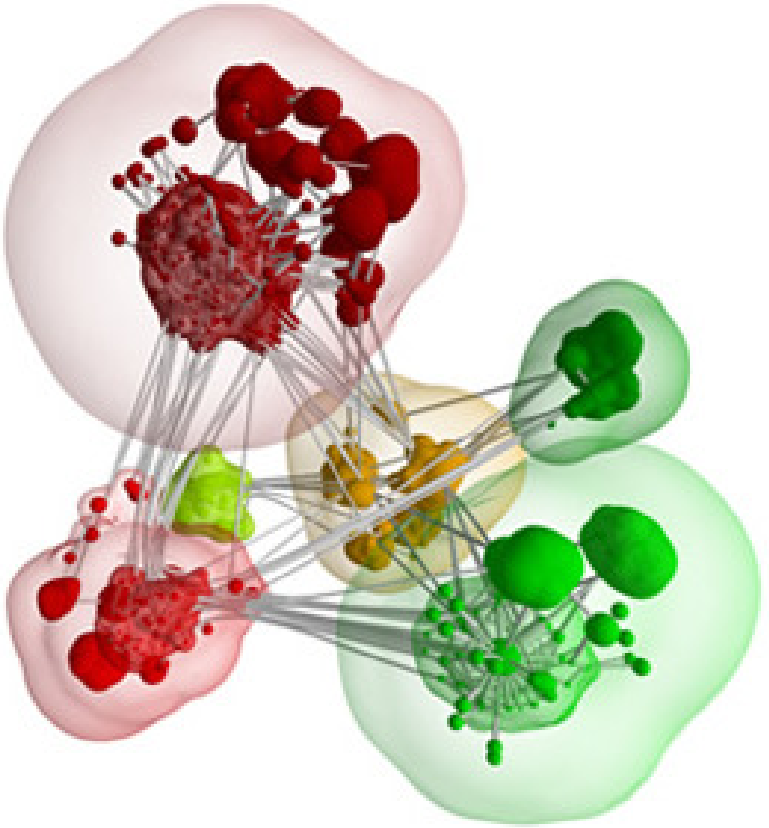
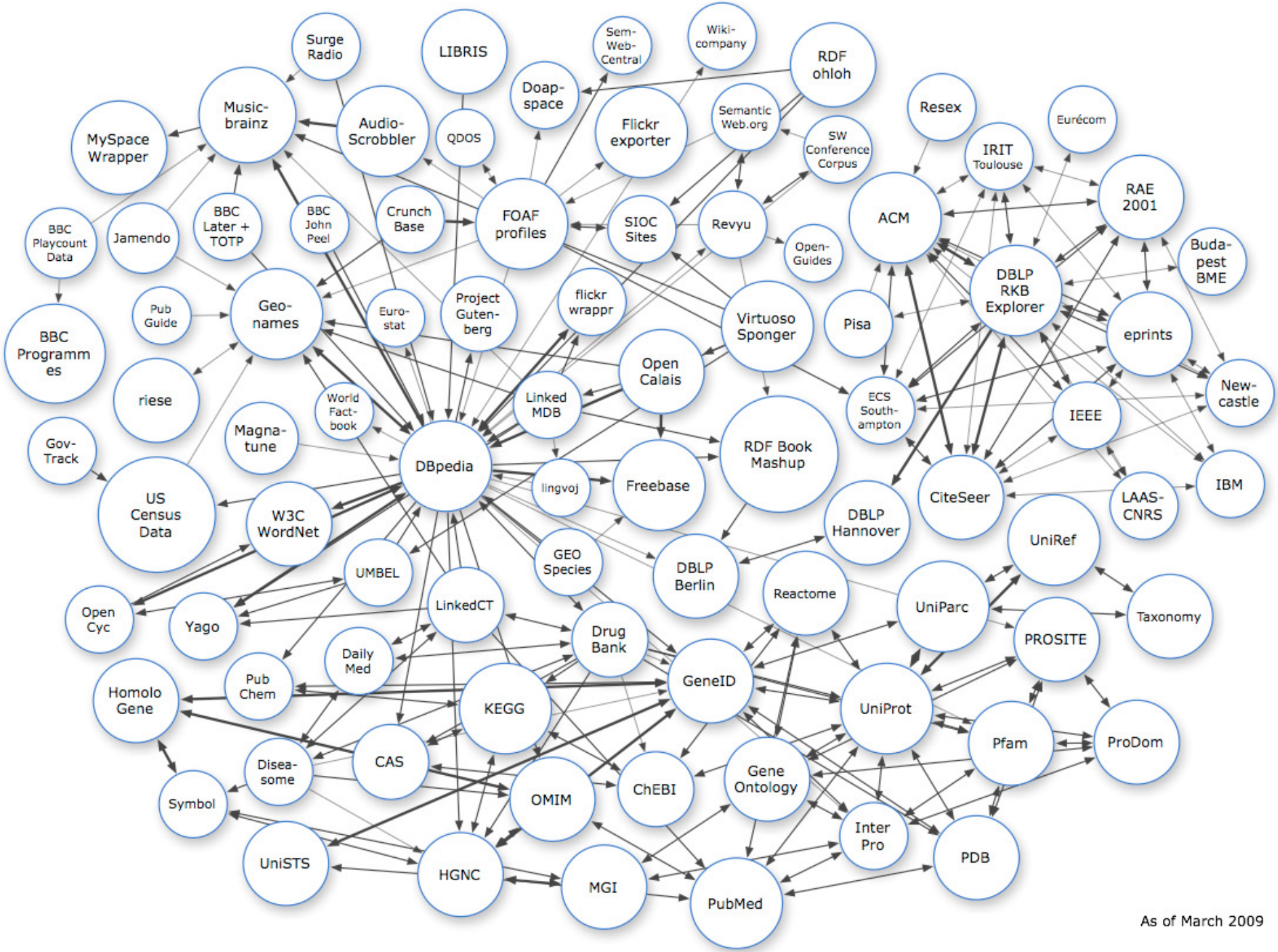
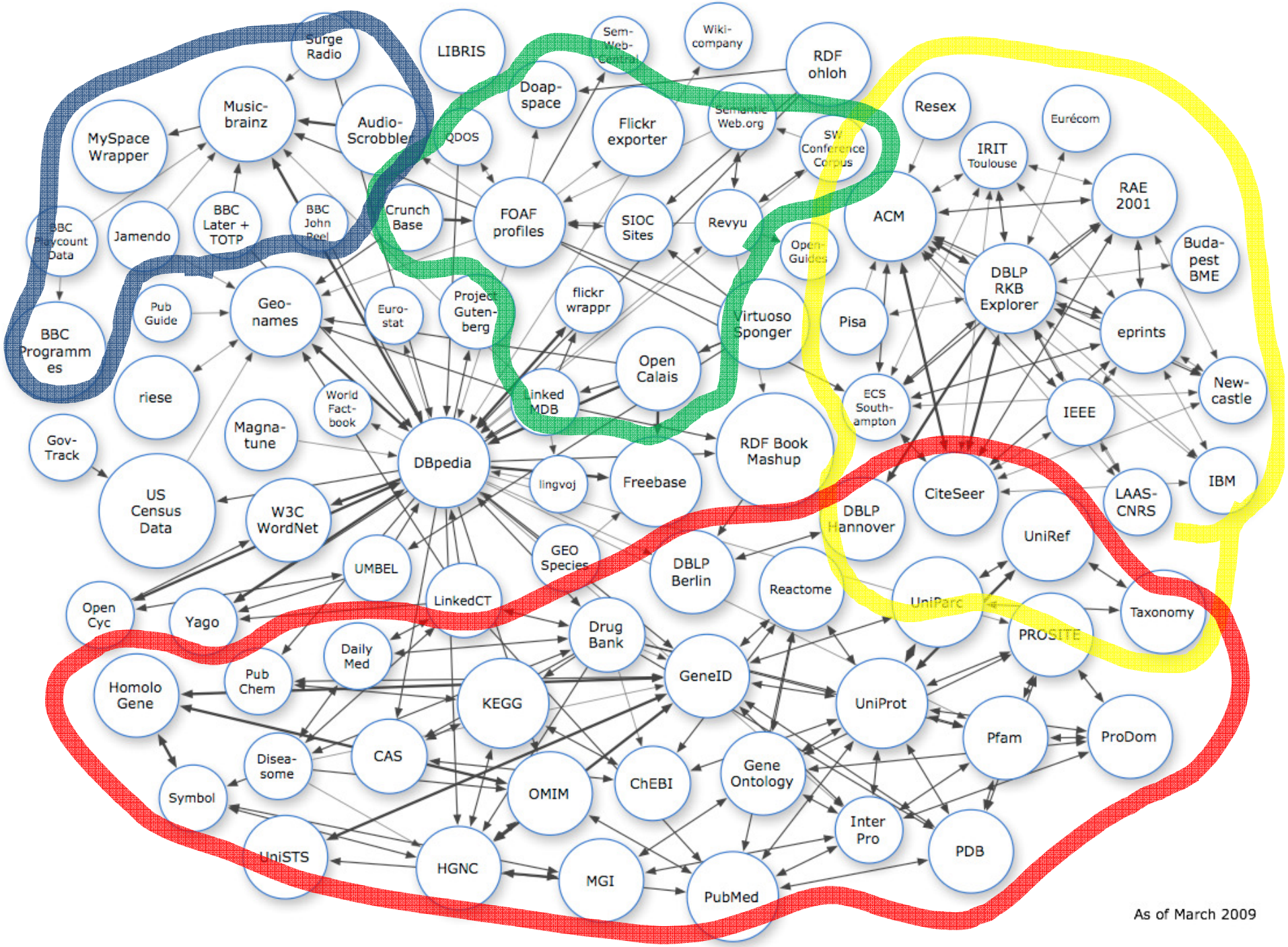
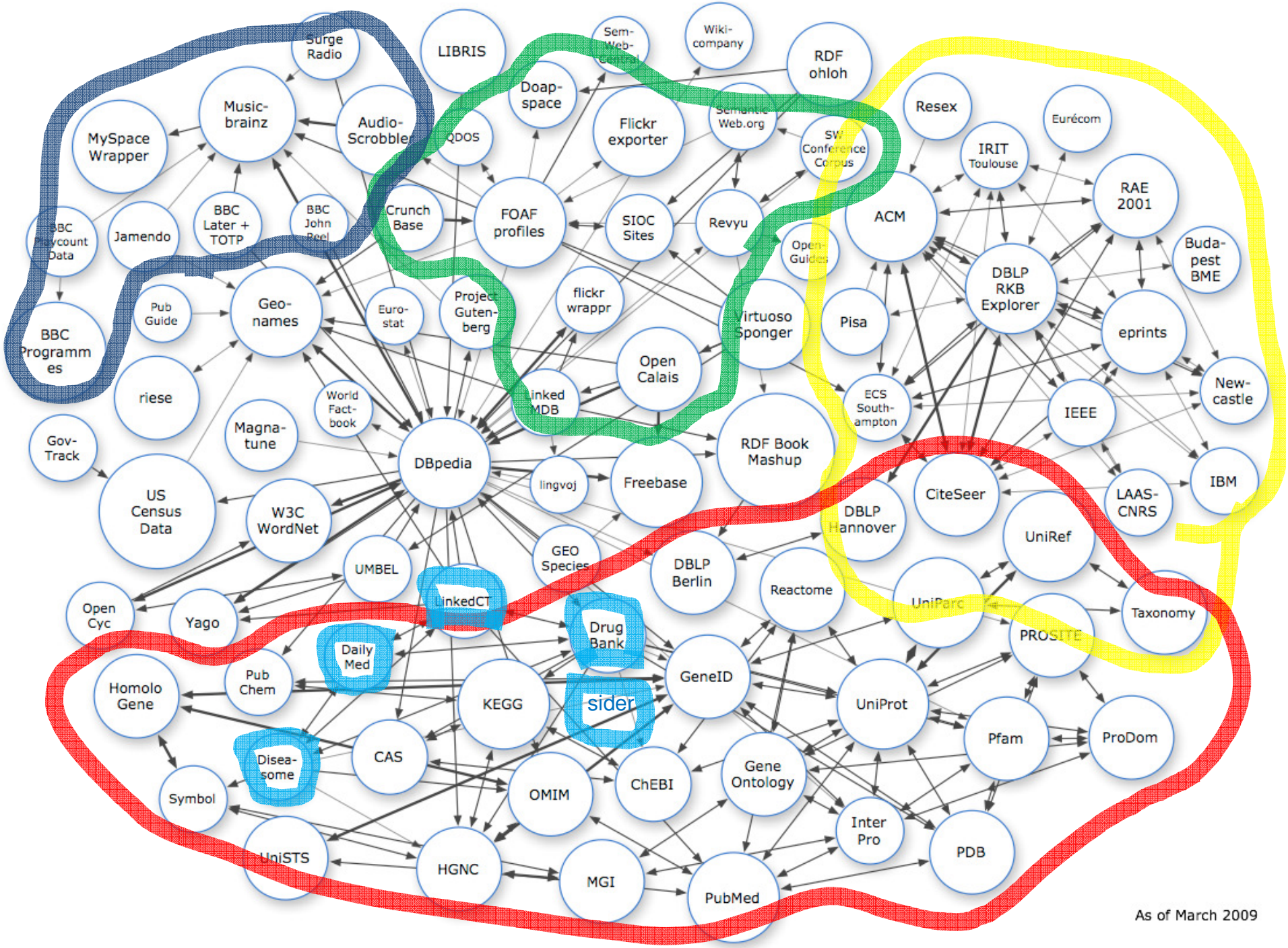
Linked Data, Life Sciences, and RDF Stores - Exploration and Demonstration



Jans Aasman, Ph.D.
CEO Franz Inc
Ja@Franz.com


Surge Radio
LIBRIS
Sem-Web-Central
Wiki-company
RDF ohloh
Doap-space
Music-brainz
Audio-Scrobbler
QDOS
Flickr exporter
Semantic Web.org
SW Conference Corpus
Resex
Eurécom
IRIT Toulouse
MySpace Wrapper
BBC Playcount Data
Jamendo
BBC Later + TOTP
BBC John Peel
Crunch Base
FOAF profiles
SIOC Sites
Revyu
Open-Guides
ACM
RAE 2001
Buda-pest BME
DBLP RKB Explorer
Pisa
eprints
BBC Programmes
Pub Guide
Geo-names
Euro-stat
Project Guten-berg
flickr wrappr
Virtuoso Sponger
Open Calais
riese
World Fact-book
Linked MDB
ECS South-ampton
IEEE
New-castle
Gov-Track
Magna-tune
DBpedia
RDF Book Mashup
US Census Data
W3C WordNet
lingvoj
Freebase
CiteSeer
LAAS-CNRS
IBM
DBLP Hannover
UniRef
UMBEL
GEO Species
DBLP Berlin
Reactome
UniParc
Taxonomy
Open Cyc
Yago
LinkedCT
Drug Bank
PROSITE
Daily Med
Homolo Gene
Pub Chem
KEGG
GeneID
UniProt
Pfam
ProDom
Disea-some
CAS
Gene Ontology
Symbol
OMIM
ChEBI
Inter Pro
UniSTS
HGNC
MGI
PubMed
PDB
As of March 2009


Surge Radio
LIBRIS
Sem-Web-
Wiki-company
RDF ohloh
Doap-space
Music-brainz
Audio-Scrobbler
MySpace Wrapper
QDOS
Flickr exporter
Semantic Web.org
SW Conference Corpus
Resex
Eurécom
IRIT Toulouse
RAE 2001
ACM
Budapest BME
BBC Wildlife
Jamendo
BBC Later + TOTP
BBC John Peel
Crunch Base
FOAF profiles
SIOC Sites
Revyu
Open Guides
DBLP RKB Explorer
BBC Programmes
Pub Guide
Geo-names
Euro-stat
Project Guten-berg
flickr wrappr
Virtuoso Sponger
Pisa
eprints
Open Calais
Linked MDB
ECS South-ampton
IEEE
New-castle
riese
World Fact-book
Gov-Track
Magna-tune
DBpedia
RDF Book Mashup
lingvoj
Freebase
CiteSeer
LAAS-CNRS
IBM
US Census Data
W3C WordNet
DBLP Hannover
UniRef
UMBEL
GEO Species
DBLP Berlin
Reactome
UniParc
Taxonomy
Open Cyc
Yago
LinkedCT
Drug Bank
PROSITE
Daily Med
Homolo Gene
Pub Chem
KEGG
GeneID
UniProt
Pfam
ProDom
CAS
Disea-some
Symbol
OMIM
ChEBI
Gene Ontology
Inter Pro
UniSTS
HGNC
MGI
PubMed
PDB
As of March 2009


Surge Radio
LIBRIS
Sem-Web-
Wiki-company
RDF ohloh
Doap-space
Music-brainz
Audio-Scrobbler
QDOS
Flickr exporter
Semantic Web.org
SW Conference Corpus
Resex
Eurécom
IRIT Toulouse
RAE 2001
MySpace Wrapper
BBC Later + TOTP
BBC John Peel
Crunch Base
FOAF profiles
SIOC Sites
Revyu
ACM
Buda-pest BME
Jamendo
Open-Guides
DBLP RKB Explorer
BBC Programmes
Pub Guide
Geo-names
Euro-stat
Project Guten-berg
flickr wrappr
Virtuoso Sponger
Pisa
eprints
Open Calais
ECS South-ampton
New-castle
riese
World Fact-book
Linked MDB
IEEE
Gov-Track
Magna-tune
DBpedia
RDF Book Mashup
lingvoj
Freebase
CiteSeer
IBM
LAAS-CNRS
US Census Data
W3C WordNet
DBLP Hannover
UniRef
UMBEL
GEO Species
DBLP Berlin
Reactome
UniParc
Taxonomy
Open Cyc
Yago
LinkedCT
Drug Bank
Daily Med
PROSITE
Homolo Gene
Pub Chem
KEGG
sider
GeneID
UniProt
Pfam
ProDom
Disea-some
CAS
Symbol
OMIM
ChEBI
Gene Ontology
Inter Pro
PDB
UniSTS
HGNC
MGI
PubMed
As of March 2009

Federated 11 linked data sets

- We took 5 public databases: Drugbank, Dailymeds, Clinical trials, Diseasome, and Sider. Entities are mostly linked together through same-as relationships.

- And using some entity extraction created some more links (and) triples
  - CT-discusses-drug, CT-discusses-side-effect
  - CT-discusses-target, CT-discusses-disease

- Alitora did some extensive NLP and entity extraction on Rheumatoid Arthritis
  - CT-mentions-genes

- And to facilitate search through schema space: Schema-connections

DrugBank

Home | Browse | Search | About | Downloads | Contact Us

Search: Search [Advanced]

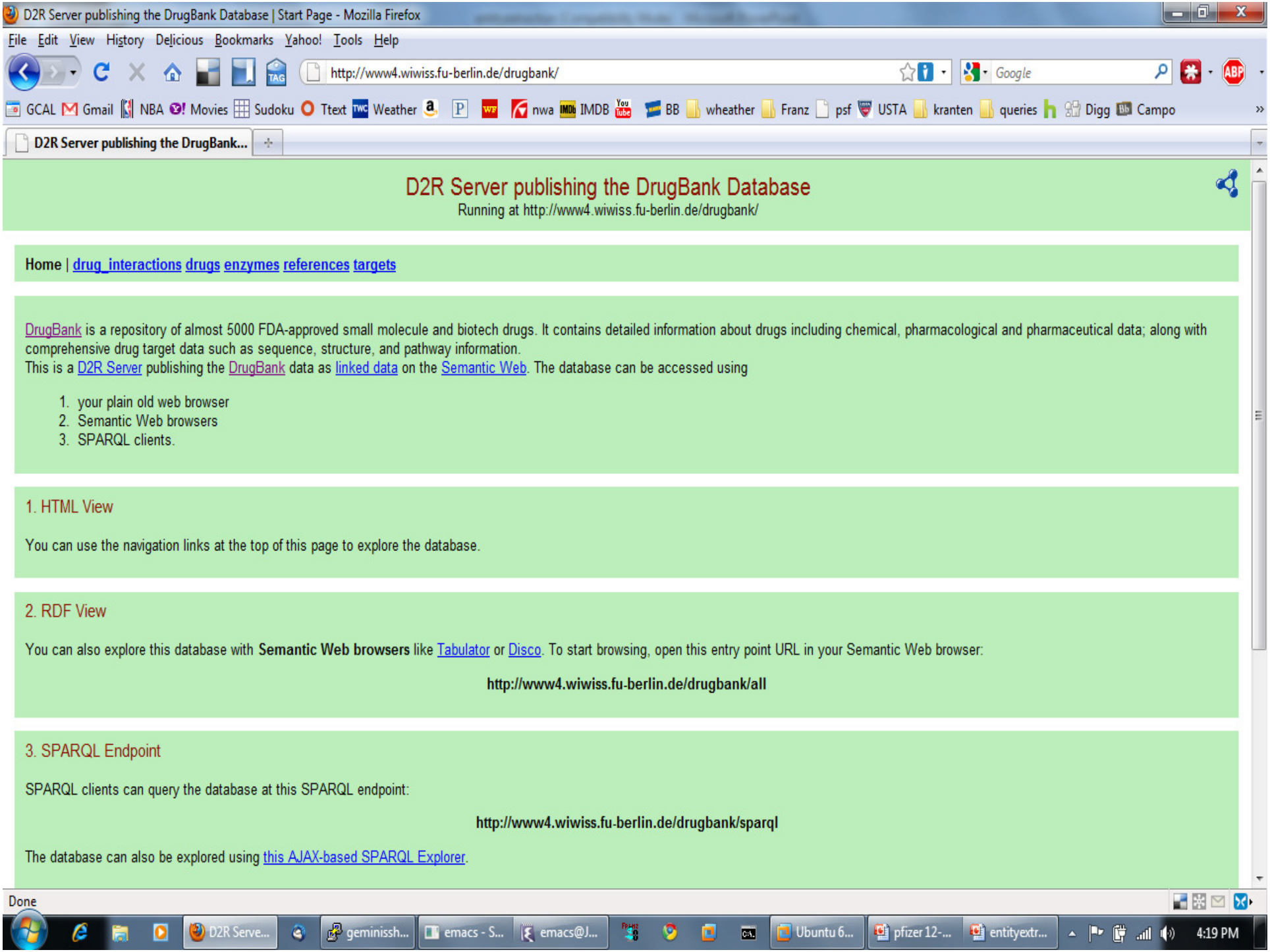
The DrugBank database is a unique bioinformatics and cheminformatics resource that combines detailed drug (i.e. chemical, pharmacological and pharmaceutical) data with comprehensive drug target (i.e. sequence, structure, and pathway) information. The database contains nearly 4800 drug entries including >1,350 FDA-approved small molecule drugs, 123 FDA-approved biotech (protein/peptide) drugs, 71 nutraceuticals and >3,243 experimental drugs. Additionally, more than 2,500 non-redundant protein (i.e. drug target) sequences are linked to these FDA approved drug entries. Each DrugCard entry contains more than 100 data fields with half of the information being devoted to drug/chemical data and the other half devoted to drug target or protein data.

DrugBank is supported by David Wishart, Departments of Computing Science & Biological Sciences, University of Alberta.

More about DrugBank

What's New?

- You can now set the maximum number of results returned from a structure search on the ChemQuery page. The limit used to be 100, which missed some important drugs in the search results.
- We have implemented the ChemAxon solution for structure searches. You can now perform similarity (tanimoto), substructure, and exact searches via the ChemQuery function. This system replaces an outdated structure search and should be faster and more accurate. We have only added the most basic features for this release, so if you would like to see more/different features added, please let us know.


D2R Server publishing the DrugBank Database | Start Page - Mozilla Firefox
File Edit View History Delicious Bookmarks Yahoo! Tools Help
http://www4.wiwiss.fu-berlin.de/drugbank/
Google
GCAL Gmail NBA Movies Sudoku Ttext Weather nwa IMDB BB wheather Franz psf USTA kranten queries Digg Campo
D2R Server publishing the DrugBank...
D2R Server publishing the DrugBank Database
Running at http://www4.wiwiss.fu-berlin.de/drugbank/
Home | drug_interactions drugs enzymes references targets
DrugBank is a repository of almost 5000 FDA-approved small molecule and biotech drugs. It contains detailed information about drugs including chemical, pharmacological and pharmaceutical data; along with comprehensive drug target data such as sequence, structure, and pathway information.
This is a D2R Server publishing the DrugBank data as linked data on the Semantic Web. The database can be accessed using
1. your plain old web browser
2. Semantic Web browsers
3. SPARQL clients.
1. HTML View
You can use the navigation links at the top of this page to explore the database.
2. RDF View
You can also explore this database with Semantic Web browsers like Tabulator or Disco. To start browsing, open this entry point URL in your Semantic Web browser:
http://www4.wiwiss.fu-berlin.de/drugbank/all
3. SPARQL Endpoint
SPARQL clients can query the database at this SPARQL endpoint:
http://www4.wiwiss.fu-berlin.de/drugbank/sparql
The database can also be explored using this AJAX-based SPARQL Explorer.
Done
4:19 PM

DrugBank

A repository of almost 5000 FDA-approved small molecule and biotech drugs.

Contains detailed information about drugs including chemical, pharmacological and pharmaceutical data; along with comprehensive drug target data such as sequence, structure, and pathway information.

DrugBank

LinkedCT: Clinical Trials

Up-to-date information for locating federally and privately supported clinical trials for a wide range of diseases and conditions.

It contains 81,571 trials sponsored by the National Institutes of Health, other federal agencies, and private industry

ClinicalTrials.gov receives over 40 million page views per month 50,000 visitors daily.

ClinicalTrials.gov
A service of the U.S. National Institutes of Health

Diseasome

Publishes a network of 4,300 disorders and disease genes linked by known disorder-gene associations for exploring all known phenotype and disease gene associations, indicating the common genetic origin of many diseases. The list of disorders, disease genes, and associations between them was obtained from the Online Mendelian Inheritance in Man (OMIM), a compilation of human disease genes and phenotypes.

DailyMed

Published by the National Library of Medicine, and provides high quality information about marketed drugs. DailyMed provides much information including general background on the chemical structure of the compound and its therapeutic purpose, details on the compound's clinical pharmacology, indication and usage, contraindications, warnings, precautions, adverse reactions, overdosage, and patient counseling.

Sider

Contains information on marketed drugs and their adverse effects. The information is extracted from public documents and package inserts.

Finding entities in Clinical Trials

- CT has too much text

- We searched for drugs, diseases, targets and side effects in Clinical trials and created new triples

  - CT100385 discusses-side-effect headache
  - CT100385 discusses-drug aspirin
  - CT100385 discusses-disease alcohol-addiction
  - CT100385 discusses-target some-protein

Combined with advanced entity extraction

- CT-mentions-drugs
- CT-mentions-genes
- CT-mentions-side-effects
- (currently only for

Interesting queries (Sparql)

- Sparql
  - Give me the title of all clinical trials that discuss the drug Lipitor and the side-effect “Diabetes type 2”
  - Give me clinical trials that discuss Rheumatoid Arthritis and give me the genes and drugs discussed
- Prolog
  - Find all clinical trials that resemble clinical trial NCT00130091 given diseases, drugs, targets, and side-effects

Finding your way in Schema Space

- Finding your way in a forest of rdf-type is non-trivial
- Many linked datasets have no range and domain definitions
- We have a tool to make it easier to explore