

Grid Prototype Project (caGRID?)

June 24, 2004

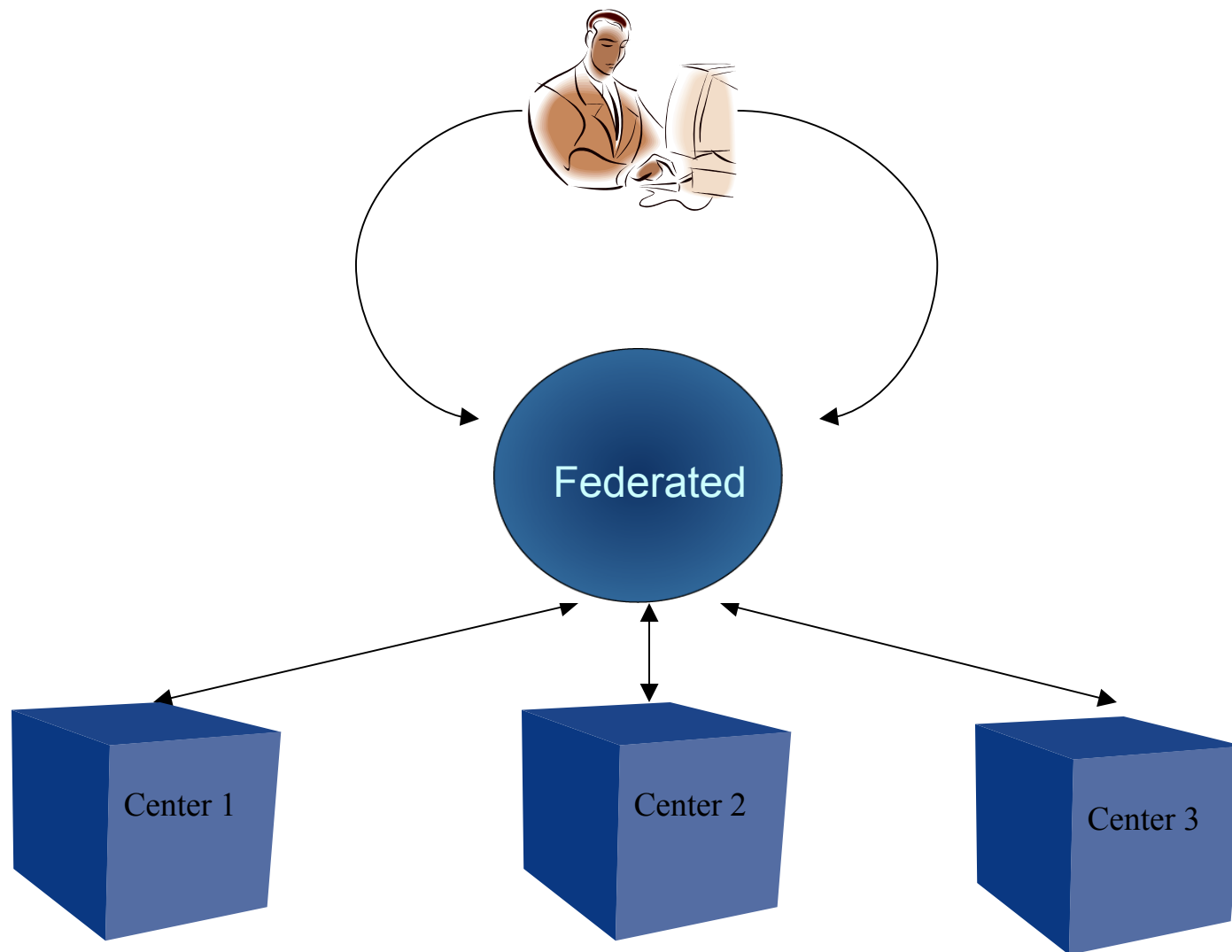
- ▶ Introduction
 - Team Members
 - Manav Kher – (SAIC)
 - William Sanchez – (SAIC)
 - Brian Gilman – (Panther Informatics)
 - Steve Poulos – (Panther Informatics)
- ▶ Requirement analysis
- ▶ Technology evaluation
- ▶ caGRID framework
- ▶ Prototype demo

Problem Statement

- ▶ Production of data is outstripping our ability to analyze it
- ▶ The research community may not be aware of other work and datasets
- ▶ Researchers may not tag data with the definition of the data they produce
- ▶ Semantic information is not often encoded nor included with data sets
- ▶ “Data Islands” or “Silo’s” of information are produced based on the problems outlined above
 - A small group of knowledgeable people transmit data amongst themselves
- ▶ “Modern” exploratory research requires the integration of disparate databases of biological information to explain results
- ▶ To elucidate the mechanism behind disease we must aggregate data from many databases

- ▶ Provide the “Grid For Cancer Research” so that we may:
 - Raise awareness of disparate datasets in the biological research community
 - Allow research groups to exchange datasets with ease
 - Allow research groups to understand the semantics of the datasets that they publish without always having to get on the phone
 - Allow for quicker publication of the analysis of integrated data

The Grid Concept



What Criteria Do We Use To Assess The Technology?

- ▶ The caGRID project team utilized the following metrics to assess the applicability of grid technologies to biological problems
 - Ability to handle large datasets
 - Robust protocols for data transmission
 - Quality of service (QOS)
 - Protocols for meta-data and semantic representations of datasets
 - Robust publication, search, and query capability
 - Mature application programming interfaces (APIs)
 - Maturity of technology
 - Change in codebase and API
 - Quality assurance best practices in place
 - Change management system in place with release notes and backward compatibility taken into account
 - Ease of installation and administration (what is the burden on system administrators/users to adopt technology)
 - Acceptance in the engineering community

What Technologies Exist To Produce a “Bio-Grid” (caGRID)

- ▶ Web Services
 - Bio-MOBY
- ▶ Grid Services
 - OGSA-DAI
 - Globus
 - SRB
- ▶ P2P
 - JXTA
- ▶ Hybrid
 - Chinook (Web services and JXTA)
- ▶ Home Grown

Related projects, technologies and standards.

OGSA-DAI
(Data grid)



(Data grid project)

Jena2
(Semantics)



STORAGE
RESOURCE
BROKER

(Data grid application)



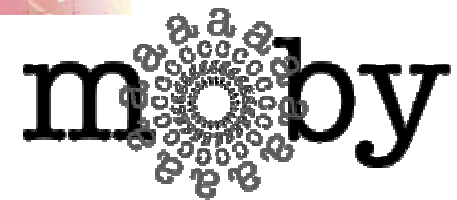
(Grid project)



(Grid infrastructure framework)



Web Services



(Web service registry for
Bioinformatics)



(Grid project)



JXTA
(P2P technology)



(Data grid application)

What is needed to make the Grid “Bio-Accessible”?

- ▶ Submit caGRID prototype project to the caBIG - Architecture Workspace for evaluation
- ▶ Increase understanding of grid computing in the biological sciences - caGRID project
- ▶ Produce a “best practices” and “lessons learned” document/presentation for the community to learn from our experiences
- ▶ Provide a forum for users to explain their requirements and pain points

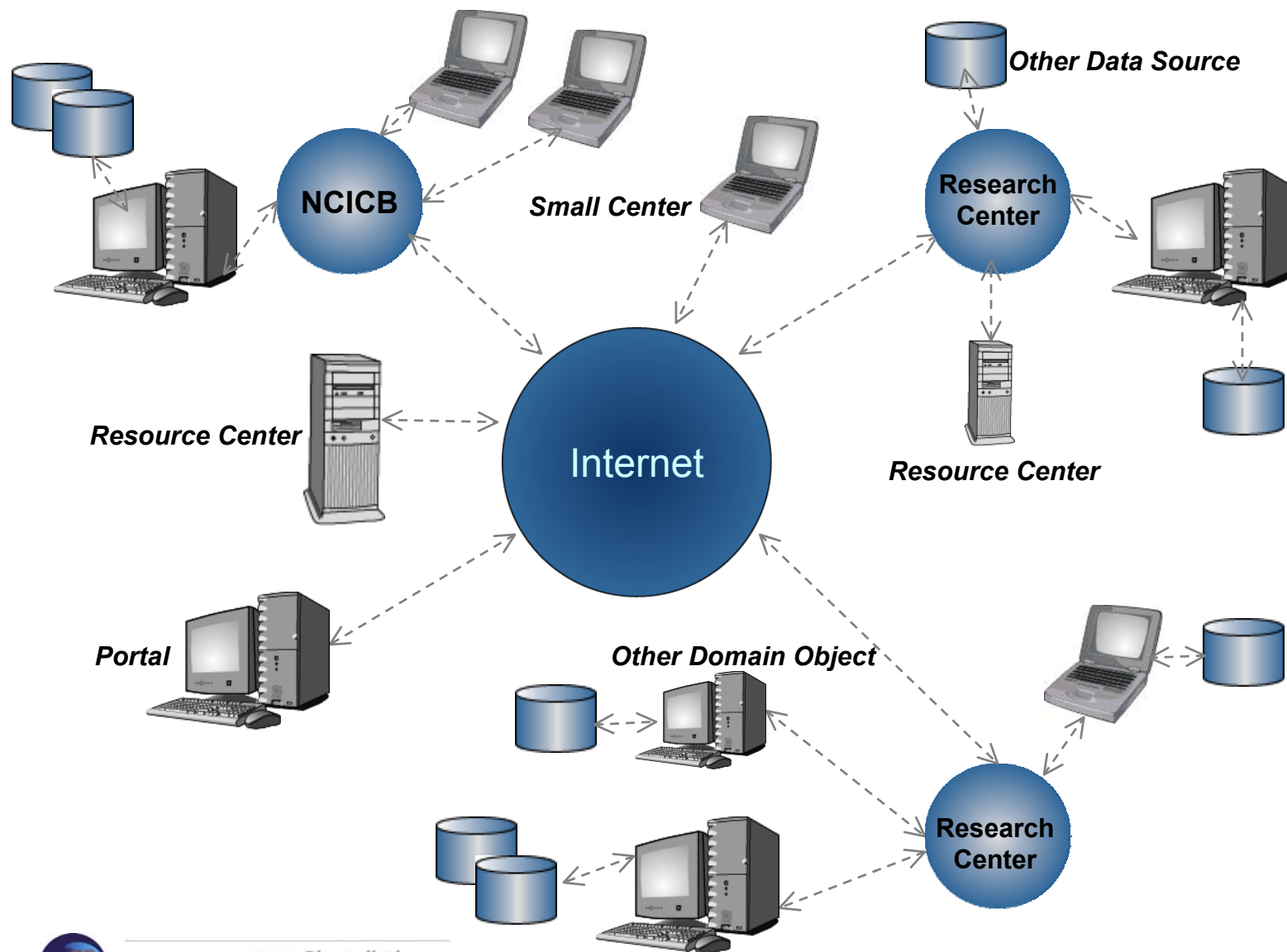
How Do We Realize a Bio-Grid?

- ▶ Provide Semantic representations and intuitive query of Datastores & Services
 - “Computer, find me all clinical trials that are in phase I and are studying neoplasia” vs. “Select * from clinical_trial_data ctd, disease d where ctd.id = d.id and d.desc like ‘%neoplasia%’”
 - Find me all genes and their aliases that are associated with neoplasia
 - Find me all micro-array experiments that have been used in studying neoplasia
 - Find me all pathways that have been implicated in neoplasia
 - Put together a compelling, publishable, “drugable” story based on these queries

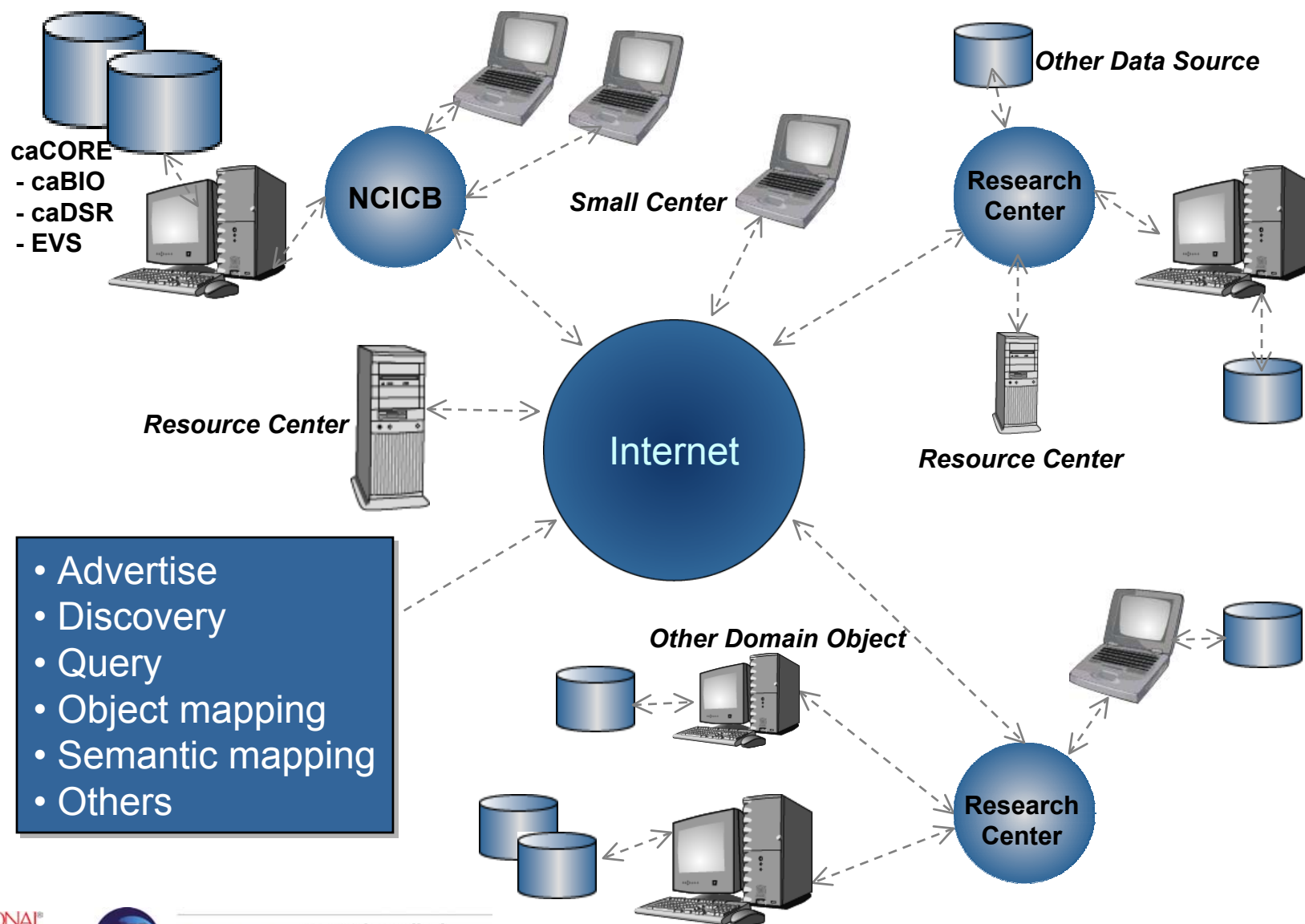
What Did We Choose And Why?

- ▶ Open Grid Services Architecture – Data Access & Integration **(OGSA-DAI)**
 - Based on Globus toolkit
 - Large development community
 - Written on top of web services technology
 - Best documentation with updated documents provided upon every release
 - Provides best tradeoff of complexity for ease of development
 - Addition of features is relatively straight forward

Grid requirements

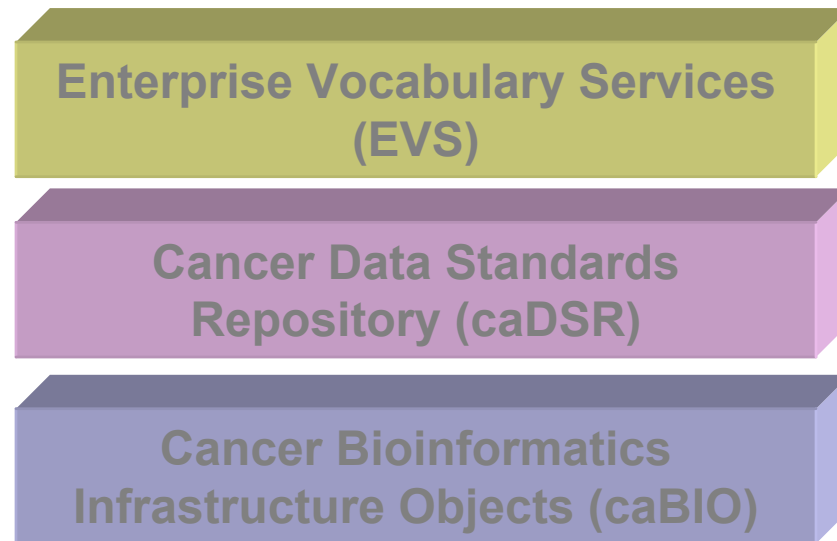


Grid requirements

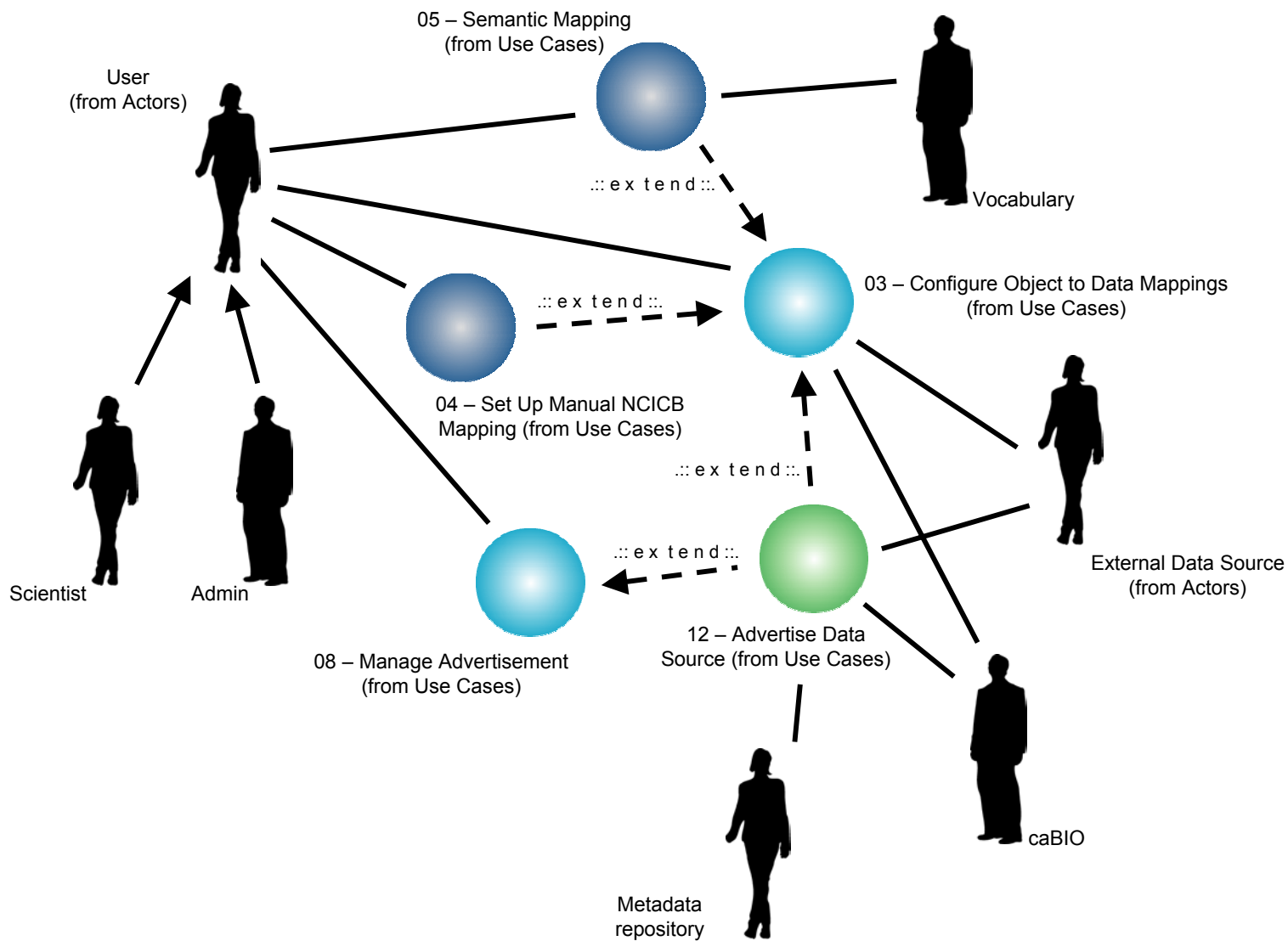


Prototype based on caCORE

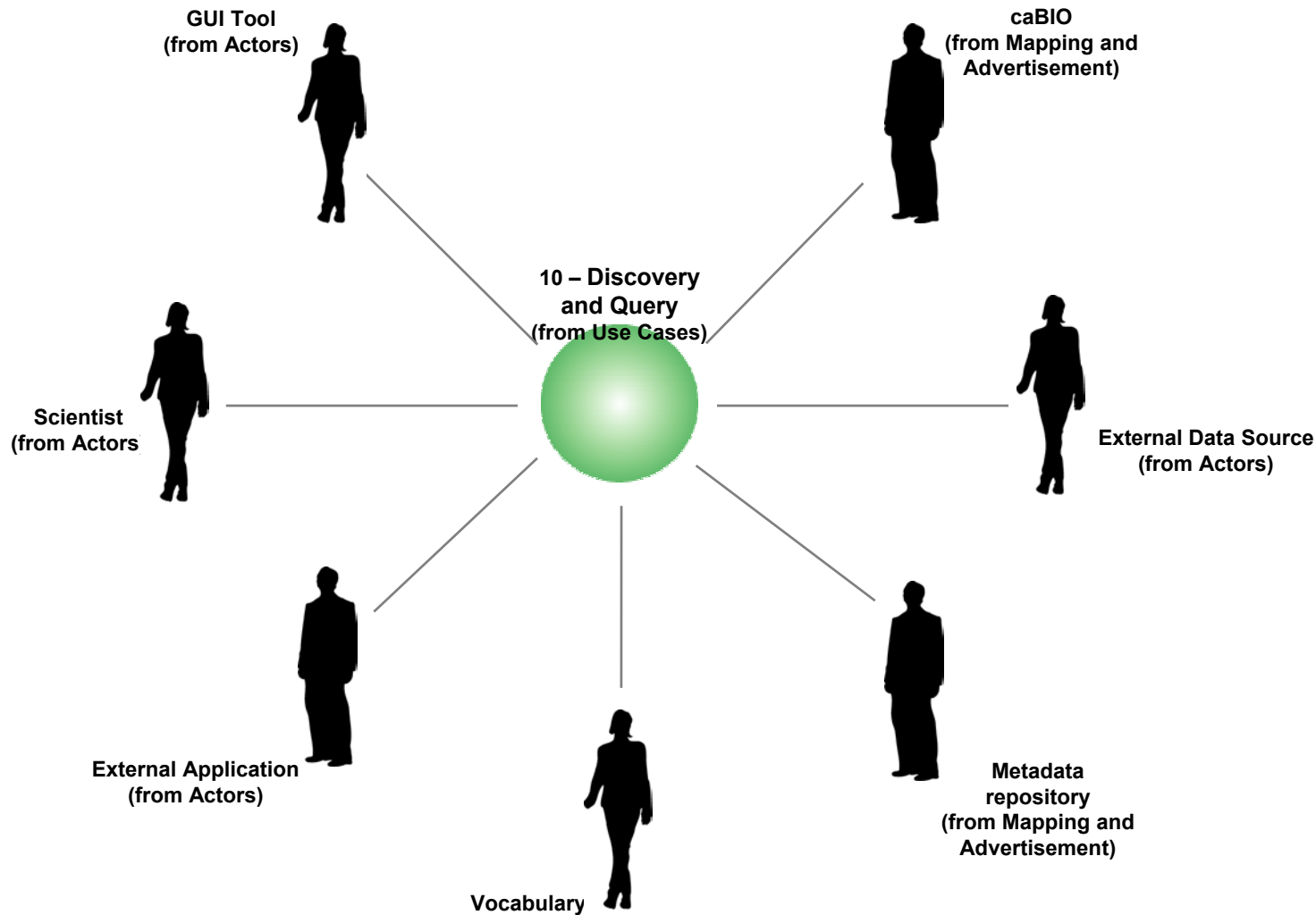
- ▶ **ca**ncer **C**ommon **O**ntologic **R**epresentation **E**nvironment (caCORE)
- ▶ caCORE is the technology stack that facilitates data integration across multiple scientific disciplines



Advertisement and Mapping



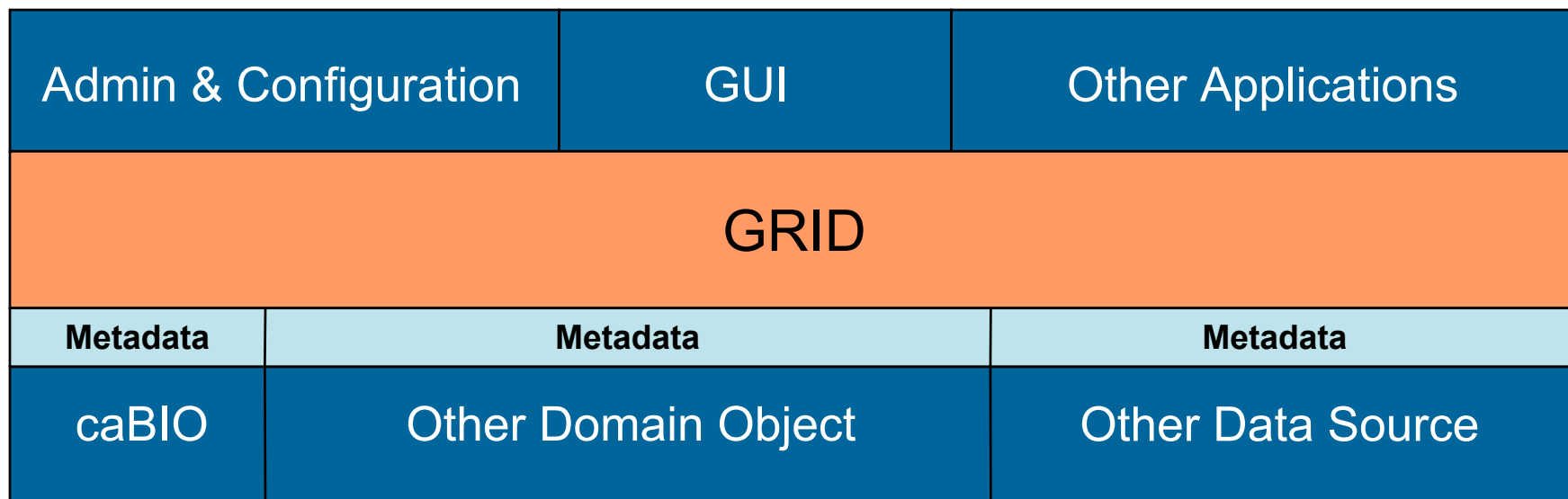
Discovery and Query



Requirement Prioritization

	Prototype
Goals	<ul style="list-style-type: none">-Query Data and Discovery Service – Full-Advertise services – Command line-Startup – Command line-Shutdown – Command line-Install – Basic
Requirement	<ul style="list-style-type: none">-Define and prioritize requirements-Perform technology evaluation (Grid, Semantics)-Define architecture-Implement prototype – caBIO-Semantics - service level-Test GRID technology/framework

Preliminary Architecture



Significant Use Case vs. Candidate Technologies

	Advertise	Discovery	Query	Object Mapping	Semantic Mapping
Web services	-UDDI -Extend UDDI. -Create a WS / server code.	-UDDI -Extend UDDI. -Create a WS / Server code.	- caBIO web services.		
Globus, OGSA-DAI, DQP	-Instantiate a grid service. -Registry new service. -Notification -Metadata Framework	-Indexing services (service data providers, data aggregators, grid service registry).	-Object model / caBIO java api. -Data bases (RDB, XML)		
SRB	- MCAP / Metadata service.	- MCAT / Metadata service.	- SRB server / FS, DB, Obj.		
Jena2	- Improve service description.	- Improve service discovery.			-RWU Ontology languages. -Representation of semantic obj.
MCS / caDSR	- Metadata service	- Metadata service			
- OJB				Customize xml representation to model other DB.	

OGSA and Globus Toolkit 3

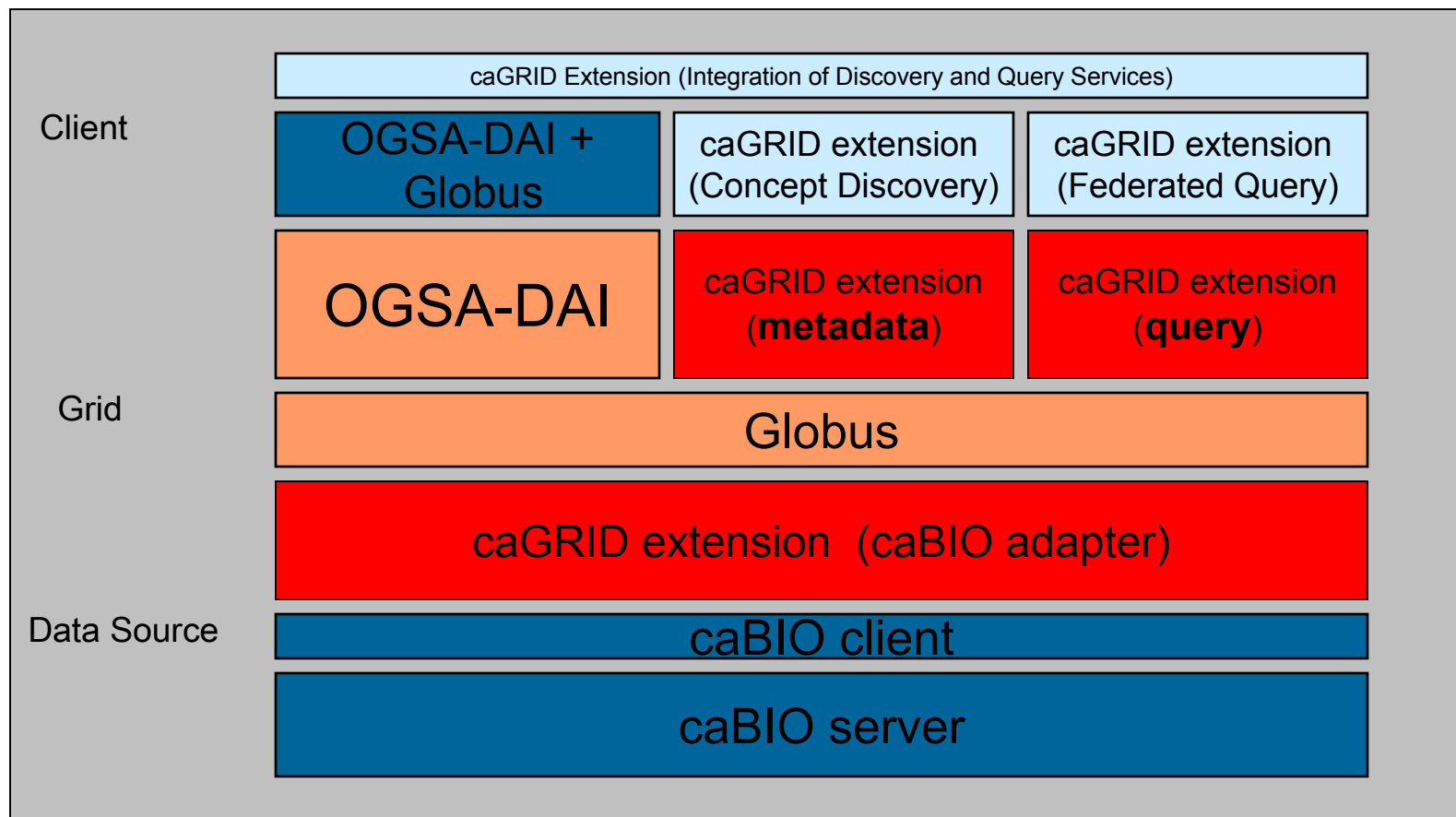
- ▶ **OGSA**: Open standard architecture for next generation grid-services to enable the creation, maintenance, and integration of grid services maintained by virtual organizations.
- ▶ **OGSI**: A core component of OGSA, which provides a uniform way to describe grid services and defines a common pattern of behavior for all grid services. In short, this architecture defines grid service behaviors, service description mechanism, and protocol binding information by using web services as the technology enabler.
- ▶ **Virtual Organizations**: Coordinated resource sharing and problem solving in dynamic, multi-institutional organizations. sharing is with direct access to computers, software, data, and other resources by a range of collaborative problem solving and resource brokering strategies. The sharing is controlled with resource providers and consumers defining what is shared, who is allowed to share and the conditions under which sharing occurs. The set of individuals and/or institutions defined by the sharing rules from the virtual organization.
- ▶ **GT3**: The Globus software technology toolkit version 3 is the major reference implementation of the OGSI standard.

The OGSA-DAI project is concerned with constructing middleware to assist with access and integration of data from separate data sources via the grid. It is engaged in identifying the requirements, designing solutions and delivering software that will meet this purpose. The project was conceived by the [UK Database Task Force](#) and is working closely with the Global Grid Forum [DAIS-WG](#) and the [Globus Team](#).

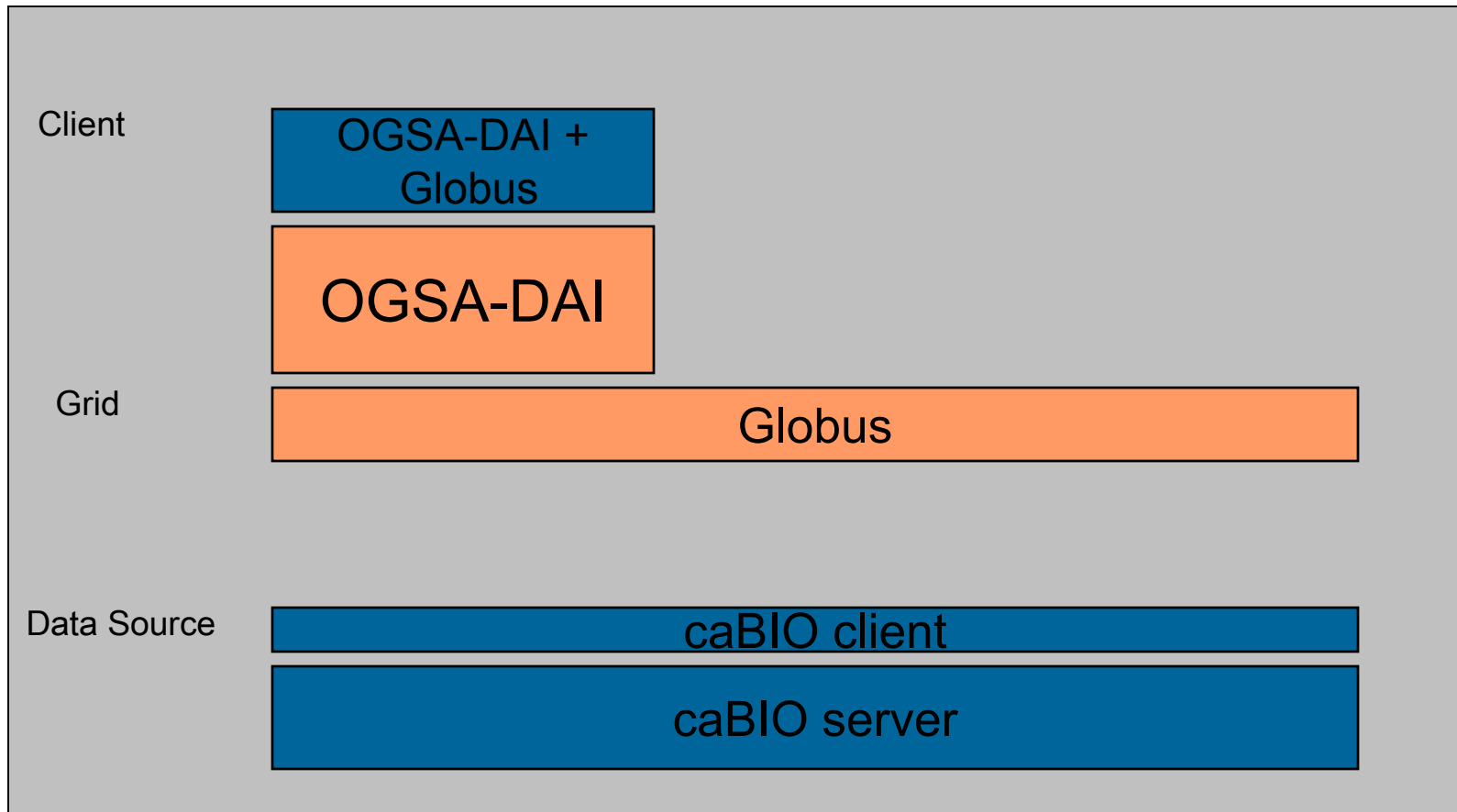
OGSA-DAI is a UK [DTI Funded](#) e-Science Grid Core Project involving the following partners:



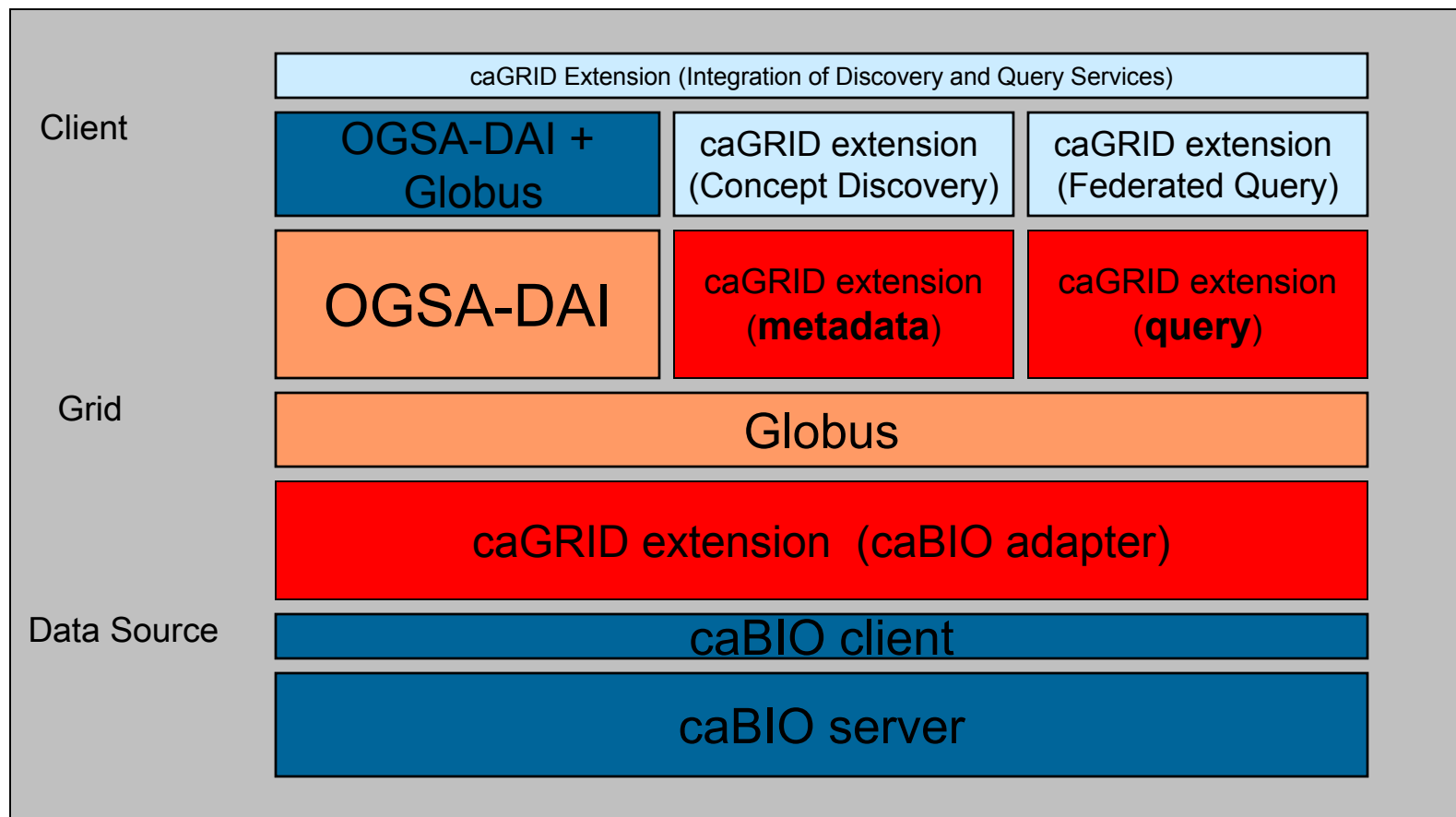
caGRID Core architecture



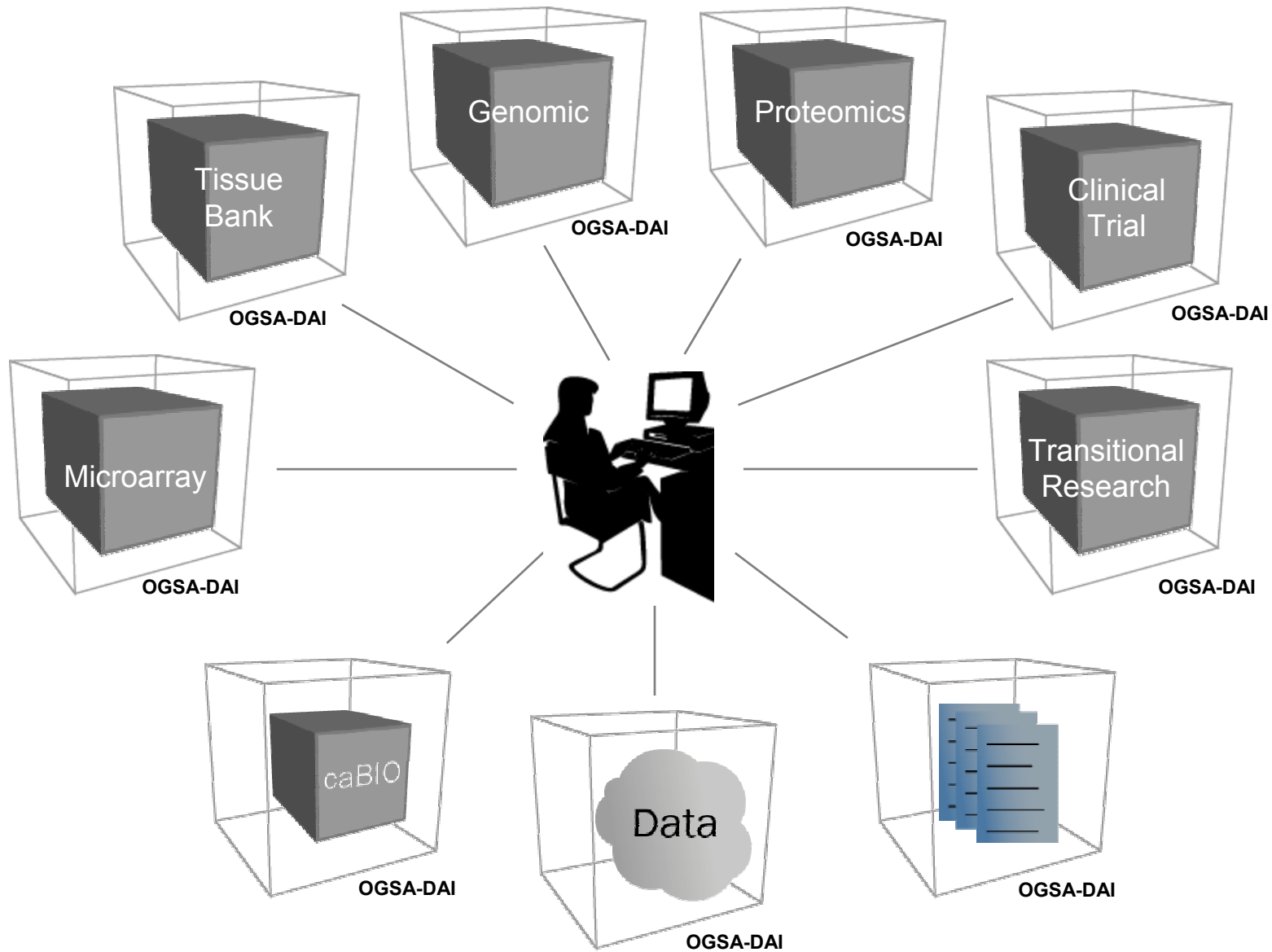
caGRID Core architecture



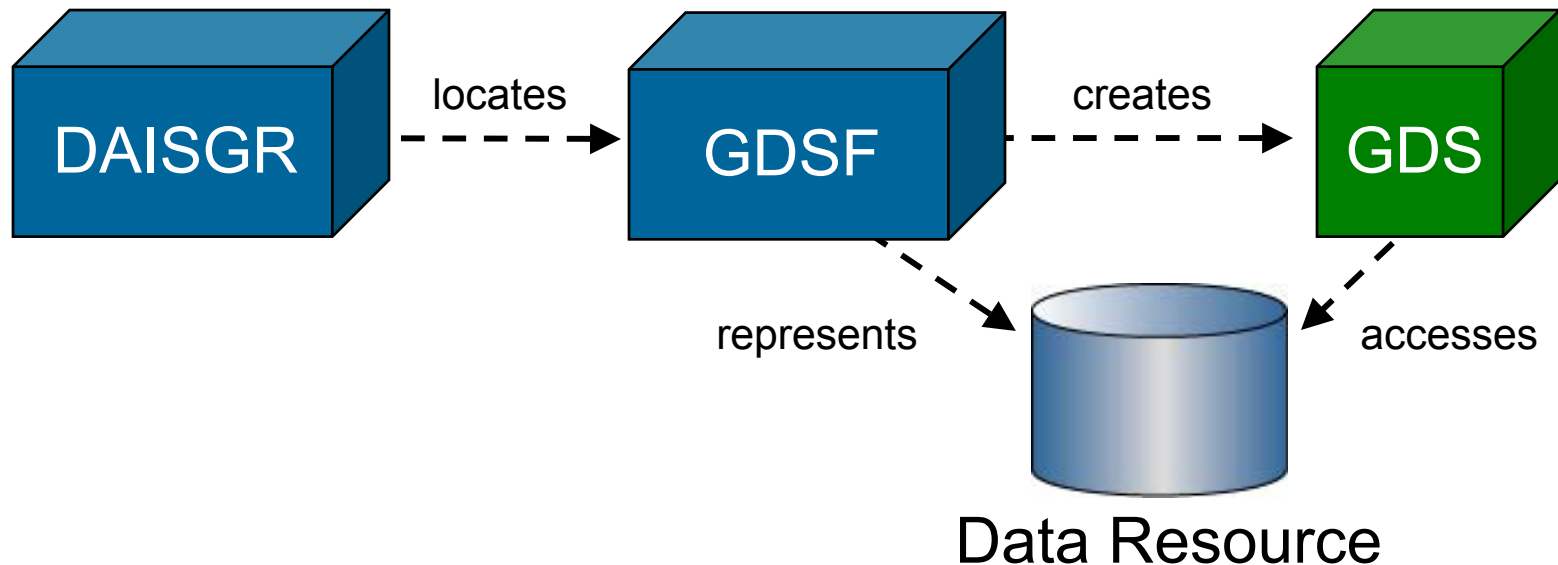
caGRID Core architecture



Facilitating data access

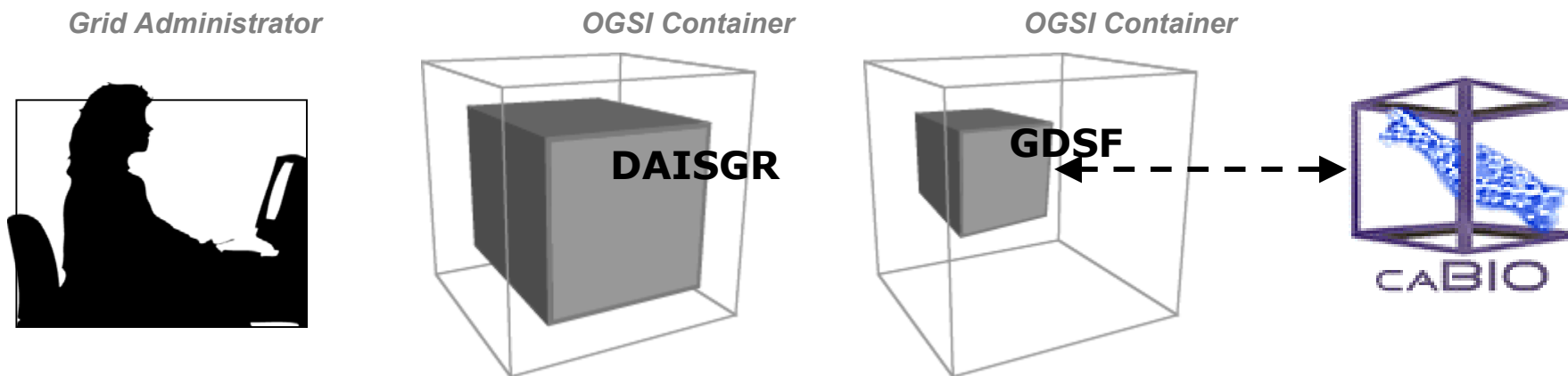


- ▶ OGSA-DAI uses three main service types
 - **Data Access & Integration Service Group Registry** (DAISGR) for discovery
 - **Grid Data Service Factory** (GDSF) to represent a data resource
 - **Grid Data Service** (GDS) to access a data resource



Interaction Model: Start up

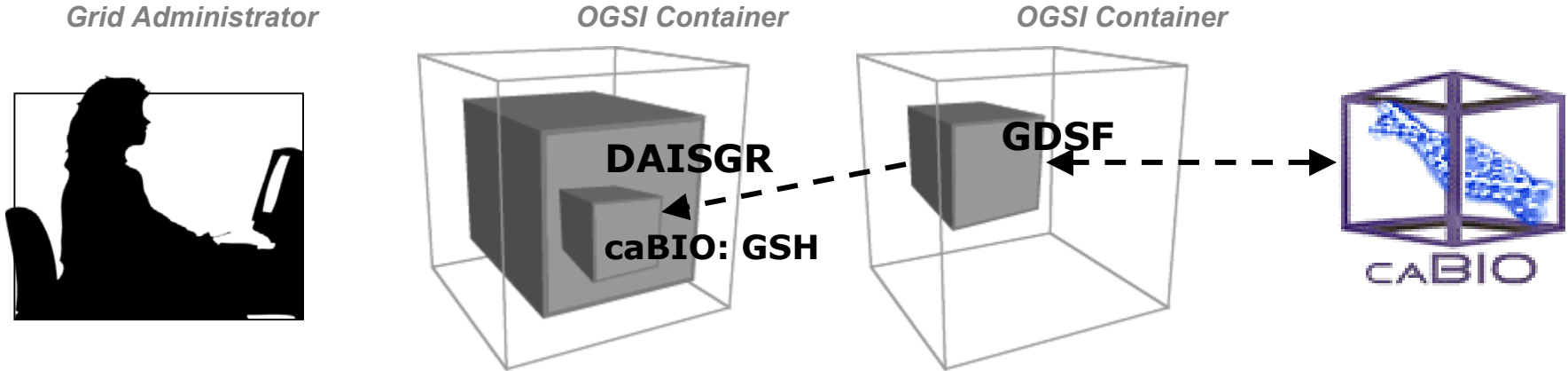
1. Start OGSi containers with persistent services.
2. Here GDSF represents caBIO database.



- DAISGR (registry) for discovery
- GDSF (factory) to represent a data resource
- GDS (data service) to access a data resource

Interaction Model: Registration

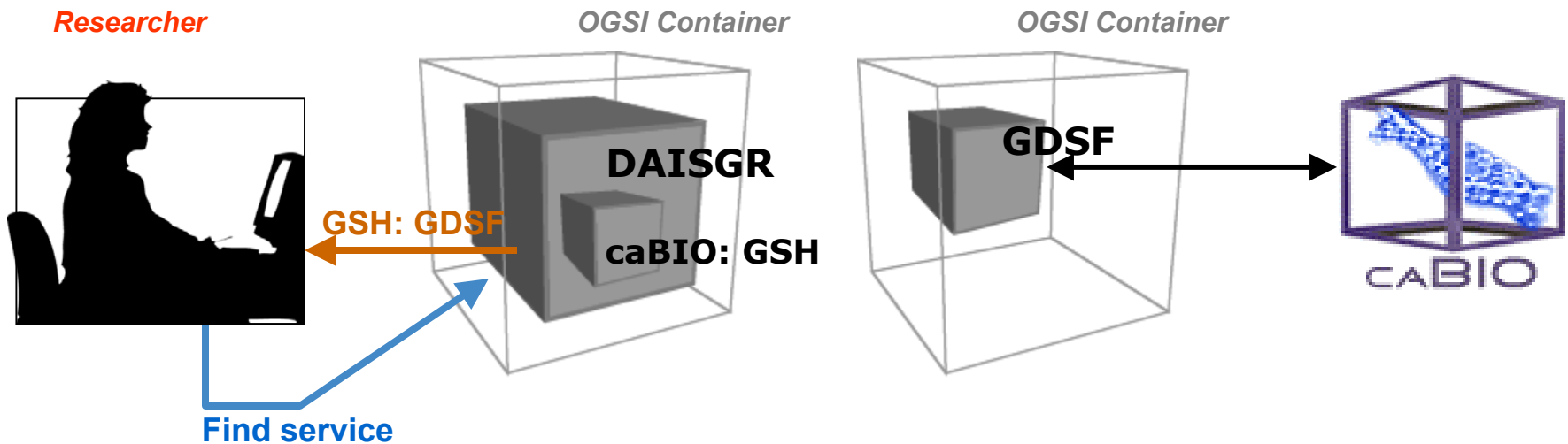
3. GDSF registers with DIASGR



- DAISGR (registry) for discovery
- GDSF (factory) to represent a data resource
- GDS (data service) to access a data resource

Interaction Model: Discovery

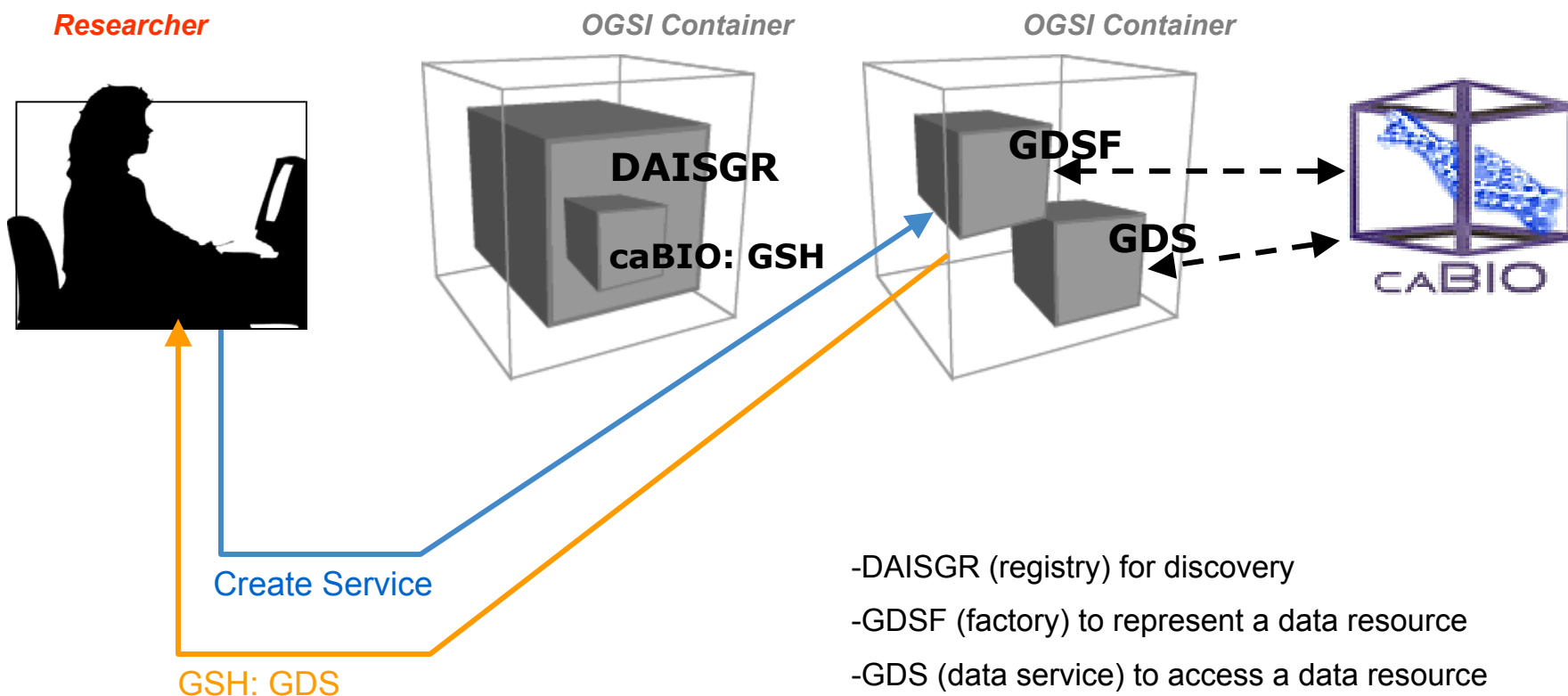
4. Client wants to know about caBIO. caBIO can...
- (i) Query the GDSF directly if known or
 - (ii) Identify suitable GDSF through DAISGR.



- DAISGR (registry) for discovery
- GDSF (factory) to represent a data resource
- GDS (data service) to access a data resource
- Grid Service Handler (GSH)

Interaction Model: Service Creation

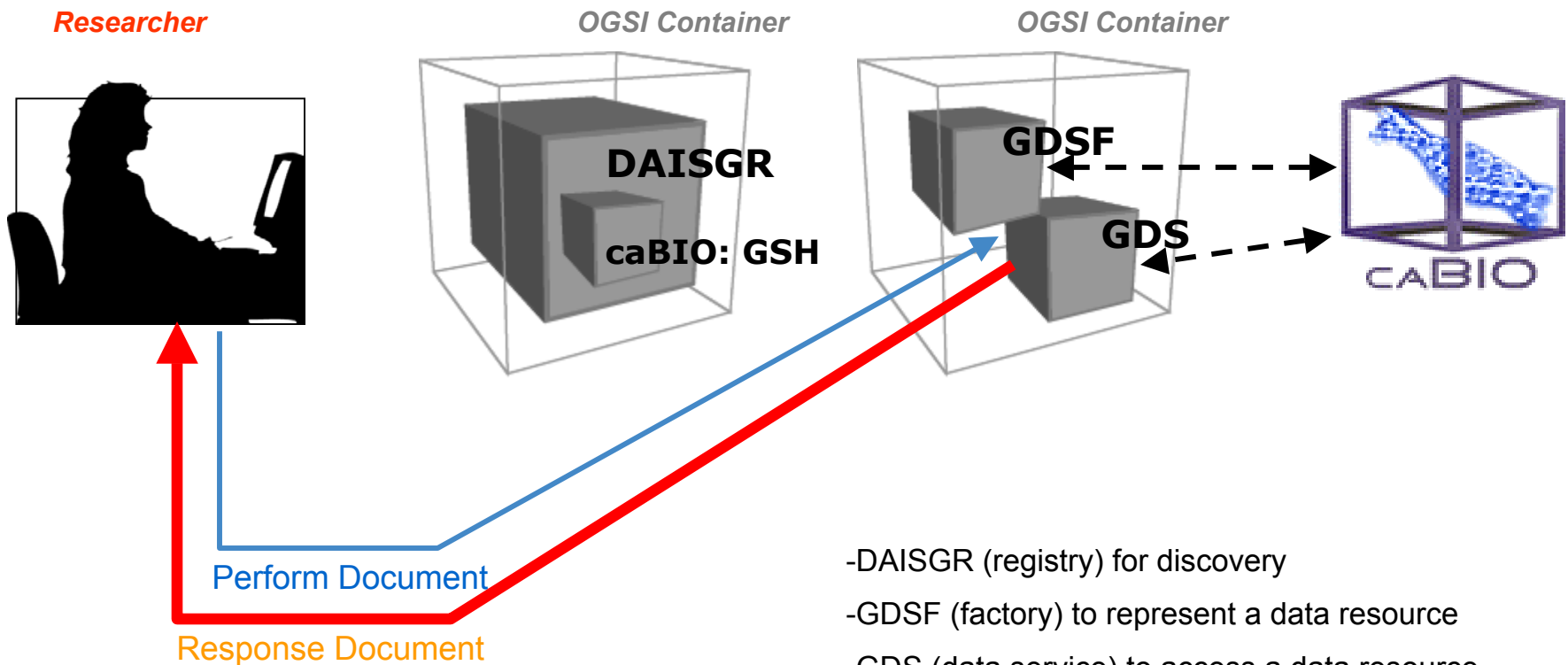
5. Having identified a suitable GDSF client asks a GDS to be created.



- DAISGR (registry) for discovery
- GDSF (factory) to represent a data resource
- GDS (data service) to access a data resource
- Grid Service Handler (GSH)

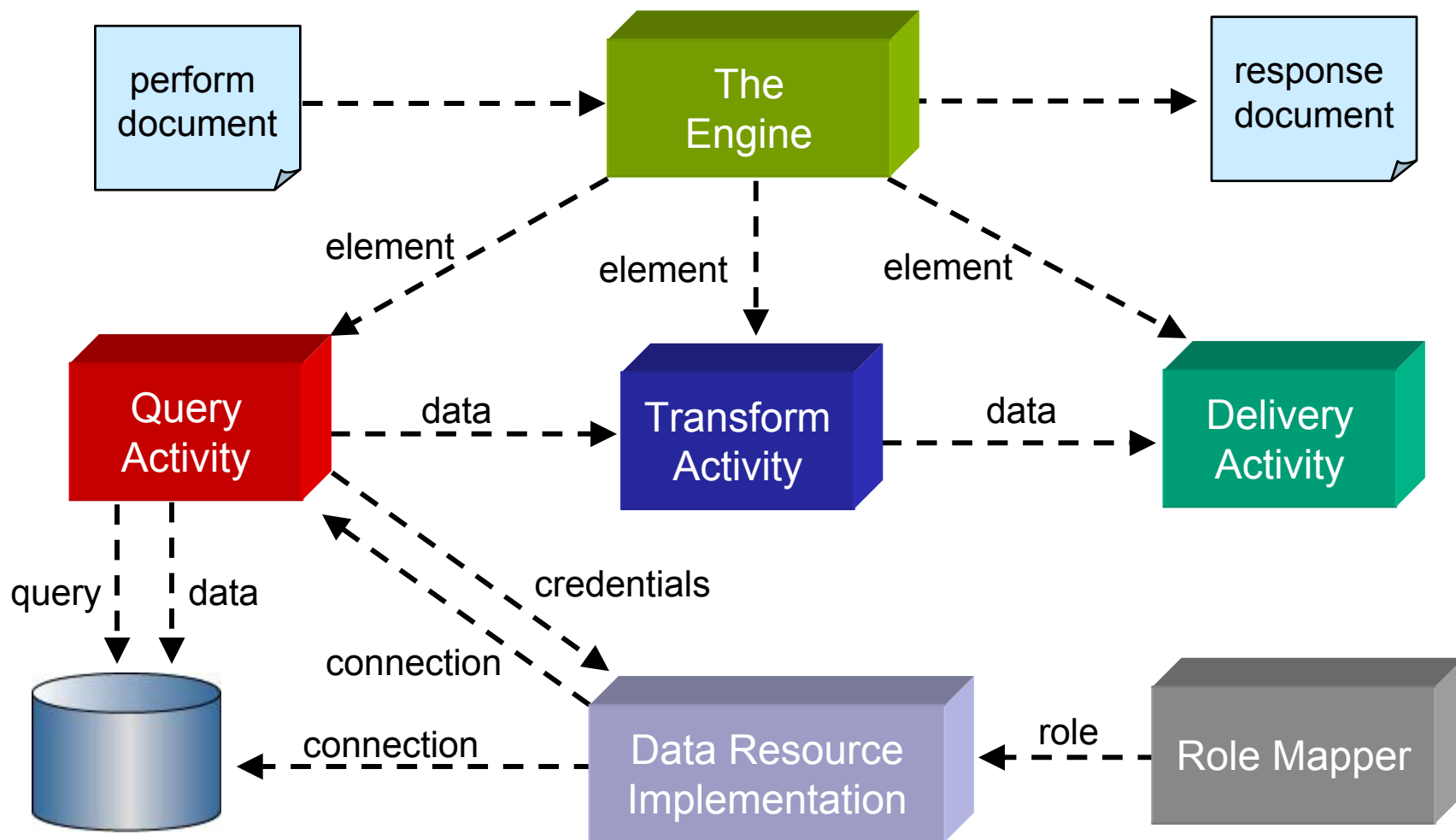
Interaction Model: Perform

6. Client interacts with GDS by sending Perform documents.
7. GDS responds with a Response document.
8. Client may terminate GDS when finished or let it die naturally.



- DAISGR (registry) for discovery
- GDSF (factory) to represent a data resource
- GDS (data service) to access a data resource
- Grid Service Handler (GSH)

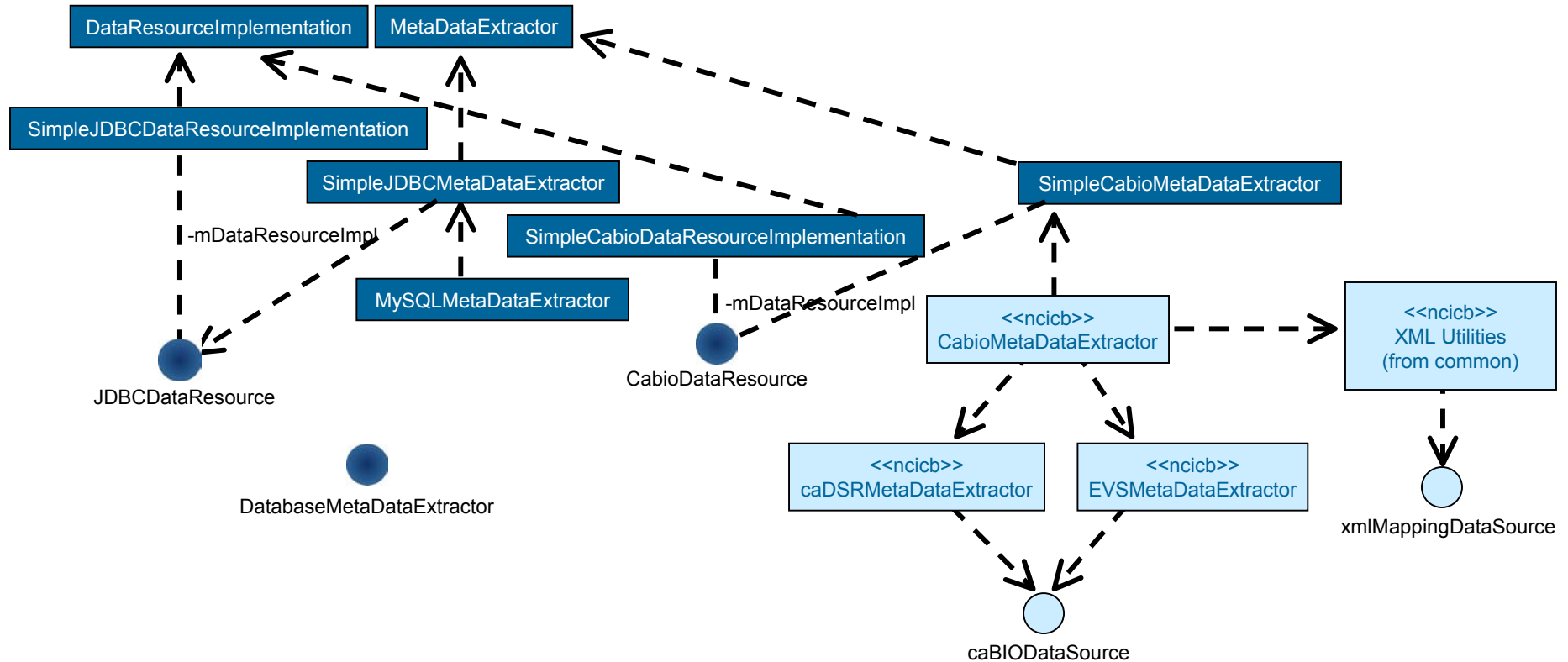
GDS Internals



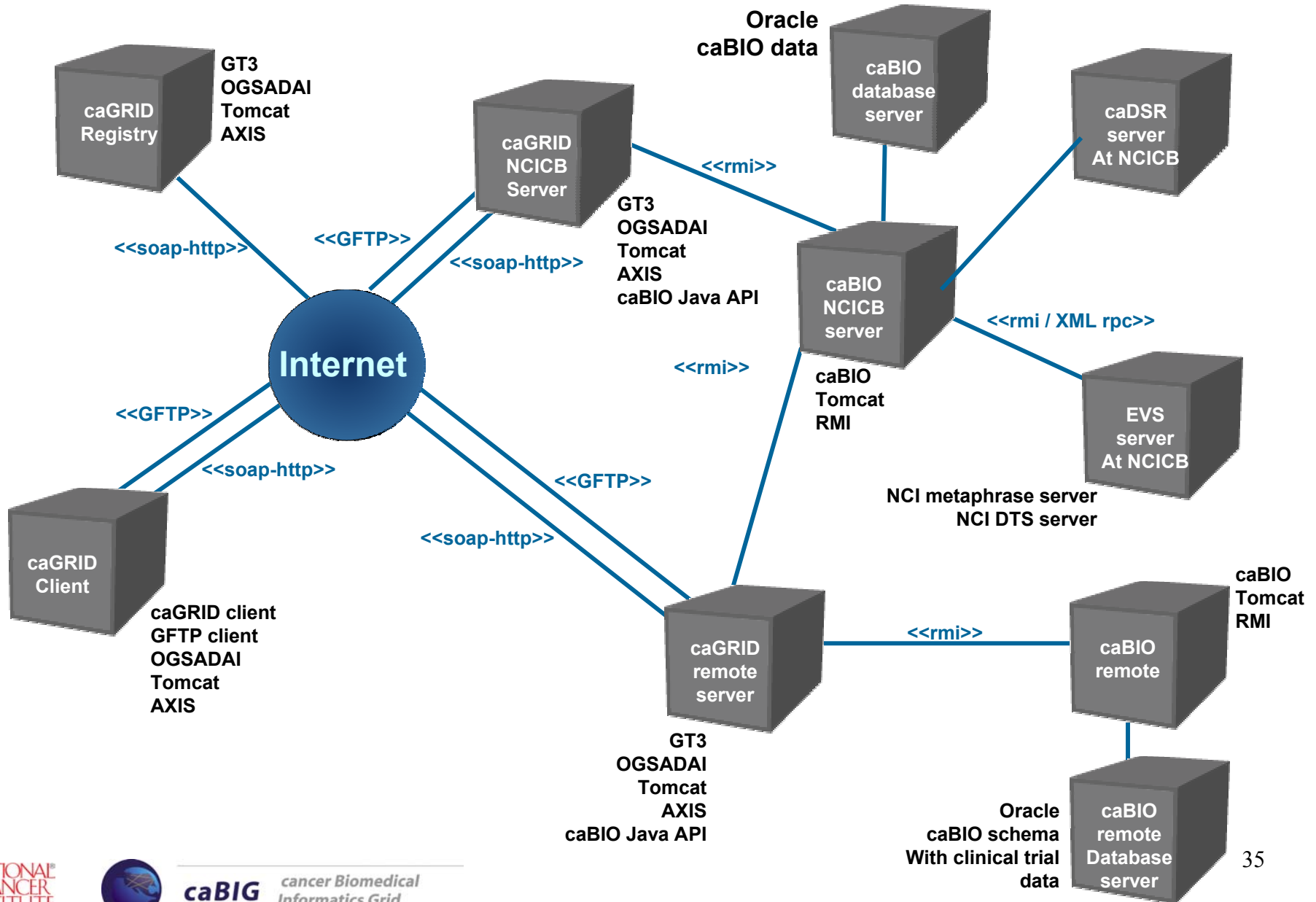
OGSA-DAI Extensions

- ▶ **New data source:** caBIO added as a new data source to the OGSA-DAI framework.
- ▶ **Activities:** Two activities that implements caBIO Java api.
- ▶ **Local caDSR:** Local xml representation of caDSR.
- ▶ **Query language:** Prototype query language with XML representation to query caBIO data source.
- ▶ **Metadata:** 2 new metadata categories added (caBIOSchema and ResearchCenterInformation). Includes the caDSR/EVS metadata extractor.
- ▶ **Concept Discovery:** Client extended to discover by concept name in the grid. The result set is the factories grouped by relationships.
- ▶ **Federated query:** Federated query by relationship (from discovery process). Send serialized queries to the factories that have advertised concept-relationship.
- ▶ **Sample Implementation:** A command line application that integrates Discovery and Query extension. The application help the user to perform discovery and federated queries.
- ▶ **Result set cached (Alpha):** The result set from a federated query will be store in a client xml database. Using another factory the result set can be query using xpath expressions.
- ▶ **Metadata query (Alpha):** ability to query metadata without using xpath expression.

Datasource – GDSF – Framework extension



Deployment Diagram (Prototype)



Prototype Demo

- ▶ Globus GUI – Grid service concept,
 - Registry
 - Metadata
 - Metadata Query
 - caBIOConfig

- ▶ OGSA-DAI GUI
 - Instantiate a factory
 - Submit perform documents
 - Show perform Documents

- ▶ Client and Client extensions
 - Add/Remove services
 - Discovery process
 - Clinical trial protocol, Diseases
 - Query services
 - Show me all phase I Clinical Trial Protocols
 - Show me all Trials for Neoplasia
 - Show me all Diseases that are associated with Phase I trials.
 - Sample using the caBIG middleware.

Demo starts here ...



Lesson Learned

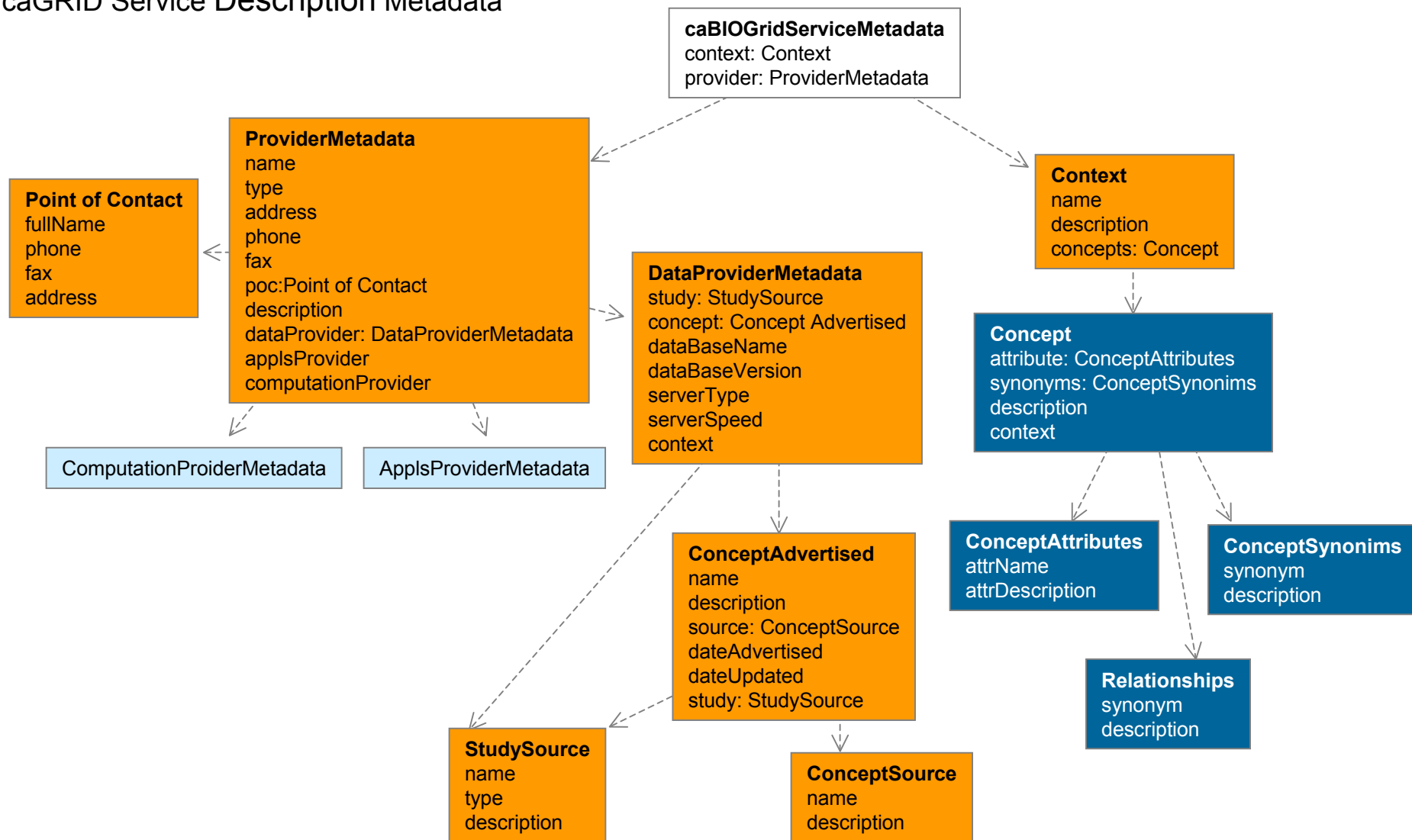
- ▶ There is an inherent learning curve in implementing grid technologies
- ▶ Grid technologies are still maturing and preparation for frequent upgrades is essential
- ▶ Common meta data structure and terminology is necessary to effectively describe services and data
- ▶ A common query language is important to support federated queries

Future Recommendations

- ▶ Map existing use cases to caBIG needs
- ▶ Continue extending caGRID to produce a robust platform. Extensions include:
 - Upgrades to the latest version of Globus in support of web services standards
 - Design and implementation of user friendly GUI query tools
 - Implementation of additional grid services including
 - Security
 - File Sharing
 - Enhance strategy for semantic web interoperability
- ▶ Document caGRID efforts
 - Draft a white paper describing evaluation and prototype efforts
 - Draft developer guides to assist in establishing grid connections

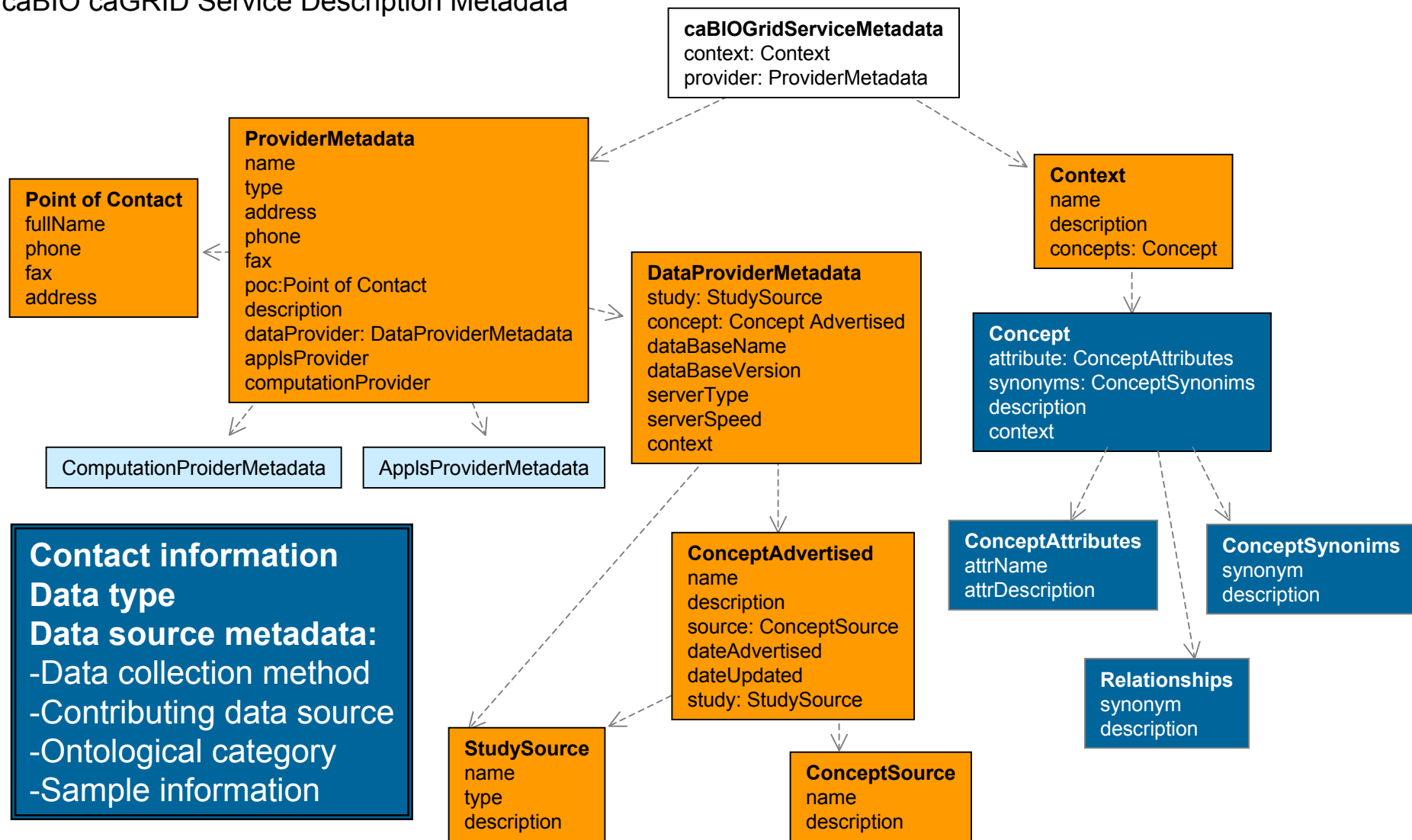
Metadata - GDSF

caGRID Service Description Metadata



Metadata - GDSF

caBIO caGRID Service Description Metadata



Server Configuration - DAISGR

```
<!-- caBIO Service Deployment -->
<service name="ogsadai/caBIODataServiceFactory" provider="Handler" style="wrapped" use="literal">
  <parameter name="ogsadai.gdsf.config.xml.file"
    value="c:/jakarta-tomcat-4.1.24/webapps/ogsa/WEB-INF/etc/caBIOConfig.xml"/>
  <parameter name="ogsadai.gdsf.registrations.xml.file"
    value="c:/jakarta-tomcat-4.1.24/webapps/ogsa/WEB-INF/etc/registrationList.xml"/>
  <parameter name="name" value="caBIO Grid Data Service Factory"/>
  <parameter name="operationProviders" value="org.globus.ogsa.impl.ogsi.FactoryProvider"/>
  <parameter name="persistent" value="true"/>
  <parameter name="instance-schemaPath" value="schema/ogsadai/gds/gds_service.wsdl"/>
  <parameter name="instance-className" value="uk.org.ogsadai.wsdl.gds.GDSPortType"/>
  <parameter name="instance-baseClassName" value="uk.org.ogsadai.service.gds.GridDataService"/>
  <parameter name="instance-operationProviders"
    value="org.globus.ogsa.impl.ogsi.NotificationSourceProvider"/>
  <parameter name="baseClassName"
    value="uk.org.ogsadai.service.gdsf.GridDataServiceFactory"/>
  <parameter name="schemaPath"
    value="schema/ogsadai/gdsf/grid_data_service_factory_service.wsdl"/>
  <parameter name="handlerClass" value="org.globus.ogsa.handlers.RPCURIPProvider"/>
  <parameter name="instance-name" value="caBIO Grid Data Service"/>
  <parameter name="className" value="uk.org.ogsadai.wsdl.gdsf.GridDataServiceFactoryPortType"/>
  <parameter name="allowedMethods" value="*" />
  <parameter name="factoryCallback"
    value="uk.org.ogsadai.service.gdsf.GridDataServiceFactoryCallback"/>
  <parameter name="activateOnStartup" value="true"/>
</service>
```

caBIO Config

```
...
<!-- caBio activities -->
<activityMap name="caBioSimpleQuery"
  implementation="uk.org.ogsadai.porttype.gds.activity.caBio.CaBioSimpleQueryActivity"
  schemaFileName="http://localhost:8080/schema/ogsadai/xsd/activities/caBIO_simple_query.xsd" />
<activityMap name="caBioQuery"
  implementation="uk.org.ogsadai.porttype.gds.activity.caBio.CaBioQueryActivity"
  schemaFileName="http://localhost:8080/schema/ogsadai/xsd/activities/caBIO_query.xsd" />
...
<!-- caGRID specific -->
<researchCenterInfo>
  <researchCenterByDataType>Genomic</researchCenterByDataType>
  <researchCenterName>Georgetown</researchCenterName>
  <researchCenterType>edu</researchCenterType>
  <researchCenterAddress>WashingtonDC</researchCenterAddress>
  <researchCenterPhone>301-xxx-xxxx</researchCenterPhone>
  <researchCenterFax>301-xxx-xxxx</researchCenterFax>
  <researchCenterPOCName></researchCenterPOCName>
  <researchCenterDescription>Georgetown University</researchCenterDescription>
  <researchCenterComments>Testing grid for caBIG</researchCenterComments>
</researchCenterInfo>

<caBioMetadata>
  <caBioSchema callback="uk.org.ogsada.porttype.gds.dataresource.CabioMetaDataExtractor"/>
</caBioMetadata>
```

The Perform Document

```
<?xml version="1.0" encoding="UTF-8"?>

<gridDataServicePerform
  xmlns="http://ogsadai.org.uk/namespaces/2003/07/gds/types"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://ogsadai.org.uk/namespaces/2003/07/gds/types
  ../../../../schema/ogsadai/xsd/activities/activities.xsd">

  <documentation>
    This example demonstrates how to parameterise an caBIO
  </documentation>

  <caBioSimpleQuery name="caBIO2.0">
    <criteria position="1" class="disease" version="1.0" id="abcd" attribute="name"
    value="neoplasia"/>
    <query class="ClinicalTrialProtocol" version="1.0" id="">ClinicalTrialProtocol</query>
    <output name="Test"/>
  </caBioSimpleQuery>

</gridDataServicePerform>
```

Within the engine ...

```
public void processBlock()
{

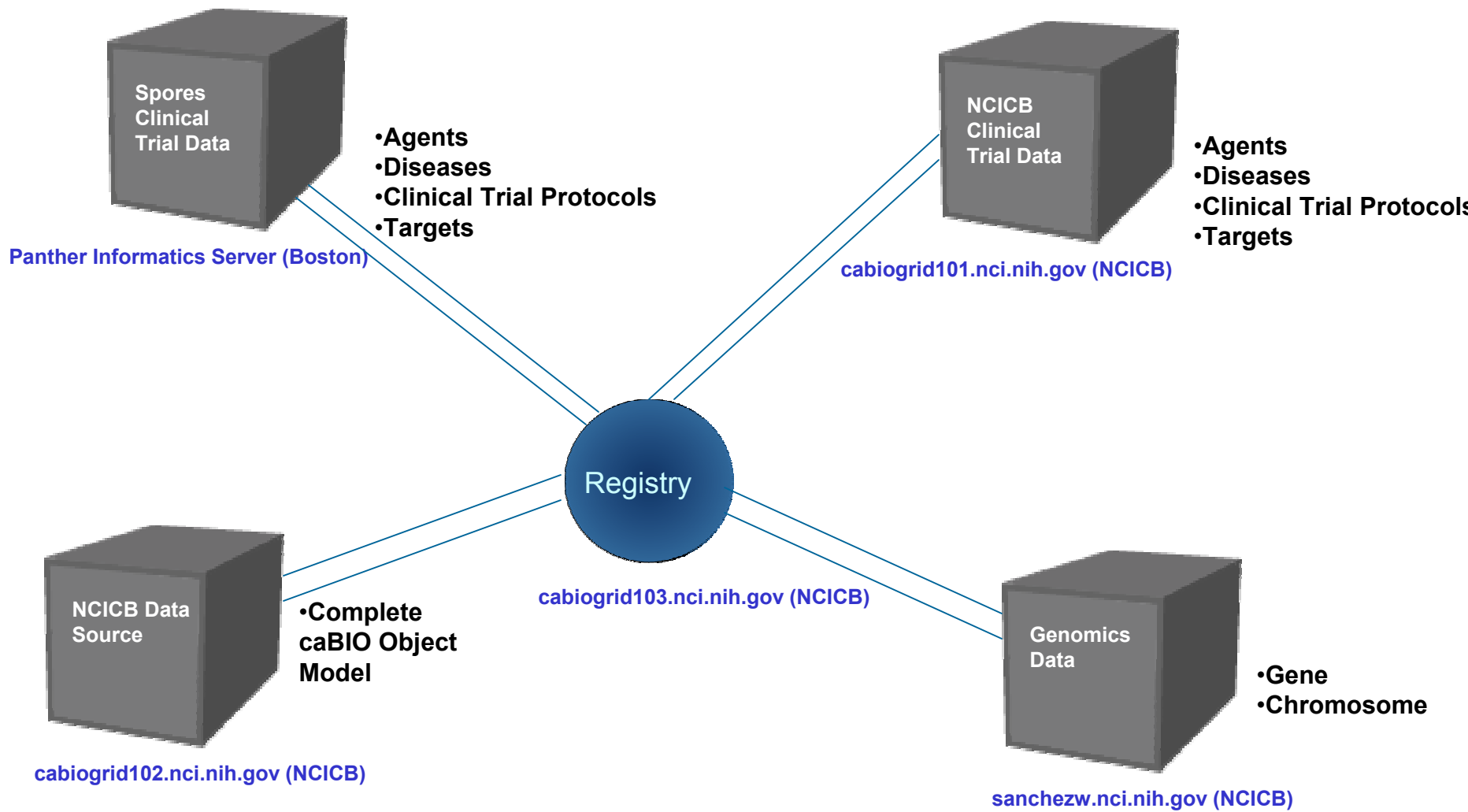
    boolean heavy = false;
    String url = this.httpRequest;

    try {

        gov.nih.nci.common.search.SearchCriteria sc = URL2SC.map( url );
        SearchResult sr = SearchCriteria.copyDownCastSR( sc.search(), SearchResult.class );
        mOutput.put(XMLUtility.makeXMLStringDoc( sr, heavy ));

    } catch (Exception exc) {
        System.out.println("Test failed in the main of GeneDemo.java: " + exc.getMessage());
        exc.printStackTrace();
    }
    mOutput.close();
    setCompleted();
}
```

Prototype Scenario Setup



Screen Shots

caGrid – Globus GUI

The screenshot shows the OGSA Service Browser interface. At the top, the address bar displays the URL: `http://cbiogrid103.nci.nih.gov:8080/ogsa/services/core/registry/ContainerRegistryService`. Below the address bar are navigation buttons: Back, Forward, New Window, Close, Refresh, and a checkbox for "Show dynamic gui". A "Go" button is next to the address bar.

The main content area is divided into sections for configuration and service inspection:

- Message Security:** Includes Authentication (radio buttons for None, GSI XML Signature, GSI Secure Conversation; dropdown for Protection: Integrity, Delegation: None) and Authorization (radio buttons for None, Host, Self, Identity).
- Grid Service:** Includes Namespace, Name, and Timeout input fields, and an XPath Expression field with an XPath Namespace Mappings field.
- Buttons:** Query, Subscribe, and Unsubscribe.
- Service Group Entry Inspection:** A table with columns Name, Handle, and State. The "Container Registry Service" row is highlighted in blue, and its "State" is "ACTIVE".

Annotations with arrows point to the "Container Registry Service" row in the table, the "caGRID Registry" text, and the "State" column.

Name	Handle	State
core/admin/AdminService	http://cbiogrid103.nci.nih.gov:8080/ogsa/services/core/admin/AdminService	INACTIVE
core/management/OgsiManagementService	http://cbiogrid103.nci.nih.gov:8080/ogsa/services/core/management/OgsiManagementService	INACTIVE
Container Registry Service	http://cbiogrid103.nci.nih.gov:8080/ogsa/services/core/registry/ContainerRegistryService	ACTIVE
core/jmsadapter/JMSAdapterFactoryService	http://cbiogrid103.nci.nih.gov:8080/ogsa/services/core/jmsadapter/JMSAdapterFactoryService	INACTIVE
core/logging/OgsiLoggingService	http://cbiogrid103.nci.nih.gov:8080/ogsa/services/core/logging/OgsiLoggingService	INACTIVE
core/notification/http/NotificationSubscriptionFactoryService	http://cbiogrid103.nci.nih.gov:8080/ogsa/services/core/notification/http/NotificationSubscriptionFactoryService	INACTIVE
DAI Registry	http://cbiogrid103.nci.nih.gov:8080/ogsa/services/ogsadai/DAIServiceGroupRegistry	ACTIVE
samples/serialization/SerializationService	http://cbiogrid103.nci.nih.gov:8080/ogsa/services/samples/serialization/SerializationService	INACTIVE
Generic Persistent Grid Service	http://cbiogrid103.nci.nih.gov:8080/ogsa/services/ogs/NotificationSubscriptionFactoryService	ACTIVE
Handle Resolver	http://cbiogrid103.nci.nih.gov:8080/ogsa/services/ogs/HandleResolverService	ACTIVE
gsi/AuthenticationService	http://cbiogrid103.nci.nih.gov:8080/ogsa/services/gsi/AuthenticationService	INACTIVE
gsi/SecureNotificationSubscriptionFactoryService	http://cbiogrid103.nci.nih.gov:8080/ogsa/services/gsi/SecureNotificationSubscriptionFactoryService	INACTIVE

Globus Grid services list

caGRID Registry

Grid service status. Active (stateful) when container starts

caGRID Registry

The screenshot shows the 'OGSA Service Browser' interface. At the top, there are navigation buttons (Back, Forward, New Window, Close, Refresh) and a 'Show dynamic gui' checkbox. Below is a search bar with the URL 'http://cbiogrid103.nci.nih.gov:8080/ogsa/services/ogsadai/DAIServiceGroupRegistry' and a 'Go' button. The main content area is divided into sections: 'Message Security' (Authentication and Authorization), 'Grid Service' (Namespace, Name, Timeout, XPath Expression), and 'Service Group Entry Inspection'. The 'Service Group Entry Inspection' section contains a table with the following data:

Name	Handle
GeorgeTown_caBIODataServiceFactory	http://156.40.48.168:8080/ogsa/services/ogsadai/GeorgeT...
Duke_caBIODataServiceFactory	http://cbiogrid101.nci.nih.gov:8080/ogsa/services/ogsadai/...
NCICB_caBIODataServiceFactory	http://cbiogrid102.nci.nih.gov:8080/ogsa/services/ogsadai/...
Panther1_caBIODataServiceFactory	http://forge.net:8080/ogsa/services/ogsadai/Panther1_caB...

Below the table are 'Auto update' and 'Refresh' buttons. At the bottom, there is a 'Registry' section with 'Unregister' and 'Register' buttons, each preceded by a 'Handle:' input field.

This factories represent the services advertised from Research centers

- Research centers can register services (factories) when their container starts or any time using Globus client.
- Service can have a lifetime.
- Services can be removed any time.
- Services can belong to different registries.

caGRID Factory

Factory

Standard way to query about services. Same queries using globus client

Query for Metadata category

productInformation (ogsadai) expanded

```
ns1:result xmlns:ns1="http://www.gridforum.org/namespaces/2003/03/OGSI" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:type="ns1:ExtensibilityType"
  ns2:serviceDataValues xmlns:ns2="http://www.gridforum.org/namespaces/2003/03/serviceData"
    ns4:productInformation xmlns:ns3="http://ogsadai.org.uk/namespaces/2003/07/gdsf/config" xmlns:ns4="http://ogsadai.org.uk/namespaces/2003/07/gdsf" xsi:type="ns3:ProductInformation"
      ns3:productName
      ns3:productVersion
      ns3:vendorName
```

Metadata

Globus
OGSADAI
caGRID

Browser window: [http://cbiogrid102.nci.nih.gov:8080/ogsa/services/ogsadai/NCICB_caBIODataServiceFactory] - OGSADAI Service B...

Address bar: http://cbiogrid102.nci.nih.gov:8080/ogsa/services/ogsadai/NCICB_caBIODataServiceFactory

Name	PortType	Type	Min	Max	Mutability	Modifiable	Nilable
interface	{http://www.gridfor...}	{http://www.w3.org/...}	1	unbounded	constant	false	false
serviceName	{http://www.gridfor...}	{http://www.w3.org/...}	0	unbounded	mutable	false	false
factoryLocator	{http://www.gridfor...}	{http://www.gridfor...}	1	1	mutable	false	true
gridServiceHandle	{http://www.gridfor...}	{http://www.gridfor...}	0	unbounded	extendable	false	false
gridServiceReference	{http://www.gridfor...}	{http://www.gridfor...}	1	unbounded	mutable	false	false
findServiceDataExtensibility	{http://www.gridfor...}	{http://www.gridfor...}	1	unbounded	static	false	false
setServiceDataExtensibility	{http://www.gridfor...}	{http://www.gridfor...}	1	unbounded	static	false	false
terminationTime	{http://www.gridfor...}	{http://www.gridfor...}	1	1	mutable	false	false
createServiceExtensibility	{http://www.gridfor...}	{http://www.gridfor...}	1	unbounded	static	false	false
productInformation	{http://ogsadai.org...}	ProductInfoType	0	1	mutable	false	false
researchCenterInformation	{http://ogsadai.org...}	ResearchCenterIn...	0	1	mutable	false	false
caBioSchema	{http://ogsadai.org...}	{http://www.gridfor...}	0	1	mutable	true	false
localCaBioSchema	{http://ogsadai.org...}	{http://www.gridfor...}	0	1	mutable	true	false
driver	{http://ogsadai.org...}	{http://ogsadai.org...}	0	1	mutable	false	false
activityType	{http://ogsadai.org...}	{http://ogsadai.org...}	1	unbounded	mutable	false	false
databaseSchema	{http://ogsadai.org...}	{http://www.gridfor...}	0	1	mutable	true	false
collectionSchema	{http://ogsadai.org...}	{http://www.gridfor...}	0	1	mutable	true	false
collectionStructure	{http://ogsadai.org...}	{http://www.gridfor...}	0	1	mutable	true	false

Values

```

ns2:serviceDataValues xmlns:ns2="http://www.gridforum.org/namespaces/2003/03/serviceData"
  ns4:researchCenterInformation xmlns:ns3="http://ogsadai.org.uk/namespaces/2003/07/gdsf/config" xmlns:ns4="http://ogsadai.org.uk/namespaces/2003/07/gdsf" xmlns:xsi="
    ns3:researchCenterBioDataType xsi:type="ns3:BioDataType"
      Genomic
    ns3:researchCenterBioDataType xsi:type="ns3:BioDataType"
    ns3:researchCenterBioDataType xsi:type="ns3:BioDataType"
    ns3:researchCenterBioDataType xsi:type="ns3:BioDataType"
    ns3:researchCenterName
      NCICB
    ns3:researchCenterType
      gov
    ns3:researchCenterAddress
      6116 Executive Blvd, Rockville
    ns3:researchCenterPhone
    ns3:researchCenterFax
    ns3:researchCenterPOCName
    ns3:researchCenterDescription
    ns3:researchCenterComments
  
```

researchCenterInformation
(caGRID extension)

Local caDSR

This category defines concepts advertised by a research center

Name	PortType	Type	Min	Max	Mutability	Modifiable	Nilable
createServiceExtensibility	{http://www.gridfor...}	{http://www.gridfor...}	1	unbounded	static	false	false
productInformation	{http://ogsadai.or...}	ProductInfoType	0	1	mutable	false	false
researchCenterInformation	{http://ogsadai.or...}	ResearchCenterInfo	0	1	mutable	false	false
caBioSchema	{http://ogsadai.or...}	{http://www.gridfor...}	0	1	mutable	true	false
localCaBioSchema	{http://ogsadai.or...}	{http://www.gridfor...}	0	1	mutable	true	false

```

ns2:serviceDataValues xmlns:ns2="http://www.gridforum.org/namespaces/2003/03/serviceData"
  ns4:caBioSchema xmlns:ns3="http://xml.apache.org/xml-soap" xmlns:ns4="http://ogsadai.org.uk/namespaces/2003/07/gdsf" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    context name="caBIO" xmlns=""
      gov.nih.nci.caDSR.bean.DataElementConcept id="AB566958-92FF-5A46-E034-0003BA12F5E7"
      gov.nih.nci.caDSR.bean.DataElementConcept id="AB55AC96-76E5-5A42-E034-0003BA12F5E7"
      gov.nih.nci.caDSR.bean.DataElementConcept id="C43EA872-F67A-12DC-E034-0003BA12F5E7"
      gov.nih.nci.caDSR.bean.DataElementConcept id="C43EA872-F6A6-12DC-E034-0003BA12F5E7"
      gov.nih.nci.caDSR.bean.DataElementConcept id="C455253F-4B71-2C96-E034-0003BA12F5E7"
      gov.nih.nci.caDSR.bean.DataElementConcept id="C455253F-4B95-2C96-E034-0003BA12F5E7"
      gov.nih.nci.caDSR.bean.DataElementConcept id="C43EA872-F624-12DC-E034-0003BA12F5E7"
      gov.nih.nci.caDSR.bean.DataElementConcept id="C43EA872-F632-12DC-E034-0003BA12F5E7"
      gov.nih.nci.caDSR.bean.DataElementConcept id="C43EA872-F63B-12DC-E034-0003BA12F5E7"
      gov.nih.nci.caDSR.bean.DataElementConcept id="C43EA872-F669-12DC-E034-0003BA12F5E7"
      gov.nih.nci.caDSR.bean.DataElementConcept id="C43EA872-F61B-12DC-E034-0003BA12F5E7"
      gov.nih.nci.caDSR.bean.DataElementConcept id="C440AAD0-DA26-464E-E034-0003BA12F5E7"
      gov.nih.nci.caDSR.bean.DataElementConcept id="C440AAD0-DA47-464E-E034-0003BA12F5E7"
      gov.nih.nci.caDSR.bean.DataElementConcept id="C440AAD0-DA51-464E-E034-0003BA12F5E7"
      gov.nih.nci.caDSR.bean.DataElementConcept id="C446B16E-0722-318F-E034-0003BA12F5E7"
      gov.nih.nci.caDSR.bean.DataElementConcept id="C446DAFB-E6AF-3180-E034-0003BA12F5E7"
      gov.nih.nci.caDSR.bean.DataElementConcept id="C44776FA-0176-32D2-E034-0003BA12F5E7"
      gov.nih.nci.caDSR.bean.DataElementConcept id="C440AAD0-DA69-464E-E034-0003BA12F5E7"
      gov.nih.nci.caDSR.bean.DataElementConcept id="C43EA872-F60C-12DC-E034-0003BA12F5E7"
      gov.nih.nci.caDSR.bean.DataElementConcept id="C16D625D-02E1-1907-E034-0003BA12F5E7"
      gov.nih.nci.caDSR.bean.DataElementConcept id="C4A2ABF7-B952-55EA-E034-0003BA12F5E7"
      gov.nih.nci.caDSR.bean.DataElementConcept id="C4A2ABF7-B9C9-55EA-E034-0003BA12F5E7"
      gov.nih.nci.caDSR.bean.DataElementConcept id="C4A2ABF7-B9D2-55EA-E034-0003BA12F5E7"
      gov.nih.nci.caDSR.bean.DataElementConcept id="C52C2C5F-5EA3-0829-E034-0003BA12F5E7"
  
```

Cbiogrid102 represents NCICB grid node which advertised all concepts from caBIO object model

List of concepts advertised

Metadata – GENE_EXPRESSION

PortType	Type	Min	Max	Mutability
{http://www.gridfor...}	{http://www.gridfor...}	1	unbounded	static
{http://ogsadai.or...}	ProductInfoType	0	1	mutable
{http://ogsadai.or...}	ResearchCenterIn...	0	1	mutable
{http://ogsadai.or...}	{http://www.gridfor...}	0	1	mutable
{http://ogsadai.or...}	{http://www.gridfor...}	0	1	mutable
{http://ogsadai.or...}	{http://www.gridfor...}	0	1	mutable

```
ns2:serviceDataValues xmlns:ns2="http://www.gridforum.org/namespaces/2003/03/serviceData"  
  ns4:caBioSchema xmlns:ns3="http://xml.apache.org/xml-soap" xmlns:ns4="http://ogsadai.org.uk/namespaces/2003/07/gdsf" xmlns:  
    context name="caBIO" xmlns=""  
      gov.nih.nci.caDSR.bean.DataElementConcept id="AB566958-92FF-5A46-E034-0003BA12F5E7"  
        preferredName  
          GENE_EXPRESSION  
        preferredDefinition  
          CL0017262 The transcription of genetic information into a message and the subsequent translation into a functional protein.  
        longName  
          Gene_Expressions  
        version  
          1.0  
        dateModified  
        dateCreated  
        publicId  
        evs-synonym  
          EXP  
        evs-synonym  
          Expression  
        evs-synonym  
          Gene Expression  
        gov.nih.nci.caDSR.bean.DataElementConcept id="AB55AC96-76E5-5A42-E034-0003BA12F5E7"  
        gov.nih.nci.caDSR.bean.DataElementConcept id="C43EA872-F67A-12DC-E034-0003BA12F5E7"  
        gov.nih.nci.caDSR.bean.DataElementConcept id="C43EA872-F6A6-12DC-E034-0003BA12F5E7"  
        gov.nih.nci.caDSR.bean.DataElementConcept id="C455253F-4B71-2C96-E034-0003BA12F5E7"  
        gov.nih.nci.caDSR.bean.DataElementConcept id="C455253F-4B95-2C96-E034-0003BA12F5E7"  
        gov.nih.nci.caDSR.bean.DataElementConcept id="C43EA872-F624-12DC-E034-0003BA12F5E7"
```

Each dataElementConcept includes:
-Concept description.
-Concept attributes
-Concept relationships

Each research center advertised the concepts desired. The concept list can be obtained from NCICB metadata, caGRID client extension and in the future using a GUI

Attributes and relationships from GENE_EXPRESSION concept (details next slide)

Another Factory

OGSA Service Browser

http://forge.net:8080/ogsa/services/ogsadai/Panther1_caBIODataServiceFactory

Name	PortType	Type	Min	Max	Mutability	Modifiable	Nillable
serviceDataExtensibility	{http://www.gridfor...}	{http://www.gridfor...}	1	unbounded	static	false	false
terminationTime	{http://www.gridfor...}	{http://www.gridfor...}	1	1	mutable	false	false
createServiceExtensibility	{http://www.gridfor...}	{http://www.gridfor...}	1	unbounded	static	false	false
productInformation	{http://ogsadai.or...}	ProductInfoType	0	1	mutable	false	false
researchCenterInformation	{http://ogsadai.or...}	ResearchCenterIn...	0	1	mutable	false	false
caBioSchema	{http://ogsadai.or...}	{http://www.gridfor...}	0	1	mutable	true	false

Values

```
ns2:serviceDataValues xmlns:ns2="http://www.gridforum.org/namespaces/2003/03/serviceData"
  ns4:caBioSchema xmlns:ns3="http://xml.apache.org/xml-soap" xmlns:ns4="http://ogsadai.org.uk/namespaces/2003/07/gdsf" xmlns:xsi="http://www.w3.org/2001/XMLSchema"
    context xmlns=""
      gov.nih.nci.caDSR.bean.DataElementConcept id="CD82C5BE-111D-349D-E034-0003BA12F5E7"
      gov.nih.nci.caDSR.bean.DataElementConcept id="CD82C5BE-10E7-349D-E034-0003BA12F5E7"
      gov.nih.nci.caDSR.bean.DataElementConcept id="CD82C5BE-10C9-349D-E034-0003BA12F5E7"
      gov.nih.nci.caDSR.bean.DataElementConcept id="CD82C5BE-10BD-349D-E034-0003BA12F5E7"
```

This factory is exposing 4 concepts

Attributes and Relationships

The screenshot shows the OGSA Service Browser interface. At the top, the browser title is "[http://jforge.net:8080/ogsa/services/ogsadai/Panther1_caBIODataServiceFactory] - OGSA Service Browser". Below the title bar are navigation buttons: Back, Forward, New Window, Close, Refresh, and a checkbox for "Show dynamic gui". A search bar contains the URL "http://jforge.net:8080/ogsa/services/ogsadai/Panther1_caBIODataServiceFactory" with a "Go" button.

The main content area is divided into three tabs: "Services", "WSDL", and "Service Data". The "Service Data" tab is active, displaying a table with the following columns: Name, PortType, Type, Min, Max, Mutability, Modifiable, and Nillable.

Name	PortType	Type	Min	Max	Mutability	Modifiable	Nillable
serviceDataExtension	{http://www.gridfor...}	{http://www.gridfor...}	1	unbounded	static	raise	raise
terminationTime	{http://www.gridfor...}	{http://www.gridfor...}	1	1	mutable	false	false
createServiceExtensibility	{http://www.gridfor...}	{http://www.gridfor...}	1	unbounded	static	false	false
productInformation	{http://ogsadai.or...}	ProductInfoType	0	1	mutable	false	false
researchCenterInformation	{http://ogsadai.or...}	ResearchCenterIn...	0	1	mutable	false	false
caBioSchema	{http://ogsadai.or...}	{http://www.gridfor...}	0	1	mutable	true	false

Below the table is a "Values" section showing a tree view of service data. The tree structure is as follows:

- gov.nih.nci.caDSR.bean.DataElementConcept id="CD82C5BE-111D-349D-E034-0003BA12F5E7"
 - preferredName
 - preferreDefinition
 - longName
 - version
 - dateModified
 - dateCreated
 - publicId
- gov.nih.nci.caDSR.bean.DataElement id="CD82C5BE-1448-349D-E034-0003BA12F5E7"
 - preferredName
 - DiseaseId (highlighted in green)
 - data Type
 - publicId
 - 2178675
- gov.nih.nci.caDSR.bean.DataElement id="CD82C5BE-144F-349D-E034-0003BA12F5E7"
 - preferredName
 - DiseaseName
 - data Type
 - publicId
- gov.nih.caDSR.bean.DataElementConceptRelationship id="CD82C5BE-169C-349D-E034-0003BA12F5E7"
 - gov.nih.nci.caDSR.bean.DataElementConcept id="CD82C5BE-10E7-349D-E034-0003BA12F5E7"
 - preferredName
 - longName
 - preferreDefinition
 - publicId
- gov.nih.nci.caDSR.bean.DataElementConcept id="CD82C5BE-10E7-349D-E034-0003BA12F5E7"
- gov.nih.nci.caDSR.bean.DataElementConcept id="CD82C5BE-10C9-349D-E034-0003BA12F5E7"

Attributes

Relationship

OGSA-DAI Client - Registry

caGrid Registry

OGSA-DAI Graphical Client Demonstrator v1.0

DAISGR Console

DAISGR Input Console

DAISGR U...

Service Data:

Registered Services:

DAISGR Output Console

```
*****
DAISGR's TerminationTime SDE:
*****
<?xml version="1.0" encoding="UTF-8"?>
<ns2:serviceDataValues
xmlns:ns2="http://www.gridforum.org/namespaces/2003/03/serviceData">
  <ns1:terminationTime ns1:after="infinity" ns1:before="infinity"
ns1:timestamp="2004-06-23T16:32:30.735Z"
xmlns:ns1="http://www.gridforum.org/namespaces/2003/03/OGSI"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:type="ns1:TerminationTimeType" />
</ns2:serviceDataValues>

Registered service URLs:

http://156.40.48.168:8080/ogsa/services/ogsadai/GeorgeTown_caBIODataServiceFactory

Registered service URLs:

http://cbiogrid101.nci.nih.gov:8080/ogsa/services/ogsadai/Duke_caBIODataServiceFactory

Registered service URLs:

http://cbiogrid102.nci.nih.gov:8080/ogsa/services/ogsadai/NCICB_caBIODataServiceFactory

Registered service URLs:

http://jforge.net:8080/ogsa/services/ogsadai/Panther1_caBIODataServiceFactory
```

© International Business Machines Corporation, 2002, 2003. © University of Edinburgh, 2002, 2003.

Services advertised

OGSA-DAI Client - Factory

OGSA-DAI Client - GDS

The screenshot displays the 'OGSA-DAI Graphical Client Demonstrator v1.0' application. The interface is divided into several sections:

- Registry DAISGR Console / Factory GDSF Console / gds GDS Console:** The active console is 'gds GDS Console'.
- GDS Input Console:** Contains a 'GDS URL' field with the value 'http://cbiogrid101.nci.nih.gov:8080/ogsa' and a 'Contact GDS' button.
- Service Data:** A text area for service data, with 'Find Service Data', 'Select All', and 'Get Service Data' buttons below it.
- Perform Docs:** A list of document paths, with the selected path being 'C:\Projects\kernel-grid\ogsadai-3.1\examples\caGRID\Scenarios\DiseasesAs'. Below the list are 'Browse...', 'Select All', and 'Perform Requests' buttons.
- GDS Output Console:** Displays the XML response from the GDS query. The output starts with 'GDS perform document results:' and contains an XML document with a root element 'gridDataServiceResponse'. The XML includes namespaces, status information, and a list of clinical trial protocols and diseases, such as 'T-cell' and 'Disease'.

Annotations on the screenshot:

- A callout box labeled 'Perform documents' points to the 'Perform Requests' button in the 'Perform Docs' section.
- A callout box labeled 'Query result set' points to the XML output in the 'GDS Output Console'.

© International Business Machines Corporation, 2002, 2003. © University of Edinburgh, 2002, 2003.

The Perform Document

```
<?xml version="1.0" encoding="UTF-8"?>
```

```
<gridDataServicePerform  
  xmlns="http://ogsadai.org.uk/namespaces/2003/07/gds/types"  
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"  
  xsi:schemaLocation="http://ogsadai.org.uk/namespaces/2003/07/gds/types  
  ../../../../schema/ogsadai/xsd/activities/activities.xsd">
```

```
<documentation>
```

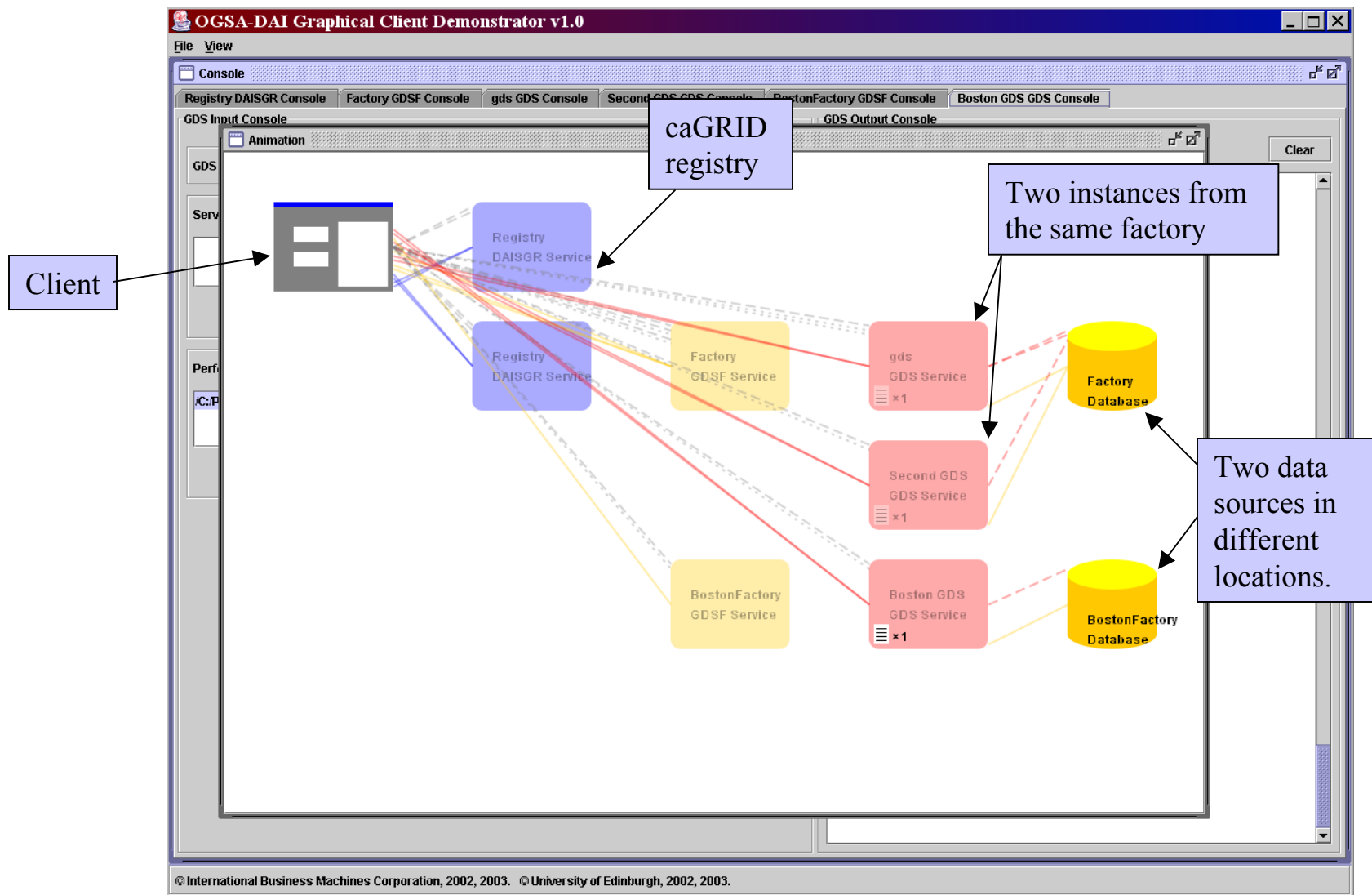
This example demonstrates how to parameterise an caBIO

```
</documentation>
```

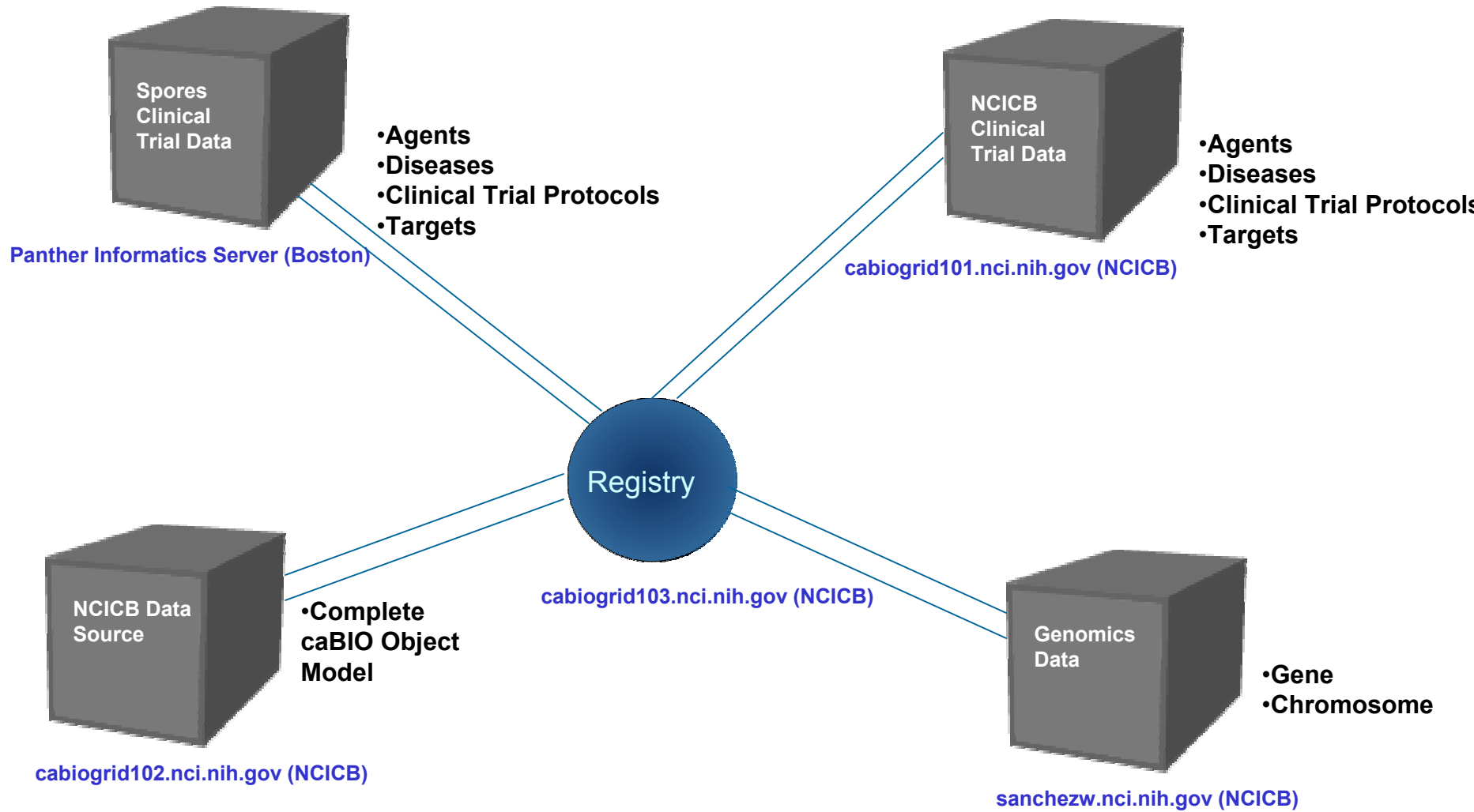
Show me all Trial
for Neoplasia

```
<caBioSimpleQuery name="caBIO2.0">  
  <criteria position="1" class="disease" version="1.0" id="abcd" attribute="name"  
value="neoplasia"/>  
  <query class="ClinicalTrialProtocol" version="1.0" id="">ClinicalTrialProtocol</query>  
  <output name="Test"/>  
</caBioSimpleQuery>  
</gridDataServicePerform>
```

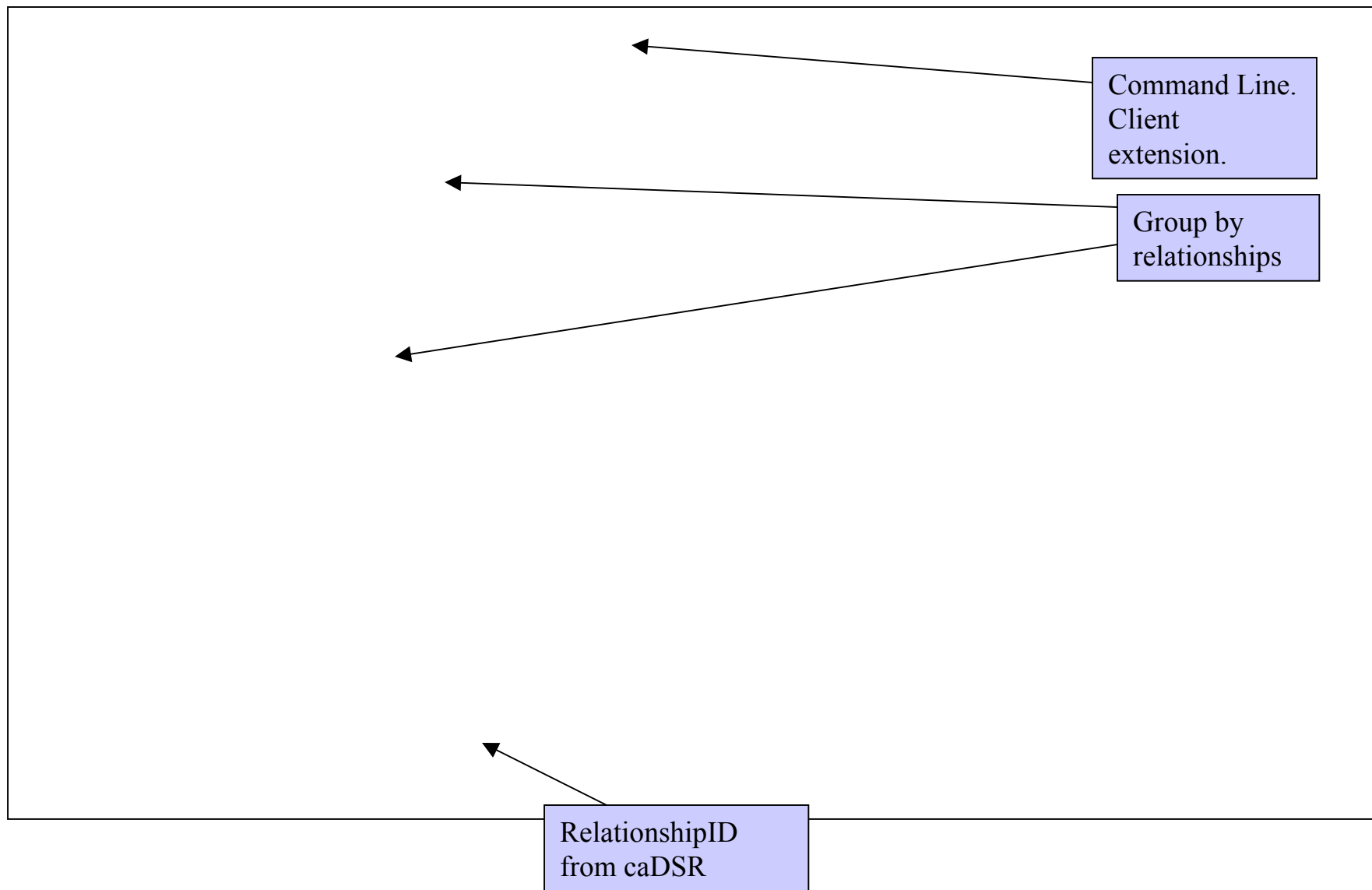
OGSA-DAI Animator – Accessing data sources



Prototype Scenario Setup – Discovery and Federated query



Discovery ClinicalTrialProtocol on the Grid



Federated query – Clinical trial protocol in Phase I

caQuery caDiscoveryResult.xml DA512A97-27D5-16D5-E034-0003BA0B1A09 ClinicalTrialProtocolPhaseISimple.xml

