CAGRID BROWSER

Overview

The caGrid browser is a web-based application that allows users to discover advertised caBIG grid resources and to query those resources for data of interest. The caGrid browser is an extension of the OGSA-DAI quick view tool and is written in Java Server Faces (JSF) and Jakarta Tiles. The caGrid 0.5 test bed has this browser software installed and is available as a web application. You can also optimally install it at your site to host the browser from local resources. For more information, see the caGrid 0.5 Technical Guide.

caBIG adopts a model-driven architecture best practice and requires that all data types used on the grid are formally described, curated, and semantically harmonized. These efforts result in the identification of common data elements, controlled vocabularies, and object-based abstractions for all cancer research domains. caGrid leverages existing NCI data modeling infrastructure to manage, curate, and employ these data models. Data types are defined in caCORE UML and converted into ISO/IEC 11179 Administered Components, which are in turn registered in the Cancer Data Standards Repository (caDSR). The definitions draw from vocabulary registered in the Enterprise Vocabulary Services (EVS), and their relationships are thus semantically described.

caBIG aims to connect data and tools from 50+ disparate cancer centers and a critical requirement of its infrastructure is to support the ability of researchers to discover these resources. caGrid enables this ability by taking advantage of the rich structural and semantic descriptions of data models and services that are available. Each service is required to describe itself using caGrid standard service metadata. When a grid service is connected to the caBIG grid, it registers its availability and service metadata with a central indexing registry service (Index Service). This service can be thought of as the "yellow pages" and "white pages" of caBIG. A researcher can then discover services of interest by looking them up in this registry. The caGrid browser is one of a series of high-level APIs and user applications for performing this lookup, which greatly facilitates the process.

As the Index Service contains the service metadata (or service data) of all the currently advertised and available services in caBIG, the expressivity of service discovery scenarios is limited only by the expressivity of the service metadata. For this reason, caGrid provides standards for service metadata to which all services must adhere. At the base is the Common Service Metadata standard that every service in caBIG is required to provide. This metadata contains information about the service-providing cancer center, such as the point of contact and the institution's name. Extending beyond this generic metadata are two standards that are specialized for the two types of community-provided services: Data Services and Analytical Services. Both of these standards leverage the data models registered in caDSR and link them to the underlying semantic concepts registered in EVS. The Data Services metadata details the domain model from which the objects being exposed by the service are drawn. Additionally, the definitions of the objects themselves are described in terms of their underlying concepts, attributes, and associations to other objects being exposed. Similarly, the Analytical Services metadata details the objects using the same format as the Data Services metadata. In addition to detailing the objects' definitions, the Analytical Services metadata defines the operations or methods the service provides. The input parameters and output of the operations are defined by referencing the appropriate object definition. In this way, both the data and analytical services fully define the domain objects they expose by referencing the data model registered in caDSR, and identify their underlying semantic concepts by referencing the information in EVS.

Launching the caGrid Browser

The caGrid browser consists of a series of web pages that are accessible from a Home page. Each page and its functionalists are described below.

The caGrid browser can be launched either from a local installation or from the caGrid browser that is part of the caGrid 0.5 test bed. To launch the test bed browser, enter the following URL:

http://cagrid-browser.nci.nih.gov

The browser opens to the Welcome page where users log in and where new features are advertised in the **What's New** and **Did You Know** sections (*Figure 1.1*).
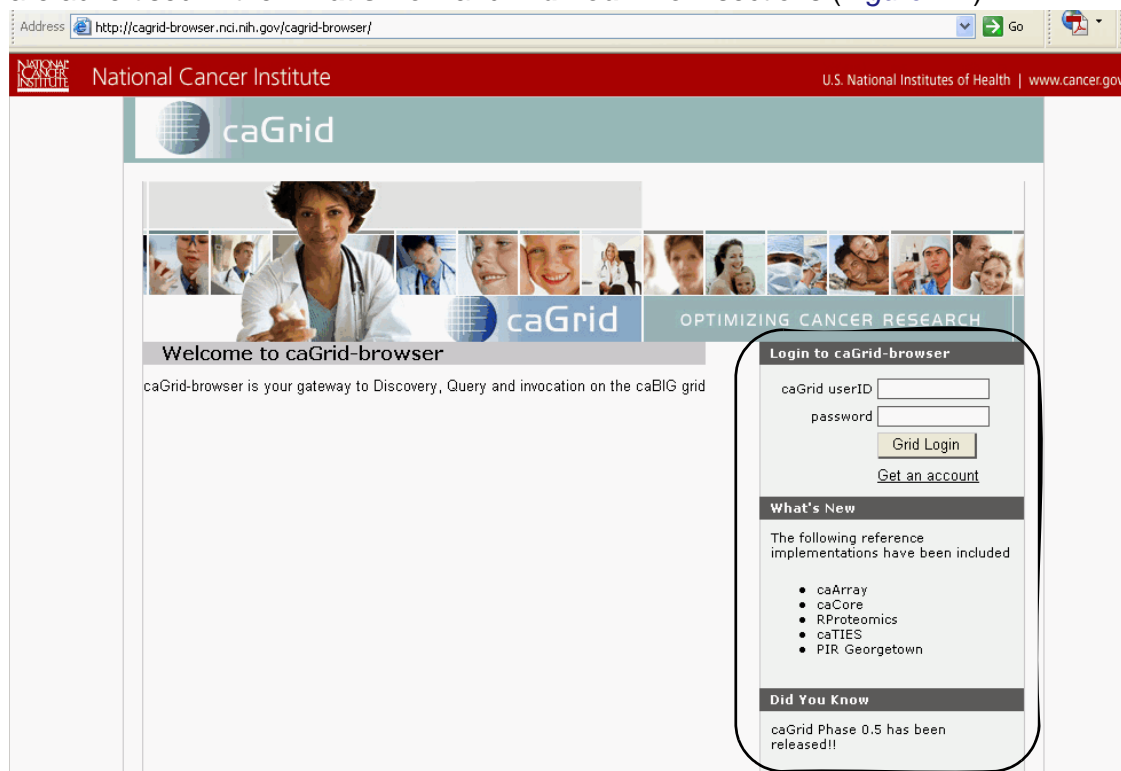

Figure 1.1 caGrid web browser opens to the Welcome page

Obtaining an Account

Before discoveries and queries can be configured from the browser, a user account and password must be obtained.

To apply for a user account to the caGrid browser, use the following steps.

1. Enter a user ID and password in the **Login to caGrid browser** fields (*Figure 1.1*) and click **Get an account**.

2. In the caGrid User Registration form that opens, enter the required information in the fields provided. In the User Information fields, the required fields are first and last name and e-mail address. In the Login fields, all fields are required. If the user/password is not valid, a red message displays adjacent to the field. Click **Get an account** at the bottom of the form.

3. Your request for registration is sent to a caGrid administrator and you will receive an email with your new account information in the next 48 hours.

Logging In

1. When you have received confirmation that your User ID and password have been established, enter them in the log in fields and click **Grid Login**. The **Home**, **Discovery**, **Help**, and **Settings** links display on the Welcome page and the Login fields change to a **Logout** button (*Figure 1.2*).

2. To logout, click the **Logout** button.

3. Your credentials are forwarded during communication with a secure grid service that requires user authentication and authorization.

Note: Each time you log in to the caGrid Browser a new session begins; logging out ends the current session. Services and data retrieved are not stored to your account between sessions.
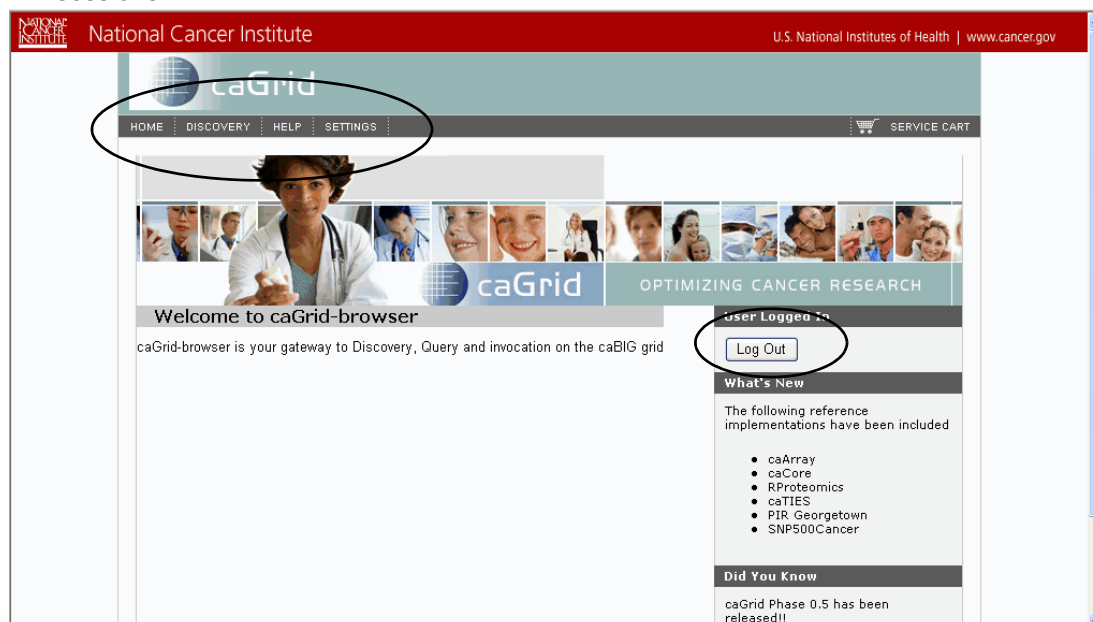
Figure 1.2 Welcome page displaying Home, Discovery, Help, and Settings links

The Home Page

The Home page is equivalent to the Welcome page and displays when you have successfully logged into the browser. It allows you to navigate to the Discovery, Settings, and Help pages by clicking on the links. You can also log out of the browser by clicking on the **Log Out** button on the Home page.

The Discovery Page

The main function of the caGrid browser is to support searching for (based on filtering criteria) and identifying caBIG resources. Discovery makes use of the rich semantic information provided by the advertised resources. Once a resource is discovered, the browser also allows a query to be performed on one or more data services on the grid to retrieve data of interest.

When a service is advertised, the owner of the caGrid service composes service metadata describing the service to the grid and publishes it to grid. The service metadata describes properties of the grid services that caGrid users and other grid services may search. For example, all services from a given cancer center can be located, data services exposing a certain domain model or objects based on a given semantic concept can be discovered, as can analytical services that provide operations that take a given concept as input.

Figure 1.3 Grid Metadata Based Discovery page

Configuring a Search

The **Discovery** link opens the Grid Metadata Based Discovery page where a search can be configured to discover caBIG-registered services (*Figure 1.3*). Keywords can be entered and metadata types and service types can be selected on this page. Once a service is identified and retrieved, one or more services can be saved to the session and a query can be performed on the data at those services. An explanation of each field and button on the Discovery page is provided below. The subsequent results and query pages are also described as is the Service Cart page.

Semantic Discovery Keyword(s)

The value entered in this field is used to narrow a search to only those services that have data related to the keyword.

Metadata Type

The metadata type checked in this field further narrows the entry in the Semantic Discovery Keyword(s) field.

- **EVS Concept** - encapsulates a concept in a vocabulary served by EVS. Enter EVS concept names here. For example, the EVS concept name for a genetic polymorphism is "C17004".
- **Object Class** - object class names as registered in the caDSR.
- **Research Center** - metadata attributes describing a Research Center.
- **All** - Searches EVS Concept, Object Class, and Research Center.

Service Type

The service type checked in this field can be either dependent on the two fields above it to further narrow the search or it can be used by itself to discover either data or analytical services.

- **Data Services** - the Data Services metadata details the domain model from which the objects being exposed by the service are drawn. Additionally, the definitions of the objects themselves are described in terms of their underlying concepts, attributes, and associations to other Objects being exposed.
- **Analytical Services** - the Analytical Services metadata details the objects using the same format as the Data Services metadata. In addition to detailing

the objects definitions, the Analytical Services metadata defines the operations or methods the service provides.

Discover Services

The **Discover Services** button performs a search for services that match the selected criteria. Matching services from a search are returned to the **Services Matching Keyword(s)** field below the **Grid Metadata Based Discovery** field. Once returned, a service can be selected by clicking on URL link to the service to perform further queries or the service can be added to a Service Cart, which saves it to the current session.

Retrieve All Services

The **Retrieve All Services** button performs a similar search as Discover Services but searches for all available caGrid-registered services at the time the search is performed. This search ignores any search criteria and/or keyword specified and retrieves all grid services registered in the caGrid index service.

Viewing Search Results

Add to Service Cart

Once services are discovered and the results displayed, the **Add to Service Cart** button saves the selected service for future use in the current session.

Note: In future releases, this functionality can be used to perform data integration or data federation among multiple data sources. Another use will be for pipelining services in a workflow process.

At any time during a current session, click the **SESSION CART** icon in the upper right corner of any page to display items in the Service Cart.

caGrid Data Service Details

Once services are discovered, each resulting URL link can be selected so that the details of that service display. The service details show the service metadata that

describes the service as well as a list of supported activities that can be performed on the service (*Figure 1.4*).

caGrid Data Service Details	
Product Information	
Product Name	caTIES
Product Version	1.0.3
Product Vendor	NCICB
Research Center Information	
Research Center Name	University Of Pittsburgh Medical Center
Research Center Type	Biomedical Research Organization
Research Center Address	UPMC Cancer Pavilion, PITTSBURGH PA, 15232
Phone #	7243615493
Fax #	4126475380
Point of Contact	Adita Nemlekar
Research Center Description	CaBIG Developer Institute
Research Center Comments	Comprehensive cancer center located in Pennsylvania and offering cancer prevention, diagnosis and treatment.
List of supported Activities	
caBIGXMLQuery caGrid XML query implementation.	
Grid Service URL	
http://128.147.244.97:80/ogsa/services/cagrid/caTIES	

Add to Cart

Object model information	
Class Name	**Description**
PathologyReport	Pathology Report
ConceptClassification	Concept Classification
Execution	Execution
Concept	Concept
Application	Computer Application
Patient	Patient
ConceptReferent	Concept Reference Object

Figure 1.4 caGrid Data Service Details

The service metadata displays by the following categories:

- Product Information - the product name, version, and vendor.
- Research Center Information - the name, address, contact number, and description of the research center hosting the service.
- Grid service URL - the URL for the registered service.
- List of supported activities - all data services support the caBIGXMLQuery (CQL) Activity. Clicking on this link opens the Query caGrid Service page, which is described below.
- Object model information - lists the class names and descriptions of each caDSR-registered object in the service's model. Clicking on a class name opens the Domain Object Details page for that class *Figure 1.5a*, which displays the

Class Attributes, Associations, and EVS Concepts. Clicking on the **Query Object** button opens the GUI Query Wizard (*Figure 1.5b*), which is explained in the next section.

a.

b.

Figure 1.5 (a) Domain Object Details page (b) GUI Query Wizard

Querying Data Services

Queries can be formulated in the caGrid Browser either by entering XML queries directly or by using a graphic-based query tool to construct the query.

Entering an XML Query

In the **List of Supported Activities** field on the service details page, the link **caGridXMLQuery** displays, which allows a query to be performed using the caGrid

XML query implementation. Clicking this link opens a form into which you can paste an XML document to submit a query to the data resource (*Figure 1.6*).

Figure 1.6 Query caGrid Service

Four sample queries are provided to four different data services; clicking on any sample query enters it in the form. The XML corresponding to the sample queries is provided in Table *1.1*. The examples can be used as a basis to form your own queries. An additional resource for writing your own query is the caBIG XML Language Schema. Once you have entered the query, click the **Send Query** button. The query is sent to the data resource, evaluated, and the results are sent back and displayed on the screen. For a

workflow example, see *Example One: Entering an Example Query in the Query Form* on page 12.

Sample Query	Description	XML
caArray	Queries the caArray database for protocol objects w/ a specific iden-tifier (1015897562219052).	``` <caBIGXMLQuery name="caArray"> <Target name="gov.nih.nci.mageom.domain.Proto- col.Protocol"> <Objects name="gov.nih.nci.mageom.search.Proto- col.Protocol"> <Property name="Id" predicate="equal" value="1015897562219052" /> </Objects> </Target> </caBIGXMLQuery> ```
caBIO	Queries the caBIO data-base for a taxon associ-ated with the gene BRCA1. (A taxon object in caBIO represents the various names (scientific, com-mon, abbreviated, etc.) for a species associated with a specific instance of a Gene, Chromosome, Pathway, Protein, or Tis-sue.)	``` <caBIGXMLQuery name="caBIO"> <Target name="gov.nih.nci.cabio.domain.Taxon" path="gov.nih.nci.cabio.domain.Taxon"> <Objects name="gov.nih.nci.cabio.domain.Gene"> <Property name="symbol" predicate="equal" value="BRCA1" /> </Objects> </Target> </caBIGXMLQuery> ```
SNP500Cancer	Queries the SNP500Cancer database for polymorphisms in the gene BRCA1	``` <caBIGXMLQuery name="SNP500Cancer"> <Target name="gov.nih.nci.pscc.snp500.domain.Po lymorphism" path="gov.nih.nci.pscc.snp500.domain.Po lymorphism"> <Objects name="gov.nih.nci.pscc.snp500.domain.Po lymorphism"> <Property name="geneSym- bol" predicate="equal" value="BRCA1" /> </Objects> </Target> </caBIGXMLQuery ```

Table 1.1 Sample XML queries

Sample Query	Description	XML

PIR	Queries the PIR protein database for protein objects w/ the gene name BRCA2.	`<caBIGXMLQuery name="PIR">`
		`    <Target name="edu.george-`
		`town.pir.domain.Protein">`
		`        <Objects name="edu.george-`
		`town.pir.domain.Gene">`
		`            <Property name="name" predi-`
		`cate="equal" value="BRCA2" />`
		`        </Objects>`
		`    </Target>`
		`  </caBIGXMLQuery>`

Table 1.1 Sample XML queries (Continued)

Building a Query from the GUI Query Wizard

Instead of entering a query in XML format, you can build a query from the metadata you are interested in from a graphical user interface. After you have selected a service and are viewing the caGrid Service Information page, scroll down to the **Object Model Information** at the bottom of the page. Click on the class for which you are interested in querying, which opens the Domain Object Details page for that class. Click the **Query Object** button at the bottom of the page. The GUI Query Wizard opens (*Figure 1.7*). The Instructions box provides a key to the contents of the top-level class folder. Object Attributes are contained in the small folders; Object Associations are contained in the larger folders. A red arrow indicates you can enter a value in a field. For a work-flow example, see *Example Two: Building an Example Query in the GUI Query Wizard* on page 14.
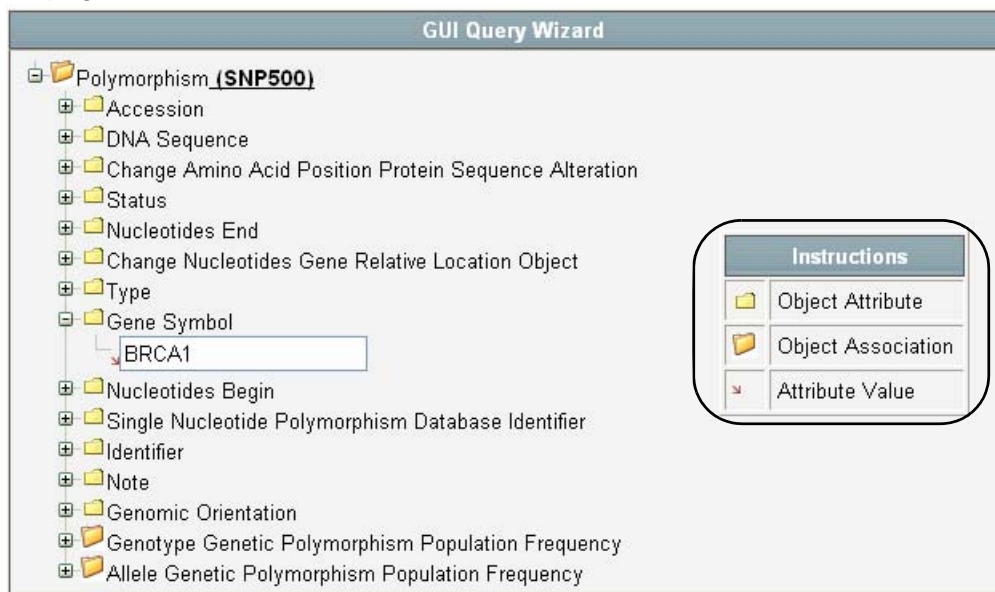
Figure 1.7 The GUI Query Wizard

The Settings Page

Setting the Index Server

The **Settings** link opens the page on which parameters are configured (*Figure 1.8*). For the 0.5 testbed browser, the only setting that can be configured on this page is the index service. The index service is a mechanism to implement virtual organizations. The caGrid browser is configured by default to use the caGrid 0.5 release's index service running at the NIH. To change to another index service, select one from the pull down menu in the **Index Server URL** field.
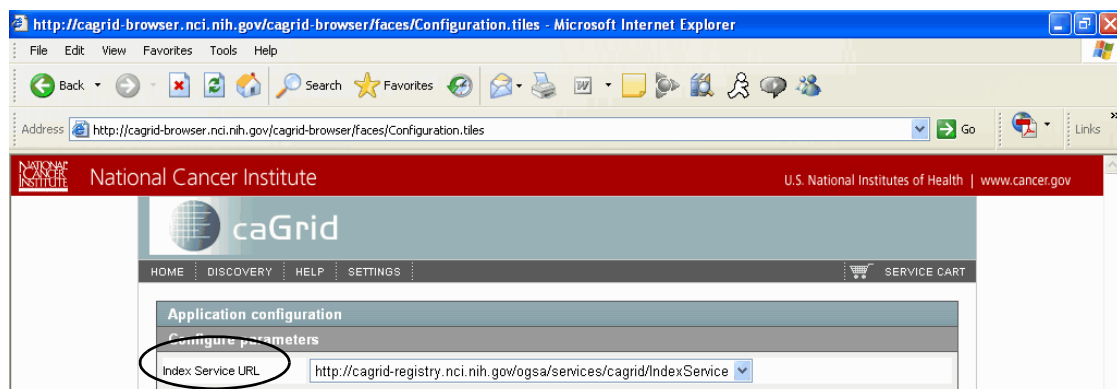
Figure 1.8 Settings page

The Help Page

The Help page provides help topics related to the caGrid Browser and provides links to the caGrid homepage, the caGrid Technical Guide, and the caBIG XML query language schema.

Workflow Examples

The following examples of using the caGrid Browser query the SNP500Cancer project, which is a registered caGrid service. The goal of this project is to resequence 102 reference samples to find known or newly discovered single nucleotide polymorphisms (SNPs) which are of immediate importance to molecular epidemiology studies in cancer.

Example One: Entering an Example Query in the Query Form

The following example illustrates using the Discover Services query to search all registered data services for the object "Polymorphism". The SNP500 Cancer Sample query is then selected, which queries the SNP500 Cancer database for polymorphisms in the gene BRCA1.

1. Configure the Grid Metadata Based Discovery page as shown in *Figure 1.9*.

Figure 1.9 Discovering services by keyword "Polymorphism"

2. The service listed in *Figure 1.10* are obtained. Click on the service hyperlink to open the service details page.

Figure 1.10 Search results

3. On the service details page, click the **caBIGXMLQuery** hyperlink to query the SNP500 Cancer data service. On the Query caGrid Service page that opens, click the **SNP500Cancer Sample Query** hyperlink. The sample query entered searches the SNP500 Cancer data service for polymorphisms in the BRCA1 gene. Click **Send Query**.

Figure 1.11 SNP500Cancer sample query

4. The results are returned in XML (*Figure 1.12*).

Figure 1.12 Results of SNP500Cancer sample query

Example Two: Building an Example Query in the GUI Query Wizard

The following example illustrates using the Discover Services query to search all registered data services for the EVS Concept ID for "Polymorphism". The SNP500 Cancer Sample query is then built in the GUI Query Wizard, which queries the SNP500 Cancer database for polymorphisms in the gene BRCA1.

1. Configure the Grid Metadata Based Discovery page as shown in *Figure 1.13*.

Figure 1.13 Discovering services by EVS concept ID for "Polymorphism"

2. The service listed in *Figure 1.14* are obtained. Click on the service hyperlink to open the service details page.

Figure 1.14 Search results

3. On the Service Details page that opens, scroll down to the Object Model Details at the bottom of the page and click on the hyperlink for **Polymorphism** (*Figure 1.15*).

Object model information	
Class Name	**Description**
PolyGenoFrequency	Genotype Genetic Polymorphism Population Frequency
Polymorphism	Genetic Polymorphism
PolyAlleleFrequency	Allele Genetic Polymorphism Population Frequency
Study	Control Group

Figure 1.15 Object Model Information

4. The Domain Object Details page opens for the class Polymorphism, listing its attributes and associated objects. Click the **Query Object** button at the bottom of the page.

5. The GUI Query Wizard opens (*Figure 1.16*). Click the "+" sign next to the Polymorphism folder to show the attributes and associated objects. Click the "+" sign next to the **Gene Symbol** attribute and enter **BRCA1** in the field adjacent to the red arrow. Click **Create Query**.

Figure 1.16 The GUI Query Wizard

6. The query is converted in the Query form in XML and is the equivalent of the SNP500Cancer Sample query used in the previous example, which searches the SNP500 Cancer data service for polymorphisms in the BRCA1 gene (*Figure 1.17*). Click **Send Query**.

Note: Multiple values can be entered in a single field and must be separated by a comma; for example, "Gene Symbol" = "BRCA1,BRCA2". Multiple attributes

and/or associations can also be entered in a single query; for example, "Gene Symbol" = "BRCA1" and "Nucleotides Begin" = "ggta"

Figure 1.17 SNP500Cancer sample query

7. The results are returned in XML (*Figure 1.18*).

Figure 1.18 Results of SNP500Cancer sample query