



# Ontological Applications at MediGRID

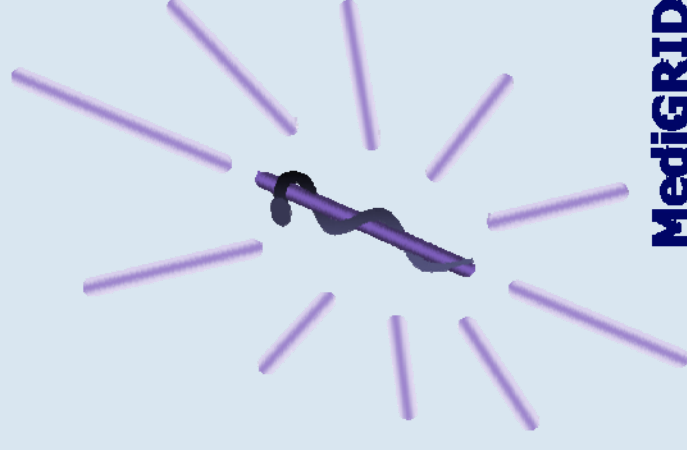
Berlin, 18.04.2007

Michael Hartung, Erhard Rahm

Interdisciplinary Centre for Bioinformatics, Leipzig

Results for "blood" in "GeneOntology":

|                   |  |
|-------------------|--|
| <b>Name</b>       | blood coagulation  |
| <b>Definition</b> | The sequential process by which the multiple coagulation factors of the blood interact, ultimately resulting in the formation of an insoluble fibrin clot. It may be divided into three stages: stage 1, the formation of intrinsic and extrinsic pathways converging principle; stage 2, the formation of fibrin; stage 3, the formation of stable form.  |
| <b>Synonyms</b>   | <ul style="list-style-type: none"> <li>blood clotting</li> <li>blood coagulation factor activity</li> </ul>  |
| <b>Hierarchy</b>  | <ul style="list-style-type: none"> <li>biological process</li> <li>biological process</li> <li>biological process</li> <li>biological process</li> </ul>   |
| <b>Parents</b>    | <ul style="list-style-type: none"> <li>GO:0007596</li> <li>GO:0007596</li> <li>GO:0007596</li> <li>GO:0007596</li> </ul>   |
| <b>Children</b>   | <ul style="list-style-type: none"> <li>GO:0007597</li> <li>GO:0007598</li> <li>GO:0007599</li> <li>GO:0007600</li> <li>GO:0007601</li> <li>GO:0007602</li> <li>GO:0007603</li> <li>GO:0007604</li> <li>GO:0007605</li> <li>GO:0007606</li> <li>GO:0007607</li> <li>GO:0007608</li> <li>GO:0007609</li> <li>GO:0007610</li> <li>GO:0007611</li> <li>GO:0007612</li> <li>GO:0007613</li> <li>GO:0007614</li> <li>GO:0007615</li> <li>GO:0007616</li> <li>GO:0007617</li> <li>GO:0007618</li> <li>GO:0007619</li> <li>GO:0007620</li> <li>GO:0007621</li> <li>GO:0007622</li> <li>GO:0007623</li> <li>GO:0007624</li> <li>GO:0007625</li> <li>GO:0007626</li> <li>GO:0007627</li> <li>GO:0007628</li> <li>GO:0007629</li> <li>GO:0007630</li> <li>GO:0007631</li> <li>GO:0007632</li> <li>GO:0007633</li> <li>GO:0007634</li> <li>GO:0007635</li> <li>GO:0007636</li> <li>GO:0007637</li> <li>GO:0007638</li> <li>GO:0007639</li> <li>GO:0007640</li> <li>GO:0007641</li> <li>GO:0007642</li> <li>GO:0007643</li> <li>GO:0007644</li> <li>GO:0007645</li> <li>GO:0007646</li> <li>GO:0007647</li> <li>GO:0007648</li> <li>GO:0007649</li> <li>GO:0007650</li> <li>GO:0007651</li> <li>GO:0007652</li> <li>GO:0007653</li> <li>GO:0007654</li> <li>GO:0007655</li> <li>GO:0007656</li> <li>GO:0007657</li> <li>GO:0007658</li> <li>GO:0007659</li> <li>GO:0007660</li> <li>GO:0007661</li> <li>GO:0007662</li> <li>GO:0007663</li> <li>GO:0007664</li> <li>GO:0007665</li> <li>GO:0007666</li> <li>GO:0007667</li> <li>GO:0007668</li> <li>GO:0007669</li> <li>GO:0007670</li> <li>GO:0007671</li> <li>GO:0007672</li> <li>GO:0007673</li> <li>GO:0007674</li> <li>GO:0007675</li> <li>GO:0007676</li> <li>GO:0007677</li> <li>GO:0007678</li> <li>GO:0007679</li> <li>GO:0007680</li> <li>GO:0007681</li> <li>GO:0007682</li> <li>GO:0007683</li> <li>GO:0007684</li> <li>GO:0007685</li> <li>GO:0007686</li> <li>GO:0007687</li> <li>GO:0007688</li> <li>GO:0007689</li> <li>GO:0007690</li> <li>GO:0007691</li> <li>GO:0007692</li> <li>GO:0007693</li> <li>GO:0007694</li> <li>GO:0007695</li> <li>GO:0007696</li> <li>GO:0007697</li> <li>GO:0007698</li> <li>GO:0007699</li> <li>GO:0007700</li> <li>GO:0007701</li> <li>GO:0007702</li> <li>GO:0007703</li> <li>GO:0007704</li> <li>GO:0007705</li> <li>GO:0007706</li> <li>GO:0007707</li> <li>GO:0007708</li> <li>GO:0007709</li> <li>GO:0007710</li> <li>GO:0007711</li> <li>GO:0007712</li> <li>GO:0007713</li> <li>GO:0007714</li> <li>GO:0007715</li> <li>GO:0007716</li> <li>GO:0007717</li> <li>GO:0007718</li> <li>GO:0007719</li> <li>GO:0007720</li> <li>GO:0007721</li> <li>GO:0007722</li> <li>GO:0007723</li> <li>GO:0007724</li> <li>GO:0007725</li> <li>GO:0007726</li> <li>GO:0007727</li> <li>GO:0007728</li> <li>GO:0007729</li> <li>GO:0007730</li> <li>GO:0007731</li> <li>GO:0007732</li> <li>GO:0007733</li> <li>GO:0007734</li> <li>GO:0007735</li> <li>GO:0007736</li> <li>GO:0007737</li> <li>GO:0007738</li> <li>GO:0007739</li> <li>GO:0007740</li> <li>GO:0007741</li> <li>GO:0007742</li> <li>GO:0007743</li> <li>GO:0007744</li> <li>GO:0007745</li> <li>GO:0007746</li> <li>GO:0007747</li> <li>GO:0007748</li> <li>GO:0007749</li> <li>GO:0007750</li> <li>GO:0007751</li> <li>GO:0007752</li> <li>GO:0007753</li> <li>GO:0007754</li> <li>GO:0007755</li> <li>GO:0007756</li> <li>GO:0007757</li> <li>GO:0007758</li> <li>GO:0007759</li> <li>GO:0007760</li> <li>GO:0007761</li> <li>GO:0007762</li> <li>GO:0007763</li> <li>GO:0007764</li> <li>GO:0007765</li> <li>GO:0007766</li> <li>GO:0007767</li> <li>GO:0007768</li> <li>GO:0007769</li> <li>GO:0007770</li> <li>GO:0007771</li> <li>GO:0007772</li> <li>GO:0007773</li> <li>GO:0007774</li> <li>GO:0007775</li> <li>GO:0007776</li> <li>GO:0007777</li> <li>GO:0007778</li> <li>GO:0007779</li> <li>GO:0007780</li> <li>GO:0007781</li> <li>GO:0007782</li> <li>GO:0007783</li> <li>GO:0007784</li> <li>GO:0007785</li> <li>GO:0007786</li> <li>GO:0007787</li> <li>GO:0007788</li> <li>GO:0007789</li> <li>GO:0007790</li> <li>GO:0007791</li> <li>GO:0007792</li> <li>GO:0007793</li> <li>GO:0007794</li> <li>GO:0007795</li> <li>GO:0007796</li> <li>GO:0007797</li> <li>GO:0007798</li> <li>GO:0007799</li> <li>GO:0007800</li> <li>GO:0007801</li> <li>GO:0007802</li> <li>GO:0007803</li> <li>GO:0007804</li> <li>GO:0007805</li> <li>GO:0007806</li> <li>GO:0007807</li> <li>GO:0007808</li> <li>GO:0007809</li> <li>GO:0007810</li> <li>GO:0007811</li> <li>GO:0007812</li> <li>GO:0007813</li> <li>GO:0007814</li> <li>GO:0007815</li> <li>GO:0007816</li> <li>GO:0007817</li> <li>GO:0007818</li> <li>GO:0007819</li> <li>GO:0007820</li> <li>GO:0007821</li> <li>GO:0007822</li> <li>GO:0007823</li> <li>GO:0007824</li> <li>GO:0007825</li> <li>GO:0007826</li> <li>GO:0007827</li> <li>GO:0007828</li> <li>GO:0007829</li> <li>GO:0007830</li> <li>GO:0007831</li> <li>GO:0007832</li> <li>GO:0007833</li> <li>GO:0007834</li> <li>GO:0007835</li> <li>GO:0007836</li> <li>GO:0007837</li> <li>GO:0007838</li> <li>GO:0007839</li> <li>GO:0007840</li> <li>GO:0007841</li> <li>GO:0007842</li> <li>GO:0007843</li> <li>GO:0007844</li> <li>GO:0007845</li> <li>GO:0007846</li> <li>GO:0007847</li> <li>GO:0007848</li> <li>GO:0007849</li> <li>GO:0007850</li> <li>GO:0007851</li> <li>GO:0007852</li> <li>GO:0007853</li> <li>GO:0007854</li> <li>GO:0007855</li> <li>GO:0007856</li> <li>GO:0007857</li> <li>GO:0007858</li> <li>GO:0007859</li> <li>GO:0007860</li> <li>GO:0007861</li> <li>GO:0007862</li> <li>GO:0007863</li> <li>GO:0007864</li> <li>GO:0007865</li> <li>GO:0007866</li> <li>GO:0007867</li> <li>GO:0007868</li> <li>GO:0007869</li> <li>GO:0007870</li> <li>GO:0007871</li> <li>GO:0007872</li> <li>GO:0007873</li> <li>GO:0007874</li> <li>GO:0007875</li> <li>GO:0007876</li> <li>GO:0007877</li> <li>GO:0007878</li> <li>GO:0007879</li> <li>GO:0007880</li> <li>GO:0007881</li> <li>GO:0007882</li> <li>GO:0007883</li> <li>GO:0007884</li> <li>GO:0007885</li> <li>GO:0007886</li> <li>GO:0007887</li> <li>GO:0007888</li> <li>GO:0007889</li> <li>GO:0007890</li> <li>GO:0007891</li> <li>GO:0007892</li> <li>GO:0007893</li> <li>GO:0007894</li> <li>GO:0007895</li> <li>GO:0007896</li> <li>GO:0007897</li> <li>GO:0007898</li> <li>GO:0007899</li> <li>GO:0007900</li> <li>GO:0007901</li> <li>GO:0007902</li> <li>GO:0007903</li> <li>GO:0007904</li> <li>GO:0007905</li> <li>GO:0007906</li> <li>GO:0007907</li> <li>GO:0007908</li> <li>GO:0007909</li> <li>GO:0007910</li> <li>GO:0007911</li> <li>GO:0007912</li> <li>GO:0007913</li> <li>GO:0007914</li> <li>GO:0007915</li> <li>GO:0007916</li> <li>GO:0007917</li> <li>GO:0007918</li> <li>GO:0007919</li> <li>GO:0007920</li> <li>GO:0007921</li> <li>GO:0007922</li> <li>GO:0007923</li> <li>GO:0007924</li> <li>GO:0007925</li> <li>GO:0007926</li> <li>GO:0007927</li> <li>GO:0007928</li> <li>GO:0007929</li> <li>GO:0007930</li> <li>GO:0007931</li> <li>GO:0007932</li> <li>GO:0007933</li> <li>GO:0007934</li> <li>GO:0007935</li> <li>GO:0007936</li> <li>GO:0007937</li> <li>GO:0007938</li> <li>GO:0007939</li> <li>GO:0007940</li> <li>GO:0007941</li> <li>GO:0007942</li> <li>GO:0007943</li> <li>GO:0007944</li> <li>GO:0007945</li> <li>GO:0007946</li> <li>GO:0007947</li> <li>GO:0007948</li> <li>GO:0007949</li> <li>GO:0007950</li> <li>GO:0007951</li> <li>GO:0007952</li> <li>GO:0007953</li> <li>GO:0007954</li> <li>GO:0007955</li> <li>GO:0007956</li> <li>GO:0007957</li> <li>GO:0007958</li> <li>GO:0007959</li> <li>GO:0007960</li> <li>GO:0007961</li> <li>GO:0007962</li> <li>GO:0007963</li> <li>GO:0007964</li> <li>GO:0007965</li> <li>GO:0007966</li> <li>GO:0007967</li> <li>GO:0007968</li> <li>GO:0007969</li> <li>GO:0007970</li> <li>GO:0007971</li> <li>GO:0007972</li> <li>GO:0007973</li> <li>GO:0007974</li> <li>GO:0007975</li> <li>GO:0007976</li> <li>GO:0007977</li> <li>GO:0007978</li> <li>GO:0007979</li> <li>GO:0007980</li> <li>GO:0007981</li> <li>GO:0007982</li> <li>GO:0007983</li> <li>GO:0007984</li> <li>GO:0007985</li> <li>GO:0007986</li> <li>GO:0007987</li> <li>GO:0007988</li> <li>GO:0007989</li> <li>GO:0007990</li> <li>GO:0007991</li> <li>GO:0007992</li> <li>GO:0007993</li> <li>GO:0007994</li> <li>GO:0007995</li> <li>GO:0007996</li> <li>GO:0007997</li> <li>GO:0007998</li> <li>GO:0007999</li> <li>GO:0008000</li> </ul> |





## Introduction

- **Module Ontology Tools in MediGRID**
  - Responsible for integration of ontologies in MediGRID
  - Tools that help to use and access ontology information in grids
- **Usage of ontologies depends on applications in MediGRID**
- **3 Pilot classes of applications**
  - Bioinformatics
  - Imaging
  - Clinical Research
- **Linking applications with currently existing ontologies**
  - Integration of further knowledge
  - Utilization of annotation databases, sources
  - First phase: No editing/construction of new ontologies



## Example of an ontology – NCI Thesaurus

- **About the thesaurus**
  - Reference Terminology for NCI
  - Broad coverage of cancer domain
    - Findings and Disorders
    - Anatomy
    - Drugs, Chemicals
  - About 38000 concepts in 20 hierarchies
- **Why**
  - Uniform conceptualization in a domain
  - Standardization, interoperability, classification
  - Enable reuse of data and information
- **Usage**
  - Annotation of medical data (images, ...)
  - Better search capabilities (categories)

### NCI\_Thesaurus Taxonomy

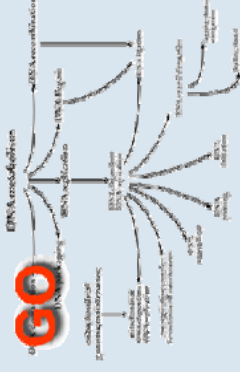
- [-] Abnormal Cell
- [-] Activity
- [-] Anatomic Structure, System, or Substance
  - [+] Body Fluid or Substance
  - [+] Body Part
  - [+] Body Region
  - [+] Body Cavity
- [-] Embryologic Structure or System
- [-] Microanatomic Structure
  - [ ] Organ
    - [ ] Biliary Tract
    - [ ] Bladder
    - [ ] Bone Marrow
    - [ ] Brain
    - [+] Breast
    - [ ] Bronchial Tree
    - [ ] Diaphragm
    - [+] Duct
    - [ ] Epididymis
    - [ ] Esophagus
    - [+] Fallopian Tube
    - [ ] Gall Bladder
    - [+] Gland
    - [+] Gonad
    - [ ] Heart



## Motivation

### Key problem: Heterogeneity between existing ontologies

- **Yet no existing ontology access system in grids**
- **Base information**
  - Terms, concepts with name, definition, synonyms
  - Relations between concepts of an ontology
  - Cross references to other ontologies, data sources
- **Different source formats of ontologies**
  - Relational databases: GeneOntology
  - Structured flat files: NCI Thesaurus
  - Standardizations: OBO-Ontologies
  - Web Standards: OWL, RDF
- **Different design of ontology information**
  - Modeling of relationships
  - Presentation of synonyms, properties



National Cancer Institute





## Goals

### Simple and transparent access middleware for ontologies in grids

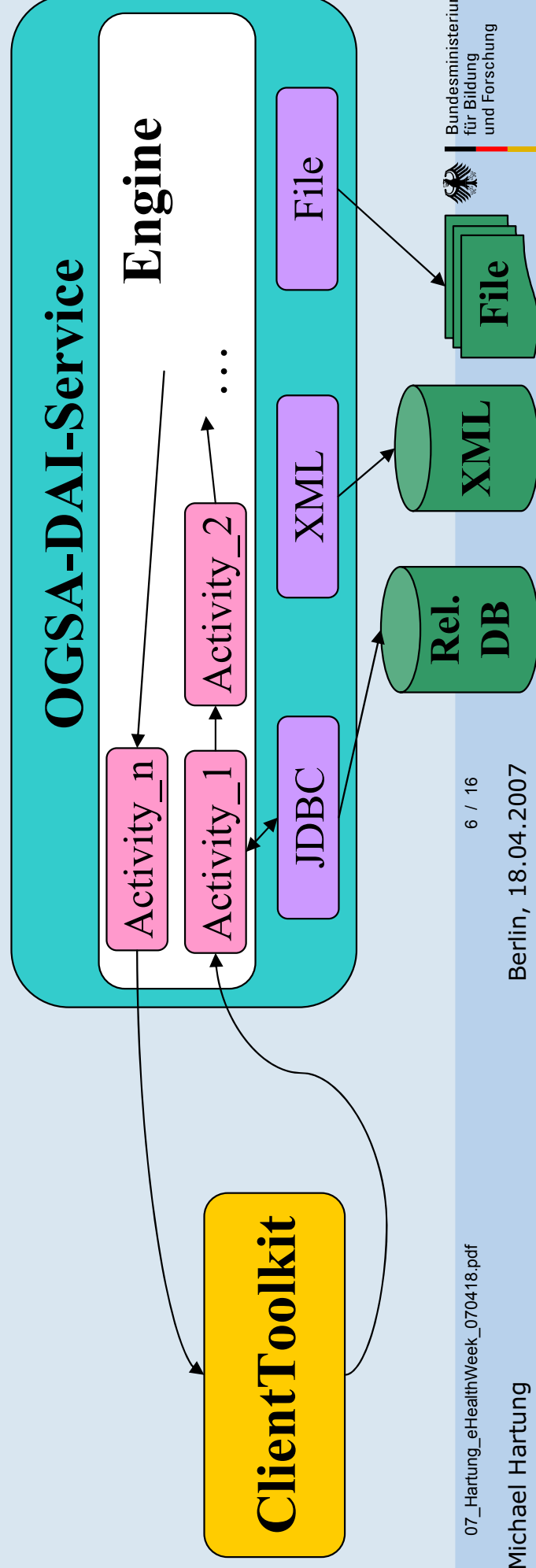
- **Functionalities**
  - Look Up of concepts/terms in available ontologies
  - Access to ontology information: definitions, synonyms, relations, cross references → uniform interface for all clients (API)
  - GUI: navigation in ontologies, graph representation of ontologies
- **Service based access to ontologies**
  - Ontology Services as middleware between applications and ontologies
  - Support for clients (MediGRID applications, other services)
  - Usage of grid: Distribution of ontologies → load balancing, failure safety
  - Extensibility: simple integration of new ontologies, extension in functionalities
  - Usage of D-Grid basis software, integration in D-Grid infrastructure



## OGSA-DAI as base infrastructure

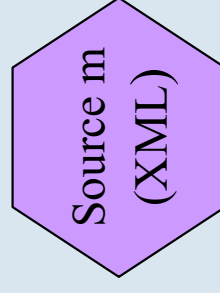
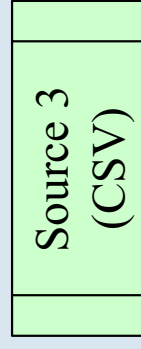
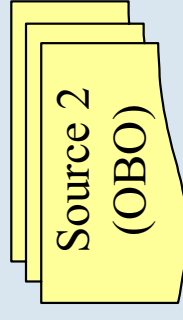
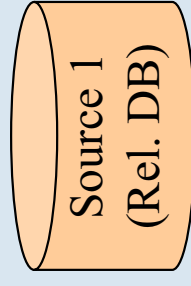
### Access to data sources in grids

- **Extensible framework for “data access” and “integration”**
  - Access to distributed and heterogeneous data sources in grids
  - Web service based access system (WSRF, Globus Toolkit 4)
  - Extensibility as an important feature (activities, resources)
  - Client Toolkit API for service interaction

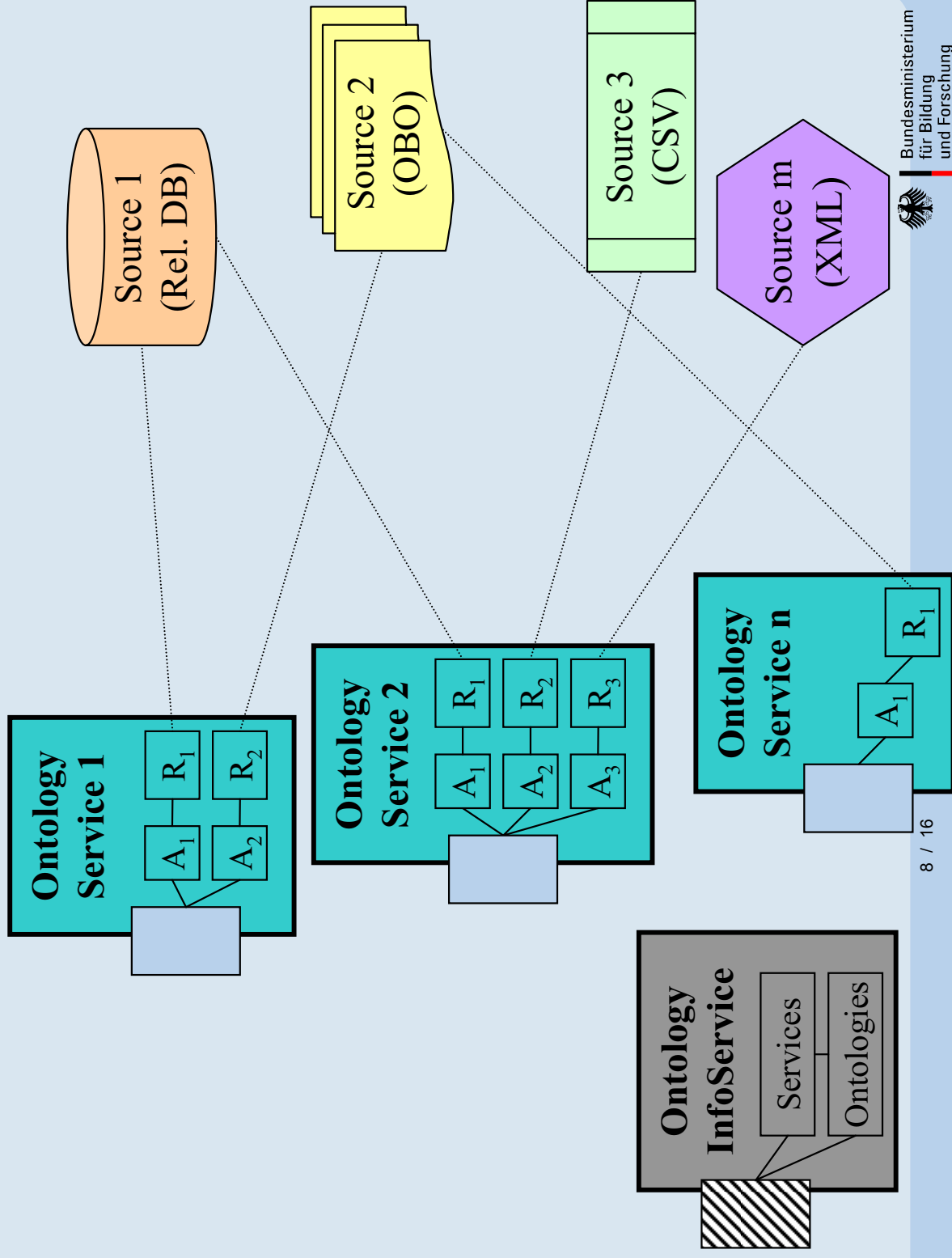




# Ontology Access Middleware

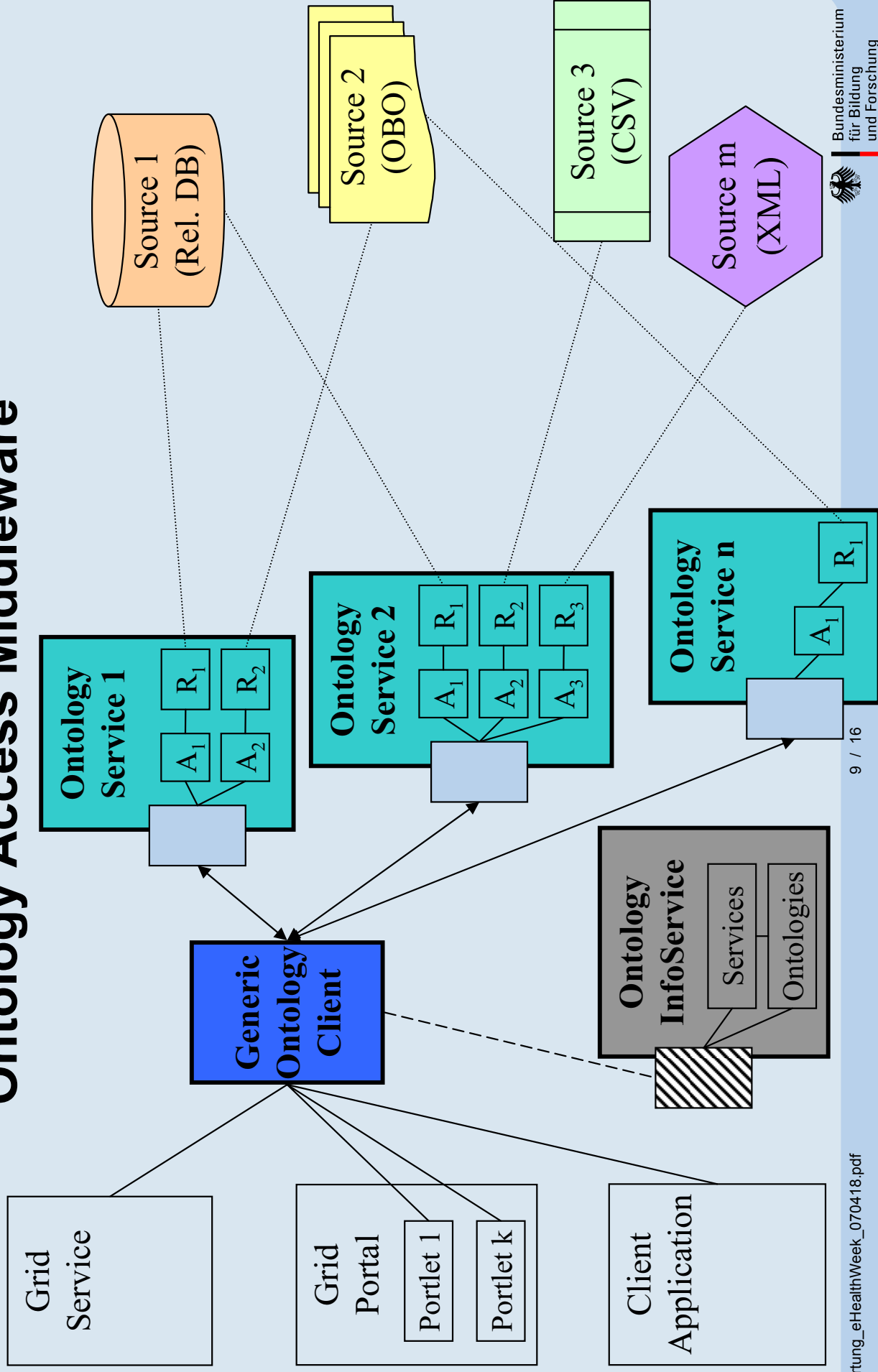


# Ontology Access Middleware





# Ontology Access Middleware





## Applications using the middleware

### MediGRID portal as central entry to grid applications

- **Advantages**
    - User friendly interaction with grid resources
    - Central access without knowledge about resource locations etc.
    - Portlets as re-usable software components
    - Graphical representation of results, workspace for users
  - **Ontology Access Middleware in the MediGRID portal**
    - Ontology Client interacts with distributed Ontology Services
    - Client is integrated in application portlets
1. Central Ontology Look Up Service
  2. AUGUSTUS gene prediction tool
  3. SNPSselection

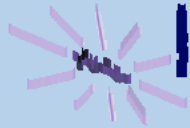


# Ontology Look Up Service (I)

Central ontology applications

Ontologies to Look Up





# Ontology Look Up Service (II)

**Results for "heart" in "GeneOntology":**

[regulation of heart contraction](#)  
 heart looping  
 negative regulation of heart contraction rate in baroreceptor response to increased blood pressure  
 decreased strength of heart contraction during baroreceptor response to increased blood pressure  
 positive regulation of heart contraction rate in baroreceptor response to decreased blood pressure

[Show term](#)

|                     |   |
|---------------------|---|
| <b>ID</b>           | GO:0008016  |
| <b>Name</b>         | regulation of heart contraction   |
| <b>Definition</b>   | Any process that modulates the frequency, rate or extent of heart contraction.  |
| <b>Comment</b>      |   |
| <b>Synonyms</b>     | <ul style="list-style-type: none"> <li>regulation of heart rate</li> </ul>  |
| <b>Hierarchy</b>    | <p>GO:0051239 regulation of organismal physiological process</p> <p>GO:0008016 regulation of heart contraction</p> <p>GO:0002026 cardiac inotropy</p> <p>GO:0002027 cardiac chronotropy</p> <p>GO:0045822 negative regulation of heart contraction</p> <p>GO:0045823 positive regulation of heart contraction</p> |
| <b>Parents</b>      | <ul style="list-style-type: none"> <li>GO:0051239; regulation of organismal physiological process (IS_A)</li> </ul>   |
| <b>Children</b>     | <ul style="list-style-type: none"> <li>GO:0002026; cardiac inotropy (IS_A)</li> <li>GO:0002027; cardiac chronotropy (IS_A)</li> <li>GO:0045822; negative regulation of heart contraction (IS_A)</li> <li>GO:0045823; positive regulation of heart contraction (IS_A)</li> </ul>                                   |
| <b>X-References</b> | <ul style="list-style-type: none"> <li>GOC:go_curators</li> <li>TIGR_TIGRFAMS:TIGR01294_phospholamban</li> </ul>  |

Search Results

Information about ontology concepts



# Ontology Look Up Service (III)

The screenshot shows a web browser window with the URL `MultiOntologyAccessPortlet`. The page title is "Überblick Ontologie- und DB-Ressourcen". A search bar contains "engbackta". Below the search bar, a list of ontology services is displayed, each with a green checkmark indicating it is available. A red location pin is placed on a map of Germany, pointing to Halle (Saale). A green callout box points to the list of services, and another green callout box points to the map.

**Monitoring of Ontology Services and corresponding ontologies**

**Locations**

Host: <http://buell.izbi.uni-leipzig.de:8081>  
 Service: [wsrf/services/ogsadai DataService](http://wsrf/services/ogsadai DataService)

- GeneOntology - GO@ducati.izbi.uni-leipzig.de - ✓
- NCI-Thesaurus - NCIThesaurusResource - ✓
- GeneOntology - GO@buell.izbi.uni-leipzig.de - ✓
- PathwayOntology - PathwayOntologyResource - ✓
- HumanDiseaseOntology - DiseaseOntologyResource - ✓
- ProteinProteinInteractionOntology - ProteinProteinInteractionOntologyResource - ✓
- CellTypeOntology - CellOntologyResource - ✓
- MultipleAlignmentOntology - MultipleAlignmentOntologyResource - ✓
- ProteinModificationOntology - ProteinModificationOntologyResource - ✓
- FlyBaseControlledVocabulary - FlyBaseControlledVocabularyResource - ✓
- MammalianPhenotypeOntology - MammalianPhenotypeOntologyResource - ✓
- MeSH - MeSHResource - ✓
- SequenceOntology - SequenceOntologyResource - ✓
- MobyObjectsOntology - MobyObjectsOntologyResource - ✓





# AUGUSTUS

The screenshot shows the AUGUSTUS web interface. The top navigation bar includes links for Welcome, Grid, Bioinformatics, Genetic Tools, Ontology Access, MedIGRID Database Access, D-GRDL, D-GRDL Admin, Monitoring, and MedIGRID Imaging Applications. The main content area displays the executed command and its output.

```

Executed command:
augustus --species=human --strand=both --singlestrand=true --genemodel=partial --codingseq=on --alternatives=false sequence.file

Result file:
Download Result file

Transformed result files:
Graph of the result: png-File
PDF of the result: pdf-File
Postscript of the result: ps-File

Output:
# This output was generated with AUGUSTUS (version 1.8.2).
# AUGUSTUS is a gene prediction tool for eukaryotes written by Mario Stanke (mstanke@gmdg.de).
# Please cite: Mario Stanke and Stephan Waack (2003) "Gene prediction with a hidden Markov model and
# a new intron submodel", Bioinformatics, Vol. 19 Suppl. 2, ii215-ii225
# No extrinsic information on sequences given.
# Initialising the parameters ...
# Looks like sequence.file is in fasta format.
# We have hints for 0 sequence and for 0 of the sequences in the input set.
# ----- prediction on sequence number 1 (length = 6483, name = HSACKI10) -----
# Constraints/Hints:
# (none)
# Predicted genes for sequence number 1 on both strands
# (none)
### gene gl
HSACKI10 AUGUSTUS 1674 5499 1 + . gl
HSACKI10 AUGUSTUS 1674 5499 0.35 + . gl.tl
HSACKI10 AUGUSTUS 1674 1676 . + 0 transcript_id "gl.tl"; gene_id "gl";
HSACKI10 AUGUSTUS 2300 0.55 + 0 transcript_id "gl.tl"; gene_id "gl";

```

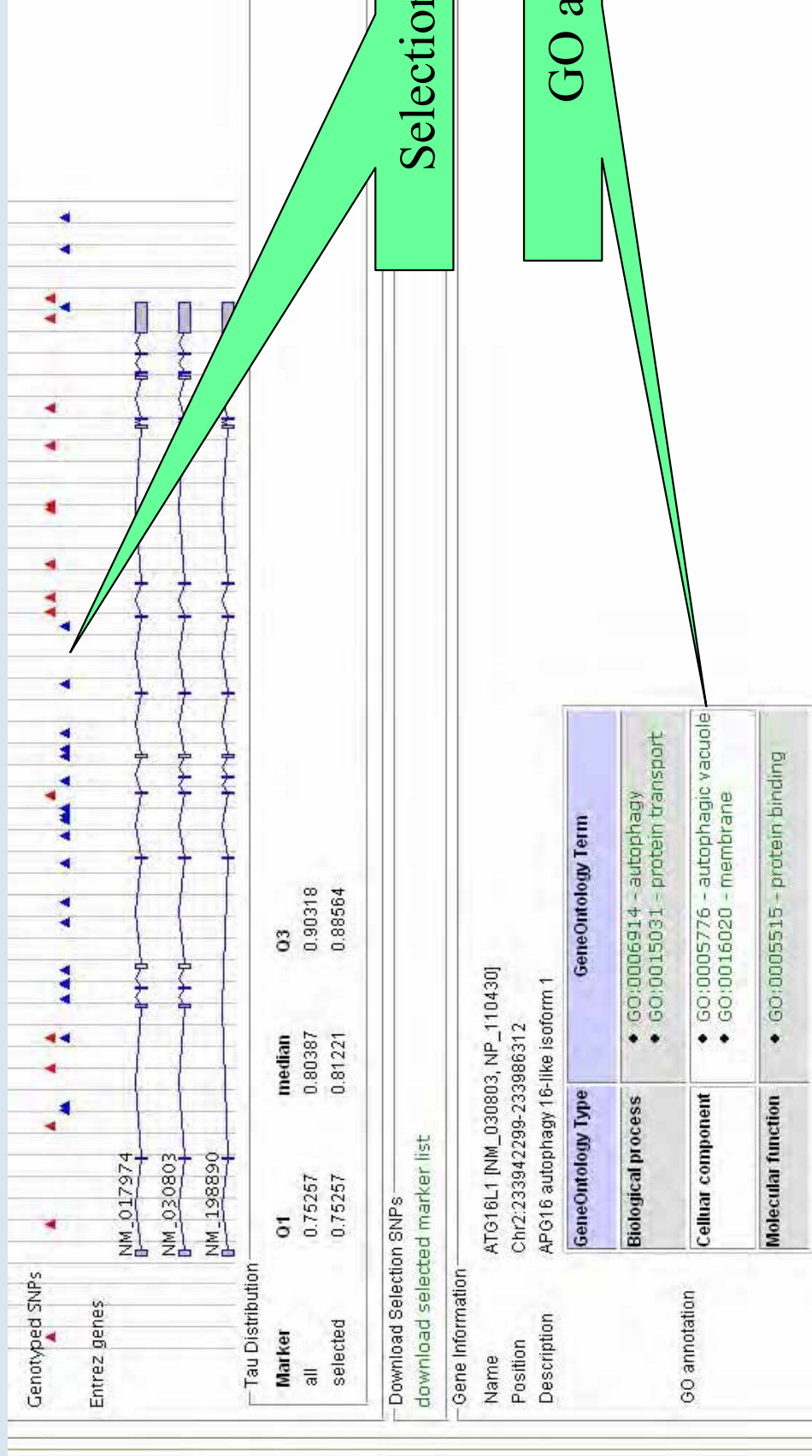
GFF result file

Information about GFF region names

- Linkage between result file and SequenceOntology (SO)



# SNP Selection



- Linkage between input data and GO concepts
- Mapping: Ensembl data source (OGSA-DAI service)



## Summary and Future

- **Service-based middleware to integrate ontologies in grids**
  - Based on grid standards (GT4, OGSA-DAI)
  - Distribution of ontologies, simple and uniform access to ontologies in grids
  - Integration in the central portal of MediGRID
  - Usage in different MediGRID applications
- **Future topics**
  - Integration of further ontologies for upcoming MediGRID applications
  - Matching of ontologies
  - Wiki-like system for collaborative editing and development of domain-specific ontologies (e.g. D-Grid initiative)