



A Grid Middleware for Ontology Access

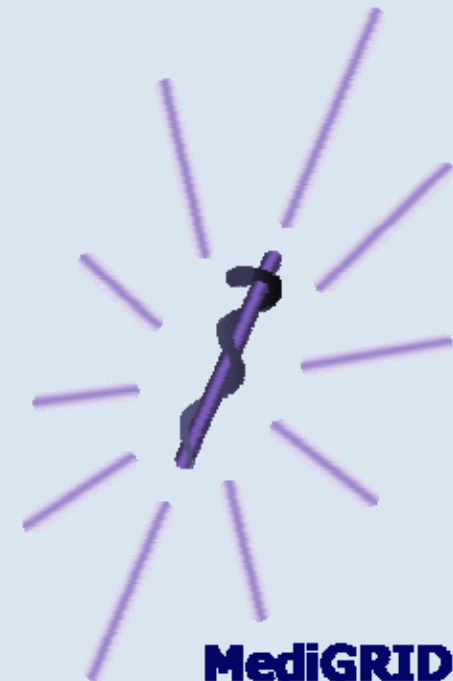
Results for "blood" in "GeneOntology":

blood vessel development	
patterning of blood vessels	
blood vessel maturation	
blood vessel remodeling	
Show more	
ID	GO:0007598
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	
Synonym	<ul style="list-style-type: none"> blood clotting blood coagulation factor activity
Hierarchy	
Parents	<ul style="list-style-type: none"> GO:0007599 hemostasis (IS_A) GO:0007597: coagulation (IS_A) GO:0043000 wound healing (PART_OF)
Children	<ul style="list-style-type: none"> GO:0007597: blood coagulation, intrinsic pathway (IS_A) GO:0007596: blood coagulation, extrinsic pathway (IS_A) GO:0007595: platelet activation (PART_OF) GO:0007595: regulation of blood coagulation (PART_OF)

German e-Science Conference
Baden Baden, May 4, 2007

Michael Hartung, Erhard Rahm

Interdisciplinary Centre for Bioinformatics, Leipzig





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



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



Bundesministerium
für Bildung
und Forschung



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

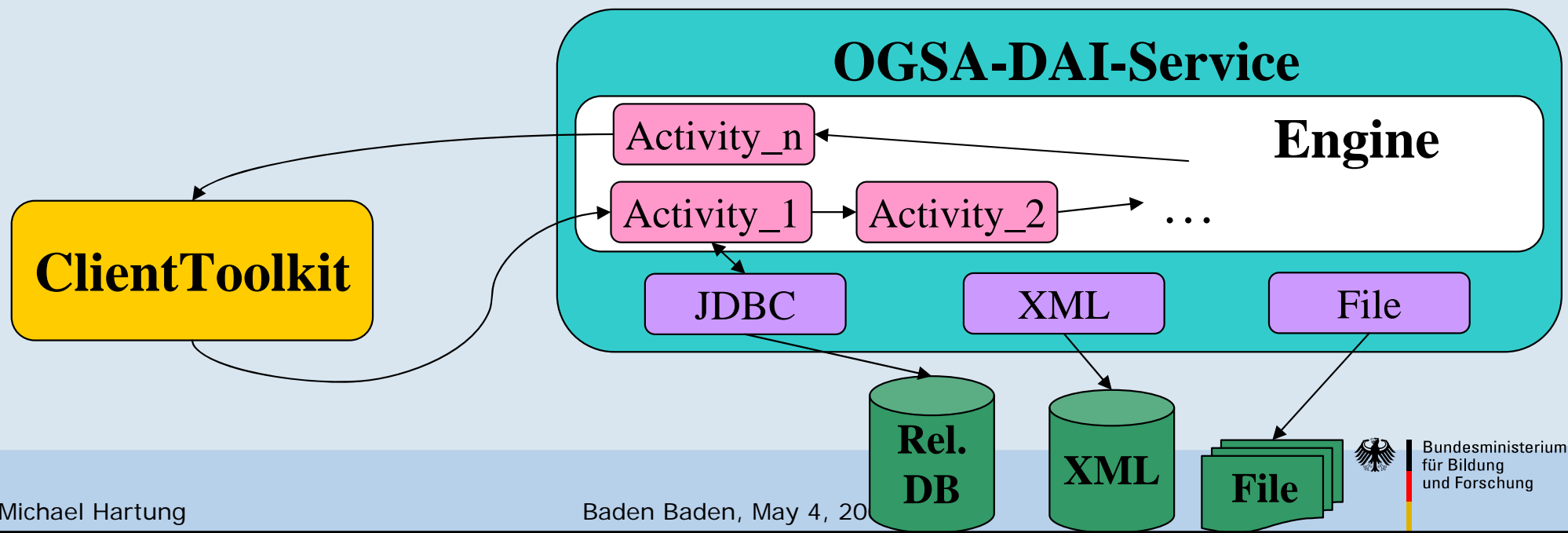


OGSA-DAI as base infrastructure



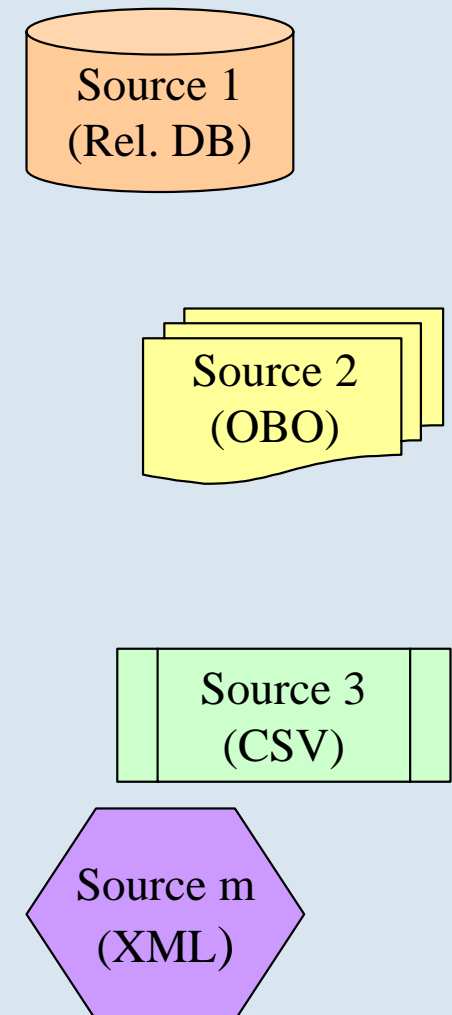
Access to data sources in grids

- **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



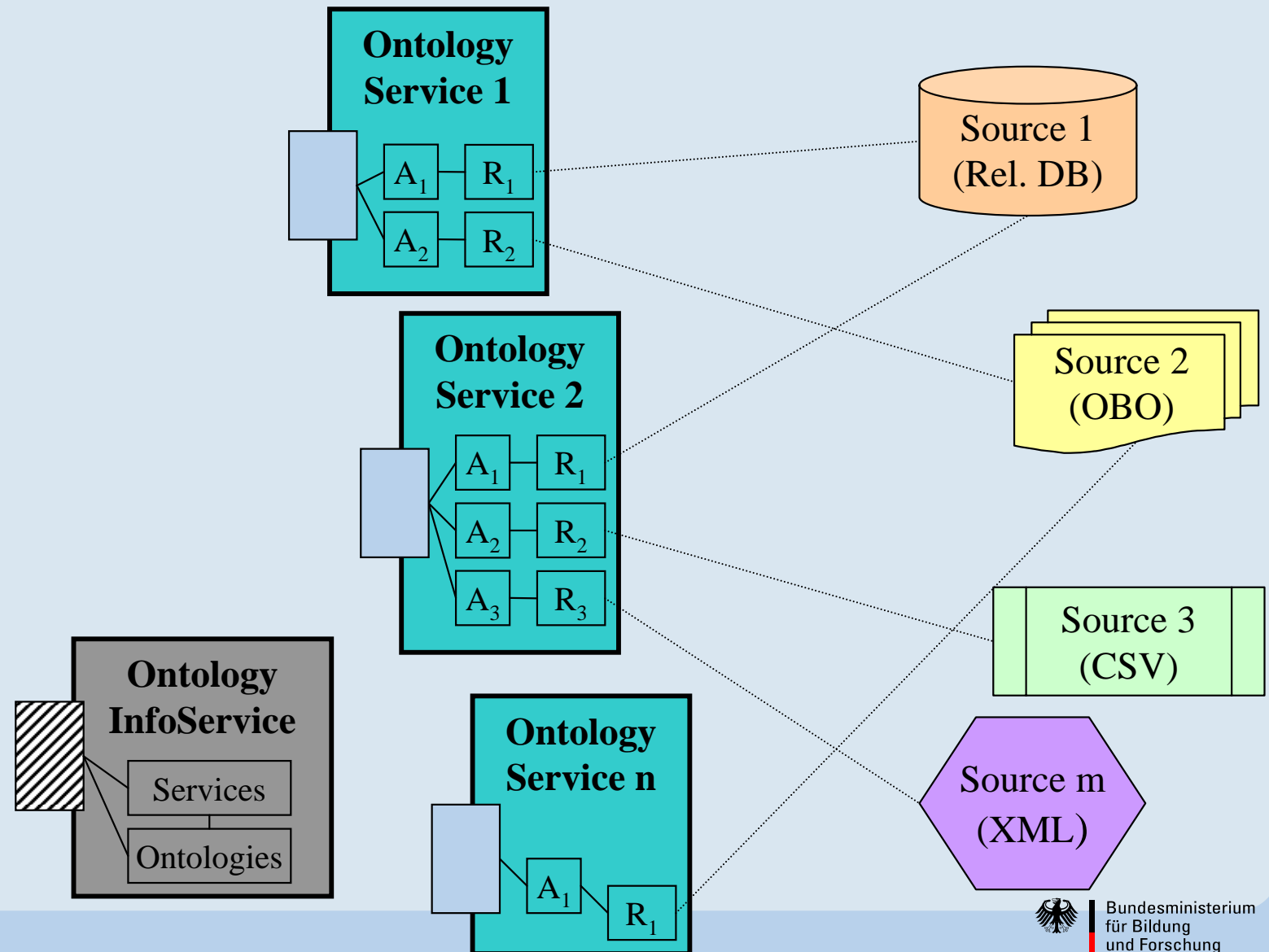


Ontology Access Middleware



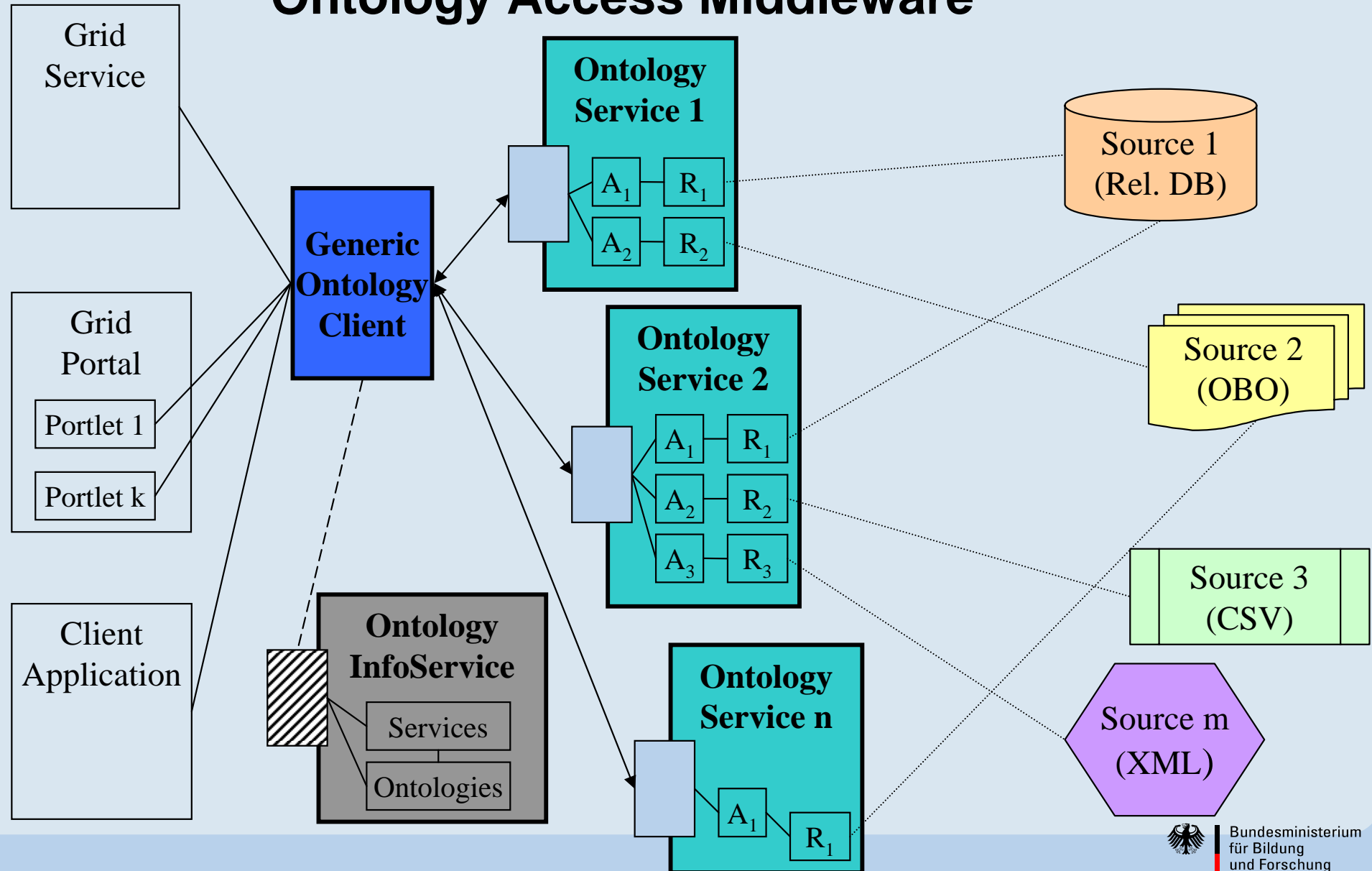


Ontology Access Middleware





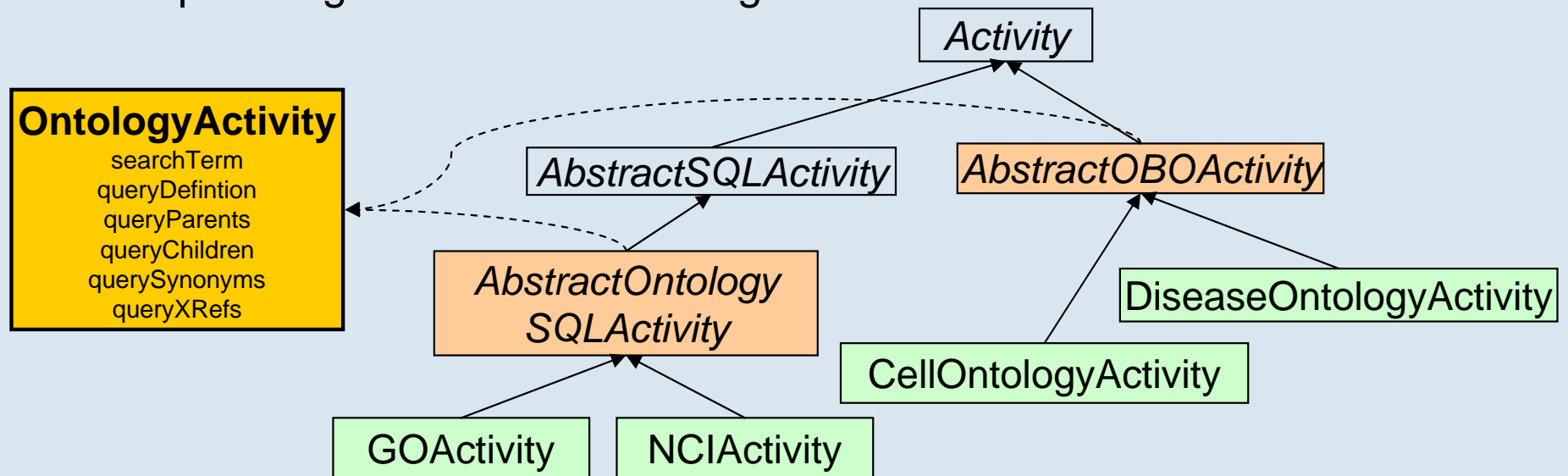
Ontology Access Middleware

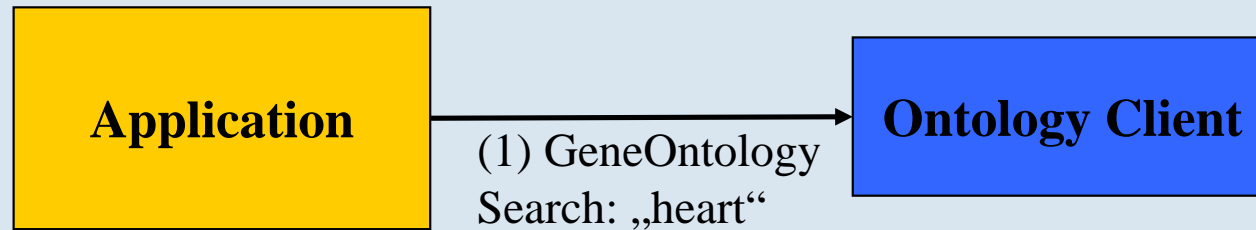


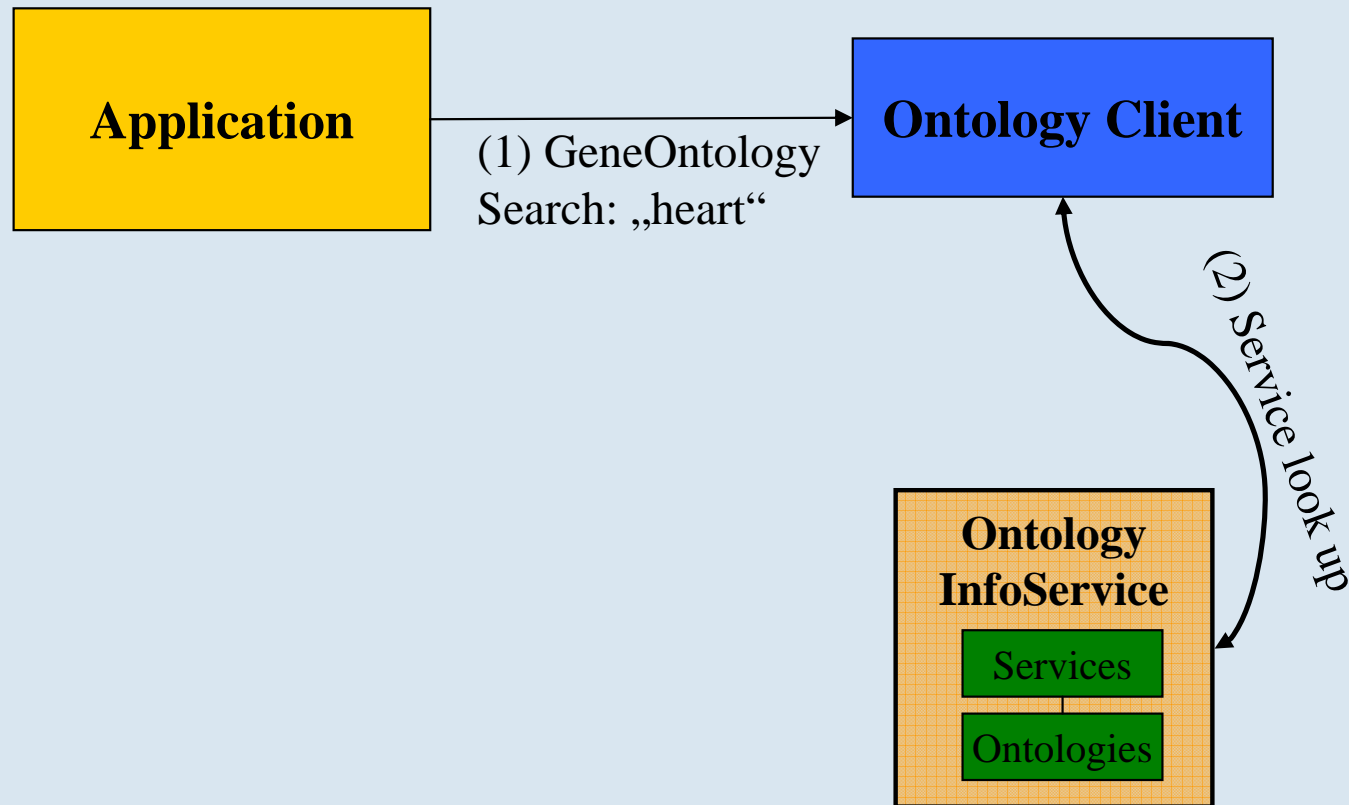
Extensions to OGSA-DAI

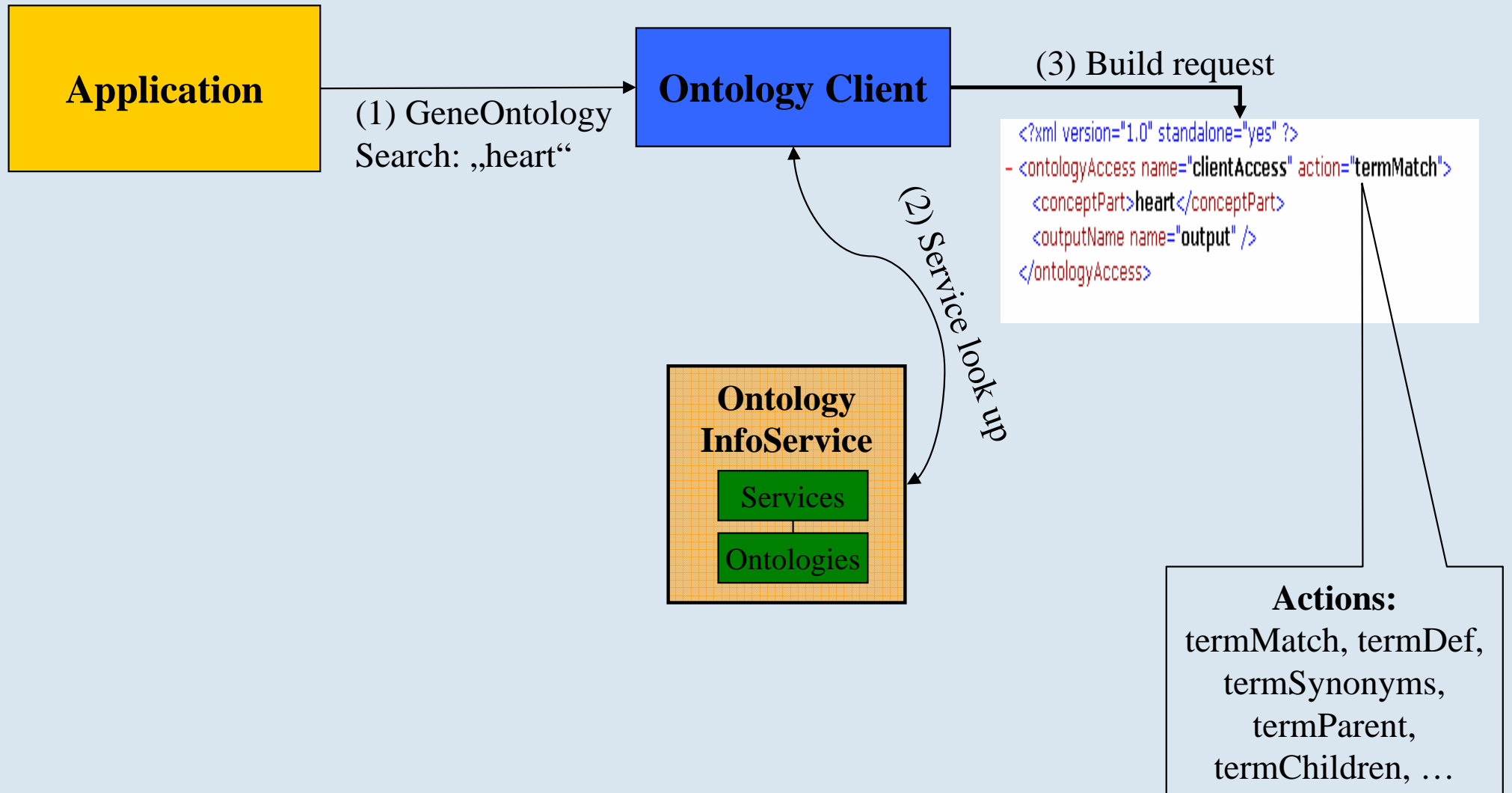
Ontology services as extension of OGSA-DAI data services

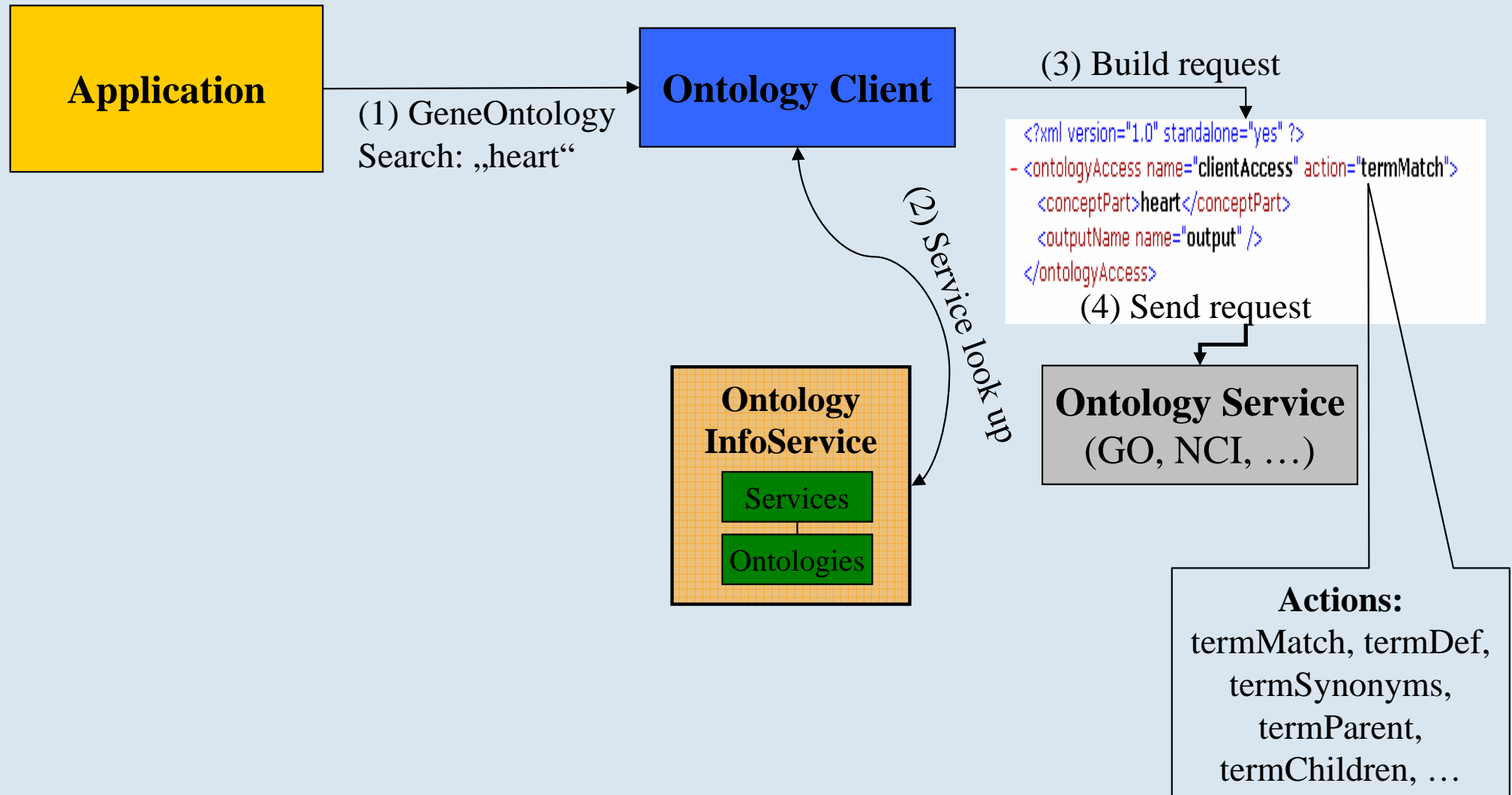
- **Activities**
 - Special ontology activities as wrapper for ontology sources
 - Central ontology interface: `OntologyActivity`
 - Hierarchy of ontology activities
 - Simple integration of new ontologies

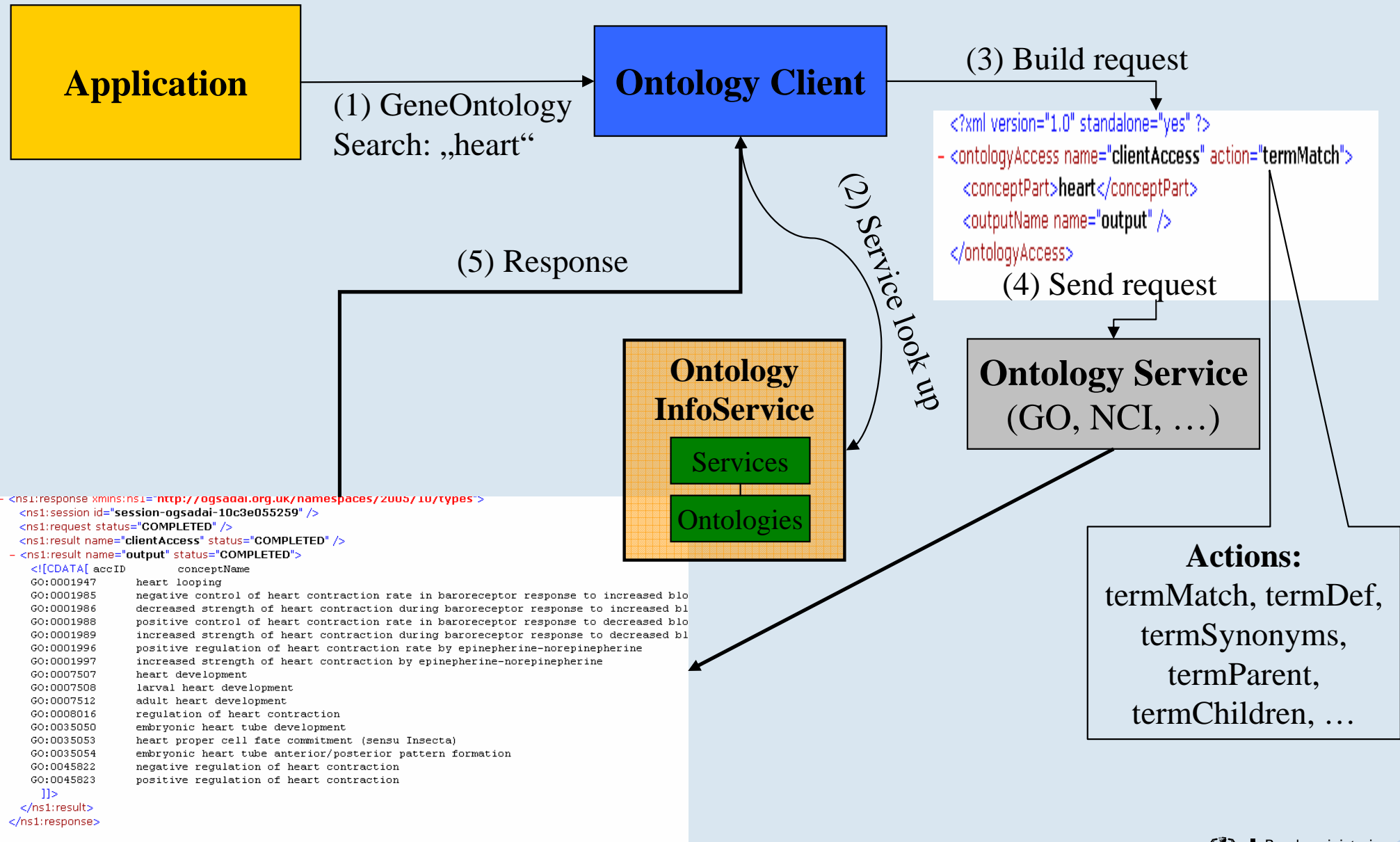


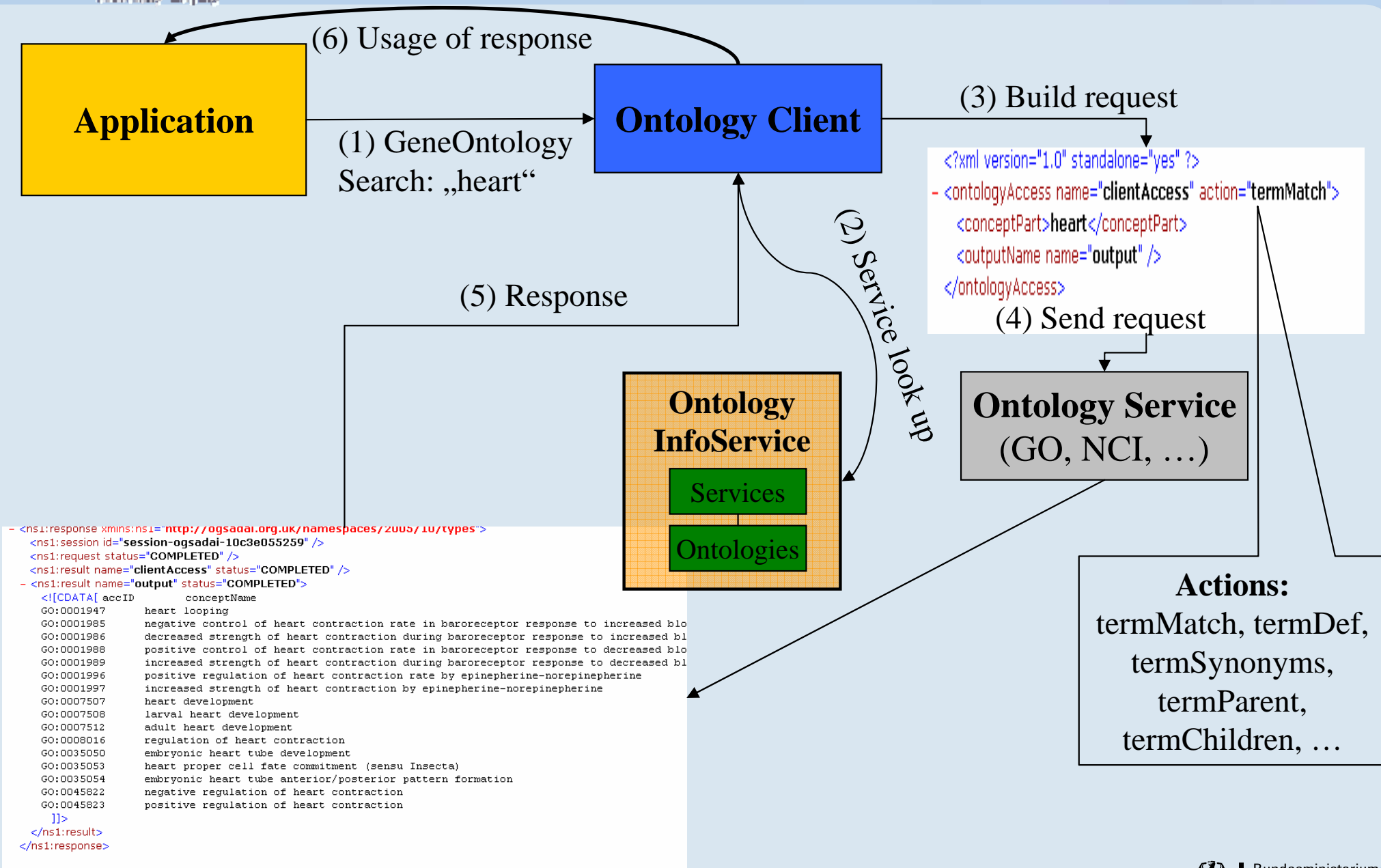












```

<?xml version="1.0" standalone="yes" ?>
- <ontologyAccess name="clientAccess" action="termMatch">
  <conceptPart>heart</conceptPart>
  <outputName name="output" />
</ontologyAccess>
  
```

```

- <ns1:response xmlns:ns1="http://ogsaai.org.uk/namespaces/2005/10/types">
  <ns1:session id="session-ogsadai-10c3e055259" />
  <ns1:request status="COMPLETED" />
  <ns1:result name="clientAccess" status="COMPLETED" />
- <ns1:result name="output" status="COMPLETED">
  <![CDATA[ accID      conceptName
GO:0001947      heart looping
GO:0001985      negative control of heart contraction rate in baroreceptor response to increased blo
GO:0001986      decreased strength of heart contraction during baroreceptor response to increased bl
GO:0001988      positive control of heart contraction rate in baroreceptor response to decreased blo
GO:0001989      increased strength of heart contraction during baroreceptor response to decreased bl
GO:0001996      positive regulation of heart contraction rate by epinephrine-norepinephrine
GO:0001997      increased strength of heart contraction by epinephrine-norepinephrine
GO:0007507      heart development
GO:0007508      larval heart development
GO:0007512      adult heart development
GO:0008016      regulation of heart contraction
GO:0035050      embryonic heart tube development
GO:0035053      heart proper cell fate commitment (sensu Insecta)
GO:0035054      embryonic heart tube anterior/posterior pattern formation
GO:0045822      negative regulation of heart contraction
GO:0045823      positive regulation of heart contraction
]]>
  </ns1:result>
</ns1:response>
  
```

Actions:
termMatch, termDef,
termSynonyms,
termParent,
termChildren, ...



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. SNPSelection



Applications using the middleware (II)

Results for "heart" in "GeneOntology":

heart looping
negative regulation of heart contraction rate in baroreceptor response to increased blood pressure
increased strength of heart contraction during baroreceptor response to increased blood pressure
positive regulation of heart contraction rate in baroreceptor response to decreased blood pressure

ID: GO:0000191
Name: regulation of heart contraction
Definition: Any process that modulates the frequency, rate or extent of heart contraction.
Comment:

Types: regulation of heart rate

Hierarchy:

- GO:0011279: regulation of organometal physiological process
- GO:0009916: regulation of heart contraction
 - GO:0060436: cardiac conduction
 - GO:0060437: cardiac electrophysiology
 - GO:0044823: negative regulation of heart contraction
 - GO:0044824: positive regulation of heart contraction

Parents: GO:0011279: regulation of organometal physiological process [P, A]

Children: GO:0000204: cardiac atrophy [P, A], GO:0000207: cardiac hypertrophy [P, A], GO:0044822: negative regulation of heart contraction [P, A], GO:0044823: positive regulation of heart contraction [P, A]

References: GO:cp:curators, TUB:TCRFAB5:TCRFAB5 phosphatidase

MediGRID Portal

Heart Map: Baden Baden May 4, 2007

Services: last time on this page: 2007-05-04

- GeneOntology - GO:0000191: regulation of heart contraction
- NCIT Thesaurus - NCIT:ThesaurusResource
- GeneOntology - GO:0060436: cardiac conduction
- FalconsOntology - FalconsOntologyResource
- HumanOntology - DiseaseOntologyResource
- ProteinInteractionOntology - ProteinInteractionOntologyResource
- CellTypeOntology - CellOntologyResource
- MultipleAlignmentOntology - MultipleAlignmentOntologyResource
- ProteinStructureOntology - ProteinStructureOntologyResource
- FalconsOntology - FalconsOntologyResource
- MammalianPhenotypeOntology - MammalianPhenotypeOntologyResource
- Medline - MedlineResource
- SequenceOntology - SequenceOntologyResource
- MicroarrayOntology - MicroarrayOntologyResource



MediGRID Portal

Computing für die Medizin und Lebenswissenschaften

Wählen Sie ein Grid Portal der MediGRID Community!

MediGRID Portal bietet einen einfachen Zugang zu den Anwendungen im Rahmen der Grid-Infrastruktur für den Bereich Medizin und

```
#!/usr/bin/perl
use strict;
my $cmd = "cat /dev/urandom | tr -dc 'a-z0-9' | fold -w 64 | xargs -n 1 md5sum | sort | md5sum";
my $result = `perl $cmd`;
my $file = "result.txt";
open(FILE, ">$file");
print FILE $result;
close(FILE);

my $input = "AT019L1.GM.G3903.NF.118433";
my $output = "AT019L1.GM.G3903.NF.118433";
my $cmd = "cat /dev/urandom | tr -dc 'a-z0-9' | fold -w 64 | xargs -n 1 md5sum | sort | md5sum";
my $result = `perl $cmd`;
my $file = "result.txt";
open(FILE, ">$file");
print FILE $result;
close(FILE);
```

Genes

AT019L1.GM.G3903.NF.118433

Position: Chr2:23942299-23946312

Description: AP016 subphage 18-likeorfam1

Gene/ontology type	Gene/ontology term
Biological process	GO:0009914 - autophagy
	GO:0015031 - protein transport
Cellular component	GO:0005776 - autophagic vacuole
	GO:0016021 - membrane
Molecular function	GO:0005915 - protein binding



Summary and Future

- **Service-based middleware to integrate ontologies in grids**
 - Based von 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)