

# CABENCH-TO-BEDSIDE v2.0

## *End User Manual*

### Document Change History

Version Number	Date	Contributor	Description
V1.0	July 12, 2007	Washington University/Persistent Systems	Reviewed
V2.0	June 9, 2008	Washington University/Persistent Systems	Reviewed



**caBIG™** cancer Biomedical  
Informatics Grid™

an initiative of the National Cancer Institute

This is a U.S. Government work.

July 14, 2008

<b>Development Teams</b>		
<b>Development</b>	<b>End User Manual</b>	<b>Program Management</b>
Chandrakant Talele <sup>1</sup>	Chandrakant Talele <sup>1</sup>	Dr. Rakesh Nagarajan <sup>2</sup>
Srinath Krishnamurthy <sup>1</sup>	Srinath Krishnamurthy <sup>1</sup>	Srikanth Adiga <sup>1</sup>
Jasbir Sachdeva <sup>1</sup>	Srikanth Adiga <sup>1</sup>	
Deepak Shingan <sup>1</sup>	Madhumita Shrikhande <sup>1</sup>	
Chetan Patil <sup>1</sup>	Madhurima Bhattacharjee <sup>1</sup>	
Hrishikesh Rajpathak <sup>1</sup>	Mukesh Sharma <sup>2</sup>	
Lalit Chand <sup>1</sup>	Jahangheer Shaik <sup>2</sup>	
Atul Jawale <sup>1</sup>		
<sup>1</sup> Persistent Systems Limited	<sup>2</sup> Washington University	

<b>Contacts and Support</b>	
Training contact	<a href="mailto:cab2b@mga.wustl.edu">cab2b@mga.wustl.edu</a>
Customer support contact	<a href="mailto:cab2b@mga.wustl.edu">cab2b@mga.wustl.edu</a>

## Model caBIG™ Open Source Software License

v.2

Release Date: January 7, 2008

**Copyright Notice.** Copyright 2008 School of Medicine, Washington University in St. Louis (“caBIG™ Participant”). ca Bench-to-Bedside was created with NCI funding and is part of the caBIG™ initiative. The software subject to this notice and license includes both human readable source code form and machine readable, binary, object code form (the “caBIG™ Software”).

This caBIG™ Software License (the “License”) is between caBIG™ Participant and You. “You (or “Your”) shall mean a person or an entity, and all other entities that control, are controlled by, or are under common control with the entity. “Control” for purposes of this definition means (i) the direct or indirect power to cause the direction or management of such entity, whether by contract or otherwise, or (ii) ownership of fifty percent (50%) or more of the outstanding shares, or (iii) beneficial ownership of such entity.

**License.** Provided that You agree to the conditions described below, caBIG™ Participant grants You a non-exclusive, worldwide, perpetual, fully-paid-up, no-charge, irrevocable, transferable and royalty-free right and license in its rights in the caBIG™ Software, including any copyright or patent rights therein, to (i) use, install, disclose, access, operate, execute, reproduce, copy, modify, translate, market, publicly display, publicly perform, and prepare derivative works of the caBIG™ Software in any manner and for any purpose, and to have or permit others to do so; (ii) make, have made, use, practice, sell, and offer for sale, import, and/or otherwise dispose of caBIG™ Software (or portions thereof); (iii) distribute and have distributed to and by third parties the caBIG™ Software and any modifications and derivative works thereof; and (iv) sublicense the foregoing rights set out in (i), (ii) and (iii) to third parties, including the right to license such rights to further third parties. For sake of clarity, and not by way of limitation, caBIG™ Participant shall have no right of accounting or right of payment from You or Your sublicensees for the rights granted under this License. This License is granted at no charge to You. Your downloading, copying, modifying, displaying, distributing or use of caBIG™ Software constitutes acceptance of all of the terms and conditions of this Agreement. If you do not agree to such terms and conditions, you have no right to download, copy, modify, display, distribute or use the caBIG™ Software.

1. Your redistributions of the source code for the caBIG™ Software must retain the above copyright notice, this list of conditions and the disclaimer and limitation of liability of Article 6 below. Your redistributions in object code form must reproduce the above copyright notice, this list of conditions and the disclaimer of Article 6 in the documentation and/or other materials provided with the distribution, if any.
2. Your end-user documentation included with the redistribution, if any, must include the following acknowledgment: “This product includes software developed by School of Medicine, Washington University in St. Louis.” If You do not include such end-user documentation, You shall include this acknowledgment in the caBIG™ Software itself, wherever such third-party acknowledgments normally appear.
3. You may not use the names “School of Medicine, Washington University in St. Louis”, “The National Cancer Institute”, “NCI”, “Cancer Bioinformatics Grid” or “caBIG™” to endorse or promote products derived from this caBIG™ Software. This License does not authorize You to use any trademarks, service marks, trade

names, logos or product names of either caBIG™ Participant, NCI or caBIG™, except as required to comply with the terms of this License.

4. For sake of clarity, and not by way of limitation, You may incorporate this caBIG™ Software into Your proprietary programs and into any third party proprietary programs. However, if You incorporate the caBIG™ Software into third party proprietary programs, You agree that You are solely responsible for obtaining any permission from such third parties required to incorporate the caBIG™ Software into such third party proprietary programs and for informing Your sublicensees, including without limitation Your end-users, of their obligation to secure any required permissions from such third parties before incorporating the caBIG™ Software into such third party proprietary software programs. In the event that You fail to obtain such permissions, You agree to indemnify caBIG™ Participant for any claims against caBIG™ Participant by such third parties, except to the extent prohibited by law, resulting from Your failure to obtain such permissions.
5. For sake of clarity, and not by way of limitation, You may add Your own copyright statement to Your modifications and to the derivative works, and You may provide additional or different license terms and conditions in Your sublicenses of modifications of the caBIG™ Software, or any derivative works of the caBIG™ Software as a whole, provided Your use, reproduction, and distribution of the Work otherwise complies with the conditions stated in this License.
6. THIS caBIG™ SOFTWARE IS PROVIDED "AS IS" AND ANY EXPRESSED OR IMPLIED WARRANTIES (INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, NON-INFRINGEMENT AND FITNESS FOR A PARTICULAR PURPOSE) ARE DISCLAIMED. IN NO EVENT SHALL THE SCHOOL OF MEDICINE, WASHINGTON UNIVERSITY IN ST. LOUIS OR ITS AFFILIATES BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS caBIG™ SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

# Contents

<b>About This Guide .....</b>	<b>1</b>
Purpose.....	1
Typical User.....	1
Topics Covered.....	1
Text Conventions Used .....	2
<b>Chapter 1 Introduction .....</b>	<b>3</b>
Conducting Research using caB2B.....	3
Getting started with caB2B.....	4
Understanding Categories and Limits .....	4
What is a Category?.....	4
Why do we need categories?.....	5
What is a Limit?.....	5
What is a Query?.....	5
Starting the Application.....	5
<b>Chapter 2 Searching Data .....</b>	<b>9</b>
Search data – The Query Wizard .....	9
Choose Search Category .....	10
Advanced Search.....	11
Define Limit.....	12
Limit Set – Diagrammatic View.....	14
DAG Operations.....	15
Connecting the Limits.....	16
The Information Panel.....	19
Reset, Previous and Next Buttons .....	20
Define Search Results View .....	20
The Next and Previous Buttons .....	21
View Search Results.....	21
Record Details.....	22
The Next, Previous and Save Condition Buttons.....	27
Save Query .....	27
Why to Save Query? .....	27
How to Save Query? .....	28
Executing a Saved Query .....	29
Viewing and Saving Data List.....	31
Operations on the Data List .....	33
<b>Chapter 3 Experiments .....</b>	<b>35</b>
What are Experiments and Projects? .....	35
Creating an Experiment.....	36
Viewing the Existing Experiments.....	37
Opening an Experiment.....	38
Working with an Experiment.....	39
The Operations panel.....	39
Visualization panel .....	39
Filtering Data .....	40
Custom Data Category.....	43
Analyzing Data.....	45
Visualizing Data .....	48
<b>Appendix A Examples of category search.....</b>	<b>52</b>

Text-based Search - For category name information .....	52
Text-based Search - For category and attribute name and description .....	53
Concept Code Based Search - For category name information .....	54
Concept code based search - For attribute name information.....	55
<b>Appendix B Define Limit: Operators and Values.....</b>	<b>57</b>
<b>Appendix C Record Details View .....</b>	<b>59</b>
Administrator defined categories .....	59
<b>Appendix D Types of Filters.....</b>	<b>61</b>
Using the Range Filter .....	61
Using the Pattern Filter .....	62
Using the Enumerated Values Filter .....	62
<b>Index .....</b>	<b>63</b>

# About This Guide

This section introduces you to caBench-to-Bedside. Topics in this section include

- Purpose
- Typical User
- A typical user of the caB2B application would be a translational research scientist. The term translational research means research which typically begins at “the bench” with basic research — in which scientists study disease at a molecular or cellular level — then progress to the clinical level, or the patient's “bedside.” As this application supports the research at the bench as well as at the bedside, scientists would be the typical users.
- Topics Covered
- Text Conventions Used

## Purpose

---

This manual provides the details of how to use **cancer Bench-to-Bedside** (caB2B) application to search data sets of interest, create an experiment from data sets, analyze data, and visualize data.

All the chapters in this user’s manual begin with an introductory statement and are followed by a list of topics included in that chapter.

## Typical User

---

A typical user of the caB2B application would be a translational research scientist. The term translational research means research which typically begins at “the bench” with basic research — in which scientists study disease at a molecular or cellular level — then progress to the clinical level, or the patient's “bedside.” As this application supports the research at the bench as well as at the bedside, scientists would be the typical users.

## Topics Covered

---

- "Chapter 1: Introduction" explains the scope and function of caB2B
- "Chapter 2: Searching Data" describes process of retrieving data by building metadata based queries
- "Chapter 3: Experiments" explains how to create and use virtual experiments using the searched data
- "Appendix A: Examples of category search"
- "Appendix B: Define Limit: Operators and Values"
- "Appendix C: Record Details View"
- "Appendix D: Types of Filters"

## Text Conventions Used

This section lists the conventions that are used in this guide. The various typefaces represent interface components, keyboard shortcuts, toolbar buttons, dialog box options, and text that you type.

<b>Convention</b>	<b>Description</b>	<b>Example</b>
<b>Bold</b>	Highlights names of option buttons, check boxes, drop-down menus, menu commands, command buttons, or icons.	Click <b>Search</b> .
<u>Underlined Blue Text</u>	Indicates a Web or e-mail address.	<a href="http://domain.com">http://domain.com</a>
text in SMALL CAPS	Indicates a keyboard shortcut.	Press ENTER.
text in SMALL CAPS + text in SMALL CAPS	Indicates keys that are pressed simultaneously.	Press SHIFT + CTRL.
<i>Italics</i>	Highlights references to other documents, sections, figures, and tables.	See <i>Figure 4.5</i> .
<i>Italic boldface monospace type</i>	Represents text that you type.	In the <b>New Subset</b> text box, enter <b><i>Proprietary Proteins</i></b> .
<b>Note:</b>	Highlights information of particular importance.	<b>Note:</b> This concept is used throughout this document.
{ }	Surrounds replaceable items.	Replace {last name, first name} with the Principal Investigator's name.



# Chapter 1 Introduction

This chapter explains the scope and function of the cancer Bench to Bedside (caB2B) application. It helps you understand how to use caB2B to satisfy diverse biomedical research needs.

This chapter contains the following topics:

- Conducting Research using caB2B
- Getting started with caB2B
- Understanding Categories and Limits
- Starting the Application

## Conducting Research using caB2B

The **N**ational **C**ancer **I**nstitute **c**ancer **B**iomedical **I**nformatics **G**rid™ (NCI caBIG™) project creates a common, extensible informatics platform that integrates diverse data types and supports interoperable analytic tools. caBIG™ is developing separate applications that will facilitate individual steps involved in micro-array analysis. These applications are also useful to bioinformaticians.

caGrid is the infrastructure for caBIG that helps integrate these applications. caGrid can be used to perform investigation involving data integration and analytical services from diverse research communities.

**caBench-to-Bedside (caB2B)** is an application that leverages these tools in a user-friendly graphical user interface (GUI). caB2B can be used by the physician scientist to perform operations such as:

- **Query any caGrid data service to obtain data**

The data service can be a single data service, multiple data services, or a combination of the two services using semantically interoperable Common Data Elements (CDEs).

For example, caB2B allows investigators to query tissue banks at multiple cancer centers, design studies that focus on very specific tumor subtypes and to target less common tumors by pooling bio-specimen resources. This alleviates the problem of small sample size in a study.

- **Collect data and create experiments**

caB2B also enables investigators to perform novel in silico experiments using microarray data that have been archived. These capabilities facilitate identifying genes that are up or down regulated in specific cancers with respect to normals and enable investigators to view the data in the context of biological pathways. Investigators gain further understanding of the complex system of cancer biology by identifying genes that are important to the development and treatment of cancer. These capabilities may lead to effective identification of novel drug targets and improved treatment strategies.

- **Perform analyses** by using various grid-enabled analytical services.

caB2B enables individual users to work on common platforms to access caGrid resources in real time from common data and analytical services.

- **Visualize analysis results** by using charts and dendrograms.

## Getting started with caB2B

---

caB2B provides a platform to bring the laboratory research to the clinical level. caB2B enables information to be shared along with the continuum from the scientific bench to the clinical bedside and back. The tool helps in querying various caGrid enabled services. For example, it is possible to query for all the up-regulated and down-regulated genes of a malignant tissue sample by using the services available in caB2B. Thus, caB2B helps bridge the gap between the bench and the bedside.

- To start caB2B client, refer to the Installation Guide.
- To search for data sets, refer to "[Search data – The Query Wizard](#)".
- To design experiments, refer to "[Experiments](#)".

## Understanding Categories and Limits

---

### What is a Category?

A category is a collection of *searchable* attributes. You can use a category to search, view, and add data to experiments. Every domain object in an application is, trivially, a category. For example, Participant is a domain object from *caTissue* Core application.

Examples:

The Participant category consists of the following attributes:


- Identifier
- First Name
- Last Name
- Gender
- Race

The Gene Annotation category consists of the following attributes:

- Entrez Gene ID
- Gene Symbol
- Gene Name
- GenBank Accession Number
- Taxonomy ID
- Chromosome

## Why do we need categories?

Categories are designed to help you to easily locate all the data elements needed for your search.


	<p>Categories are created by the administrator.</p> <p>Your administrator can create new categories based on your requirements. If there are queries that you perform regularly, discuss them with your administrator.</p>
---	--

## What is a Limit?

A limit is a condition (or criterion or predicate or condition) on a category. It is the smallest unit of a query. For example, *female participants*

## What is a Query?

The collection of limits is a query. For example, *female participants who have DNA specimens collected from their breast*. This query has two limits: *female participants*, *DNA specimens from breast*. There are two categories involved in this query: *Participant and Specimen*.

	<p>Your administrator can create a new category that contains relevant data elements from Specimens and Participants. This enables entering all the limits in a single category.</p>
---	--

## Starting the Application

For details on installing and starting the caB2B client, refer to the *Client Installation* chapter of the Installation Manual.

When caB2B application is launched, you see the caB2B Login page as shown in Fig. 1.1.



*Figure 1.1. caB2b Login Page*

To access the caB2B client, the user must be a valid caGrid user. Currently, all the training grid users and the production grid users having a Dorian Identity Provider are supported. To access the caB2B application, the following steps must be followed:

1. Type the **User Name** and **Password**.
2. Select an **ID Provider** from the dropdown.
3. Click **Login** to start the application.

If the user provided credentials are authentic, access is granted to the application else access to the application is denied and an error message is displayed as shown in Figure 1.2



*Figure 1.2. Login Error*

When the user is authenticated and the caB2B application is launched, you see the caB2B Home page as shown in Figure 1.3.

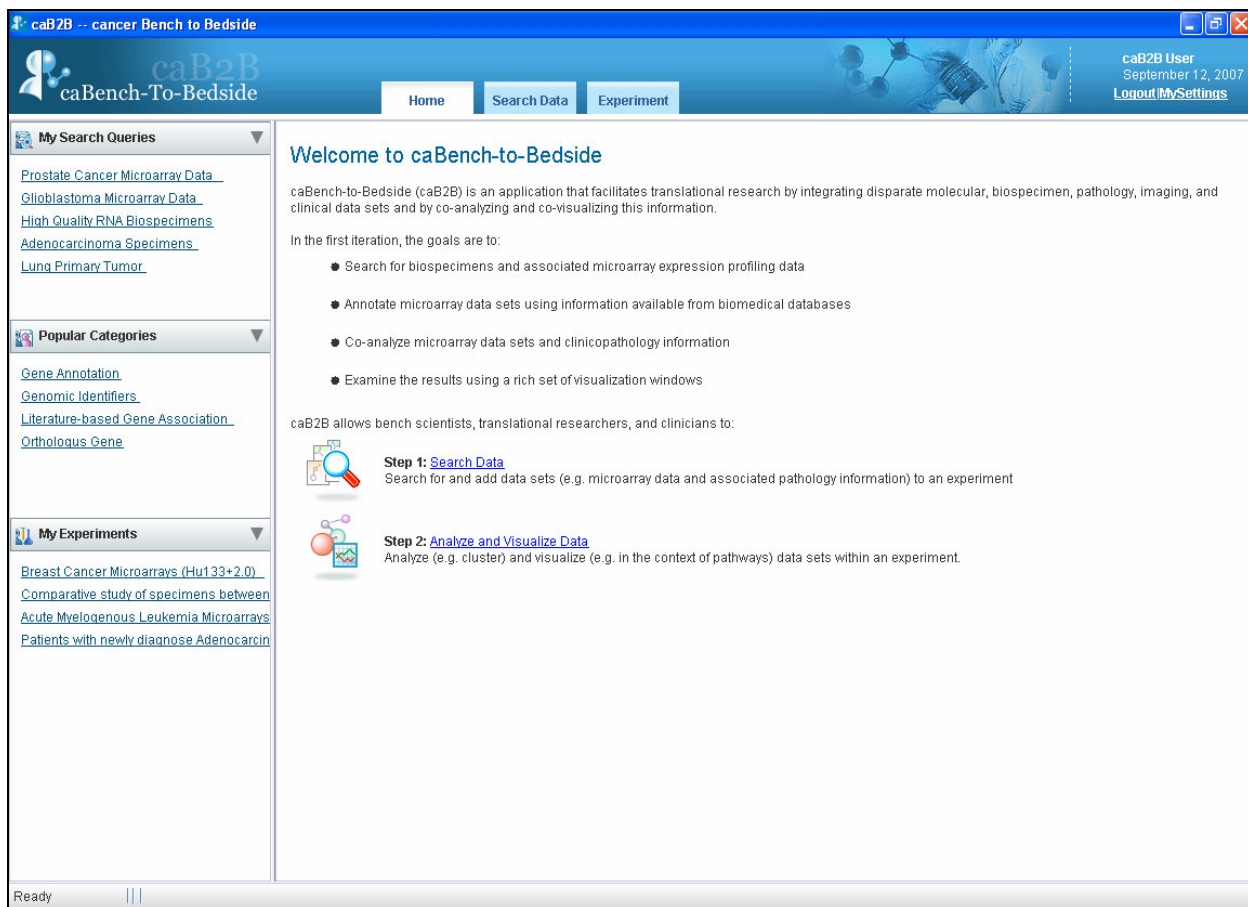


Figure 1.3. caB2B Home page

As mentioned in the Conducting Research using CaB2B section, a caB2B end user typically performs the following steps:

1. Execute queries to obtain the dataset(s) of interest. Details of this step are provided in the “Searching Data” chapter.
2. Analyze and visualize the data. Details of this step are provided in the “Experiments” chapter.

---

## Chapter 2 Searching Data

This chapter describes how caB2B can be used to fetch data from available caGrid data services like caArray, GeneConnect, caFE, caTissue etc.

The chapter is broadly divided into the following subtopics:

- Search data – The Query Wizard
- Choose Search Category
- Define Limit(s)
- Define Search Results View
- View Search Results
- Save Query
- Viewing and Saving Data List

### Search data – The Query Wizard

---

This section details the steps required to obtain data. Click the **Search Data** tab on the **Home** page to display the **Search data for Experiment** wizard. The wizard begins at the Choose Search Category tab. (See *Figure 2.1. Begin search for data*).

The wizard guides you through the following steps

Step one	Step two	Step three	Step four	Step five
Choose Search Category	Add Limits	Choose an Output Category (Save Query)	View Search Results (Save Query)	View Data List

To search data:

1. Choose Search Category: Search for Category to apply limits.
2. Add Limits: Add limits on the selected Category
3. Choose Output Category (Optional): Select the output format (that is, define the results view). User can also save the query in this step
4. View Search Results: View the results view in a Google-like search results page and add data to the data list (or shopping cart). User can also save the query in this step
5. View Data List: View the data that you have added in the data list and create an Experiment.

## Choose Search Category

This step helps you to search for the category on which you want to define the limits. For example, if you have a query with limits on participants, you can use this feature to first find the Participant category.

To find the categories, do the following:

1. Type the search string in the text box (near the top of the screen). Gene, gender, RNA, experiment are examples of search strings.
2. Click **Search** or press ENTER.
3. Categories that match the entered string are displayed as hyperlinks along with a short description.
4. Once a Category of your choice is found, click the desired search result to proceed to the next step.

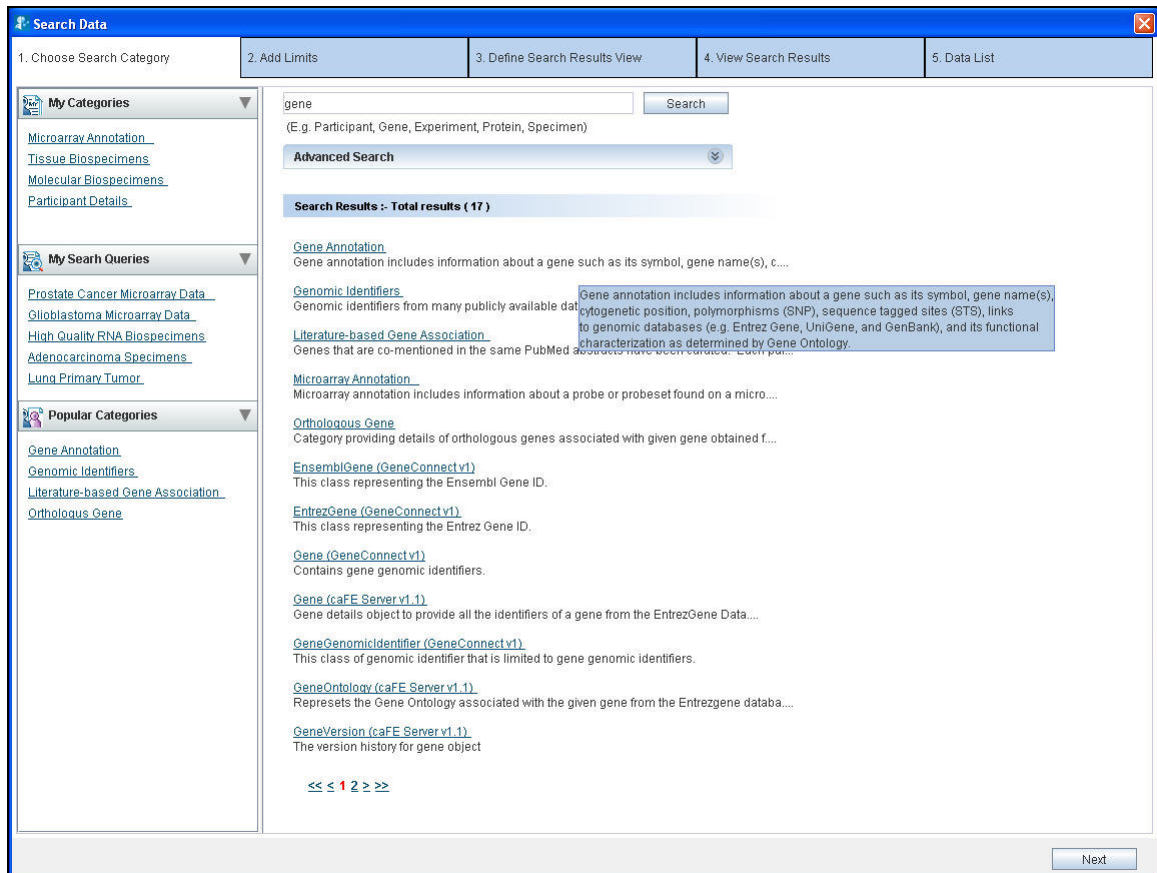


Figure 2.1. Begin search for data

More on searching:

- The search performed is partial and is not case-sensitive. Hence, the search for terms such as *Gene*, *gen* or *GENE* returns the same results.



- If a category description is long, you can view the complete description as a tool-tip by positioning the mouse over that description.
- If the number of matching categories is large, the system displays the results in multiple pages. Click the page numbers at the bottom to navigate through different pages.
- You can separate the multiple search terms by spaces or tabs. The system searches for each term individually and displays a collective set of matching categories. For example, if you enter the search string *Participant Gene*, the categories that match either *Participant*, *Gene*, or both will be found.

## Advanced Search

Expanding the **Advanced Search** box displays the advanced options that help you refine the category search. The Figure 2.2 depicts the default selection of the various options

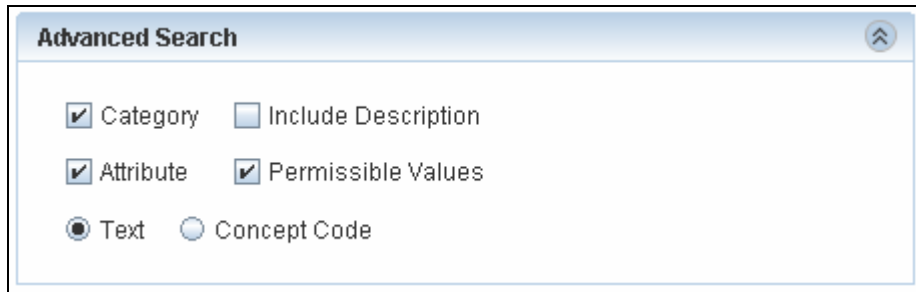


Figure 2.2. Advanced Search

If the system returns too many results for a Category search, you may use the Advanced Search to narrow your search.

Table 2.1 Advanced Search Options (Checkboxes and radio buttons)

Option	Description
<b>Check Boxes</b>	
Category	Includes the category names in the search.
Attribute	Attribute (also known as data element) is the lowest unit of the searchable items. A category contains many attributes. Select this check box to include the individual attribute names in the search. For example: Select <i>attribute</i> and search for <i>gender</i> . You will find the category Participant since it contains an attribute by name <i>Gender</i> .
Permissible value	Some attributes contain a fixed set of values. For example, <i>Gender can be Male, Female or Unspecified. Organ can be heart, lung, breast, and so on.</i> This fixed set is known as <b>Permissible Values</b> . Select this check-box to include the permissible values in the search. For example, if you select the permissible value as <i>Male</i> the system

	displays the Participant category since it contains an attribute <i>Gender</i> which contains a permissible value <i>Male gender</i> .
Include Description	Includes the description text in the search. <b>Note: Include Description</b> check-box is enabled only for a <b>text based search</b> when you select the <b>Category</b> check box, the <b>Attribute</b> check box, or both.
<b>Option Buttons</b>	
Text	The system matches the search string that you type with the name of the category, its attributes, or the permissible values based on the text. This is simple, most commonly used and the default way of searching for a category.
Concept Code	The system treats the search string as an <b>Enterprise Vocabulary Services (EVS)</b> concept code and matches it with the concept code of the category, its attributes, or the permissible values thereby showing the matching categories.  More details on the concept code can be found at <a href="http://nciterms.nci.nih.gov/NCIBrowser/Dictionary.do">http://nciterms.nci.nih.gov/NCIBrowser/Dictionary.do</a> by connecting to <b>NCI_Thesaurus</b> and <a href="http://ncimeta.nci.nih.gov/MetaServlet/">http://ncimeta.nci.nih.gov/MetaServlet/</a>

Additional refinements that may be made with these options are illustrated in the *Appendix A: Examples of category search*

## Define Limit

A limit is a condition (or criterion or predicate) on a category and multiple limits together form a query. Table 2.2 outlines two example queries and the limits added to them.

Table 2.2

<b>Example</b>	<b>Category</b>	<b>Limit</b>
Get the details about all the African male participants	Participant	Gender equals Male Race equals African
Biospecimens from prostate adenocarcinoma	Specimen Characteristics	Tissue Site equals prostate
	Specimen Collection Group	Clinical diagnosis equals adenocarcinoma

Screen for define limit is shown below. It has 3 parts,

1. Category search is present on the left hand part. It is explained in section [choose search category](#)
2. *Define Limit* present at right hand side top is explained in section [Steps for defining a limit](#)
3. *Limit set* present at right hand-side bottom is explained in section [Limit Set – Diagrammatic View](#)

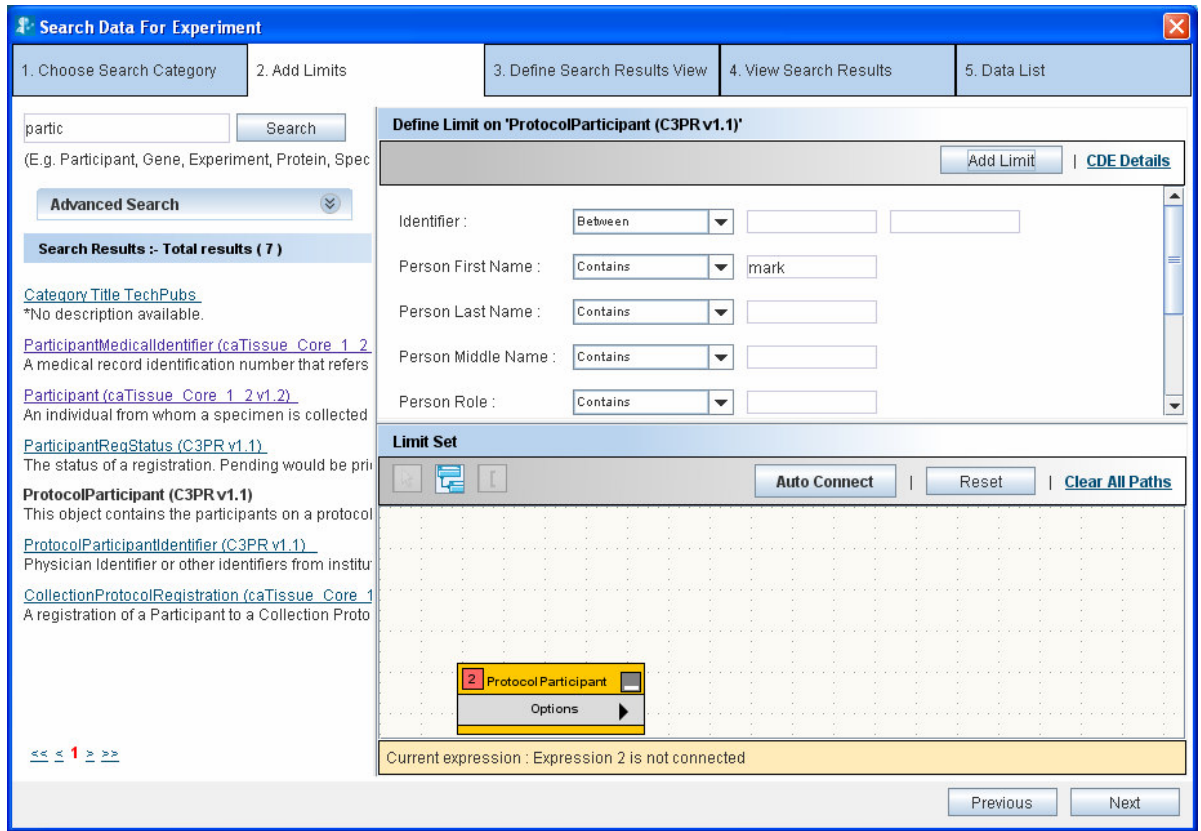



Figure 2.3 Define Limit

### Steps for defining a limit

Perform the following steps to define a limit:

1. Choose an operator from the list of operators adjacent to the attribute you wish to constrain.
2. Specify the constraining values.

	<p>For detailed explanation of various operators and how to specify values, refer to <i>Appendix 2: Define Limit: Operators and Values</i>.</p>
---	---

Example:

Single constraint – Show all male participants

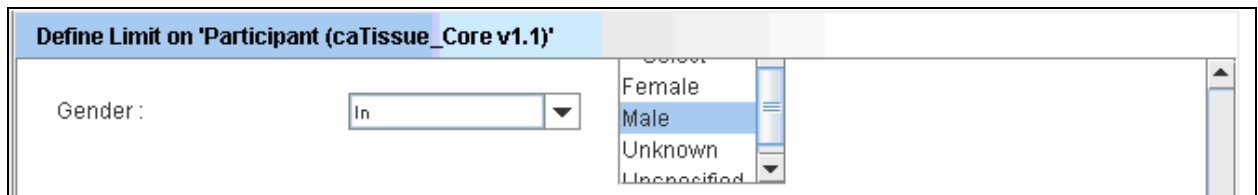


Figure 2.3. Male participants

Example:

Multiple constraints within one limit: male participants who are alive. Multiple conditions can be defined on an attribute by pressing CTRL key while selecting different options: Specimens which are either Serum or Plasma.

Figure 2.4. Male participants who are alive

**WARNING!**

Ensure that you click **Add Limit** at the top of the **Define Limits** screen before proceeding. In case you do not click **Add Limit**, the constraints that you have specified are not saved.

## Limit Set – Diagrammatic View

Once the limits have been set on the category, clicking on the Add Limit button adds the category in the form of a yellow rectangle (also called a node) into the Limit Set Panel. The **Limit Set** panel consists of the **DiAGrammatic** view (DAG view) panel. DAG view is like a canvas where all the categories on which limits have been added appear in the form of nodes (Figure 1.3). Whenever a new limit is added, the system adds a yellow rectangle that represents the limit being added to the DAG view. As you continue to add limits, the system adds corresponding rectangles to the DAG view.

These rectangles in the DAG view provide a graphical view of all the limits that have been added.

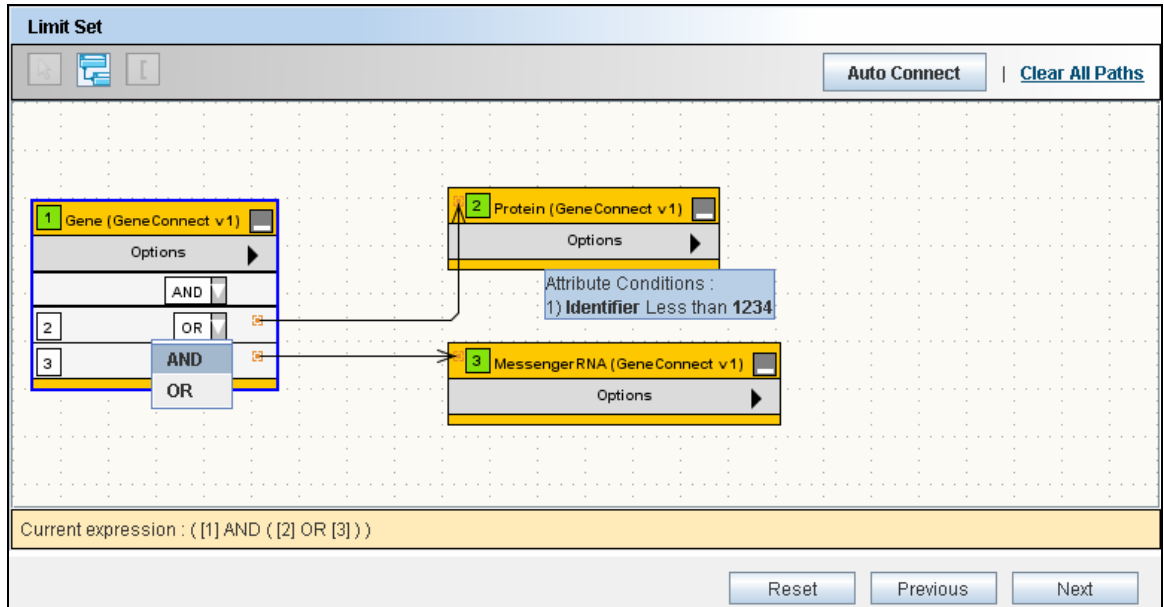
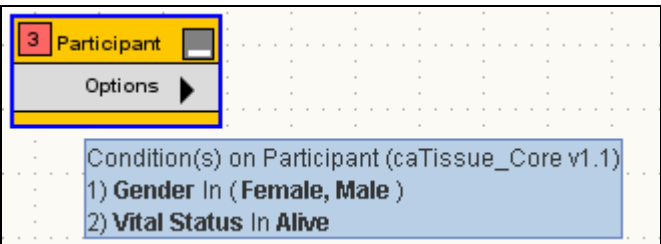
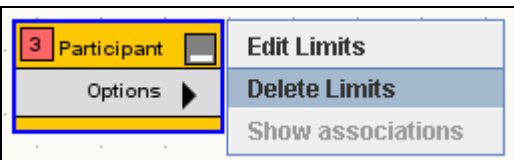


Figure 2.5. Limit Set Panel

## DAG Operations

You can perform the following operations within the DAG view.

Table 2.3

View details of the limit	<p>Position your mouse over a node to view the details of that limit.</p> 
Edit the limit	<p>Click Edit Limits from the Options menu of the node whose limits you wish to edit. Alternatively, double click the title of the node.</p>  <p>This opens the limit in the Define Limit panel in edit mode. Edit the constraints in the <b>Define Limit</b> panel, and click <b>Edit Limit</b>.</p>
Delete the limit	<p>Select the nodes that you wish to delete and press DELETE.</p> <p>Alternatively, you can delete a single node by selecting <b>Delete Limits</b> from the <b>Options</b> menu of the node as shown above.</p>

## Connecting the Limits

You must connect all the limits before proceeding to the next step.

When multiple limits are specified, you need to connect these limits to form an expression. Two limits may be connected by using either the **AND** or the **OR** operator. Following is an example of how the DAG looks when the limits are connected.

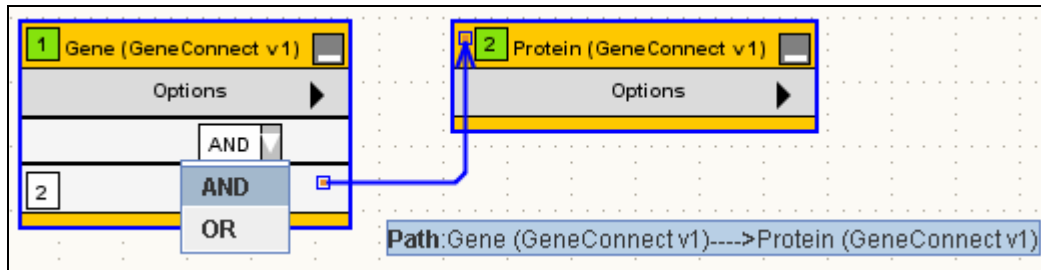


Figure 2.6 Connected Nodes

If there are only two limits in the query, you have to connect only those two rectangles

To connect the limits, you need to specify the following:

- The **path** to be taken in connecting the nodes
- The logical operator (**AND**, **OR**) in the dropdown box in a node

The rest of this section focuses on various ways in which you can specify the path(s)

### What is a Path?

A path tells the system how two categories are to be connected. A path is a way to travel from one category to another. There could be multiple ways to connect two categories and the results of the query differ based on the path chosen.

For example, consider the two categories *Specimen* and *Site* (from the caTissue Core application). A *Specimen* is associated with the *Site* category in two ways: 1) Site where the specimen was collected 2) Site where the specimen is *stored*.

That means there are two paths between *Specimen* and *Site*: 1) Through Specimen Collection Group for the collection site. 2) Through Storage Container

Query example 1: Show all the available specimens collected in Barnes Jewish Hospital. In this query, as you are specifically interested in a particular collection site, you should choose the path through Specimen Collection Group.

Query example 2: Show all the available specimens stored in Barnes Jewish Hospital. In this query, as you are specifically interested in a particular storage site, you must choose the path through Storage Container.

Therefore it is important to choose the right path for your query.

## Connecting Limits

Click **Auto Connect**. The system automatically connects to all the boxes based on the settings provided by your administrator.

However, the administrator might not have configured the path between all the categories. In such cases, you can manually connect the limits as described below.

Example query: For all the male participants, show all the available specimens collected in Barnes Jewish Hospital.

1. After adding three limits on the Participant, Specimen and Site the DAG view looks as shown in the figure below:

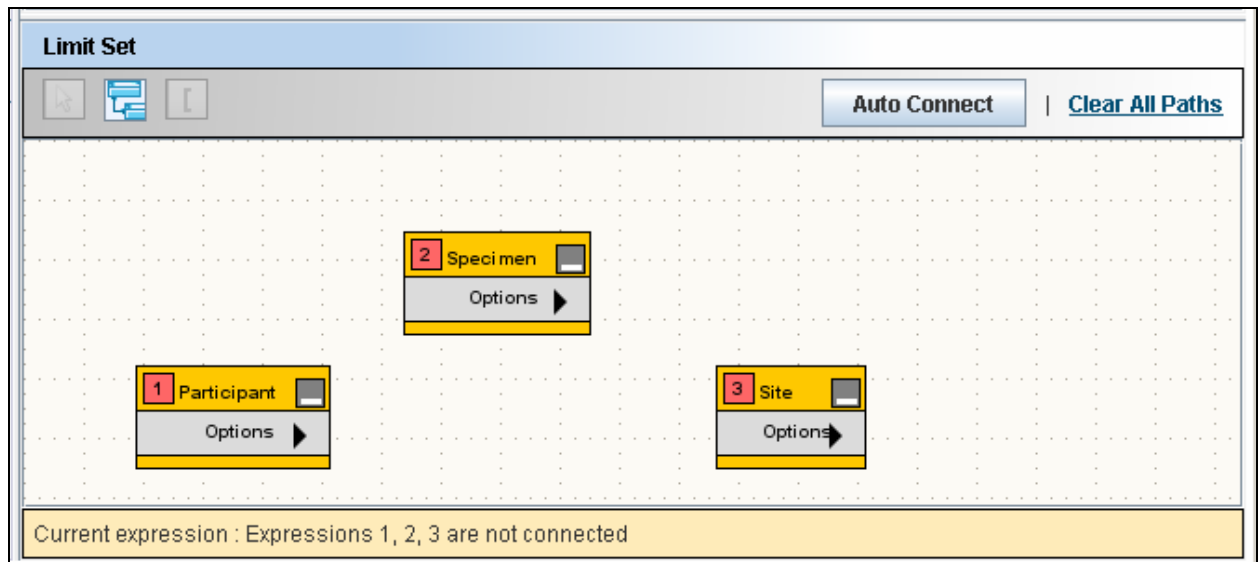



Figure 2.7. Limit Set

2. Click the Participant and the Specimen rectangles to select them.
3. Click **Connect Limits**  to connect the two nodes. The resulting DAG view is shown in *Figure.2.9*.

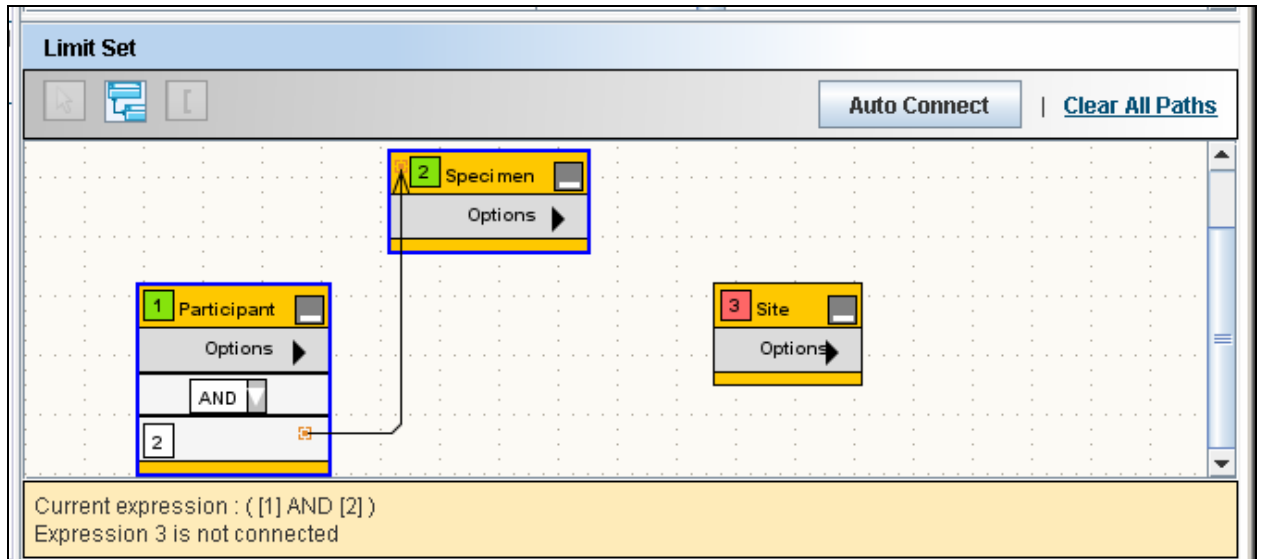



Figure 2.8. Limit Set with Two Connected Limits

**Note:** There is only one path between *Participant* and *Specimen*, therefore the system automatically connects it. Default operator is AND, it can be changed to OR using operator dropdown.

4. Select the *Specimen* and *Site* rectangles and click **Connect Limits** .
5. Since there are more than one path between *Specimen* and *Site*, the system cannot connect the limits on its own. So it displays the **Ambiguity Resolver** window. The **Ambiguity Resolver** window is shown in *Figure 2.9. Path Ambiguity Resolver*

The screenshot shows a 'Path Ambiguity Resolver' window with a blue header and a close button. It has two tabs: 'General Path' (selected) and 'Curated Path'. Below the tabs is a table with three columns: 'Select', 'Paths', and 'PathPopularity'. There are two rows of data, both with 'Select' checkboxes.

Select	Paths	PathPopularity
<input type="checkbox"/>	Specimen >> (specimenCollectionGroup) >> SpecimenCollectionGroup >> (site) >> Site	50 %
<input type="checkbox"/>	Specimen >> (storageContainer) >> StorageContainer >> (site) >> Site	50 %

Figure 2.9. Path Ambiguity Resolver

6. Select the first path to find out the collection site.
7. Once the path(s) is determined, the system creates a connection between the two nodes (see figure below). You can position the mouse over the connection to view the path(s) that you chose.



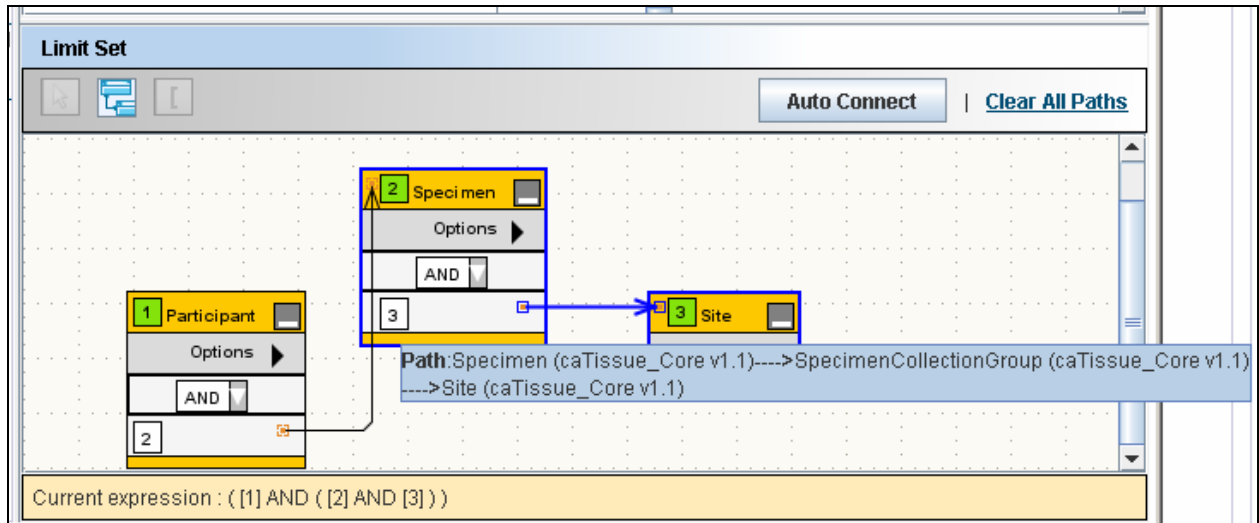


Figure 2.10. Limit Set with Three Connected Limits

8. Specify the operator (**AND** or **OR**) using the dropdown box in the node. For above query the operator will be **AND**
9. Click **Next**.

### Deleting a Connection

You can delete a connection by clicking the connection arrow between the two nodes and pressing the DELETE key.

### The Information Panel

The **Information** panel is located at the bottom of the **Limit Set** panel.

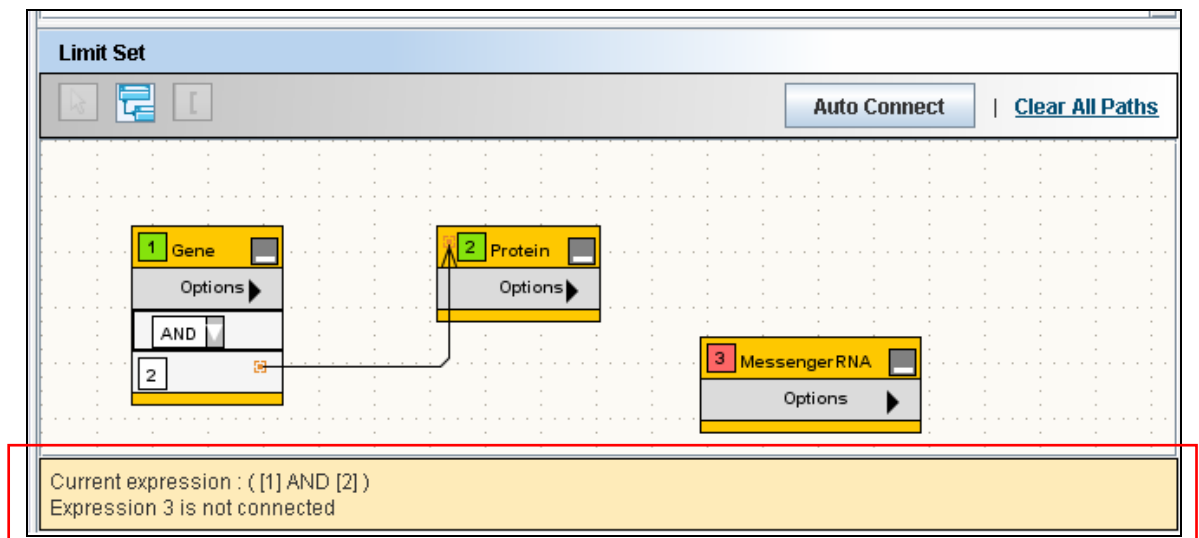


Figure 2.11. The Information Panel

The **Information** panel provides the following information:

- The expression formed as a result of how the limits are connected.

- Limits that not yet connected to any other limit
- To execute the query, there should not be any unconnected rectangles.

## Reset, Previous and Next Buttons

These buttons are located on the right hand bottom of the screen, see Figure 2.5. Limit Set Panel.

- The **Reset** button clears all the specified limits.
- The **Previous** button takes you to **Choose Search Category** step of the wizard.
- The **Next** button takes you to the **Choose Output Category** step of the wizard. A valid query allows you to proceed to the next step. If the query is invalid, an appropriate error message is displayed.

## Define Search Results View

After specifying the limits, user proceeds to choose the category for the data. For example, “show all the participants who had biospecimens collected from their breast”. *Participant* is the category that you want to view, based on a limit on the specimens.

The **Define Search Results View** tab allows you to select the category to be shown in the output of the query. It is also called as output category for the query.

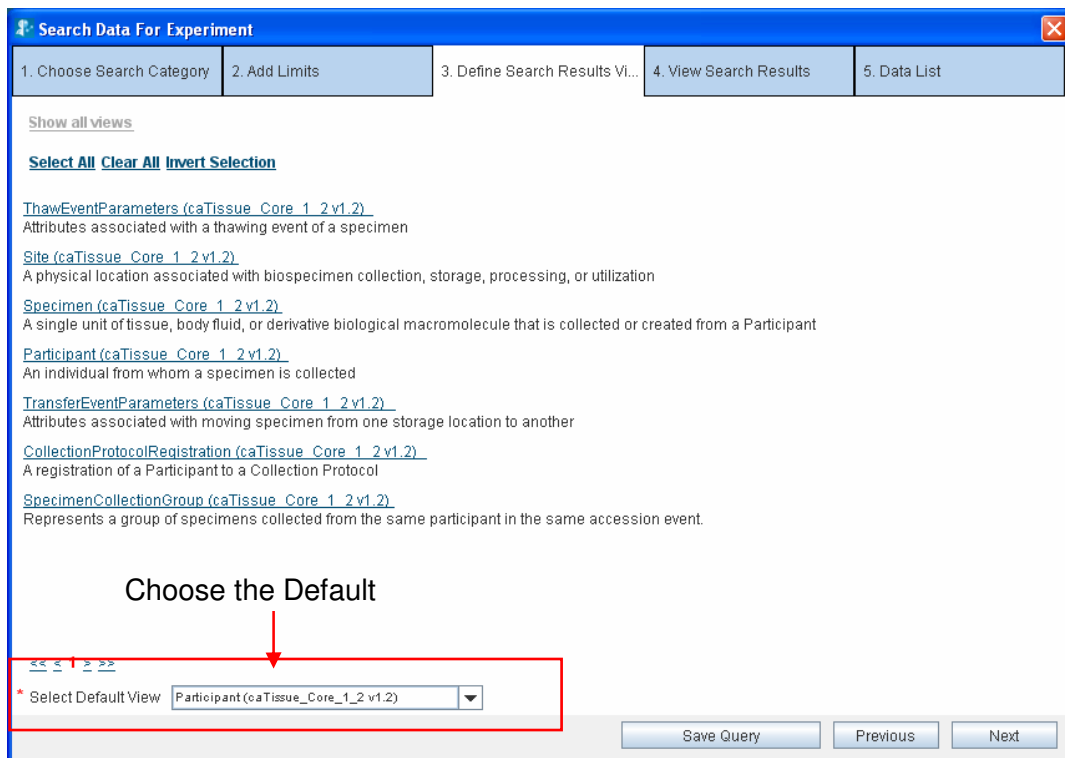



Figure 2.12. Define Search Result View.

---

All the categories with specified limits are shown in the **Select Default View** dropdown box at the bottom of the screen as shown in Figure 2.12. From the **Select Default View** dropdown box, select the category whose data you wish to obtain.

	Ignore the other contents of this screen such as the check boxes and hyperlinked categories. Those are intended to allow you to fetch data for multiple related categories in one go. However, this feature is not supported in this release of caB2B.
---	--

## The Next and Previous Buttons

- To edit the limits, click **Previous** and go to the Define Limit step
- Click **Next** to execute the query that you have created

<b>WARNING!</b>	On clicking the Next button, the system executes the query and fetches the data from appropriate data services over the caGrid. This step could be time-consuming. It may take few minutes to couple of hours based on the network speed and complexity of the query.
-----------------	--

## View Search Results

The **View Search Results** tab displays the results obtained by executing the query defined in the previous steps. The concept of viewing results in caB2B is similar to viewing results after a Google search. The following are the similarities:

- caB2B returns a set of records that match your limits.
- Some details of the record are displayed along with the record.
- Click on one of the records to view more details.
- Records provide additional information, for example, from Participants you can see the Collection Protocol in which they registered. This is similar to Google where, after you click one of the hyperlinks, you can traverse into other web pages by clicking the hyperlinks within that page.
- As in Google, the results are divided into multiple pages for easy navigation.

If you find any data of interest, you can add it to the **Data List**.

The following image displays a typical results view. The **View Search Results** tab displays the records that match the limits and also the data that you have already added to the data list.

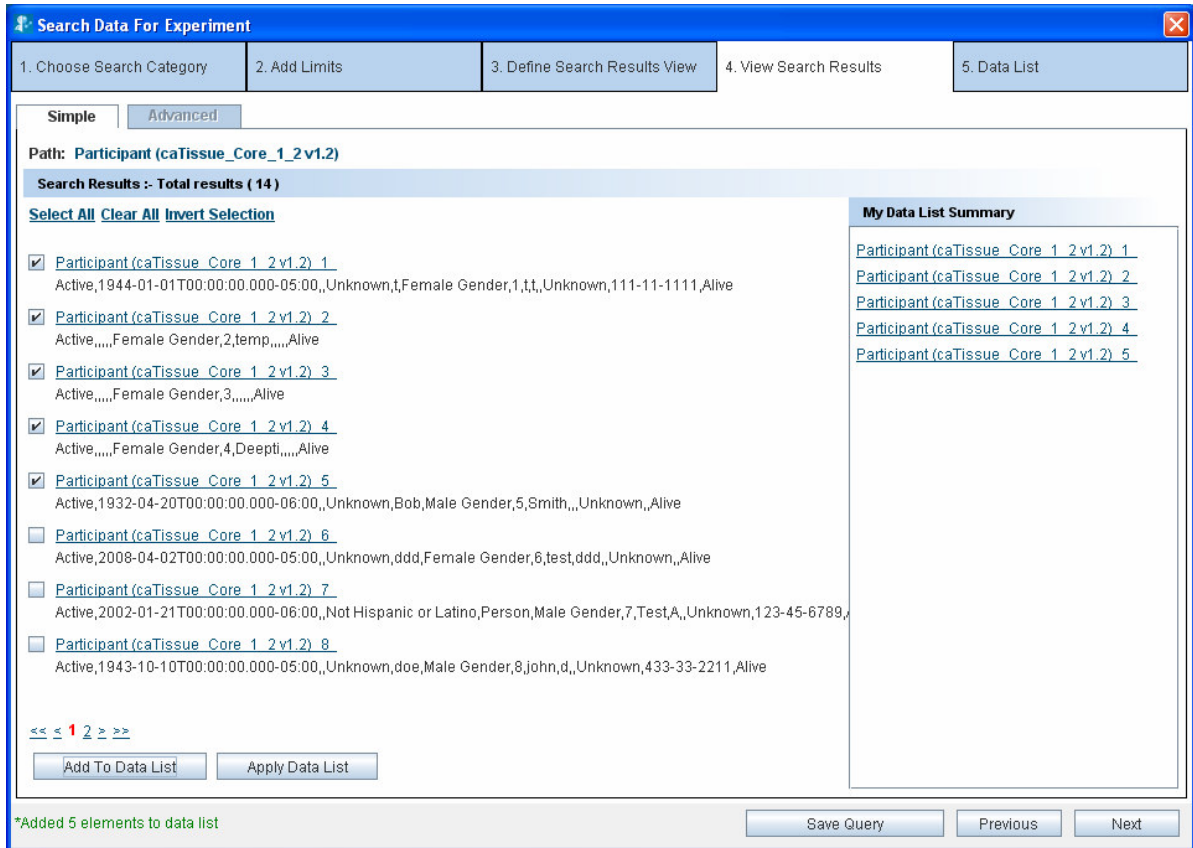


Figure 2.13. View Search Results

## Record Details

Click on the hyperlink of a record on the **View Search Results** to take you to the **Record Details** page. Here you can view the values of all the attributes of the category.

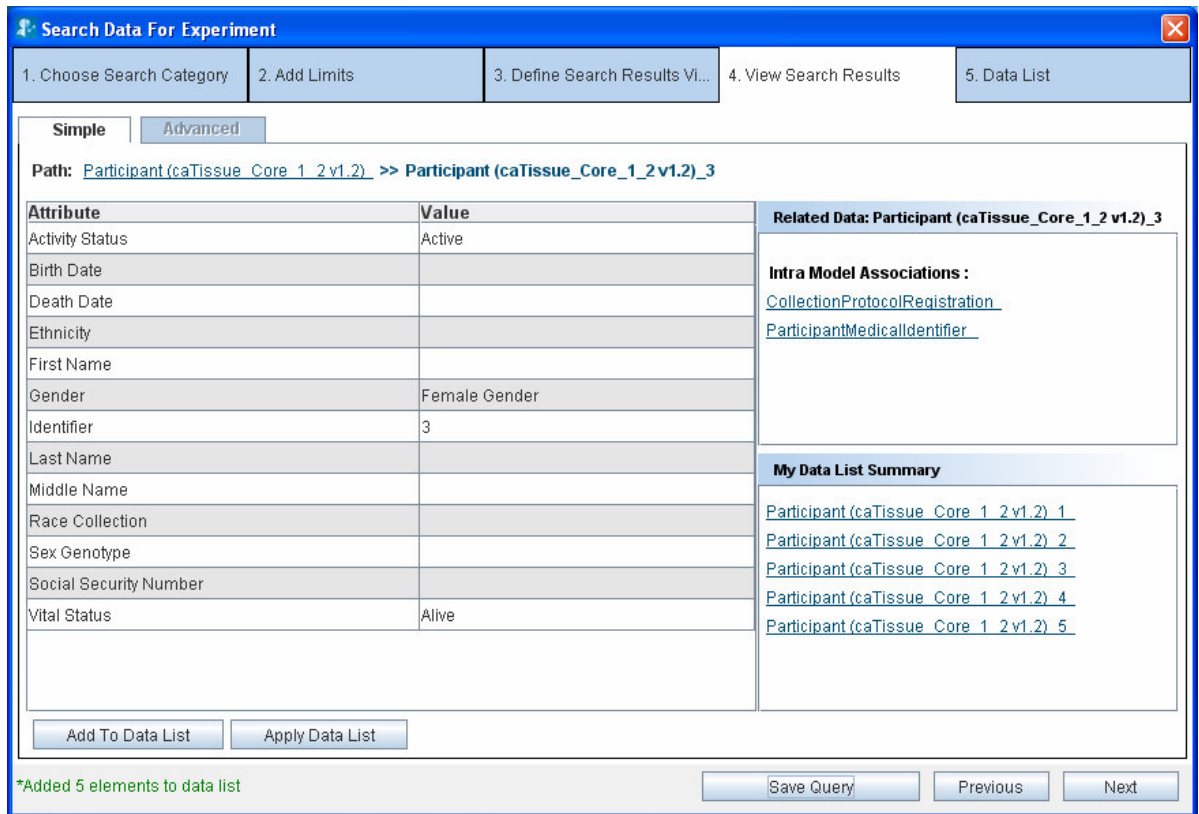


Figure 2.14. Results' Details

A view of the Results Details page is shown in *Figure 2. 14* and shows the following sections:

- Value of all the data elements in the record
- Links to related data
- Summary of the records that you have already added to the data list

## Related Data

As shown in Figure 2.14 above, the **Related Data** box on the right displays the categories that are related to the current category. Click on any of the categories to display records for that category. For example, if you are currently viewing records of *Participant (caTissue)*, the following categories are shown in the related data box:

- GenomicIdentifierSet (GeneConnect)
- Protein (GeneConnect)
- MessengerRNA (GeneConnect)
- Gene (caFE Server)

If you now click *Protein*, you would see the Proteins related to the Gene you are currently viewing.

## Breadcrumbs

Breadcrumbs or a breadcrumb trail is a navigation technique used in user interfaces. Its purpose is to give users a way to keep track of their location within programs or documents. (definition from Wikipedia entry)

As you traverse the results view, breadcrumbs are displayed at the top of the page. For example, if you are viewing a Protein record related to a Gene, you would see the path as

Gene >> Gene \_1 >> Protein >> Protein\_5

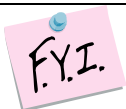
This means that you originally queried for Gene. Then you fetched related proteins for Gene\_1. From the proteins obtained, you are currently viewing details of Protein\_5.

Click any of the links in the breadcrumbs to navigate back to the corresponding page. For example, you can click *Gene* in the above path to go back to the page where you were shown the list of Genes.

## Adding data to the Data List

A data list is like a shopping cart which enables you to add interesting data to the shopping cart as you traverse through the results. You can add data to the data list in two ways:

- Click **Add to Data List**
- Click **Apply Data List**



As you add the data to the data list, they are reflected in the **My Data List Summary** box on the right hand side.

### Add to Data List

Select the record that you want to add to the data list and click **Add to Data List**. You can either add one record at a time or add multiple records by selecting the checkboxes placed along with the record summary.

### Using Apply Data List

Once you obtain the related data for a particular record, **Apply Data List** allows you to fetch similar related data for several other records easily. We outline an example of this approach below.

A typical user query might do the following:

1. Fetch the data for the *Gene* category.
2. Fetch related data from the category *Chromosome* for a record of *Gene*, for example, *Gene\_1*.
3. Add the *Chromosome* record to the data list.

The resulting data list is shown in Figure 2.15 (below).

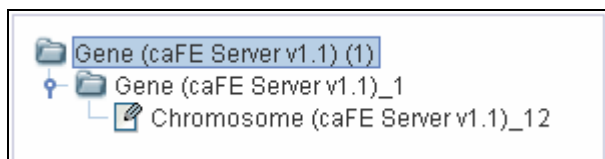


Figure 2.15. Data List

If you wish to obtain the *Chromosome* data for other *Gene* records, you have to fetch related *Chromosome* for one *Gene* record at a time. The **Apply Data List** functionality allows you to fetch the *Chromosome* for several *Gene* records in a single operation. To do this, perform the following steps:

1. Fetch the data for the *Gene* category.
2. Fetch the related data from the category *Chromosome* for a record of *Gene*, for example, *Gene\_1*.
3. Add the *Chromosome* record to the data list.
4. From the **View Results** screen for the *Gene* records, (see Figure 2.13. *View Search Results*) select *Gene* records for which you want the related *Chromosome* records.
5. Click **Apply Data List**, it will fetch *Chromosome* for all selected *Gene* records.
6. A message is displayed in the status bar when the operation is complete. Click **Next** to see the data list. . Figure. 2.16 show the resulting data list.

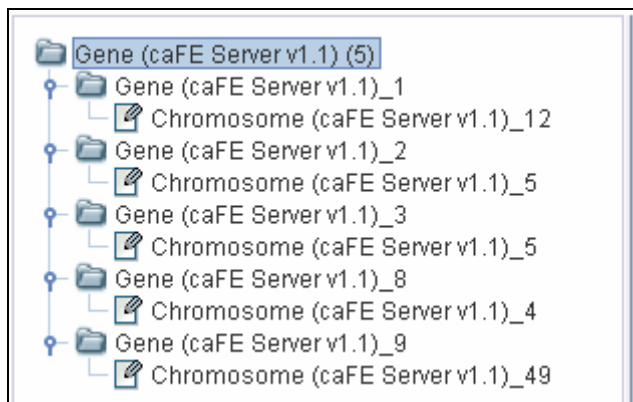


Figure 2.16. Result of Apply Data List to genes with id 2, 3, 8, 9

### How Apply Data List functionality works

To understand this functionality better, consider a slightly more complicated data list below:

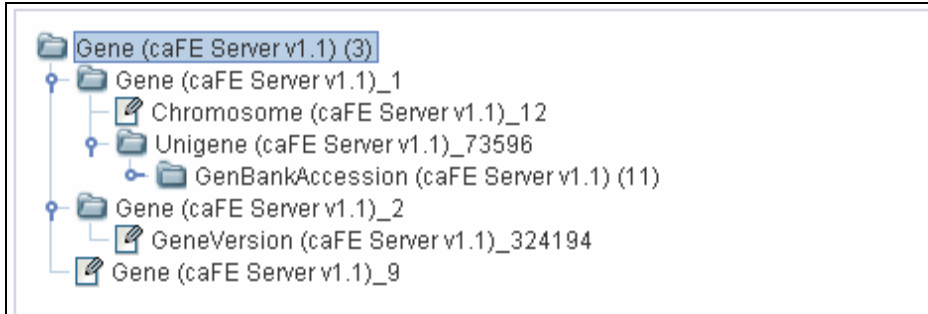


Figure 2.17. Example 3- Data List

### Example 3:

The following records have been added to the data list

- For *Gene\_1*, related *Chromosome*, *Unigene*, and *GenBankAccession* records are present.
- For *Gene\_2*, related *GeneVersion* is present.
- For *Gene\_9*, there is no other related data.

Now, **Apply Data List** to the genes *Gene\_1*, *Gene\_2* and *Gene\_9*. The resulting data list is shown in Figure 2.18.

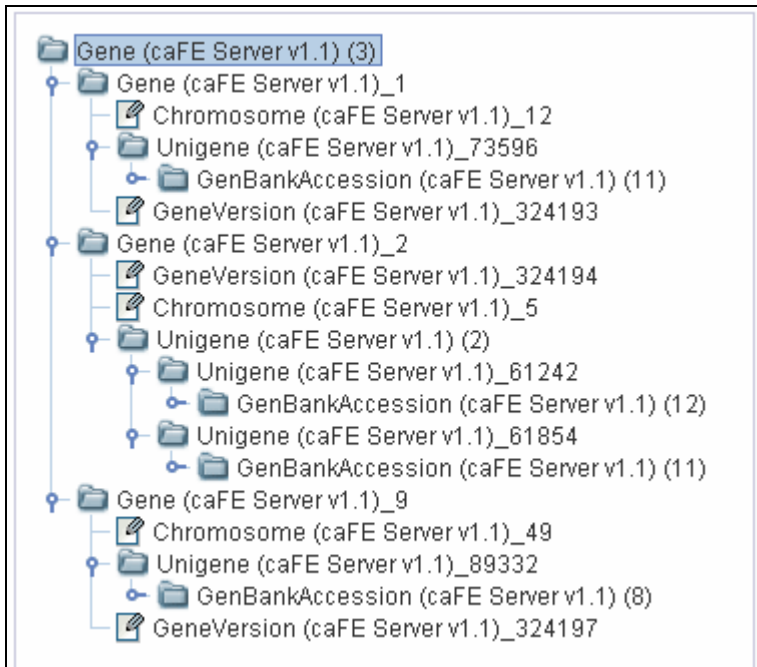


Figure 2.18. Example 3- Apply Data List

As seen in Figure 2.18, the following related new data has been fetched:

- For *Gene\_1*, the system fetches *GeneVersion* (as *Gene\_2* had *GeneVersion*).



- 
- For Gene\_2, the system fetches Chromosome, Unigene, and GenBankAccession, as these were the related data for Gene\_1.
  - For Gene\_9, the system fetches all of GeneVersion, Chromosome, Unigene, and GenBankAccession.

As a result, each gene contains the related data for *GeneVersion*, *Chromosome*, *Unigene*, and *GenBankAccession*.

Thus, the **Apply Data List** operation results in a homogenous structure in the data list. The system fetches related data for all the other records such as the related data present for one record.

## The Next, Previous and Save Query Buttons

- Click **Next** to view the contents of the data list. It will take you to *Data List* tab.
- Click **Previous** to go back to Choose Search Category, which is *Define Search Result View* tab.
- Click **Save Query** will popup a window to save the query. This is explained in detail in section below.

## Save Query

---

This option is used to save the user executed query. It saves all the conditions on the various classes as defined in the query.

### Why to Save Query?

When the user saves the executed query, he can simply view the result of the query by executing already saved query. This saves considerable amount of time as the user does not have to form the query again. The saved query can also be parameterized by choosing to save conditions on attributes which form the parameters (described further below). The user can execute a saved query with or without changing the conditions defined in the saved query.

### Advantages of Save Query can be listed as follows:

- Saving complex queries with multiple nodes
- Availability of saved queries to all the users to aid the users who might not know how to query
- Setting up parameterized queries that allow a user to use the same query for a variety of conditions on a single attribute
- Queries can be saved in two ways:
  - Non-parameterized: No parameters have been defined while saving the query. It means that the saved query cannot be modified at the time of execution. The conditions defined on attributes while querying cannot be changed during execution of the saved query.

- Parameterized: Parameters, that is, certain attributes are saved while saving the query. For example, during the execution of the query, the condition on the attributes can be changed to generate a new query.

## How to Save Query?

Once the user has formed the query, that is, added certain categories and connected them via the available paths, the query can be saved. After forming the query, user navigates to the Define Search Results View tab and then to the View Search Results tab, both these tabs provide the option to save a query. The **Save Condition** button at the bottom of both these tabs can be used to save the query.

### How to save a non-parameterized query?

1. Click **Save Condition** (as shown in *Figure 2.12. Define Search Result View*).
2. Enter the **Title** and **Description** for the query to be saved, in the save query panel (see the figure below).
3. To save the query, click **Save**.

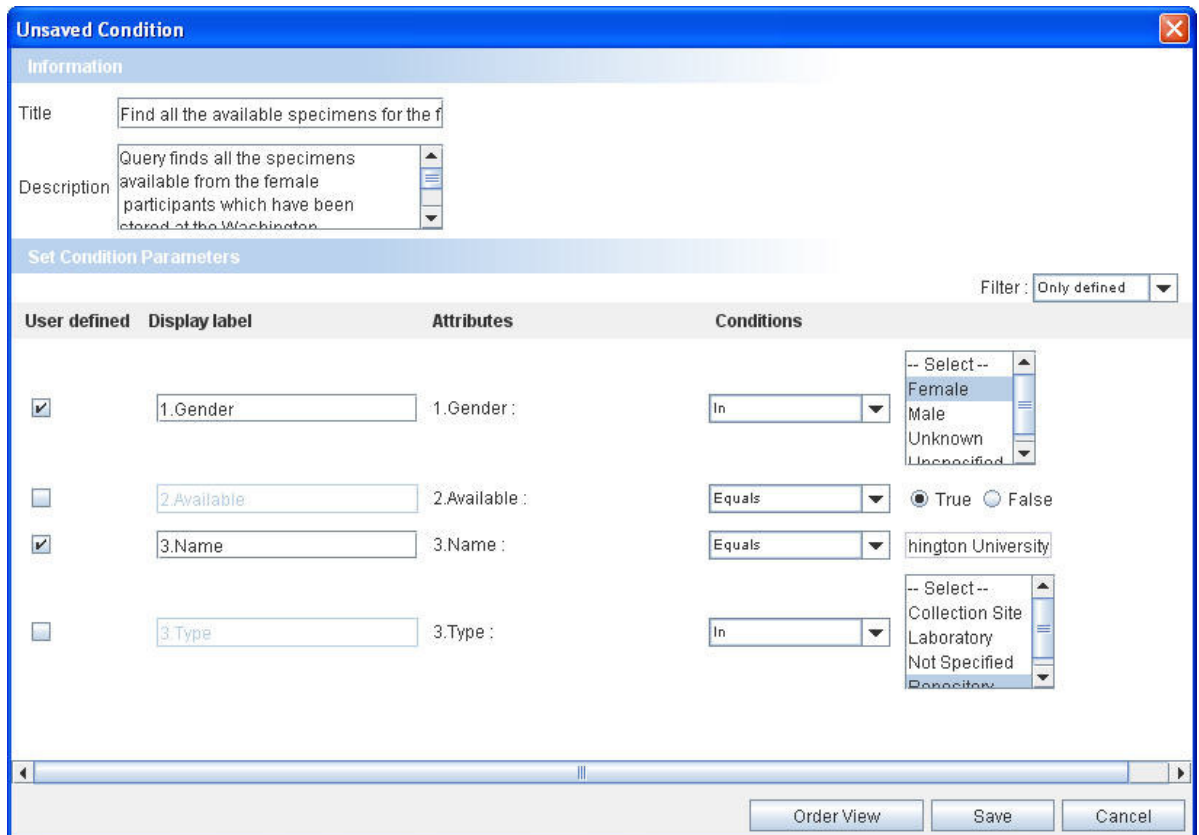
The screenshot shows the 'Unsaved Condition' dialog box. The 'Information' section has a 'Title' field with the text 'Find all the serum specimens from all th' and a 'Description' field with the text 'Query helps to find the serum specimens from the alive female participants'. Red arrows point from the labels 'Title' and 'Description' to these fields. The 'Set Condition Parameters' section has a table with columns 'User defined', 'Display label', 'Attributes', and 'Conditions'. Below the table are three dropdown menus for selecting conditions: 'In', 'Equals', and 'In'. The 'Equals' dropdown has radio buttons for 'True' and 'False'. At the bottom of the dialog are buttons for 'Order View', 'Save', and 'Cancel'.

Figure 2.19. Save Query Panel-Non Parameterized Query

### How to save a parameterized query?

1. Click **Save Condition** (as shown in *Figure 2.12. Define Search Result View.*).
2. Enter the **Title** and **Description** for the query to be saved, in the save query panel (see the figure below).
3. Select the **User Defined Check Box** for the attribute you wish to select different values for in the future.
4. Select the order of attributes in the Save Query Panel by changing it by **Order View** button.
5. Use **Filter** option to define condition on any of the attribute of the category other than the attributes against which the conditions were initially defined.
6. Click **Save** to save the query.

	Click <b>Cancel</b> to exit without saving the query.
---	---



User defined	Display label	Attributes	Conditions
<input checked="" type="checkbox"/>	1.Gender	1.Gender :	In
<input type="checkbox"/>	2.Available	2.Available :	Equals
<input checked="" type="checkbox"/>	3.Name	3.Name :	Equals
<input type="checkbox"/>	3.Type	3.Type :	In

Figure 2.20. Save Query Panel-Parameterized Query

## Executing a Saved Query

All the saved queries are visible in the **Home Page** in the **My Search Queries** stack box in the left hand side panel. To execute any query, click on the query to view the results.

To execute a saved query, perform the following steps:

1. Click on the title of the query to be executed

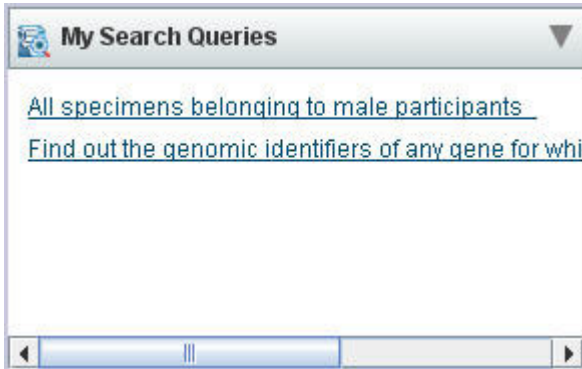


Figure 2.21. Saved Query Stack box

The **Saved Query Panel** opens.

2. For non-parameterized queries, click **Show Results** on the execute query panel

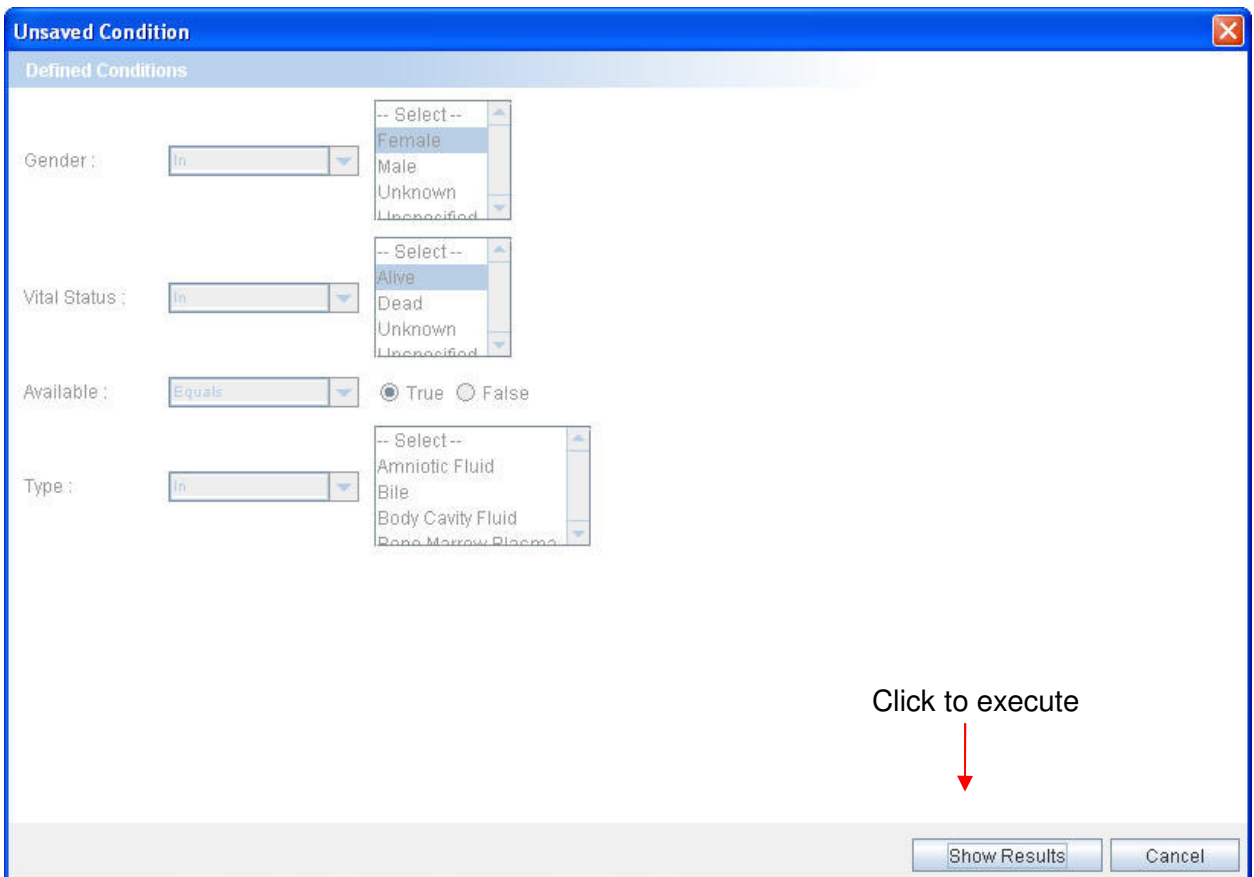


Figure 2.22. Execute Query Panel- Non Parameterized Query

3. For parameterized queries,

- Change the conditions in the parameterized attributes
- Click **Show Results** to view the results

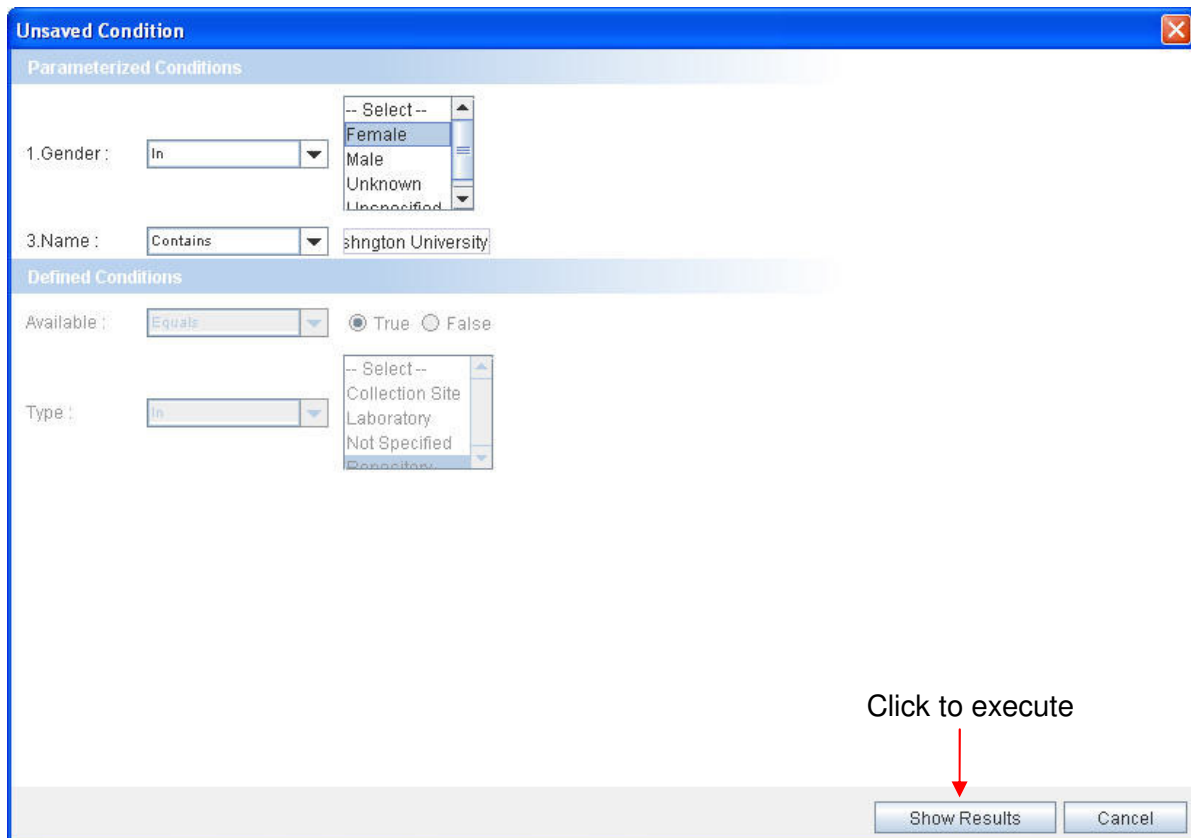


Figure 2.23. Execute Query Panel-Parameterized Query

The save query option provides an easy alternative way of executing the queries without actually forming the query again. The parameterized query allows changing the conditions on attributes thereby allowing flexibility to the saved queries.

## Viewing and Saving Data List

After adding the records of interest to the data list, you can get the consolidated view of the data list that shows all the records added to the list in a hierarchical tree fashion. These include all the records as well as the details of a particular record. You can also export the data list into a Comma Separated Value (CSV) file or create an Experiment by using the contents of the data list.

The following figure shows the data list tab:

The screenshot shows a software window titled "Search Data For Experiment" with a progress bar at the top indicating five steps: 1. Choose Search Category, 2. Add Limits, 3. Define Search Results View, 4. View Search Results, and 5. Data List. The left pane displays a tree view of search results, with "Gene (GeneConnectv1) (15)" selected. The right pane shows a table with 15 rows of data. The table columns are: #, Ensembl G..., Ensembl G..., Entrez Gen..., Entrez Gen..., Identifier, Unigene As..., and Unigene Cl... The data rows are as follows:

#	Ensembl G...	Ensembl G...	Entrez Gen...	Entrez Gen...	Identifier	Unigene As...	Unigene Cl...
1	true	ENSG0000...	true	115098	1	true	Hs.100043
2	true		true	115098	2	true	Hs.100043
3	true	ENSG0000...	true	140901	3	true	Hs.100057
4	true		true	140901	4	true	Hs.100057
5	true	ENSG0000...	true		5	true	Hs.100057
6	true	ENSG0000...	true	10570	6	true	Hs.100058
7	true		true	10570	7	true	Hs.100058
8	true	ENSG0000...	true		8	true	Hs.100058
9	true	ENSG0000...	true	57165	9	true	Hs.100072
10	true		true	57165	10	true	Hs.100072
11	true	ENSG0000...	true		11	true	Hs.100072
12	true	ENSG0000...	true	752	12	true	Hs.100217
13	true		true	752	13	true	Hs.100217
14	true	ENSG0000...	true	3842	14	true	Hs.100235
15	true	ENSG0000...	true		15	true	Hs.100235

Figure 2.24. Data List

The left-hand side pane shows all the records that you have added to the data list in a tree format. Each node in the tree represents either a category or a record of a category.

- For a node representing a category:
  - The number of records for that category is shown in parentheses. For example Gene (3) means that you have added 3 records for the Gene category.
  - Clicking on the node updates the right hand side panel to show a spreadsheet containing the records of that category (see figure above) with the columns depicting the attributes and the rows are depicting the records (Figure 2.17. Example 3- Data List).
- For a node representing a record
  - The unique identifier for that record is displayed. For example, Gene\_1 represents the record with identifier 1.

- Click the node updates on the right hand side pane to show the details of that record.

The right hand side panel shows the record details. The details shown in the panel depends on the record chosen in the left hand side panel.

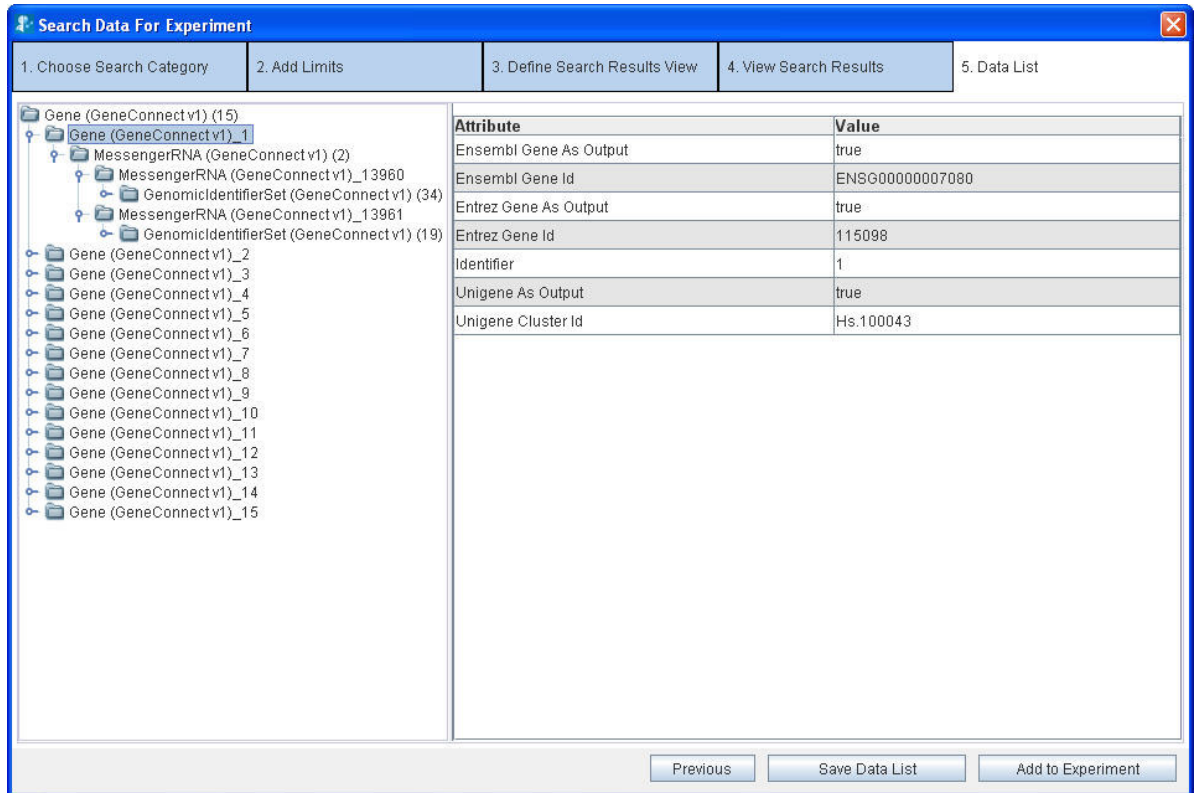


Figure 2.25. Record Details for a single record in Data List

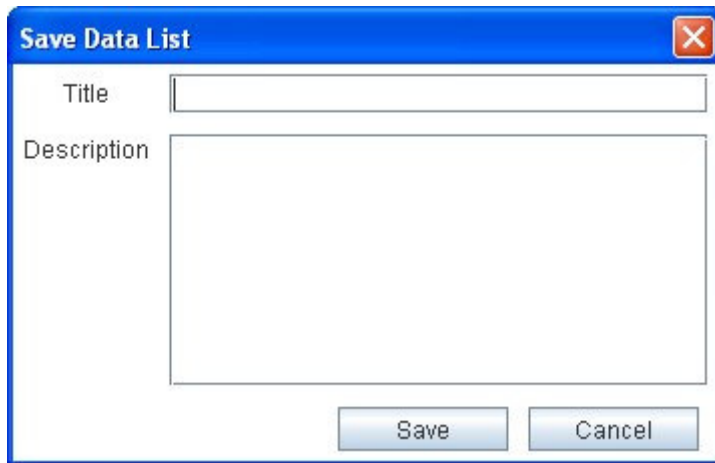
## Operations on the Data List

You can perform the following operations on the data list:

- Save the data list – Click **Save Data List**.
- Export the data list into a CSV file – Click **Export**. (see Figure 2.24. Data List)
- Add the data list to an Experiment – Click **Add To Experiment** (see Chapter 3 for additional information).

To save the data list:

- Click **Save Data List**.
- Type the title and description for the data list.



*Figure 2.26. Save Data List Panel*

- Click **Save**. The system displays a message in the status bar indicating that the data list is saved. The data list gets saved in the database.



---

## Chapter 3 Experiments

After you search and add some data of interest to the data list, you can create an experiment using that data. This chapter explains the various operations that you can perform on the data sets collected.

This chapter covers the following topics:

- What are Experiments and Projects?
- Creating an Experiment
- Viewing the Existing Experiments
- Working with an Experiment
- Filtering Data
- Analyzing Data

### What are Experiments and Projects?

---

An experiment is defined as collection of data. After an experiment is created, you can perform the following:

- Perform various analyses on the data using the grid-enabled analytical services
- Visualize analysis results using various viewers such as bar charts and scattered plots.
- Create new data sets by applying filters on the existing datasets

A project is a collection of experiments. A project can contain many experiments or other sub-projects. The grouping of experiments into projects helps you to manage multiple experiments easily.

Example of an experiment

**Aim:** Find micro-array data for non-malignant brain tissue. Perform **Comparative Marker Selection (CMS)** on this data to determine the expression values that are most closely correlated with a class template. View the correlation among the expression values as a scattered plot.

You need to perform the following steps using caB2B

<b>Step no.</b>	<b>Description</b>	<b>Related Section</b>
1	Search for micro array data	Search data – The Query Wizard
2	Create experiment with this data	Creating an Experiment
3	Open the experiment you have created	Opening an Experiment
4	Filter the data further if desired	Filtering Data
5	Save the filtered data	How to save filtered data
6	Perform CMS analysis on selected data	Analyzing Data

7	Plot analysis result as a scattered plot	Visualizing Data
---	--	------------------

## Creating an Experiment

To create an experiment, you need to have a saved data list with the required data. See *Viewing and Saving Data List* for more details.

After you have saved the data list, do the following to create an experiment:

1. Click **Add to Experiment** (*Figure 2.15. Data List*). This opens the **Create New Experiment** window.

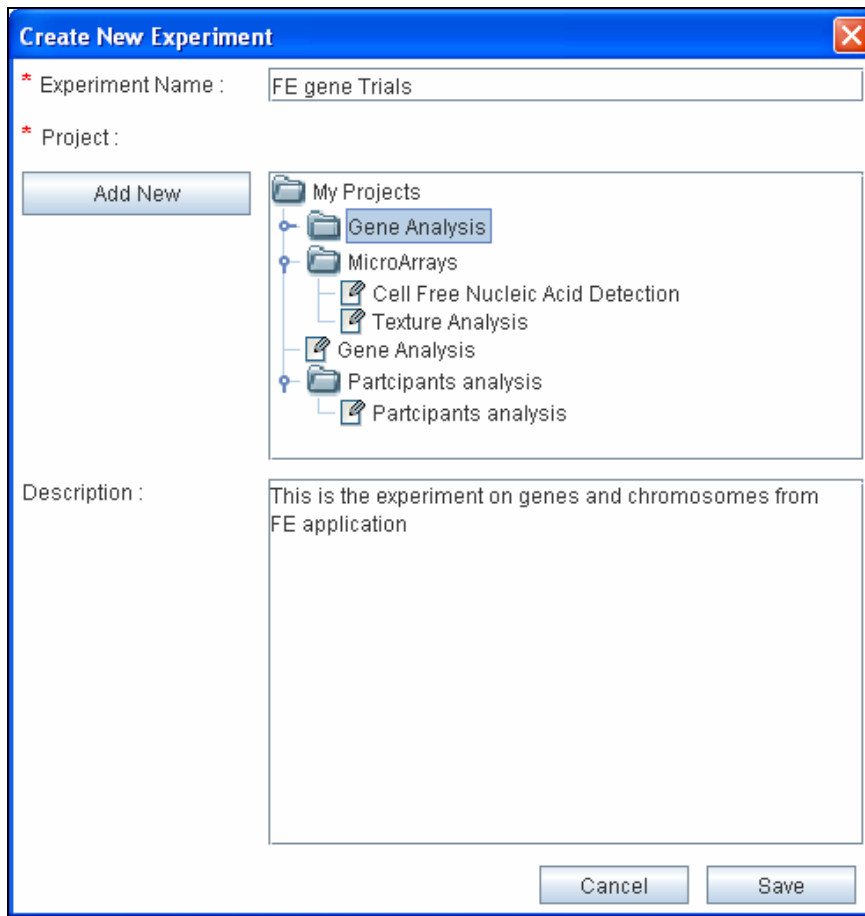



Figure 3.1. Create New Experiment

2. Give an appropriate name to the experiment.
3. Select the project in which you wish to create this experiment (see note below).
4. Provide a description for the experiment.
5. Click **Save**.

You can now close the **Search Data** wizard to return to the **Home** page and to open the experiment.



F.Y.I.

- To create a new project, perform the following:
  1. Select the parent project under which you wish to create a new project
  2. Click **Add New**. This adds a new sub project under the selected project with the default name– New Project.
- To rename the new project
  1. Select the project.
  2. Press F2.
  3. Type the desired name.
  4. Press ENTER.

## Viewing the Existing Experiments

On the caB2B Home page, click the **Experiment** tab. This opens the My Experiments page.

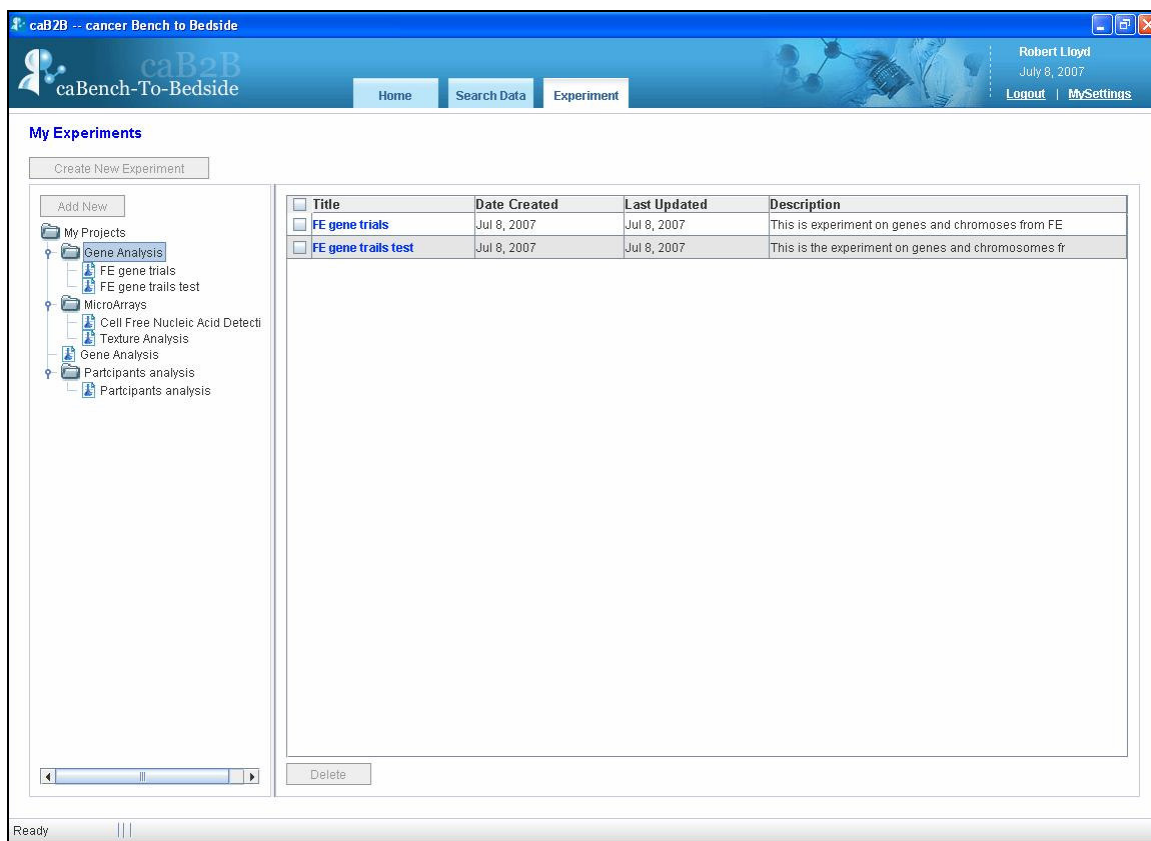


Figure 3.2. My Experiments

You can see all the projects and experiments that you created in the left panel.

When you select a project in the left hand side panel, the system displays a list of experiments in the project in right hand side panel.

## Opening an Experiment

To open an experiment, go to the My Experiments page and click on the experiment title seen in the right panel (see *Figure 3.2. My Experiments*).

When an experiment is opened, you see the **Experiment Details** page. The experiment details page shows the data categories added to experiment. It gives the user an option to modify the experiment by adding a new column, adding new data, exporting data etc. All these options are also available on the right click menu.

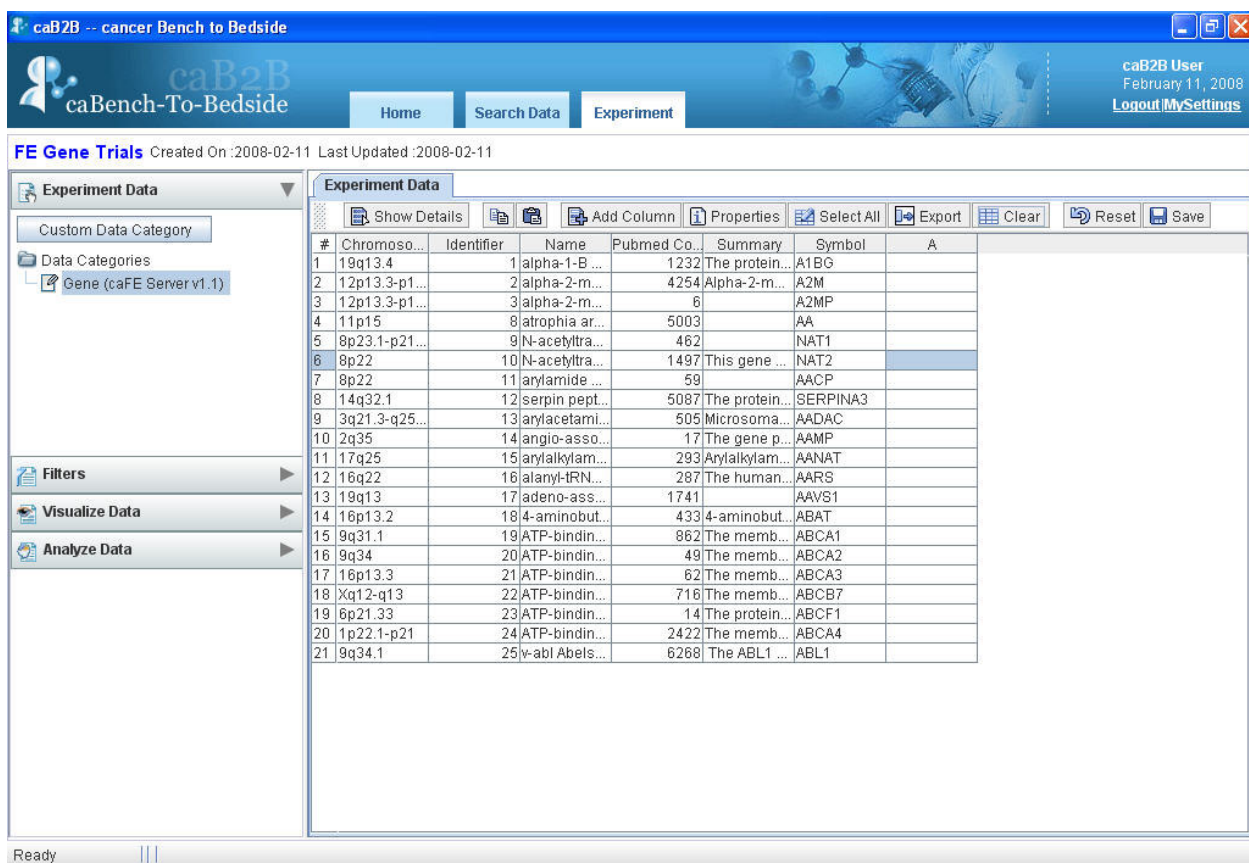

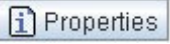

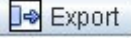

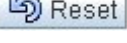
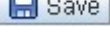


Figure 3.3. Experiment Details

The various options provided by the spreadsheet on the experiment details page are listed in the table below:

Button	Description
Show Details	Shows details of the particular record.
	Copies data of the selected cells.
	Pastes the data to the specified cells.

<b>Button</b>	<b>Description</b>
 Add Column	Adds a new column to the spreadsheet.
 Properties	Shows the properties of the spreadsheet, like which attributes are visible, what filter is applied.
 Select All	Selects all the records.
 Export	Exports data to a .csv file.
 Clear	Clears the user added data from the selected cell.
 Reset	Undo's all the changes and shows the original data.
 Save	Creates a data category.

## Working with an Experiment





The **Experiment Details** page consists of the following sections:

- The **Operations** panel on the left hand side
- The **Visualization** panel on the right hand side

### The Operations panel

The **Operations** panel encompasses the following collapsible boxes.

Table 3.1. Operations on the Experiment Page

<b>UI Elements</b>	<b>Description</b>
 Experiment Data ▶	Displays a list of all the data categories present in the experiment. See How to save filtered data in this chapter for more information.
 Filters ▶	Shows all the filters that are currently applied to the data of the selected data-category. See Filtering Data in this chapter for more details.
 Visualize Data ▼	Shows the visualization options available for the selected data-category. See Visualizing Data in this chapter for more details.
 Analyze Data ▼	Shows the links to call analytical services that are applicable to the selected data-category. See Analyzing Data in this chapter for more details.

### Visualization panel

This is the right hand side panel on the **Experiment Details** page (see Figure 3.3. Experiment Details). This panel can contain several pages at a time (see figure

below). You can switch between the pages by clicking on the **tabs** at the top of this panel.

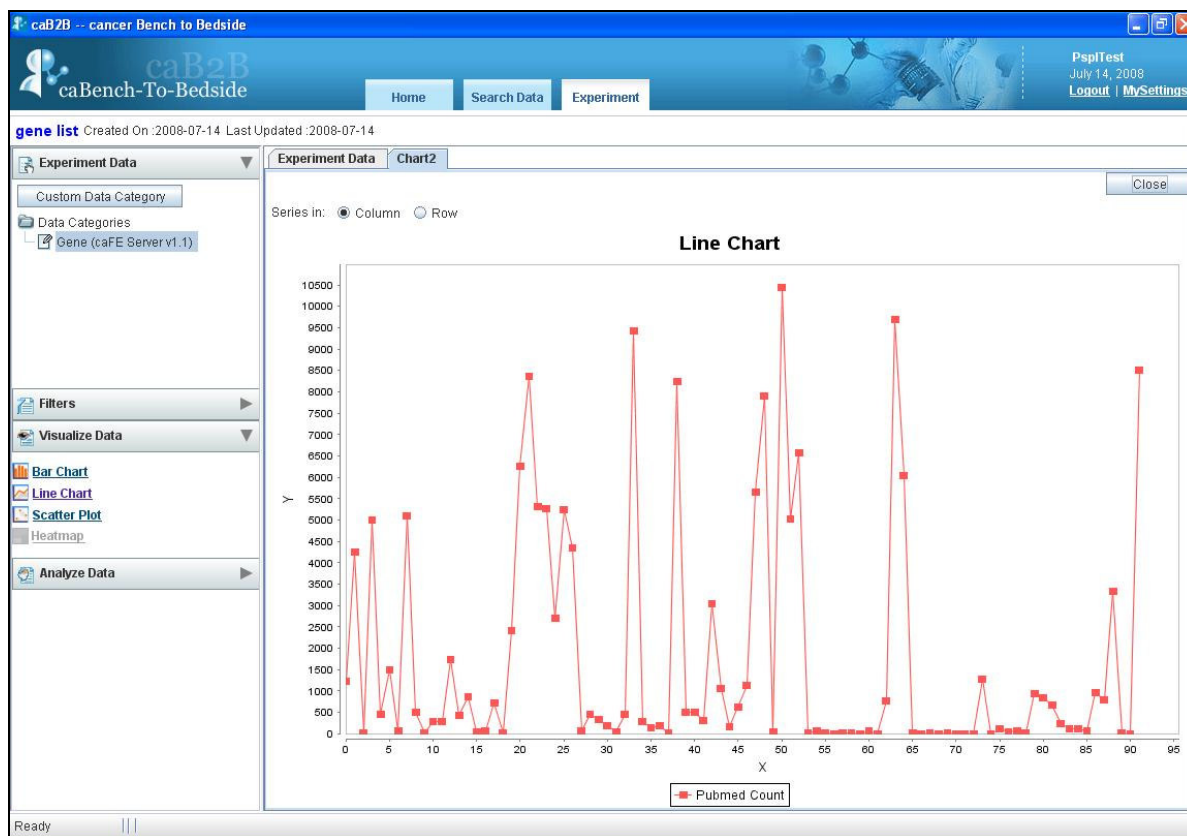


Figure 3.4. Visualization Panel

The visualization panel is updated based on the operations you perform in the operations panel. For example,

<b>Operations pane</b>	<b>What happens in Visualization pane</b>
Click a data-category in the <b>Select Data Category</b> box	The <b>Experiment Data</b> tab is updated to show the records of that data category
Choose a visualization from the <b>Visualize Data</b> panel	A new tab is created to display the desired visualization

## Filtering Data

The purpose of creating experiments is to perform analysis on the data. To aid the analysis, caB2B provides features to filter the data so that you can zero in on the data of interest.

After filtering the data, you can save the filtered data as a **data-category**. For example, if you have created an experiment with all female Participants, you can now create different data categories based on age less than 50, age between 50 and 65 and age greater than 65.

## How to filter data

1. Select the column whose data is to be filtered, select the filter from the **Filter** stack box, (located in the left hand side panel)

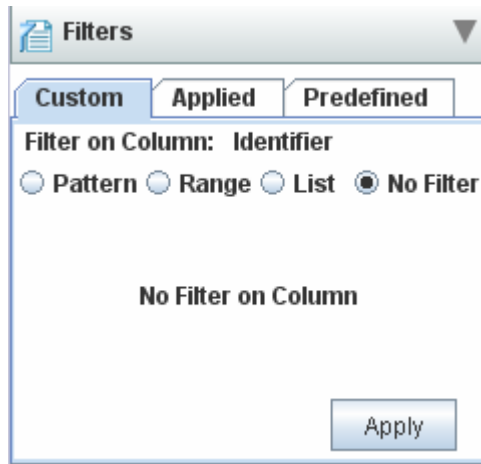


Figure 3.5: Filter Stack box

2. System displays appropriate filter controls.
3. Choose one of the available filters and specify the conditions that the data must satisfy (see Appendix D: Types of Filters).
4. Click **Apply**.
5. The data in the **Experiment Data** tab is updated to show the filtered data.
6. The filter that you have applied is also shown in the **Applied** tab of the **Filter stack box** in the panel on the left hand side.

You can now save the filtered data as a new data-category (more information about this is provided below) and use it to perform analyses or visualization.

	<p>You can filter on multiple columns by successively applying filters on those columns. For example, to obtain <i>female participants between age 30 and 40</i></p> <ol style="list-style-type: none"><li>1. Apply the filter “Gender = female”.</li><li>2. Then apply the filter “age between 30 and 40”.</li></ol> <p>For details of the various filter controls and how to use them, see. <b>Appendix D: Types of Filters.</b></p>
--	--

## How to edit a filter

1. Choose the column whose filter you wish to edit.

An appropriate filter control is shown in the Filter stack box with the values that you previously entered.

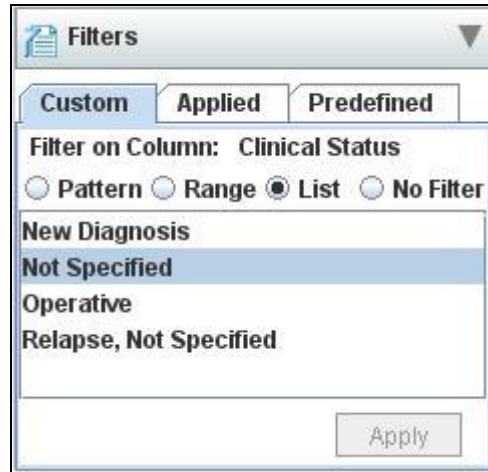


Figure 3.6. Edit Filter

2. Specify desired values of the filter (see Appendix D: **Types of Filters**).
3. The data in the **Experiment Data** tab is updated to show data based on the new filter.
4. The filter criteria shown in the **Applied tab** of **Filter stack box** as shown in Figure 3.6 in the operations panel is also updated.

#### How to remove a filter

1. Choose the column whose filter you wish to remove.
2. An appropriate filter control is shown with the values previously entered.
3. Select the **No Filter** option.
4. Click **Apply**.
5. The data in the **Experiment Data** tab is updated to show data after removing the filter.
6. The filter is also removed from **Applied** tab the **Filter Data** stack box in the operations panel.

#### How to save filtered data

Once you apply filters to a data set, save the filtered data as a **data category**. You can create as many data-categories as required. The data-categories can be further used for data analysis.

To do this,

1. Click **Save** (located in the toolbar of the right hand side panel of the experiment details page *Figure 3.3.Experiment Details*).
2. The **Save as Data category** panel is displayed.



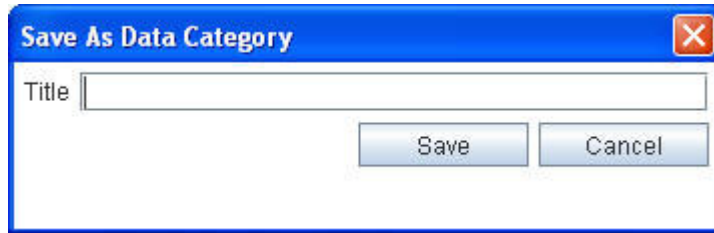


Figure 3.7. Save Data Category Panel

3. In the **Title** field, type the title of the new data-category.
4. Click **Save**.

The new data-category will appear under **Custom Data Categories** section in **Experiment Data** box of the operations pane.

## Custom Data Category

**Data Category** is used to define a class which is a subset of other classes. In an experiment there are various data categories, like participant, specimen etc. Typically, user wants to use data from various categories, to create a new category. This new category shows the information which makes sense to the end user, **Custom Data Category** (CDC), is a way to bring data from various categories under a single category. For example, end user wants to see the specimen details and participant information in a single view. For this he can create a CDC with the desired attributes from the participant and the specimen class and create a third category.

### How to create CDC

1. Click **Custom Data Category** (located at the top of the Experiment Data box on the left hand side panel of the experiment page).

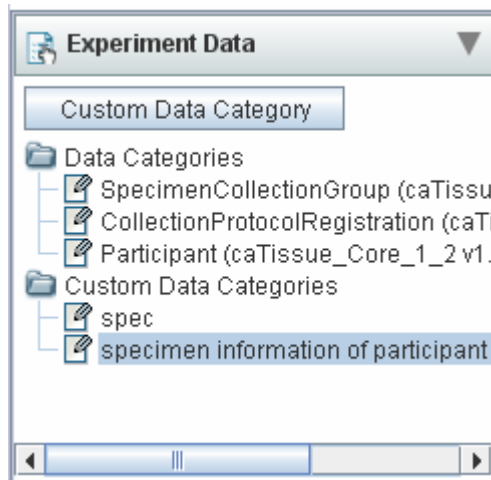


Figure 3.8. Custom Data Category

The **Create Custom Data Category Panel** is displayed (Figure 3.8 below).

2. In the **Custom Data Category Title** field, type the title of the new custom data-category.
3. In the **Data List** field, select the data list to create the category.
4. In the **Categories** field, select the category to be customized.
5. Select attributes from **Available Attributes** to be present in CDC.
6. Click **Add/Add All** to add to **Selected Attributes**.
7. Select attributes from the **Selected Attributes** and click **Remove/Remove All** button to remove the attributes from CDC.
8. Click **Save**.

The screenshot shows a software interface for creating a custom data category. At the top, there's a title bar 'Custom Data Category'. Below it, a text field contains 'Specimen Information of e'. Underneath are two dropdown menus: 'Data List' (set to 'Participants Analysis') and 'Category' (set to 'edu.wustl.catissuecore.domain.Participant'). The main area is split into two columns: 'Available Attributes' on the left and 'Selected Attributes' on the right. Between these columns are four buttons: 'Add', 'Add All', 'Remove', and 'Remove All'. The 'Available Attributes' list includes various fields like 'Participant : Sex Genotype', 'Participant : Vital Status', 'Participant : Birth Date', 'Participant : Ethnicity', 'Participant : Identifier', 'Participant : Social Security Number', 'Participant : Activity Status', 'Participant : Death Date', 'CollectionProtocolRegistration : Registration Date', 'CollectionProtocolRegistration : Activity Status', 'CollectionProtocolRegistration : Identifier', 'SpecimenCollectionGroup : Identifier', 'SpecimenCollectionGroup : Activity Status', 'SpecimenCollectionGroup : Clinical Diagnosis', 'SpecimenCollectionGroup : Clinical Status', 'SpecimenCollectionGroup : Name', 'TissueSpecimen : Comment', 'TissueSpecimen : Identifier', 'TissueSpecimen : Available', and 'TissueSpecimen : Position Dimension One'. The 'Selected Attributes' list includes 'Participant : First Name', 'Participant : Middle Name', 'Participant : Last Name', 'Participant : Gender', 'CollectionProtocolRegistration : Protocol Participant Id', 'TissueSpecimen : Label', 'TissueSpecimen : Pathological Status', and 'TissueSpecimen : Lineage'. At the bottom right, there are 'Save' and 'Cancel' buttons.

Figure 3.9. Custom Data Category Panel

The CDC created appears under **Custom Data Categories** section in **Select Data Category** box of the operations panel as shown below.

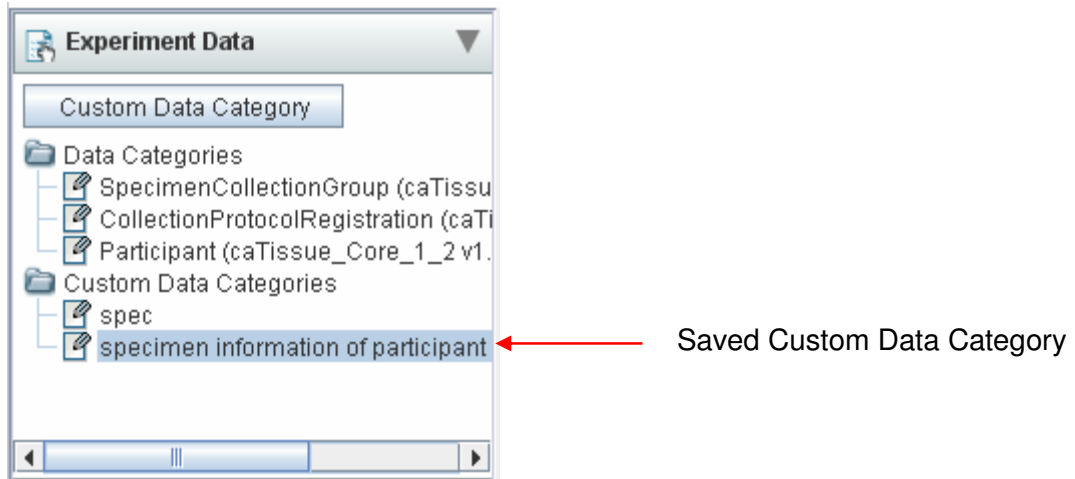


Figure 3.10. Saved Custom Data Category

## Analyzing Data

After you searched for the data of interest and created experiments from it, you can analyze that data. To do this, you can use various available analytical services. Typically they accept data, perform analyses on it and return the result. You can also pass configuration parameters along with the data to the analytical services.

When you are viewing the data, the services that can operate on that data are shown in the **Analyze Data** box. In the current version one analytical service **Comparative Marker Selection**, from gene pattern is available.

### How to analyze data?

1. Select the data category from **Select Data Category** box, for example Bioassay.
2. In the **Analyze Data** box, click on the link to call the service, for example **Comparative Marker Selection** link.
3. It will open a popup that accepts configuration parameters for the service. *Figure 3.11. Configuration parameters for CMS.*
4. Enter the **Title** for this analysis and the values for various parameters
5. Click **Finish** to submit this analysis.

Figure 3.11. Configuration parameters for CMS

You can see the new **Analysis** tab added in the right hand side panel. It contains following fields

<b>Field</b>	<b>Description</b>
Data Category	The data category on which you performed the analysis.
Analysis Title	The title you provided for the analysis. It is a link to the result of the analysis.
Date	The date on which you performed the analysis.
Status	Indicated whether analysis is <b>completed</b> or <b>pending</b> .

You can view the result on **completed** analysis by clicking the link in Analysis **Title** field.

---

Experiment Data		Analysis	
<b>Analysis performed for 'FE gene trails test'</b>			
Data Category	Analysis Title	Date	Status
Gene (caFE Server v1.1)	<a href="#">CMS for Hu95Av2</a>	12-Jul-2007	Completed

*Figure 3.12. List of analysis performed for the experiment*

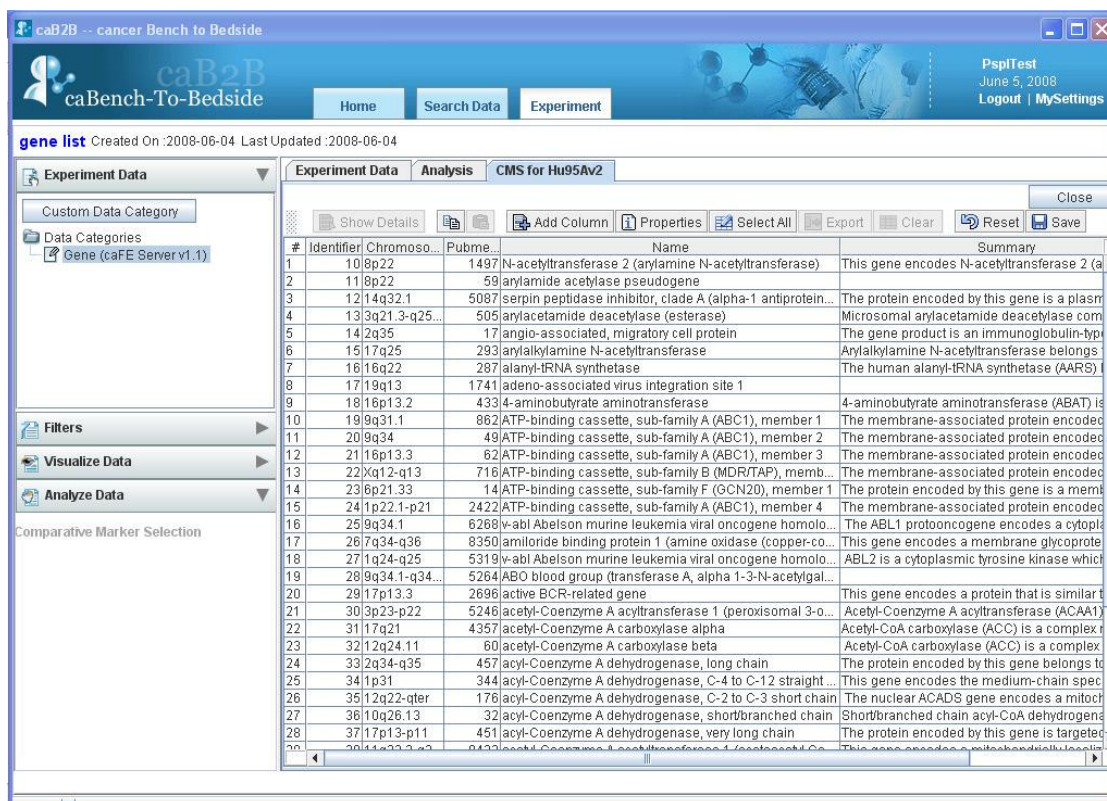


Figure 3.13. Result of CMS Analysis

## Visualizing Data

You can visualize the data in the experiment by using various graphical components such as charts and scatter plots. These options provide a graphical summary of the data allowing comparison, variation and co-relational analysis of the data.

Some terms used:

**Data series:** Related data points plotted in a chart that originate from data rows or data columns. Each data series in a chart has a unique color or pattern. This is represented by a *Legend*. You can plot one or more data series in a chart.


**Legend:** A small color box that identifies each series in the chart.

The current version of caB2B supports following visualization options in **Visualize Data** box.

Visualize option	Description
<b>Bar Chart</b>	This is a chart with rectangular bars of lengths proportional to the data value it represents.
<b>Line Chart</b>	This is a chart with a line joining the points that represents data value
<b>Scatter Plot</b>	A scatter plot will show a linear or non-linear relationship between the variables (column selected). It also helps in finding outliers in the data.

Figure 3.14. Types of charts supported

## How to plot data series

1. Select the data that you want to plot from **Experiment tab**.
2. Click on the link for type of the chart you wanted from the **Visualize Data** Stack Box, for example  [Bar Chart](#)
3. It will open the chart in a new tab in right hand side panel.
4. By default data series are from columns on the **Experiment Data** tab. This can be changed by toggling the radio buttons in the upper left corner of the chart area.

Following screen shots represents the various charts for **PubMed count** for few Genes from FE server.

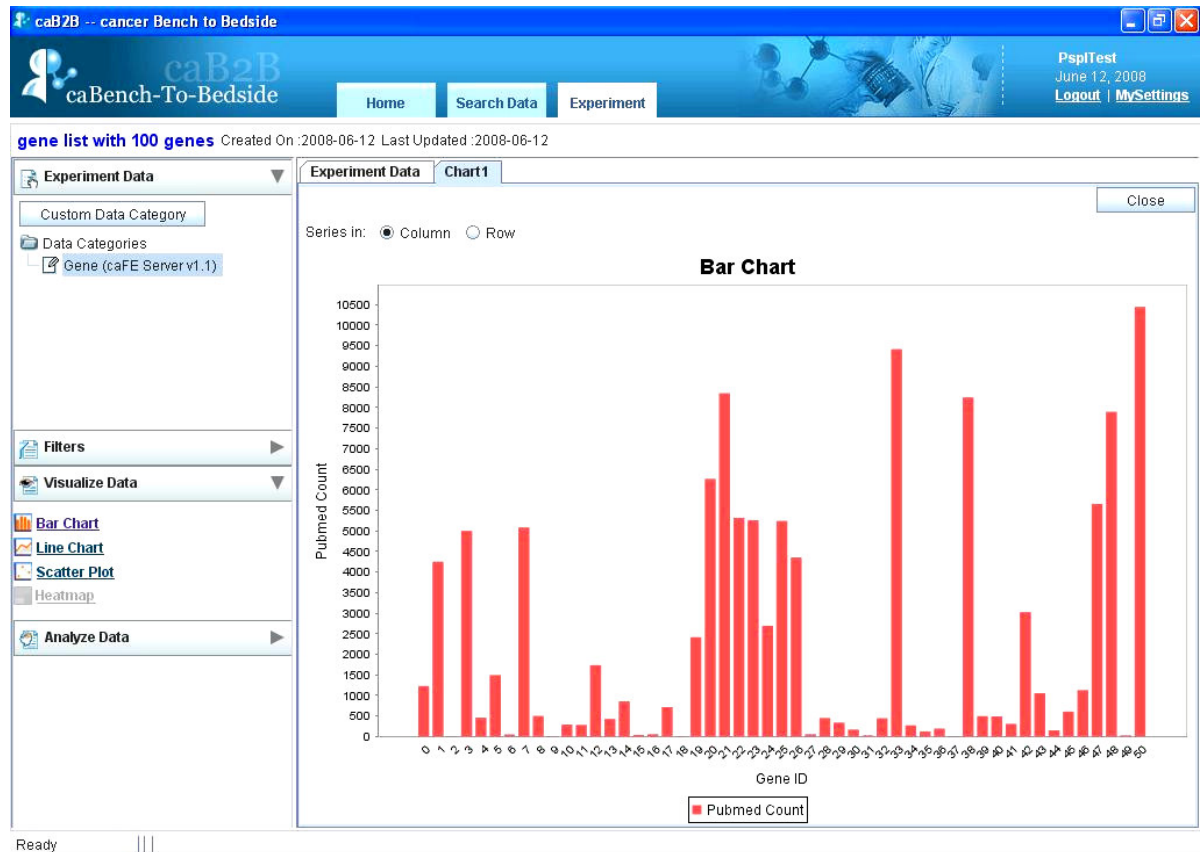


Figure 3.15. Bar chart example

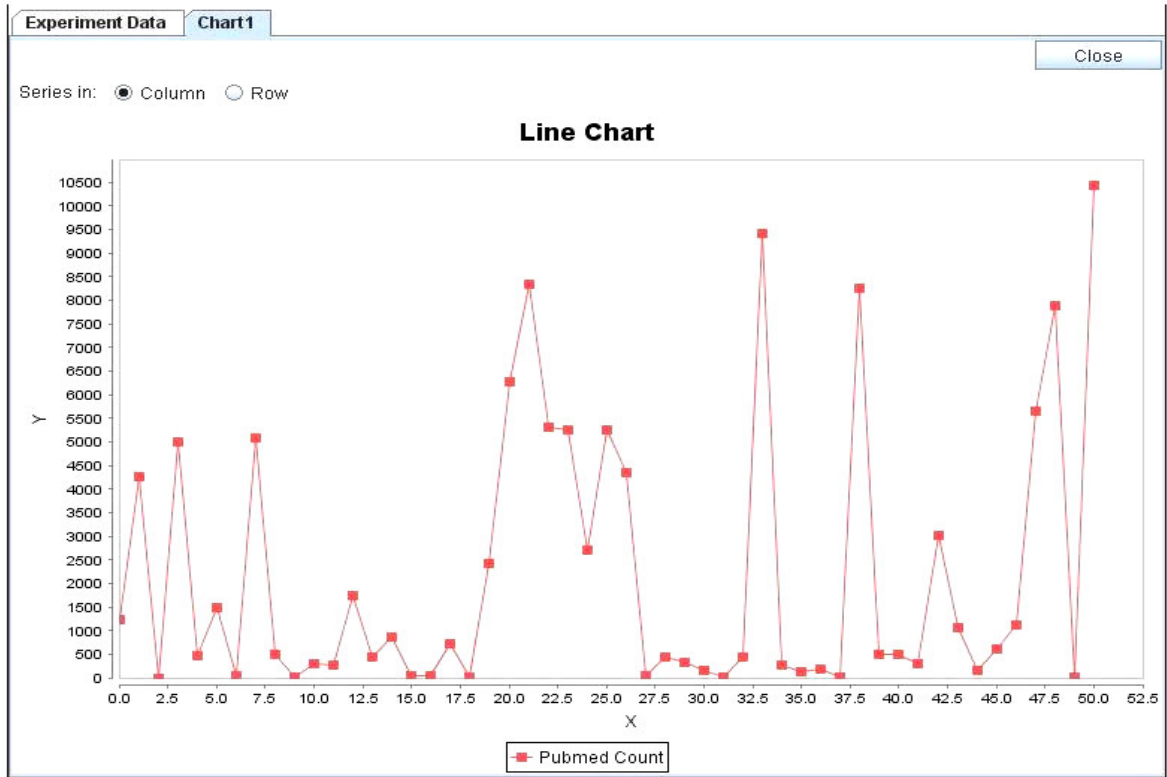


Figure 3.16. Line chart example

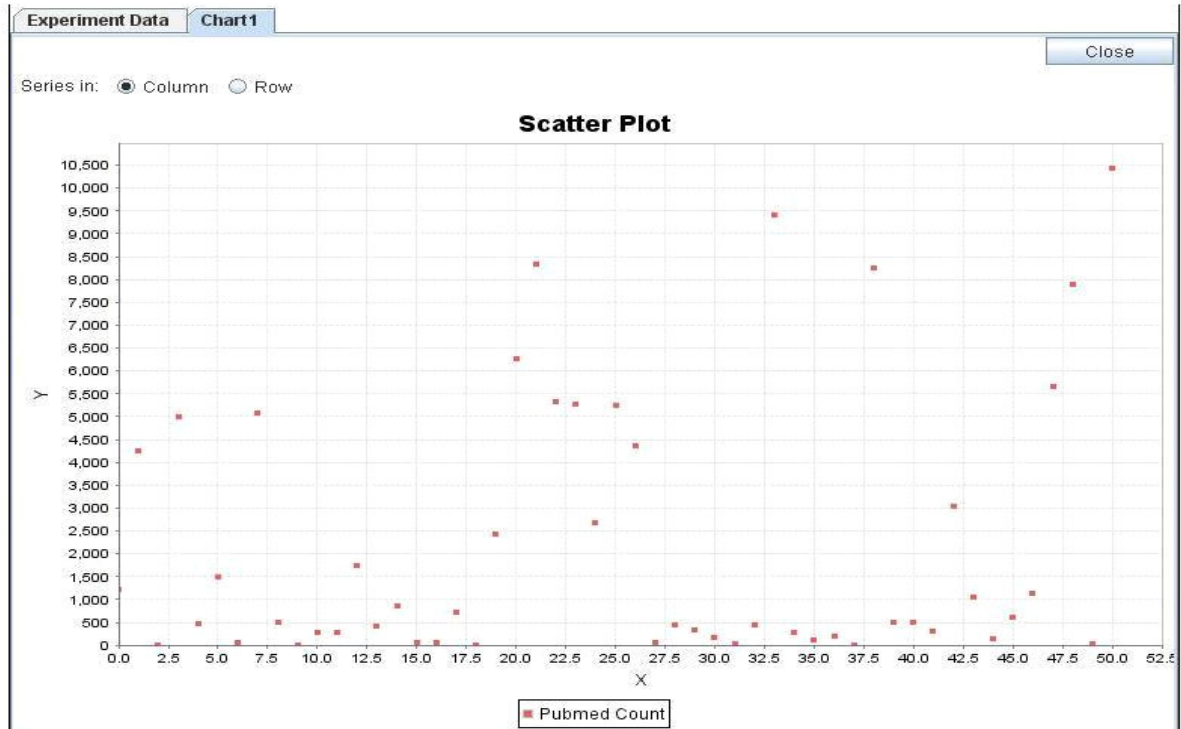


Figure 3.17. Scatter Plot example



## How to change data series from rows to columns and vice versa?

You can configure whether data series are from rows or columns by selecting appropriate **Series In** options.

- To have the data columns to represent the data series, select  Column option.
- To have the data rows to represent the data series, select  Row option.

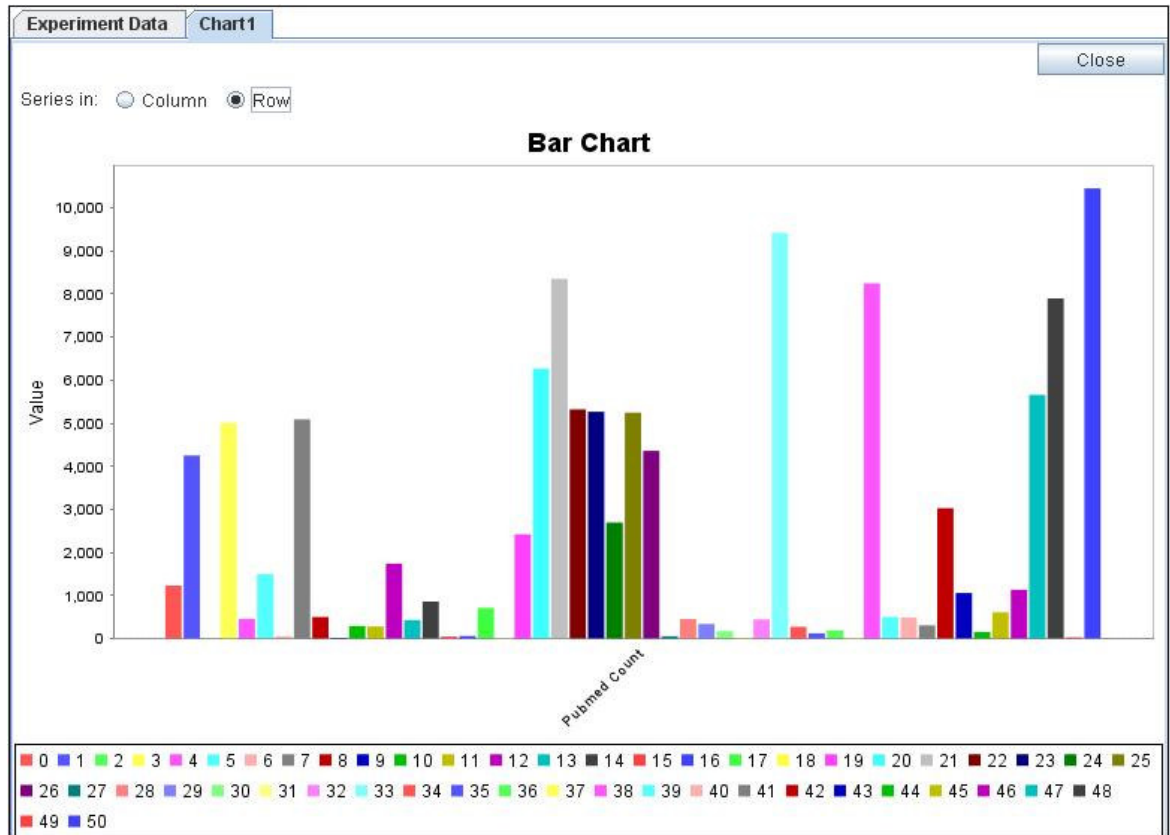


Figure 3.18. Bar chart with data series from rows



You can change the type of the chart that you are currently viewing by selecting required chart link from **Visualize Data** box.

You can create multiple charts for the data.

The name of the tab added for each chart is “chart” followed by a unique number for that chart. For example chart1, chart2, etc.

# Appendix A Examples of category search

This appendix provides examples of the advanced search strategies that are possible in caB2B, while searching for categories.

## Text-based Search - For category name information

To search categories having “gene” in their name do the following:

1. In the **Search** box, type **gene**.
2. Expand the Advanced Search box
3. Select the **Category** check box.
4. Click **Text**.
5. Click **Search**.

The results of this search are as shown below. All the categories shown below have *gene* in their names.

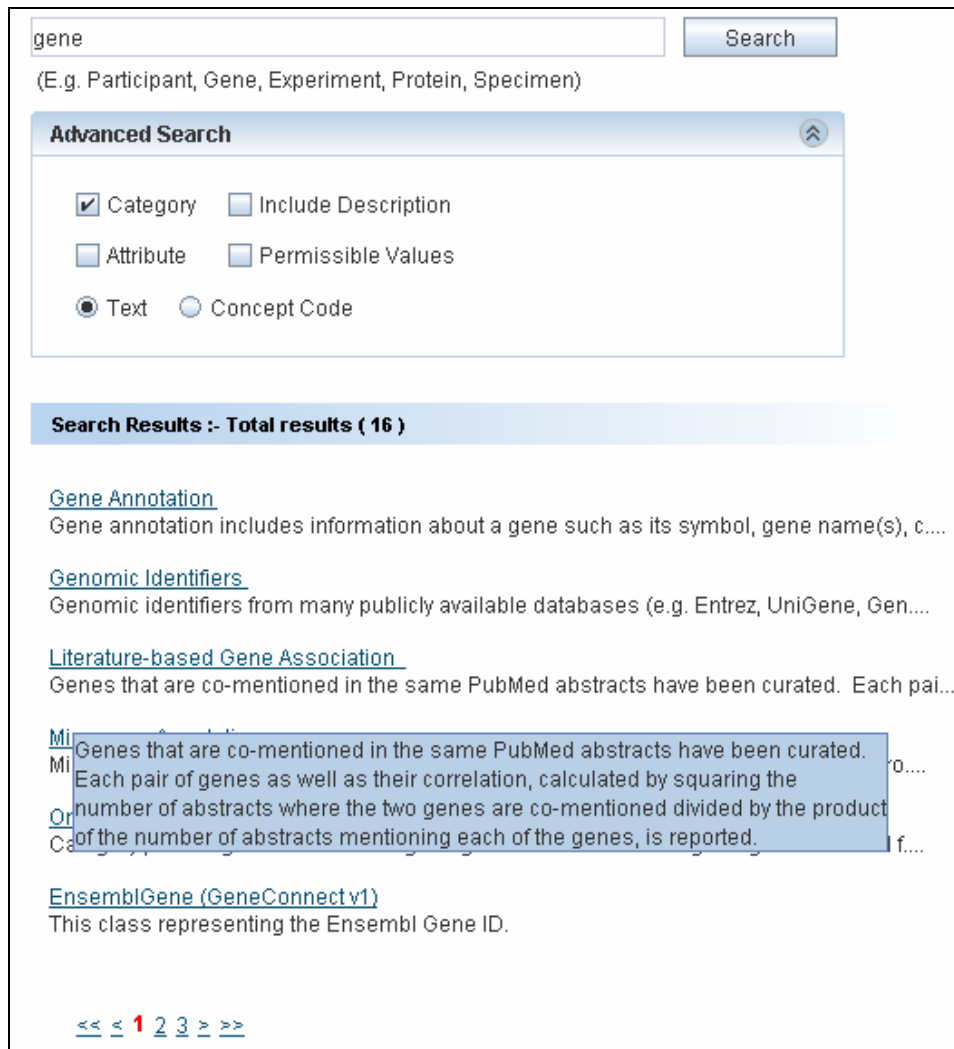


Figure A.1. Search Result of Example 1

## Text-based Search - For category and attribute name and description

To search the categories having “Protein” in their name or description or to search the attributes having “Protein” in their name or description, you need to do following:

1. In the **Search** box, type **Protein**
2. **Open Advanced Search box.**
3. Select the **Category** check box.
4. Select the **Attribute** check box.
5. Select the **Include Description** check box.
6. The **Text** option button is selected by default.
7. Click **Search**.

The result of this search is as shown in the screenshot below.

protein

(E.g. Participant, Gene, Experiment, Protein, Specimen)

**Advanced Search** ⤴

Category     Include Description

Attribute     Permissible Values

Text     Concept Code

**Search Results :- Total results ( 11 )**

[Gene Annotation](#)  
Gene annotation includes information about a gene such as its symbol, gene name(s), c...

[Genomic Identifiers](#)  
Genomic identifiers from many publicly available databases (e.g. Entrez, UniGene, Gen...

[Microarray Annotation](#)  
Microarray annotation includes information about a probe or probeset found on a micro...

[BioMaterial \(caArray v1.1\)](#)  
BioMaterial is an abstract class that represents the important substances such as cel...

[BioSequence \(caArray v1.1\)](#)  
A BioSequence is a representation of a DNA, RNA, or protein sequence. It can be repr...

[GenBankAccession \(caFE Server v1.1\)](#)  
The GenBank Accession for a Gene

<< < 1 2 > >>

Figure A.2. Search Result of Example 2

## Concept Code Based Search - For category name information

To search categories having concept code "c29867":

1. In the **Search** box, type **c29867**.
2. Open the advanced search box.
3. Select the **Category** check box.
4. Select the **Concept Code** option button.
5. Click **Search**.

The result of this search is as shown in the screenshot below.

---

(E.g. Participant, Gene, Experiment, Protein, Specimen)

**Advanced Search** ⌆

Category     Include Description  
 Attribute     Permissible Values  
 Text     Concept Code

**Search Results :- Total results ( 3 )**

[CollectionProtocolRegistration \(caTissue Core v1.1\)](#)  
A registration of a Participant to a Collection Protocol

[Participant \(caTissue Core v1.1\)](#)  
An individual from whom a specimen is collected

[ParticipantMedicalIdentifier \(caTissue Core v1.1\)](#)  
A medical record identification number that refers to a Participant

*Figure A.3. Search Result of Example 3*

### Concept code based search - For attribute name information

To search categories whose attribute has concept code "C42614":

1. In the **Search** box, type **C42614**.
2. Select the **Attribute** check box.
3. Select the **Concept Code** option button.
4. Click **Search**.

The result of this search is as shown in the screenshot below.

C42614

(E.g. Participant, Gene, Experiment, Protein, Specimen)

**Advanced Search**

Category    Include Description  
 Attribute    Permissible Values  
 Text    Concept Code

---

**Search Results :- Total results ( 110 )**

[Gene Annotation](#)  
Gene annotation includes information about a gene such as its symbol, gene name(s), c....

[Literature-based Gene Association](#)  
Genes that are co-mentioned in the same PubMed abstracts have been curated. Each pai...

[Microarray Annotation](#)  
Microarray annotation includes information about a probe or probeset found on a micro....

[Orthologous Gene](#)  
Category providing details of orthologous genes associated with given gene obtained f...

[Array \(caArray v1.1\)](#)  
The physical substrate along with its features and their annotation

[ArrayDesign \(caArray v1.1\)](#)  
Describes the design of an gene expression layout. In some cases this might be virtu....

<< < **1** 2 3 4 5 > >>

Figure A.4. Search Result of Example 4


# Appendix B Define Limit: Operators and Values

This appendix explains the various operators and how to specify values while defining limits on a category.

The following table explains the different operators and provides examples as to how values are to be specified.

Operator	Data types for which applicable	Sample condition	Expected Results
Equals	Numeric Date String	Name <b>equals</b> "John"	Returns only John and not john.
Not equals	Numeric Date String	Identifier <b>not equals</b> 32	Returns all records except those having identifier = 32
Less than	Numeric Date	Participant's birth date <b>less than</b> 2000/1/1	Returns all Participants born before January 1st, 2000
Less than or equal	Numeric Date	Signal value <b>less than or equals</b> to 0.5	Returns all records for which signal value is not more than 0.5
Greater than	Numeric Date	Cell count is <b>greater than</b> 10000	Returns all records with Cell count > 10000
Greater than or equals	Numeric Date	Specimen's collection-date <b>greater than or equal to</b> 2005/5/1	Returns all specimens collected on or after May 1st, 2005
Between	Numeric Date	Patient's admission date <b>between</b> 2000/1/1 and 2000/12/31	Returns all Patients admitted in the year 2000
Is null	Numeric String Date	Patient's contact number is <b>null</b>	Returns all patients who do not have any contact number
Is not null	Numeric Date	Participant' middle name is <b>not null</b>	Returns all Participants for whom middle name is specified

In	Numeric String Enumerated	<ul style="list-style-type: none"> <li>• Signal value <b>in</b> 0.1,0.5,0.7,</li> <li>• Name <b>in</b> "John", "john"</li> </ul>	<ul style="list-style-type: none"> <li>• Returns all records whose signal value is 0.1 or 0.5 or 0.7</li> <li>• Returns both John and john.</li> </ul>
Not in	Numeric String Enumerated	Participant's gender <b>not in</b> male, unspecified	Returns all participants whose gender is either not specified or gender is not male.
Starts with	String	Participant's first name <b>starts with</b> Jo	Returns all participants whose first name starts with Jo like John, Jona.  Does not return the participant with name Scott, Henry, Jen
Ends with	String	Participant's last name <b>ends with</b> n	Returns all participants whose last name ends with n like John, Jon, and Jen. Does not return participant with name Scott, Henry
Contains	String	Participant's first name <b>contains</b> ann	Returns all participants whose first name is Joanna, Anna, and Johanna. Does not return participant with name Jen, Denny

- For attributes with enumerated values and Boolean attributes, clear a value by clicking on the value with the CTRL key pressed.
- If you use the **In** operator for a non-enumerated attribute, you can specify multiple values as follows:
  - Multiple values should be separated using a comma. For example, to specify the constraint **Identifier IN 2, 4, 5, 7**, you should type the value as **2,4,5,7**.
- If a value itself contains a comma, you can enclose the value in double quotes. For example, to specify the constraint **Name IN "Einstein, Albert", "Newton, Isaac"** the value that you should enter is **"Einstein, Albert", "Newton, Isaac"**.
- For attributes that requires a date value, you can click the  icon to launch a date selector. You can also type a date string directly in the text box. The format of this string must be **YYYY/MM/DD**. A valid date string becomes green whereas an invalid date string turns red. An invalid date string will be ignored.



# Appendix C Record Details View

The details of a record usually consist of the values of the attributes of the category. See *Figure 2.14. Results' Details* for an example.

This section provides examples of special categories that result in a slightly different view of the record's details.

## Administrator defined categories

These categories are built by manually grouping relevant attributes from one or more existing categories.

For each record of such a category the following information is shown:

- Values of the attributes in a table
- Related information as a tree. The leaf nodes of this tree are attributes; you can see the values of those in the **value** column.

Path: [Genomic Identifiers](#) >> [Genomic Identifiers\\_18667](#)

Attribute	Value
Ensembl Gene ID	ENSG00000120738
Entrez Gene ID	1958
Uni Gene Cluster ID	Hs.326035
Associated Information	
<ul style="list-style-type: none"> <li>[-] MessengerRNA (3)           <ul style="list-style-type: none"> <li>[-] MessengerRNA_6174</li> <li>[-] MessengerRNA_6175</li> <li>[-] MessengerRNA_198617               <ul style="list-style-type: none"> <li>[-] GenBank Nucleotide Accession Number</li> <li>[-] RefSeq mRNA Accession Number: NM_001964</li> <li>[-] Ensembl Transcript ID</li> </ul> </li> </ul> </li> <li>[-] Protein (5)           <ul style="list-style-type: none"> <li>[-] Protein_372670               <ul style="list-style-type: none"> <li>[-] Ensembl Peptide ID: ENSP00000239938</li> <li>[-] UniProtKB Primary Accession: P18146</li> <li>[-] GenBank Protein Accession Number</li> <li>[-] RefSeq Protein Accession Number: NP_001955</li> </ul> </li> <li>[-] Protein_372672</li> <li>[-] Protein_3435049</li> <li>[-] Protein_5676098</li> <li>[-] Protein_5852919</li> </ul> </li> </ul>	

Figure C.1. One record of an administrator defined category (Genomic Identifiers)



## Appendix D Types of Filters

A filter control is shown when you choose to filter the data of a data-category (see **How to filter data**).

The type of filter control upon the column based on which you wish to filter data.

Table D.1: Column Type and Applicable Filter

<b>Data type of the column</b>	<b>Applicable filter</b>
Numeric	Range filter
Text	Pattern filter
Column with permissible values	Enumerated filter
Boolean	Enumerated filter

### Using the Range Filter

The system shows a **Range Filter** if the column selected as a basis for the filter has numeric values. It allows you to select the desired minimum and maximum values of the column.

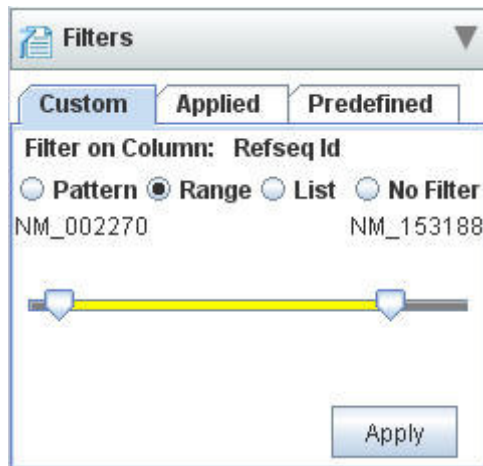


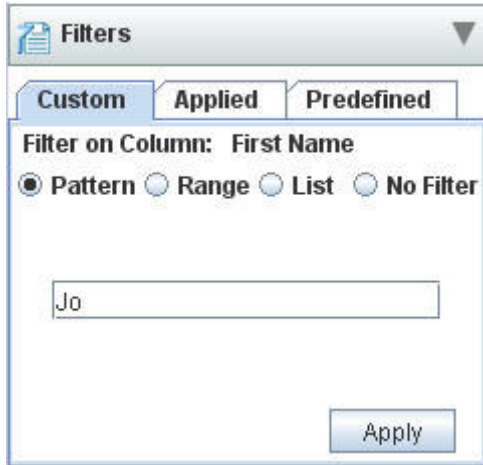
Figure D.1. Set Filter for Range

There are two sliders, the left slider and the right slider. The value denoted by each slider is shown at the two ends.

1. Position the left slider to the minimum value desired.
2. Position the right slider to the maximum value desired.
3. Click **Apply** to apply the filter.

## Using the Pattern Filter

The system shows a **Pattern Filter** if the column selected as a basis for the filter has text values. It allows you to specify a pattern that the values of that column should satisfy.



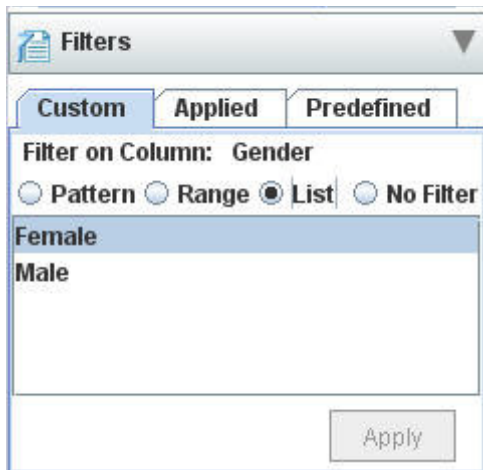
The screenshot shows a dialog box titled "Filters" with a dropdown arrow. It has three tabs: "Custom", "Applied", and "Predefined". The "Custom" tab is selected. Below the tabs, it says "Filter on Column: First Name". There are four radio buttons: "Pattern" (selected), "Range", "List", and "No Filter". Below the radio buttons is a text input field containing the text "Jo". At the bottom right of the dialog is an "Apply" button.

Figure D.2. Pattern Filter

In the **Enter Pattern** field, type the desired pattern and click **Apply**.

## Using the Enumerated Values Filter

The system shows an **Enumerated Values Filter** if the column selected as a basis for the filter has a set of permissible values. It allows you to specify the values that the column should satisfy.



The screenshot shows a dialog box titled "Filters" with a dropdown arrow. It has three tabs: "Custom", "Applied", and "Predefined". The "Custom" tab is selected. Below the tabs, it says "Filter on Column: Gender". There are four radio buttons: "Pattern", "Range", "List" (selected), and "No Filter". Below the radio buttons is a list box containing the text "Female" and "Male". At the bottom right of the dialog is an "Apply" button.

Figure D.3. Enumerated Values Filter for the column gender

Select the desired permissible values and click **Apply**.

---

# Index

- Administrator defined category, 60
- Advanced search, 11
- Ambiguity resolver, 18
- Analyze Data, 36, 40, 46
- Apply Data List, 24
- Breadcrumbs, 24
- caB2B, 3
- caBench-to-Bedside. *See* caB2B
- Categories, 4
- CDC, 43
- Concept code, 12
- Creating Experiments, 36, 37
- Custom Data Category
  - creating, 44
  - definition, 43
- Data
  - analyzing, 36, 40, 46
  - filtering, 36, 40, 41
  - visualizing, 37, 40, 48
- Data List, 9, 32, 37
  - Saving, 34
- Define Search Results View, 9, 20
- Editing filters, 42
- Executing saved query, 30
- Experiment Details page, 39
  - Operations panel, 40
  - Visualization panel, 40
- Experiments
  - creating, 36, 37
  - definition, 36
  - example, 36
  - opening, 36, 39
  - viewing, 38
- Filters
  - editing, 42
  - enumerated value, 64
  - pattern, 64
  - range, 63
  - removing, 43
  - saving, 36, 40, 43
  - types, 63
- Limits
  - connecting, 16
  - DAG, 14
  - defining, 9, 12, 15, 21
  - definition, 5
  - deleting, 15
  - editing, 15
  - information panel, 19
- Login, 3, 5
- My Data List Summary, 24, 29
- Opening Experiments, 36, 39
- Path, 16
- Plotting data series, 49
- Projects, 36
- Query
  - definition, 5
  - executing, 30
  - saving, 9, 27
- Query Wizard, 9, 36
- Related data, 23
- Removing filters, 43
- Save query
  - advantages, 27
  - executing, 30
  - non-parameterized, 28
- Search categories, 53
- Spreadsheet view, 39
- View Search Results, 9, 21
- Viewing Experiments, 38
- Vizualise data, 37, 40, 48
- Vizualize data
  - bar chart, 48, 49
  - line chart, 48, 50
  - scatter plot, 48, 50