



UNIVERSITÄT LEIPZIG

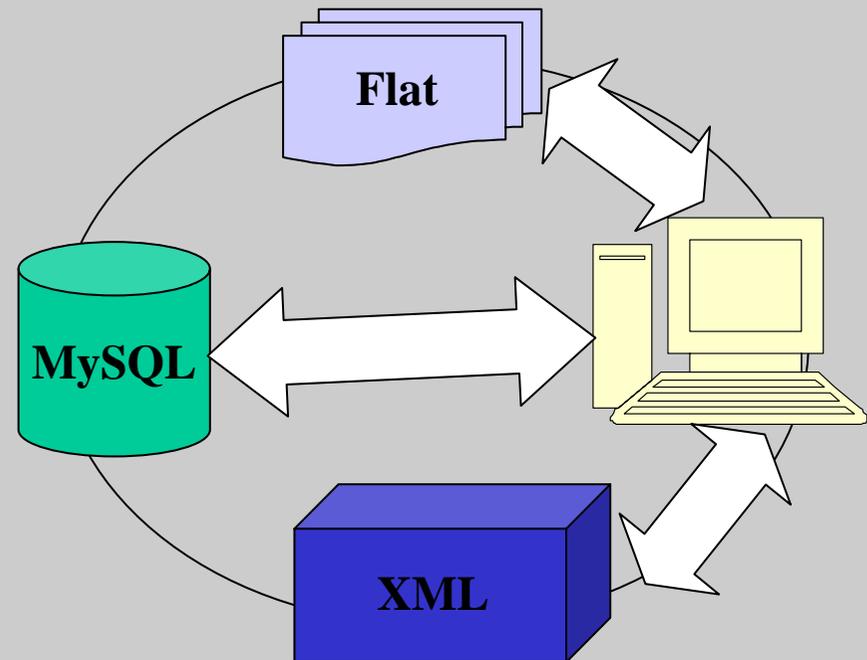
Nutzung und Integration von Ontologien in MediGRID

Results for "blood" in "GeneOntology":

blood coagulation
blood vessel development
patterning of blood vessels
blood vessel maturation
blood vessel remodeling

ID: GO:0007596
Name: blood coagulation
Definition: 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 prothrombin converting principle; stage 2, the formation of thrombin; stage 3, the formation of stable fibrin polymers.
Comment:
Synonyms: blood clotting, blood coagulation factor activity
Hierarchy:
Parents: GO:0007599: hemostasis (IS_A), GO:0050817: coagulation (IS_A), GO:0042050: wound healing (PART_OF)
Children: GO:0007597: blood coagulation, intrinsic pathway (IS_A), GO:0007598: blood coagulation, extrinsic pathway (IS_A), GO:0030160: platelet activation (PART_OF), GO:0030192: regulation of blood coagulation (PART_OF)

Hartung, Michael
IZBI Leipzig
Göttingen, März 2007



Überblick

- **Einführung**
 - Ausgangssituation in MediGRID
 - Ziele
- **OGSA-DAI als Basis**
 - Überblick OGSA-DAI
 - Erweiterungsmöglichkeiten
- **Ontologiezugriff in MediGRID**
 - Architektur, Szenario
 - Integrierte Ontologien
- **Anwendungen in MediGRID**
 - LookUp-Service
 - Bioinformatik-Anwendungen
- **Ausblick**
 - Erweiterungen, Semantisches Wiki



MediGRID

Ausgangssituation (I)

Bedarf an Ontologien ist vorhanden, abhängig von der jeweiligen Applikation

MediGRID-Anwendungen:

- Bioinformatik
- Bildverarbeitung
- Klinische Forschung

Verknüpfung mit verfügbaren Ontologien:

- Einbringung von zusätzlichem Wissen
- Nutzung von Annotationsdatenbanken
- Vorerst keine Erstellung neuer Ontologien (Aufwand)

Ausgangssituation (II)

Heterogenität zwischen den verschiedenen Ontologien

Unterschiedliche Quelltypen:

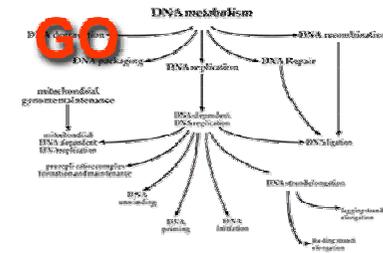
- Relationale Datenbanken: GeneOntology (GO)
- Flat files: NCI-Thesaurus
- OBO-Format
- XML und weitere

Unterschiedliche Formate:

- Modellierung von Relationen (is-a, part-of)
- Repräsentation von Synonymen

Grundlegende Information:

- Terme, Konzepte mit Name, Definition, Synonyme
- Relationen zwischen den Konzepten einer Ontologie
- Cross references



Ziele

Einfaches, transparentes Zugriffssystem für Ontologien in Grids

- **Servicebasierter Zugriff auf Ontologien**
 - ‘Ontology services’ als ‘Middleware’ zw. Anwendungen und Ontologien
 - Support für Clients (z.B. andere Services, MediGRID-App.), einfache Zugriffsmöglichkeiten
 - Verteilung der Ontologien: Ausfallsicherheit, Lastbalancierung, Ausnutzung des Grids
 - Erweiterbarkeit
- **Funktionalitäten**
 - Suche nach Termen/Konzepten in Ontologien
 - Zugriff auf Ontologieinformationen: Definitionen, Synonyme, Relationen, cross references
 - Navigation, graphische Präsentation der Ontologien

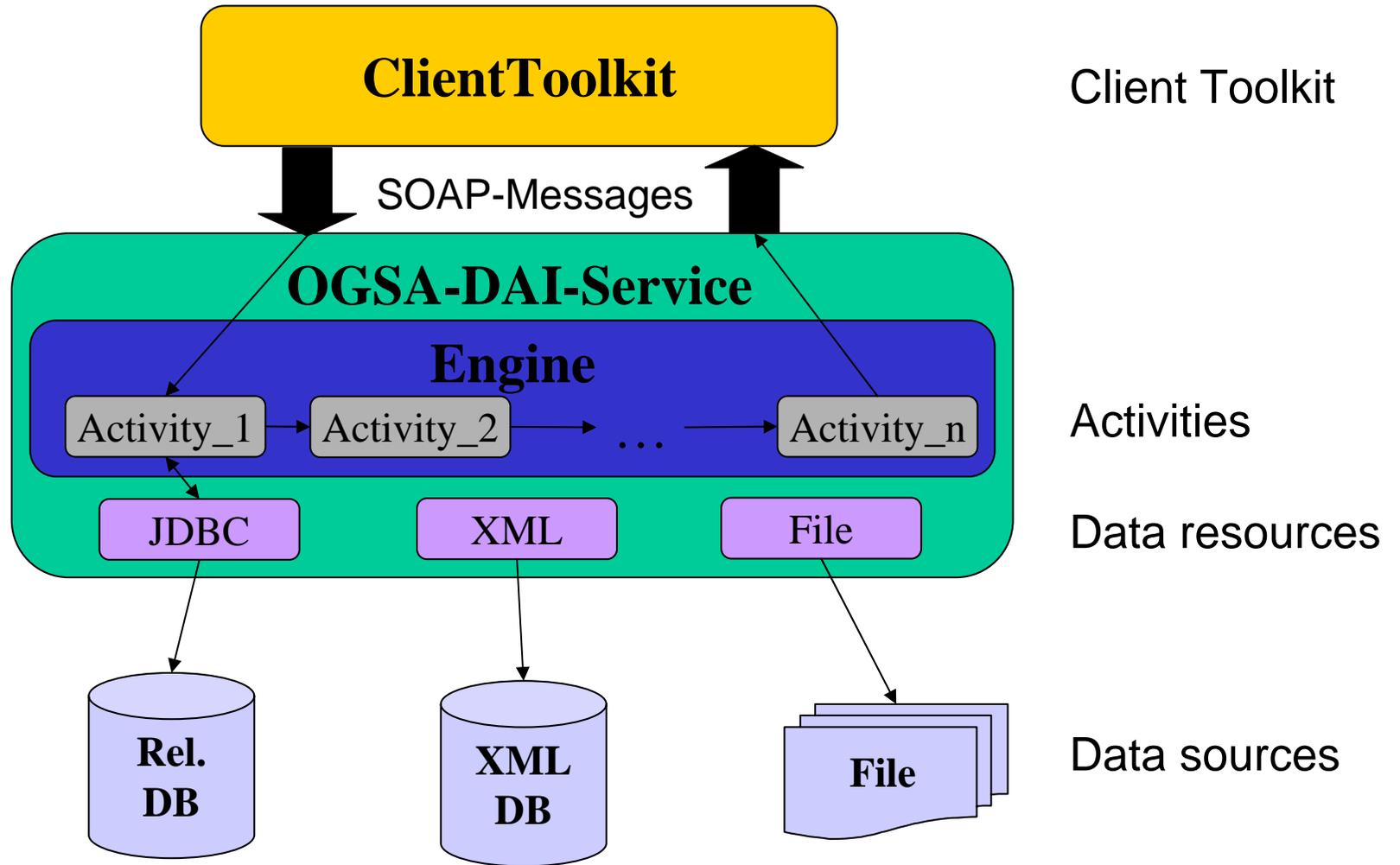


OGSA-DAI

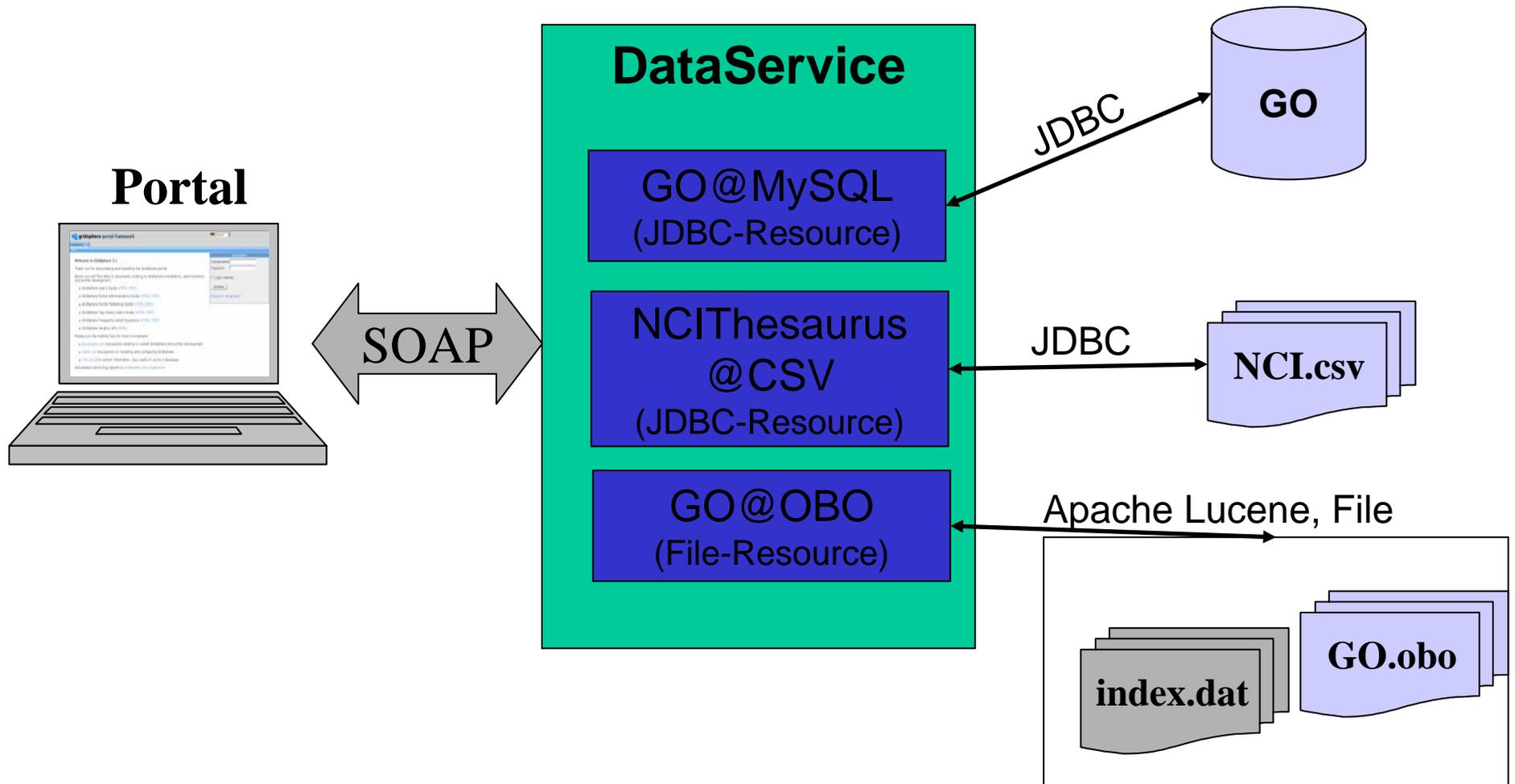
Zugriff auf Datenquellen in Grids

- **Erweiterbares Framework für “data access and integration” in Grid Systemen**
- Zugriff auf verteilte and heterogene Datenquellen im Grid
- Web Service-basiert (WSRF, Globus Toolkit 4)
- Erweiterbarkeit als wichtiges Merkmal
- Client Toolkit: API für Serviceinteraktion
- **Pro:**
- Einfacher Zugriff auf Datenbanken, XML und Flat files
- Quasi-Standard in Grid Systemen
- **Contra:**
- Performance Aspekte: WebRowSet (Web Service) → für Ontologien weniger relevant (kleinere Datenmengen)
- Keine echte Transparenz, Nutzer benötigt Kenntnis über Schema und Typ der Quelle

OGSA-DAI Architektur



Erster Schritt: Direkter Zugriff auf Ontologien

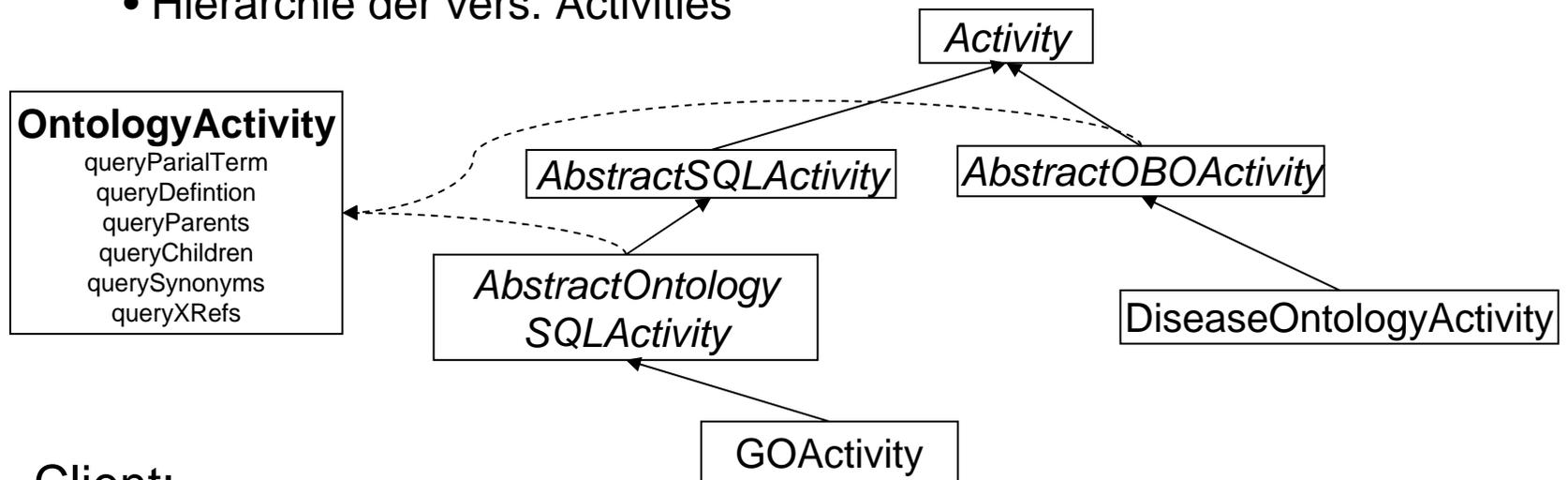


Zweiter Schritt: Erweiterung von OGSA-DAI

Ontology Services als Erweiterung der OGSA-DAI Data Services

Activities:

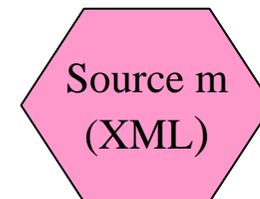
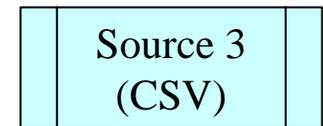
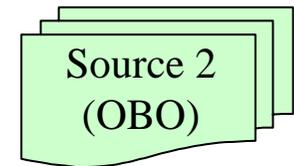
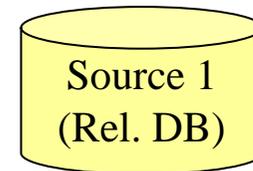
- Spezielle “ontology activities” als Wrapper für Ontologien
- Hierarchie der vers. Activities



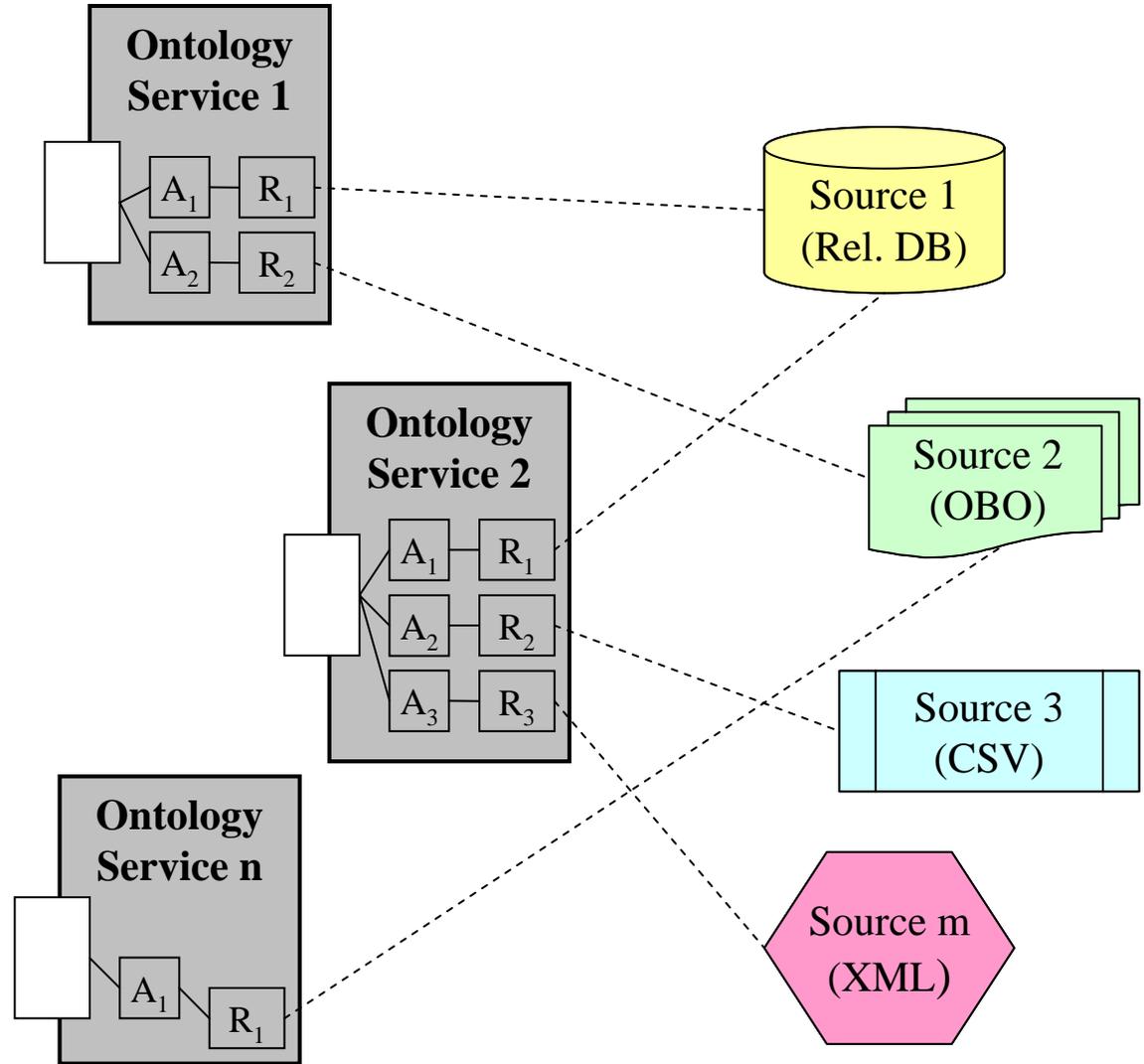
Client:

- Transparenter Zugriff auf Ontologien über Services
- Interface für das Erstellen und Absenden von Requests
- Handling der “SOAP messages” im Hintergrund

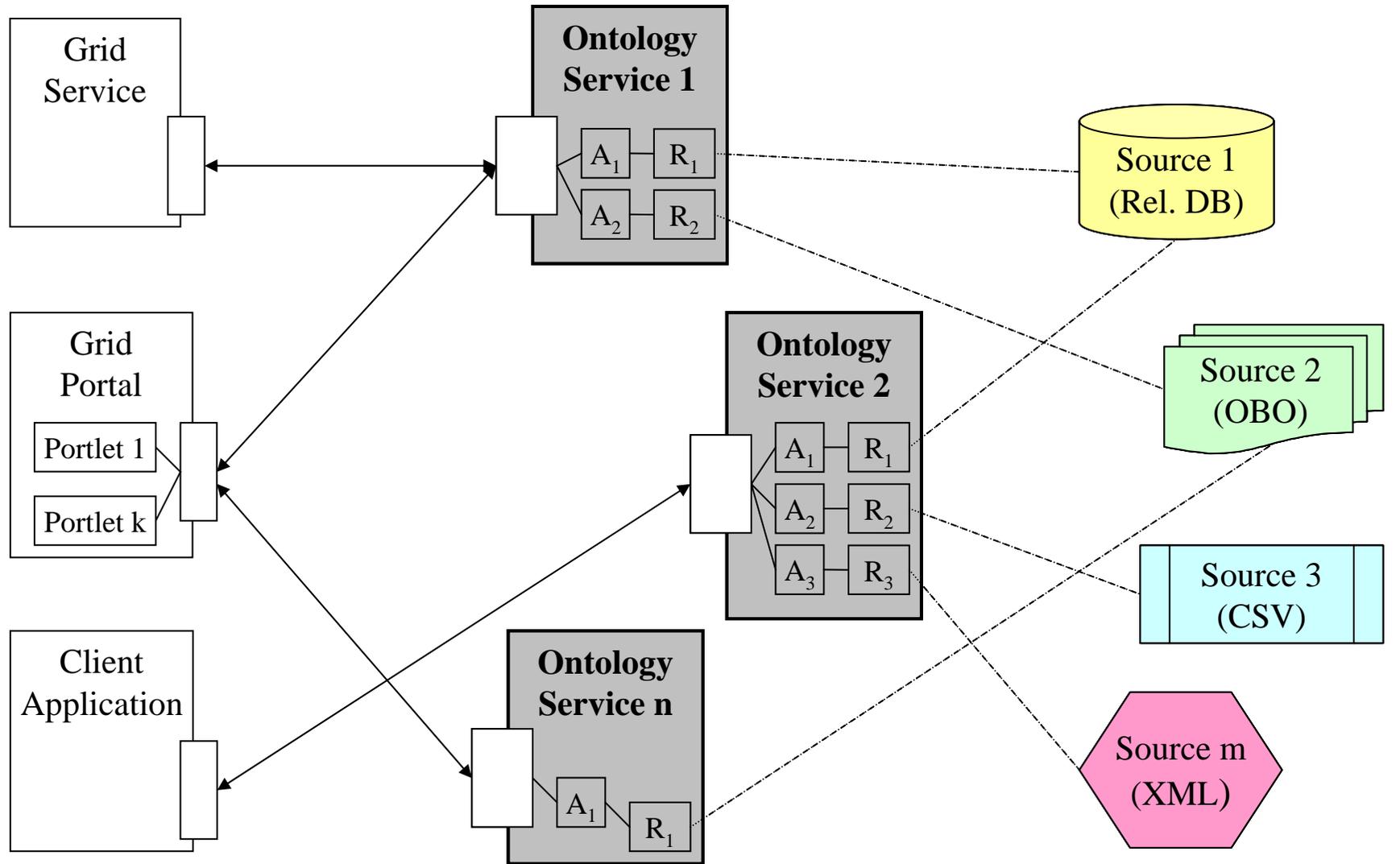
Architektur



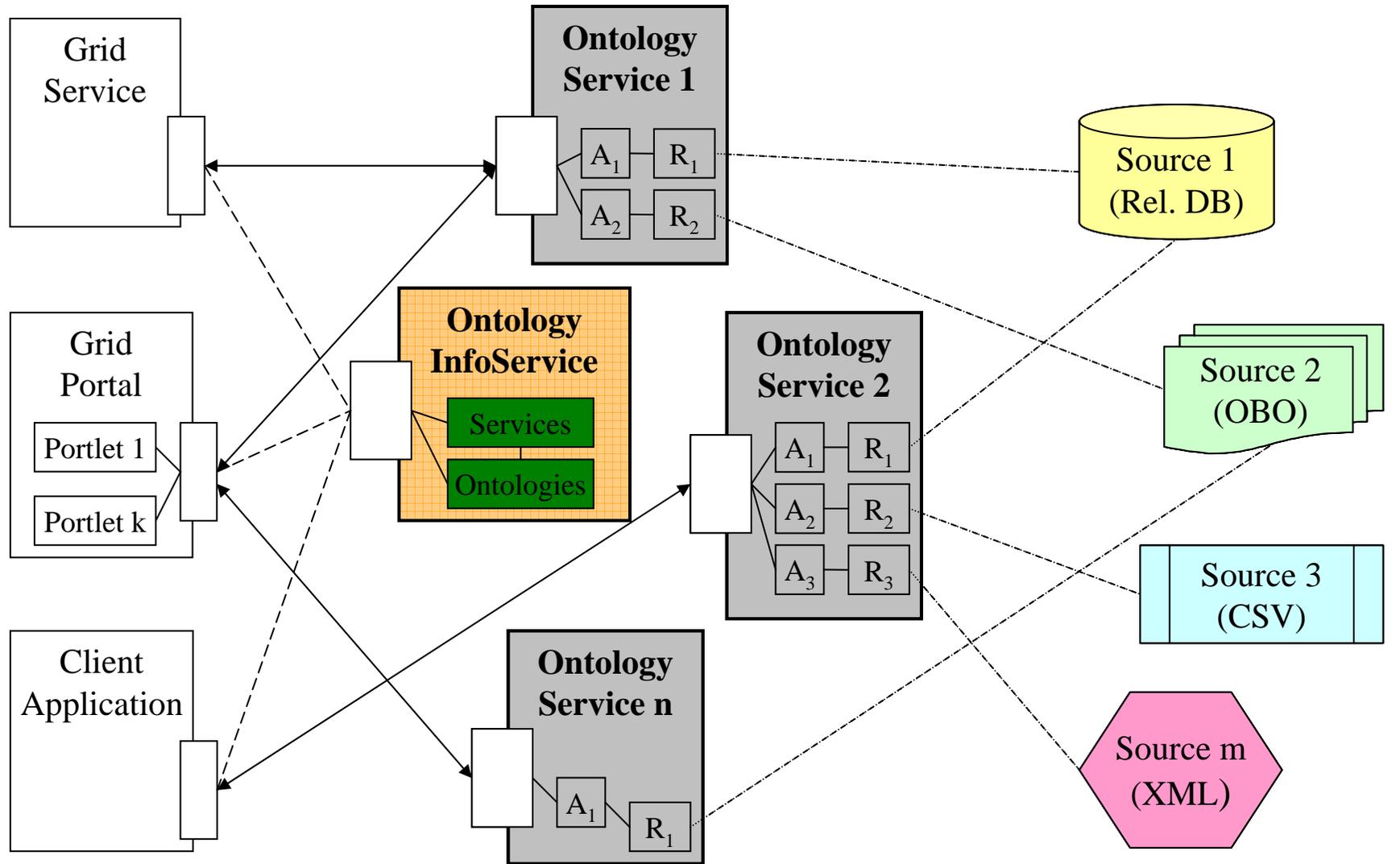
Architektur



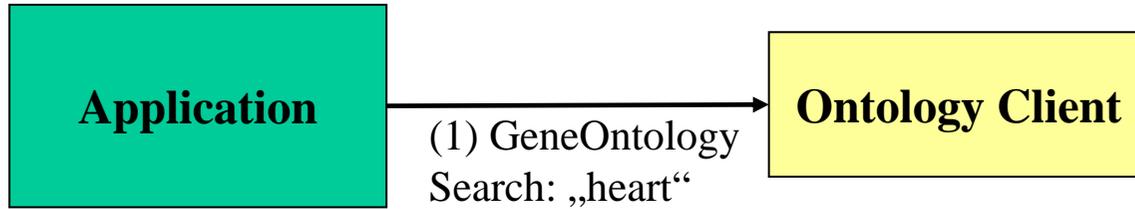
Architektur



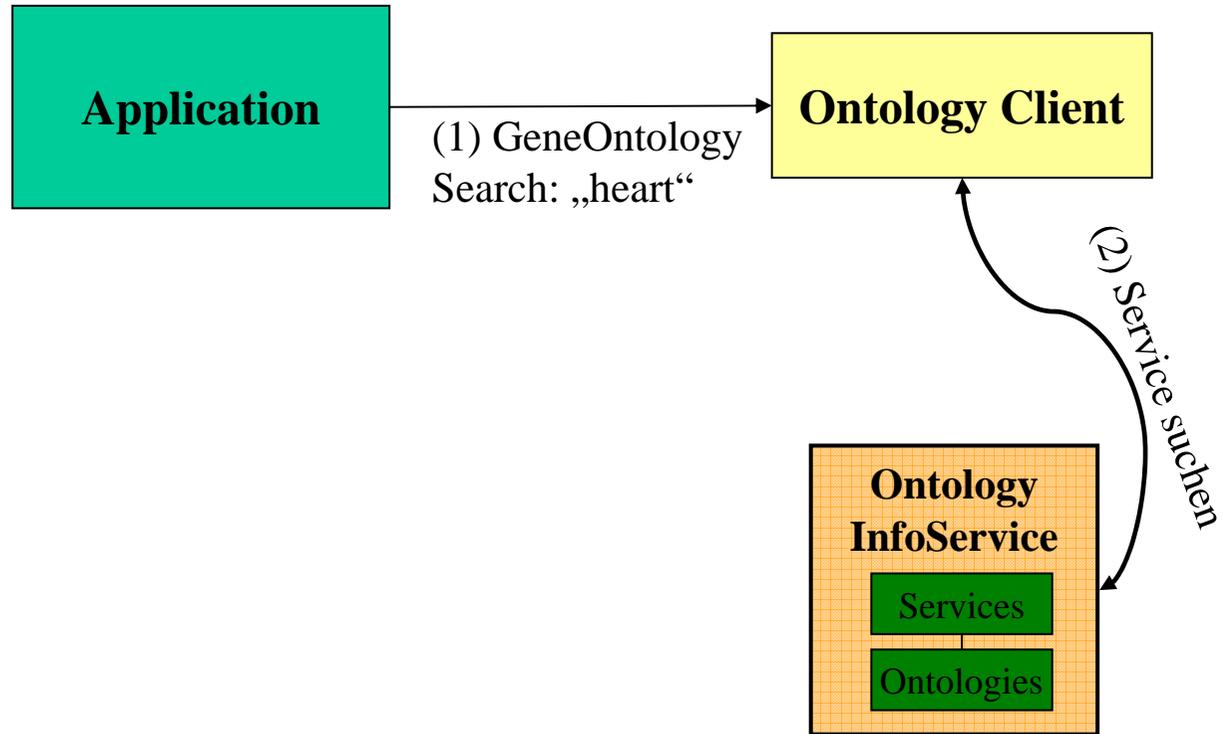
Architektur



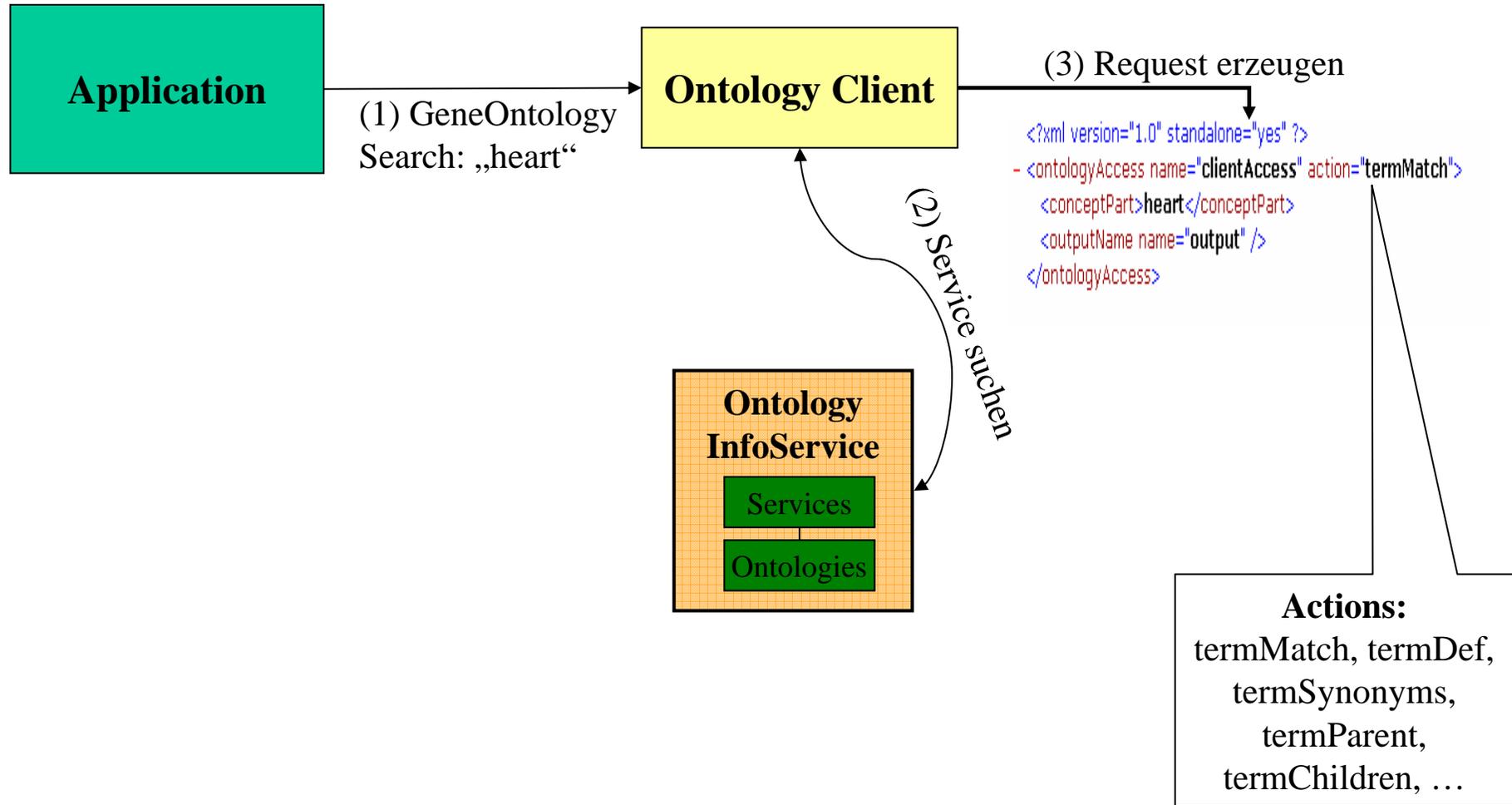
Ontologiezugriff Szenario



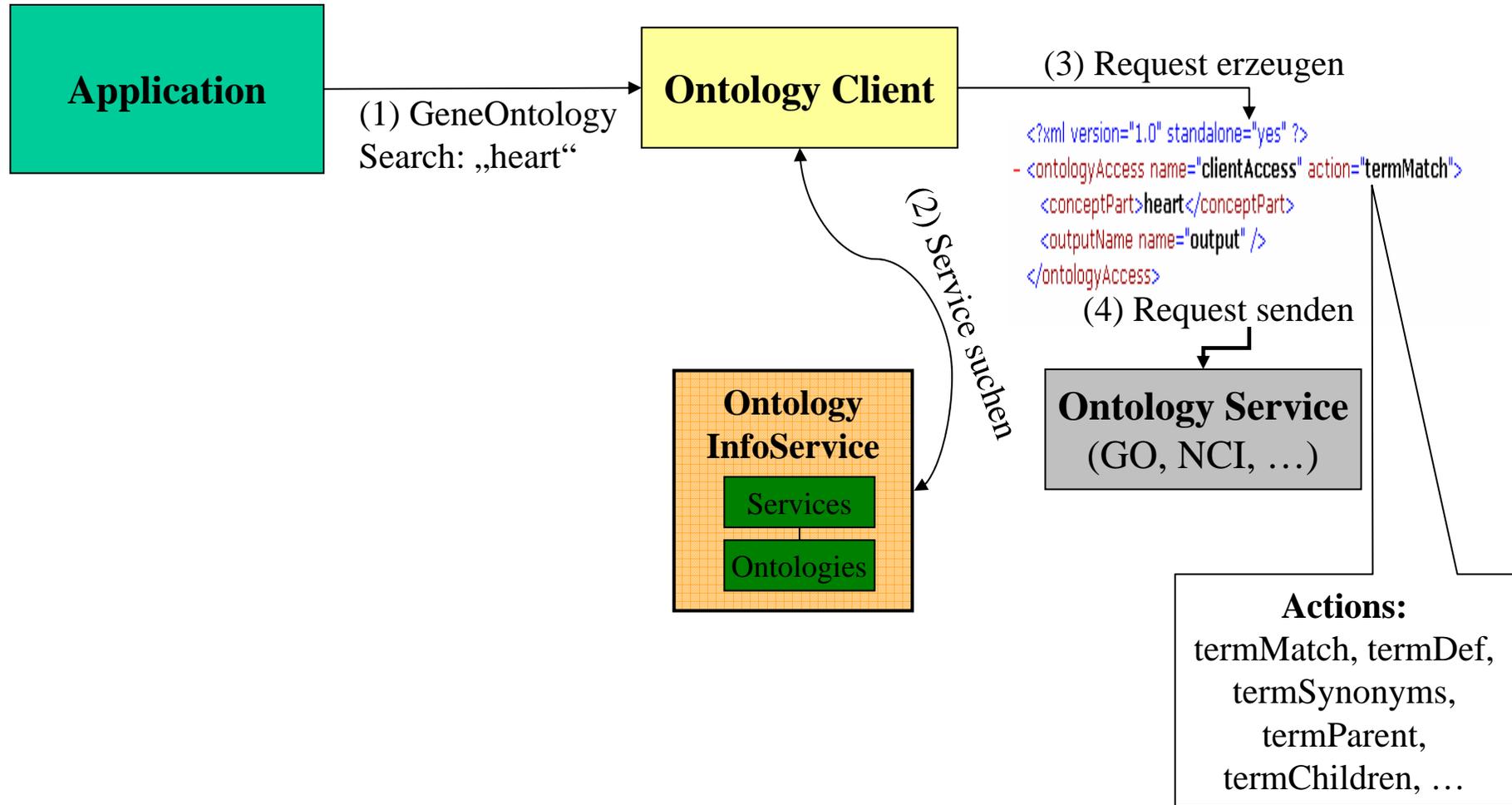
Ontologiezugriff Szenario



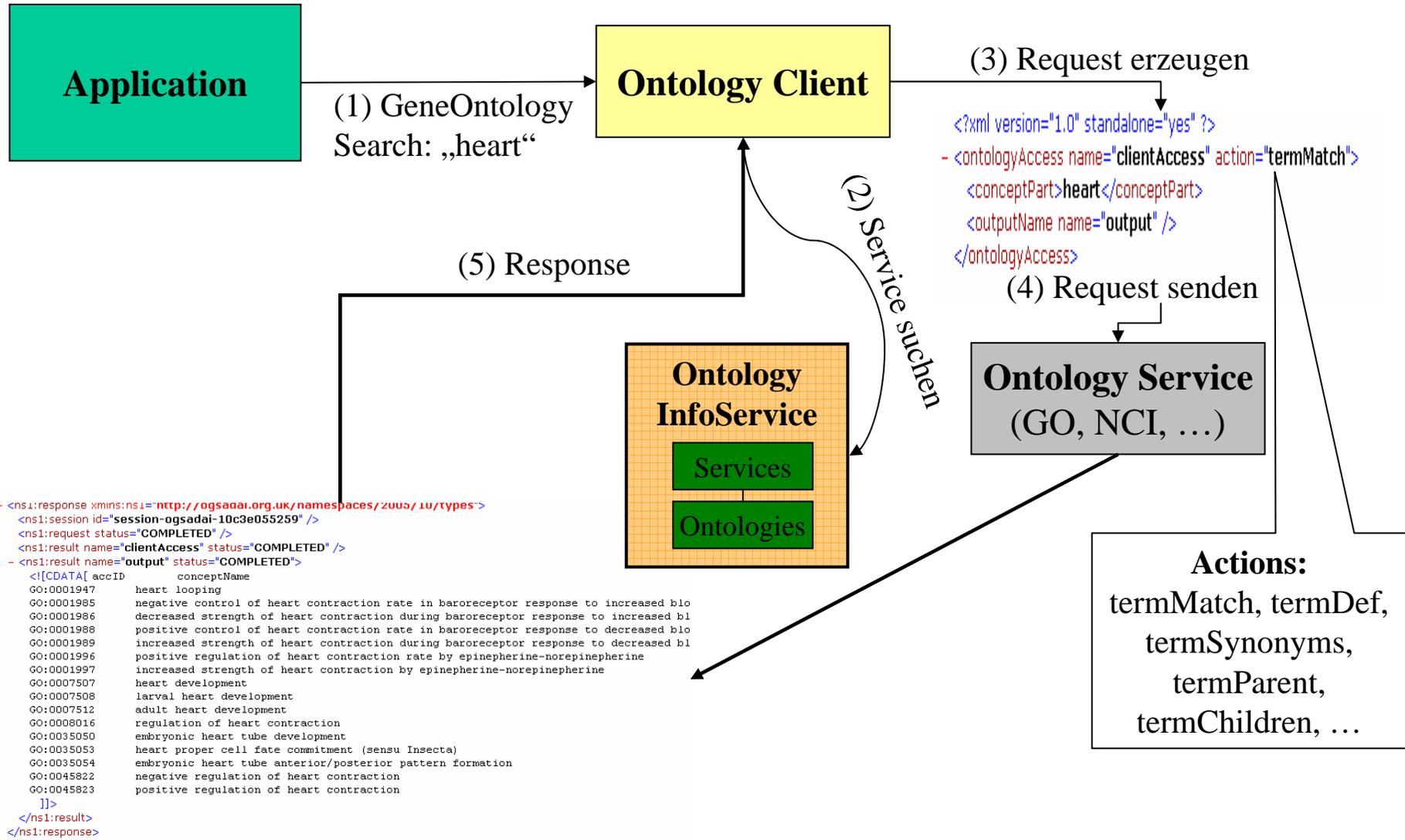
Ontologiezugriff Szenario



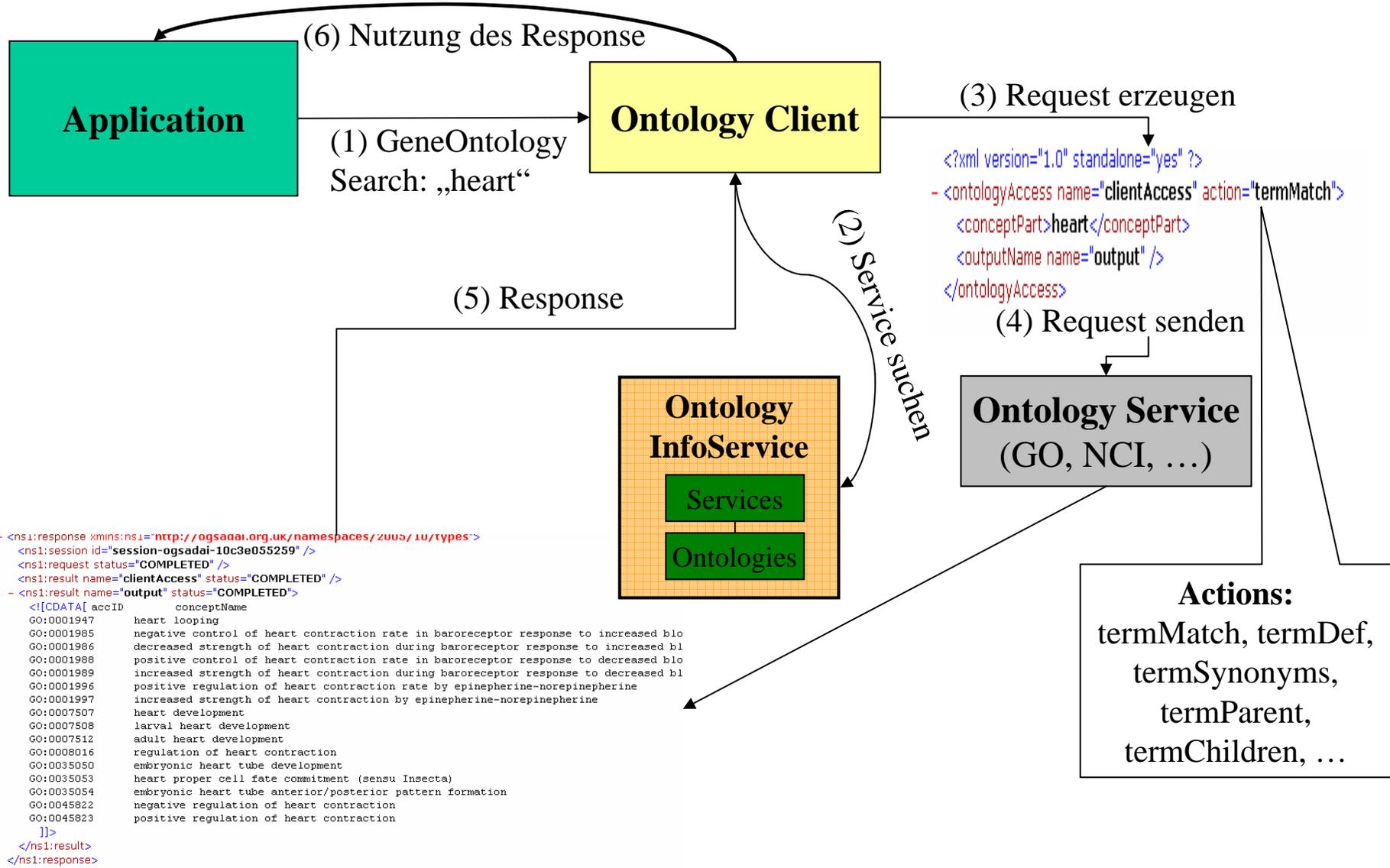
Ontologiezugriff Szenario



Ontologiezugriff Szenario



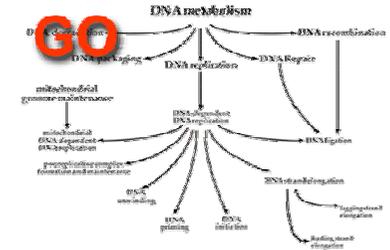
Ontologiezugriff Szenario



Integrierte Biomedizinische Ontologien

GeneOntology:

- Zur Beschreibung von Genen und Genprodukten in verschiedenen Organismen
- 3 Sub-Ontologien: molecular function, cellular component, biological process



NCI-Thesaurus:

- Terminologie genutzt und erzeugt am NCI (Krebsforschung)
- 20 Hierarchien: Medikamente, Organismen, Gene, Standards, ...



National Cancer Institute

OBO-Ontologien:

- Strukturiert kontrollierte Vokabulare
- Open-Source, biomedizinische Domäne
- “attribute – value flat files”, OBO-Format Syntax
- GO, DiseaseOntology, ProteinModification, ProteinInteraction, SequenceOntology, ...



Applikationen im MediGRID-Portal

Probleme von Grid-Applikationen in der Medizin

- Schwierige Interaktion: Syntax, Schnittstellen
- Nutzerunfreundliche Interaktion mit Gridressourcen
- Z.B. Web Service-Interfaces, Kommandozeilentools

Zentrales Portal als Einstieg

- Integration der Gridressourcen, Applikationen in das Portal
- Zentraler Zugriff ohne großes Hintergrundwissen (Transparenz)
- Portlets als “re-useable” Softwarekomponenten

Ontologien im MediGRID-Portal

- LookUp-Service über alle verfügbaren Ontologien
- AUGUSTUS-Anwendung (Göttingen)
- SNPSelection-Anwendung (Kiel)

LookUp-Service (I)

UNIVERSITÄT LEIPZIG

Zentrale Ontologie-
Anwendungen

The screenshot displays the web interface of the LookUp-Service. At the top, there is a navigation menu with items: Welcome, Grid, Bioinformatics, Genetic Tools, **Ontology Access**, MediGRID Database Access, D-GRDL, D-GRDL Admin, Monitoring, and MediGRID Imaging Applications. Below this is a secondary menu with MultiOntologyAccess, GridWiki, Gene Ontology, and NCI Thesaurus. The main content area is titled "MultiOntologyAccessPortlet" and contains a search interface. The search interface has tabs for "Search", "Show Results", "OntologyInfo", and "ResourceInfo". Under the "Search" tab, there is a section titled "Available ontologies:" with a list box containing: GeneOntology, NCI-Thesaurus, HumanDiseaseOntology, ProteinProteinInteractionOntology, and CellTypeOntology. To the right of the list box is a "Search" button. Below the list box is a "Refresh" button. The interface also includes a "Logout" link and a welcome message "Welcome, Michael Hartung" in the top right corner.

Auswahl von
Ontologien für Suche

LookUp-Service (II)

Suchergebnisse

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

ID	GO:0008016
Name	regulation of heart contraction
Definition	Any process that modulates the frequency, rate or extent of heart contraction.
Comment	
Synonyms	♦ regulation of heart rate
Hierarchy	<pre>graph TD; A[GO:0051239 regulation of organismal physiological process] --> B[GO:0008016 regulation of heart contraction]; B --> C[GO:0002026 cardiac inotropy]; B --> D[GO:0002027 cardiac chronotropy]; B --> E[GO:0045822 negative regulation of heart contraction]; B --> F[GO:0045823 positive regulation of heart contraction];</pre>
Parents	♦ GO:0051239: regulation of organismal physiological process (IS_A)
Children	♦ GO:0002026: cardiac inotropy (IS_A) ♦ GO:0002027: cardiac chronotropy (IS_A) ♦ GO:0045822: negative regulation of heart contraction (IS_A) ♦ GO:0045823: positive regulation of heart contraction (IS_A)
X-References	♦ GOC:go_curators ♦ TIGR_TIGRFAMS:TIGR01294 phospholamban

Informationen zu ausgewählten Konzepten



LookUp-Service (III)

Status der Ontologieressourcen

Host: <http://buell.izbi.uni-leipzig.de:8081>

Service: </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 - ✓

Standorte

AUGUSTUS

UNIVERSITÄT LEIPZIG

The screenshot shows the Augustus web interface. At the top, there is a navigation menu with options like 'Welcome', 'Grid', 'Bioinformatics', 'Genetic Tools', 'Ontology Access', 'MediGRID Database Access', 'D-GRDL', 'D-GRDL Admin', 'Monitoring', and 'MediGRID Imaging Applications'. Below the menu, the 'Info / Help' section is active, displaying the executed command: `augustus --species=human --strand=both --singlestrand=true --genemodel=partial --codingseq=on --alternatives=false sequence.file`. It also provides links for downloading the result file and transformed result files (Graph, PDF, Postscript). The 'Output' section shows the following text:

```
# This output was generated with AUGUSTUS (version 1.8.2).
# AUGUSTUS is a gene prediction tool for eukaryotes written by Mario Stanke (mstanke@gwdg.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 gene 1674 5499 1 + . gl
HSACKI10 AUGUSTUS transcript 1674 5499 0.35 + . gl.tl
HSACKI10 AUGUSTUS start_codon 1674 1676 . + 0 transcript_id "gl.tl"; gene_id "gl";
HSACKI10 AUGUSTUS initial 1674 2300 0.55 + 0 transcript_id "gl.tl"; gene_id "gl";
```

Ergebnis-Protokoll

Zusatzinformation über Terme aus Ontologie (SO)

Genvorhersage-Tool

- Input: DNA-Sequenzen, Constraints (Hints), ...
- Output: GFF-File mit Inf. über die Genstruktur (Annotationen aus SO)



SNPSelection

UNIVERSITÄT LEIPZIG

Define Parameter:

SNP Amount: 50

Population: CEU

Vendor: HapMapRel20

Region: (for Example: Chr2:233942... 992987000 or ATG16L1 or NM_017974)

Run SNPSelection Example

March 23, 2007

Parameter-Eingabe

SNP-Selection:

- Auswahl optimaler SNP-Daten aus öffentlichen Datenquellen (Genetic markers)
- Input: Population, Region (direkt, indirekt: Gename, RefSeq-ID)
- Ontologie-Kopplung: Mapping zwischen RefSeq-ID und GeneOntology-Termen

SNPSelection (II)

Mapping zwischen Eingabe und GO:

- Ausnutzung der Bioinformatik-DB „Ensembl“
- Gen- und Proteininformationen für vers. Species
- Zahlreiche Annotationen, Links: RefSeq, GO, Interpro, PubMed, ...
- Erstellung des Mappings: RefSeq – GO über OGSA-DAI-Ressource

Genotyped SNPs

Entrez genes

NM_017974

NM_030803

NM_198890

Tau Distribution

Marker	Q1	median	Q3
all	0.75257	0.80387	0.90318
selected	0.75257	0.81221	0.88564

Download Selection SNPs

[download selected marker list](#)

Gene Information

Name ATG16L1 [NM_030803, NP_110430]

Position Chr2:233942299-233986312

Description APG16 autophagy 16-like isoform 1

GeneOntology Type	GeneOntology Term
Biological process	<ul style="list-style-type: none">♦ GO:0006914 - autophagy♦ GO:0015031 - protein transport
Cellular component	<ul style="list-style-type: none">♦ GO:0005776 - autophagic vacuole♦ GO:0016020 - membrane
Molecular function	<ul style="list-style-type: none">♦ GO:0005515 - protein binding

GO annotation

Marker-Selektion

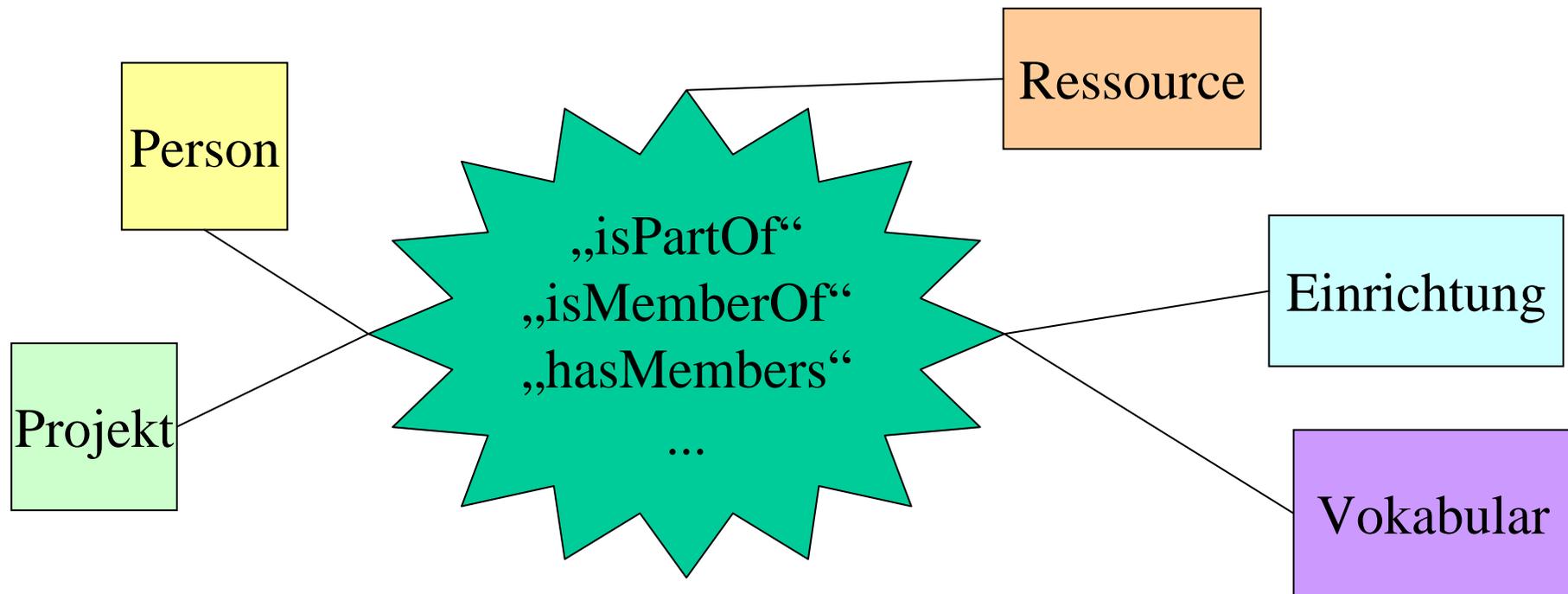
GO-Annotation

Zusammenfassung/Ausblick (I)

- **Servicebasierte Middleware für den Zugriff auf Ontologien**
 - Basierend auf D-Grid Basis Software (GT4, OGSA-DAI)
 - Verteilung von Ontologien, Schaffung eines einfachen Zugangs zu Ontologien
 - Integration ins zentrale MediGRID-Portal
 - Nutzung in diversen Anwendungen
- **Zukünftige Aufgaben**
 - Security-Aspekte: Delegation von Credentials
 - Verbesserungen an Interfaces (GUI)
 - Ontology-Matching
 - Unterstützung weiterer MediGRID-Anwendungen

Zusammenfassung/Ausblick (II)

- **Semantisches Wiki für die Darstellung der D-Grid-Initiative**



- Welche Personen ist im Projekt 'X' involviert?
- Welche Subprojekte besitzt Projekt 'X'? Wer ist mit welchen Begriffen vertraut?
- Erleichterung des Einstiegs in die D-Grid-Initiative: z.B. neue Projekte aus Call II

***Vielen Dank
für Ihre Aufmerksamkeit !***