
- dbpedia-owl:Mountain
- <http://schema.org/Mountain>
- yago:MountainsOfItaly
- yago:MountainsOfSwitzerland
- yago:Mountain109359803
- dbpedia-owl:NaturalPlace
- dbpedia-owl:Place
- <http://schema.org/Place>
- yago:MountainsOfTheAlps
- gml:_Feature
- yago:MountainsOfAostaValley
- yago:MountainsOfValais
- <http://umbel.org/umbel/rc/Mountain>
- http://umbel.org/umbel/rc/Location_Underspecified
- yago:InternationalMountainsOfEurope
- yago:PennineAlps

rdfs:label

- Matterhorn
- Cervino
- Matterhorn
- Cervin
- Cervino
- マッターホルン
- Matterhorn
- Matterhorn
- Matterhorn
- Matterhorn
- Matterhorn
- Matterhorn
- Маттерхорн
- 馬特洪峰

geo:geometry

- POINT(7.65833 45.9764)

geo:lat

- 45.976387 (xsd:float)

geo:long

- 7.658333 (xsd:float)



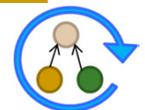
Regel 4

- Links zu weiteren Ressourcen („semantische Umgebung“)
 - In Datenquelle selbst

<code>dbpedia-owl:firstAscentPerson</code>	<ul style="list-style-type: none">■ <code>dbpedia:Edward_Whymper</code>■ <code>dbpedia:Douglas_Robert_Hadow</code>■ <code>dbpedia:Lord_Francis_Douglas</code>■ <code>dbpedia:Charles_Hudson_(climber)</code>■ <code>dbpedia:Michel_Croz</code>
<code>dbpedia-owl:firstAscentYear</code>	■ <code>1865-01-01 00:00:00</code> (xsd:date)
<code>dbpedia-owl:locatedInArea</code>	■ <code>dbpedia:Italy</code>
<code>dbpedia-owl:mountainRange</code>	■ <code>dbpedia:Pennine_Alps</code>
<code>dbpedia-owl:nationalTopographicSystemMapNumber</code>	■ <code>Swisstopo 1347 Matterhorn</code>
<code>dbpedia-owl:parentMountainPeak</code>	■ <code>dbpedia:Weisshorn</code>

- Zu anderen LOD Datenquellen

<code>owl:sameAs</code>	<ul style="list-style-type: none">■ http://linkedgedata.org/triplify/node/26863664#d■ <code>fbase:Matterhorn</code>■ http://sws.geonames.org/2659729/
-------------------------	---



Zusammenfassung

- Beispielontologien
 - Top-Level vs. Domänenontologien
- Ontologien zur Annotation
 - Semantisch eindeutige Beschreibung von Objekten
 - Zahlreiche Vorteile (z.B. Suche, Wissensaustausch, ...)
- Semantische Suche
- Linked Data
 - Ontologien zur Gestaltung eines Web of Data
 - Designprinzipien

