caArray 2.0

User's Guide

NATIONAL® CANCER INSTITUTE

Center for Bioinformatics

Revised May 7, 2008

CREDITS AND RESOURCES

caArray Development and Management Teams			
Development	***Quality Assurance***	***Documentation***	***Project and Product Management***
Eric Tavela[2]	Tom Boal[5]	Eric Tavela[2]	Mervi Heiskanen[1]
Bill Mason[2]	Ron Keene[5]	Brent Gendleman[2]	Anand Basu[1]
Todd Parnell[2]	Xiaopeng Bian[1]	Todd Parnell[2]	Brent Gendleman[2]
Dan Kokotov[2]		Paul Duvall[4]	Juli Klemm[1]
Rashmi Srinivasa[2]		Levent Gurses[4]	
Scott Miller[2]		Jill Hadfield[1]	
Systems and Application Support		***Training***	
Wei Lu[3]		Don Swan[3]	
Vanessa Caldwell[3]			
Michael Gomes[3]			
[1] National Cancer Institute Center for Bioinformatics (NCICB)		[2] 5AM Solutions	[3] Terrapin Systems
[4] Stelligent	[5]NARTech		

Contacts and Support	
NCICB Application Support	http://ncicbsupport.nci.nih.gov/sw/ Telephone: 301-451-4384 Toll free: 888-478-4423

TABLE OF CONTENTS

USING THE CAARRAY USER'S GUIDE

This chapter introduces you to the *caArray 2.0 User's Guide* and suggests ways you can maximize its use.

Topics in this chapter include:

Introduction to the caArray User's Guide

The *caArray 2.0 User's Guide* is the companion documentation to the caArray software application. The *caArray User's Guide* includes information and instructions for the end user about using caArray.

Organization of this Guide

The *caArray 2.0 User's Guide* contains the following chapters:

Using the caArray User's Guide—This chapter introduces you to the *caArray 2.0 User's Guide* and suggests ways you can maximize its use.

Chapter 1 About caArray—This chapter provides a caArray overview as well as a description of the relationship of caArray to caBIG™.

Chapter 2 Getting Started in caArray—This chapter provides details about launching caArray, navigating the interface, accessing online help and other links in the interface.

Chapter 3 Navigating and Searching in caArray—This chapter describes the process for creating and working with Contacts, namely individual and group contacts in caArray.

Chapter 5 Curation Tools—This chapter describes curation tasks available to all logged in users that relate to array designs, protocols and vocabulary terms.

Chapter 6 Creating and Managing Experiments—This chapter details instructions for creating and working with Experiments in caArray.

Chapter 7 Submitting Data to an Experiment—This chapter describes the processes for uploading, validating and importing array content and annotation data files into a caArray Experiment.

Chapter 8 Extracting Data from caArray—This chapter describes the processes for downloading data from the caArray repository.

Chapter 9 User Account Management—This chapter describes the process for creating user accounts and collaboration group accounts and managing the group accounts of caArray.

Appendix A References—This appendix includes descriptions and links to references closely related to caArray technology and bioscience.

Index—This section of the guide provides a complete index.

User's Guide Text Conventions

Table *2.1* illustrates how text conventions are represented in this guide. The various typefaces differentiate between regular text and menu commands, keyboard keys, toolbar buttons, dialog box options and text that you type.

Convention	*Description*	*Example*
Bold & Capitalized Command **Capitalized command > Capitalized command**	Indicates a Menu command Indicates Sequential Menu commands	**Admin > Refresh**
TEXT IN SMALL CAPS	Keyboard key that you press	Press ENTER
TEXT IN SMALL CAPS + TEXT IN SMALL CAPS	Keyboard keys that you press simultaneously	Press SHIFT + CTRL and then release both.
`Monospace type`	Used for filenames, directory names, commands, file listings, and anything that would appear in a Java program, such as methods, variables, and classes.	`URL_definition ::= url_string`
Icon	A toolbar button that you click	Click the **Paste** button () to paste the copied text.
Boldface type	Options that you select in dialog boxes or drop-down menus. Buttons or icons that you click.	In the Open dialog box, select the file and click the **Open** button.
Italics	Used to reference other documents, sections, figures, and tables.	*caCORE Software Development Kit 1.0 Programmer's Guide*
`Italic boldface monospace type`	Text that you type	In the New Subset text box, enter ***`Proprietary Proteins.`***
Note:	Highlights a concept of particular interest	**Note:** This concept is used throughout the installation manual.

Table 2.1 caArray Guide Text Conventions

Convention	*Description*	*Example*
Warning!	Highlights information of which you should be particularly aware.	**Warning!** Deleting an object will permanently delete it from the database.
{ }	Curly brackets are used for replaceable items.	Replace `{root directory}` with its proper value, such as `c:\cabio`

Table 2.1 caArray Guide Text Conventions (Continued)

CHAPTER 1
ABOUT caARRAY

This chapter provides a caArray overview as well as a description of the relationship of caArray to caBIG™.

Topics in this chapter include:

caArray Overview

caArray is an open-source, user-driven, role-based, web and programmatically accessible data management system. caArray guides the annotation and exchange of array data using a federated model of local installations whose results are sharable across the cancer Biomedical Informatics Grid (caBIG™) (https://cabig.nci.nih.gov/). Identified primarily as a data service on the Grid, caArray furthers translational cancer research through acquisition, dissemination and aggregation of semantically interoperable array data to support subsequent analysis by tools and services on and off the Grid. As array technology advances and matures, caArray will extend its logical library of assay management.

The following services are provided by caArray:

- Browsing and searching across experiments. See *Chapter 3 Navigating and Searching caArray*
- Creating and managing array experiments. See *Chapter 4 Creating and Managing Experiments*
- Annotating experiments. See *Chapter 6 Submitting Data to an Experiment*
- Managing array designs, protocols and vocabulary terms. See *Curation Tasks*
- Uploading, validating, and importing array data. See *Chapter 6 Submitting Data to an Experiment*

- Extracting data from caArray. See *Chapter 7 Extracting Data from caArray*
- Managing collaboration groups. See *Chapter 8 User Account Management*

Relationship of caArray to caBIG™

The National Cancer Institute (NCI) has launched the caBIG™ (cancer Biomedical Informatics Grid™) initiative to accelerate research discoveries and improve patient outcomes by linking researchers, physicians, and patients throughout the cancer community.

The mission of caBIG™ is to provide an infrastructure for creating, communicating and sharing bioinformatics tools, data and research results, using open data standards and shared data models. This supports the development of new types of analysis within and across experiments and allows new forms of collaboration, enabling the sharing of data sets and a range of analytical tools.

The primary goal of caArray is to further translational cancer research through acquisition, dissemination and aggregation of high quality array data to support subsequent analysis. The opportunity for caArray use among the cancer centers and their collaborators through caBIG™ will ultimately benefit the cancer community.

caArray development continues to proceed with an open architecture and supportive documentation to allow for future enhancements, particularly with regard to interfacing with additional analysis tools. The goal is to create an extensible array data management system that is non-platform-specific and potentially customizable, enabling development that will continue to expand the vision of caBIG™.

CHAPTER

2

GETTING STARTED IN CAARRAY

This chapter introduces you to the caArray interface and its navigation as well as to global operations used in all of the caArray viewing windows.

Topics in this chapter include:

caArray Fundamentals

caArray 2.0 supports the following browsers:

- PC's running Microsoft XP: Firefox 2.0+ and Internet Explorer 6.0+
- Apple Macs running OS X: Firefox 2.0+ and Internet Explorer 5.5

If you have questions about this, contact NCICB Application Support, http://ncicb.nci.nih.gov/NCICB/support.

The caArray application can be accessed from NCICB using the following URL: https://array.nci.nih.gov.

For instructions about downloading and installing caArray 2.0 at your site, see your local administrator and/or refer to the *caArray Local Installation Guide* that can be downloaded from the caArray download page: http://ncicb.nci.nih.gov/download/downloadcaarray.jsp.

Note: If you are using a local installation of caArray, contact your Principal Investigator/Laboratory Manager/System Administrator for the correct URL for your use.

Browsing and Searching caArray

Browsing and Searching Before Login

Once you are at the caArray Portal Welcome login page, without being logged in you can browse caArray public data or you can perform a search of caArray public projects. (*Figure 2.1*).

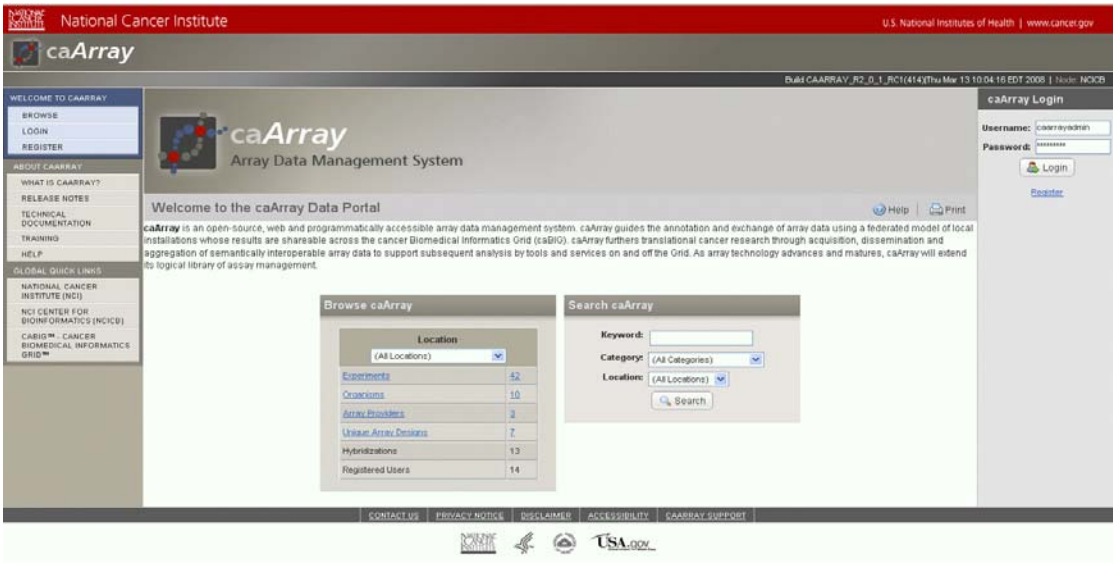


Figure 2.1 Prior to login, you can use these dialogs to browse or search the caArray database

The Browse caArray dialog box on the left center of the page lists the number of experiments saved in the database in each of the listed categories. The text categories shown in blue hypertext format can be clicked to open additional pages that display experiments with attributes in that category.

Note: The pages that open from the Browse dialog box list all caArray Experiments that have not been explicitly removed from visibility. For more information on the visibility options for Experiments, see *Managing Experiment Visiblity* on page 49.

- For more information about browsing the caArray database, see *Browsing the caArray Repository* on page 19.
- The Search caArray dialog on the right center of the page allows you to launch a search of the caArray database for public objects. For more information about executing a caArray search, see *Searching the caArray Repository* on page 22.
- Options allowing you to submit, view, modify, and add microarray experiment data to caArray are dependent on your user privileges, once you are logged in. See Table *2.2* for more information.

Note: You must obtain a user account in order to log in. For more information, see *Requesting a User Account* on page 10.

Browsing and Searching After Login

Once you have logged into caArray, the Browse and Search features are available from any page.

- To start the Browse function after login, click the **Browse** option on the left sidebar (*Figure 2.2*).

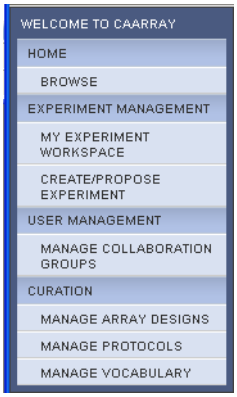


Figure 2.2 Browse options on left sidebar

caArray opens the Welcome to caArray Data Portal page, where you can launch a browse through the system, and as described in in the previous section.

- To start a Search after login, enter search query parameters in the Search text box in the upper right corner of the user interface (*Figure 2.3*).

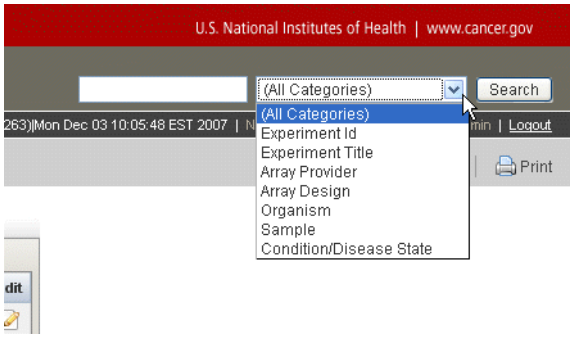


Figure 2.3 Search text box displays in every browser page

For more information about launching a search, see *Chapter 3 Navigating and Searching caArray*.

caArray New User Account and Login

From the Welcome to caArray Data Portal page, you can register as a new user or log in if you already have a user account (*Figure 2.4*):

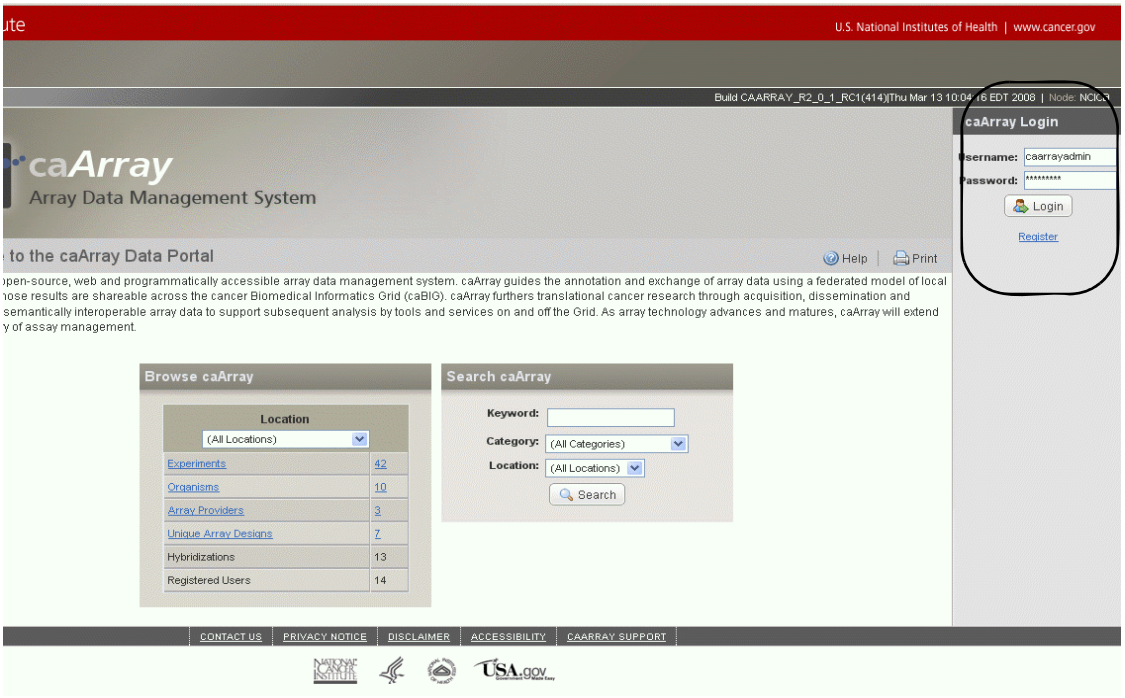


Figure 2.4 caArray login page

- To register for a new user account, see the following section for more information.

To log in, follow these steps:

1. Navigate to the caArray home page. Use the URL to the NCICB instance http://array.nci.nih.gov (*Figure 2.4*), or contact your System Administrator for the URL to your local instance of caArray.
2. Enter your user ID and password in the upper right corner of the Welcome to caArray Data Portal page
3. Click **Login**.

After caArray verifies your credentials, the application opens to the caArray workspace, providing access to all features allowed by the permissions granted to you. For more information about finding your way around caArray, see *Navigating the caArray User Interface* on page 14.

Requesting a User Account

To request a caArray user account, you must complete steps 1 through 4.

1. Go to the NCICB caArray login page http://array.nci.nih.gov or use the URL provided by your System Administrator for the caArray instance at your institution.

2. Click the **Register** hypertext link, either in the left sidebar or under the caArray Login section in the upper right of the page. This opens the account registration form (*Figure 2.5*).

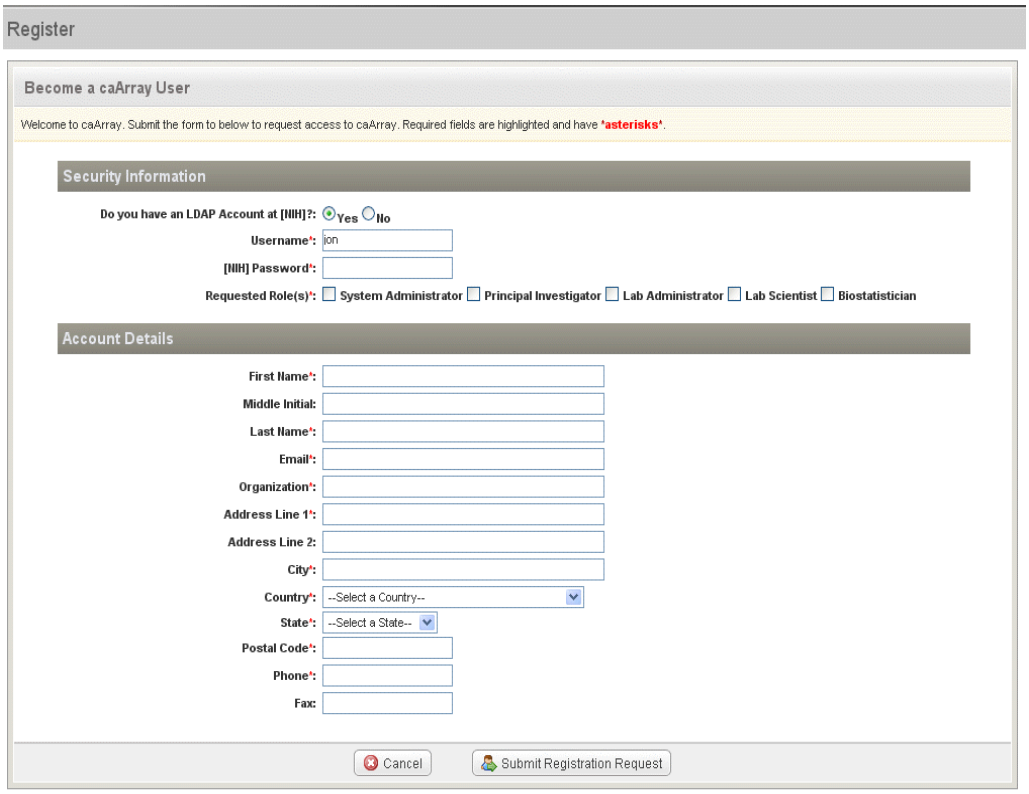


Figure 2.5 New user account registration form

3. In the Become a caArray User form, enter the appropriate information[1].

   - **Security Information**

     — **Do you have an LDAP account** [a user profile with your institution] at [NCICB or your institution]**?**

     If **Yes**, enter your username and case-sensitive password for the purposes of verifying that it is correct. After you submit your request, you can continue to use caArray without an account to browse and search available experiments and download data while your account is verified and activated.

     --Username*

     --Password*

     --**Requested role(s)***--Select one or more of the roles. Roles are described in Table *2.1*.

     If your LDAP profile is not validated, caArray indicates that the LDAP credentials do not check out. You are asked to reenter them, but you can choose to answer no, and the System Administrator will manually

1. Items with an asterisk or highlight are required.

ensure you don't get a duplicate LDAP account during provisioning. You can **Cancel** or talk with your System Administrator about the problem.

If you select **No** [you do not have an LDAP account], the text boxes for entering the LDAP account information disappear. You must indicate the role you would like to be assigned in caArray, and continue entering the appropriate information in the **Account Details** section.

Role	*Description*	*Permissible 2.0 Actions*
System Administrator	Person responsible for the effective operation of caArray	Manages users
Principal Investigator [PI]	Owns experiments and studies and/or projects	Submit data Write Experiment designs Submission of annotation Submission of array data
Lab Administrator	Same as PI in caArray 2.0	Same as PI in caArray 2.0
Lab Scientist	Same as PI in caArray 2.0	Same as PI in caArray 2.0
Biostatistician	Same as PI in caArray 2.0	Same as PI in caArray 2.0
Note: In subsequent versions of caArray, role-specific features will be implemented that will limit certain actions to specific users.		

Table 2.1 caArray role descriptions

- **Account Details**
  - **First Name***
  - **Middle Initial**
  - **Last Name***
  - **Email [address]***
  - **Organization***
  - **Address [Lines 1* and 2}**
  - **City***
  - **State/Province***
  - **Postal [or Zip] Code***
  - **Country*--Select from the drop-down list**
  - **Phone***
  - **Fax**

4. Click **Submit Registration Request** to execute the request, or click **Cancel** to abort the registration.

Once the request for a new account is sent to NCICB, it takes 24-72 hours to process. (The process time for a local installation at your institution may differ.) You will receive an email response when the account has been activated.

Note: Once you register, you can continue to use caArray without an account to browse and search available experiments and download data while your account is activated.

When your account is registered, the UserID and password you are assigned determines your access rights for the software.

Using caArray Online Help

The caArray online help explains how to use all of the features of the caArray portal.

Note: You can open online help without being logged into caArray.

To access online help in caArray, click the **Help** icon/menu options available in the upper right corner of the user interface (

) and on the left sidebar under the **About caArray** section.

Online help opens with two display panels:

1. The left panel displays the Table of Contents (TOC), and also offers access to the Index and Search features of online help. The TOC can be expanded. All topics listed in the TOC and index are hypertext links to the referenced topics.
2. The right panel displays the Welcome to caArray Online Help page and other topic contents.

The following features facilitate your navigation of online help:

- The breadcrumb trail at the top of the page shows the relative location of the current help topic relative to neighboring topics. Click a breadcrumb link to display that help topic.
- Click the **Back** or **Forward** links at the top of the page to display help topics you have previously viewed.
- Follow hypertext links or the **Related Topics** buttons in the help topics to open other closely related topics. If the current help page has related topics associated with it, you can also view them by clicking the **Related Topics** button () at the top right of the help page.
- Locate topics using the table of contents that displays in the left pane of the online help project or the **Index** tab that displays at the top of the Table of Contents pane.
- Perform word searches of Help by entering query text in the search text box.
- Print the current topic by clicking the **Print** button () at the top right of the help page.

Note: caArray 2.0 does not have context sensitive help.

Navigating the caArray User Interface

The caArray provides a user-friendly interface to display options relating to login, accessing documentation or performing tasks for proposed or existing experiments.

Elements in the caArray User Interface

The terms in Table *2.2* identify and describe elements of the caArray user interface.

Term	*Definition*
Left Vertical Navigation Task Menu	Hypertext links associated with the caArray application, caArray documentation and Global Quick Links.
[Online] Help	This caArray documentation accessed from your computer screen provides help and answers for questions about using the software. Help icon/menu options are available in the upper right corner of the user interface and on the left sidebar under the About caArray section. **Note:** caArray 2.0 does not have context sensitive online help. You can open online help and use the TOC, index or perform a text search.
Print	A **Print** icon displays on each browser interface. This prints the current page.
Browse caArray	The **Browse** dialog lists database categories and the number of public experiments in each. Click each hypertext link to browse details of the experiment categories.
Search {caArray database objects}	The Search caArray dialog box, available pre- and post login to all users, allows you to launch a search of public experiments. Enter keywords or select a category. (The **Location** dropdown lists only the current caArray instance of the users.) For more information about caArray searches, see *Searching the caArray Repository* on page 22.
Work Area Tabs	Located across the top of many of the caArray user interface windows. Work area tabs represent the tasks/annotations you create as components of an experiment
Work Queue/ Public Experiments	caArray task tabs in the My Experiment Workspace. These comprise sets of Experiment information and annotations performed on or associated with array experiments. Experiments displayed on the Work Queue tab are those with which you are associated. Experiments displayed on the Public Experiments tab are those with which you are associated and that have been made public. For more information, see *Managing Experiment Visiblity* on page 49.

Table 2.2 Elements of the caArray browser window

caArray Welcome Page Navigation Menu

The left sidebar of the caArray Welcome page provides links to an array of information.

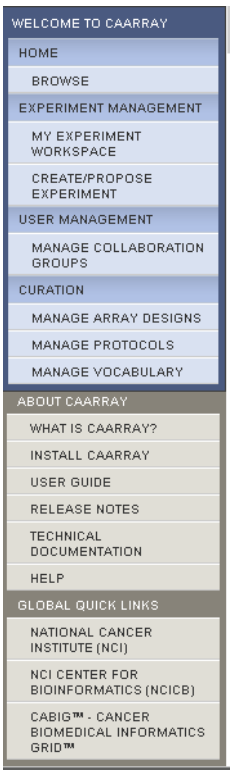


Figure 2.6 The caArray Welcome page navigation or left sidebar menu

The **Welcome to caArray** links direct you to Welcome caArray pages:

- These three options appear only before login:
  - **Login**--the Welcome/Login page
  - **Register**--the form for requesting a new account
- These four categories appear only after login:
  - **Home**--Click **Browse** to display the Browse and Search dialog boxes
  - **Experiment Management**--Open the My Experiment Workspace and the page to begin creating an Experiment.
  - **User Management**--Allows you to create and manage collaboration users groups

    Note: Only System Administrators can create users. For more information, see *Managing User Accounts* on page 80.
  - **Curation**--Includes functions related to managing Array Designs and Protocols and creating and editing vocabulary terms and protocols.

The **About caArray** links direct you to various components of caArray documentation:

- **What is caArray**--Opens a brief introduction to caArray
- **Install caArray**--Opens the caArray Installation Guide pdf

- **User Guide**--opens the caArray User's Guide pdf
- **Release Notes**--opens Release Notes for caArray
- **Technical Documentation**--opens the caArray Technical Guide pdf
- **Help**--Opens the full online help project

The **Global Quick Links** provide sources for caArray-related bioinformatics information on the Internet. These include links to the following websites:

- **National Cancer Institute (NCI)** (http://www.cancer.gov/)
- **NCI Center for Bioinformatics (NCICB)** (http://ncicb.nci.nih.gov/)
- **caBIG™ Cancer Biomedical Informatics Grid™** (https://caarraydb.nci.nih.gov/caarray/)

User Interface Footer

Options available in the footer are described as follows:

- **Contact Us--**Contact information for NCICB
- **Privacy Notice**--NIH Web Privacy Notice
- **Disclaimer--**NIH Disclaimers
- **Accessibility--**NCI Web Accessibility Feedback Form
- **User Support**--Contact information for NCICB Application Support (http://ncicb.nci.nih.gov/NCICB/support)

My Experiment Workspace

Note: This section describes in limited detail the elements of the experiment user interface. For more information about working with experiments, see *Chapter 3 Navigating and Searching caArray* and *Chapter 4 Creating and Managing Experiments* in this guide.

The three most common uses of caArray are described as follows:

- Submit new data, or modify existing data (with appropriate permissions)
- Search public or private data already in the caArray system
- User and group management (with appropriate permissions)

Upon login, in the user interface, the My Experiment Workspace displays tabs and options relating to draft, proposed, or public experiments with which you are associated.

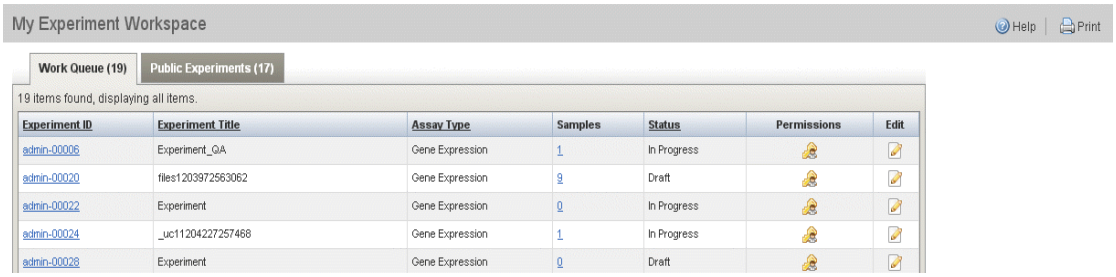


Figure 2.7 caArray My Experiment Workspace

The My Experiment workspace displays two tabs:

1. **Work Queue**--This page lists non-public experiments with which you are associated, either in Draft or In Progress status.
2. **Public Queue**--Lists all experiments in caArray with which you are associated that have been made public.

For more information, see *Managing Experiment Visiblity* on page 49.

The listed experiments on each tab are visible in table format, according to the permissions assigned to you. The column headings display fields populated with metadata for each corresponding experiment.

Note: Column headings that are underlined are sortable by clicking on the heading.

- **Experiment ID**--The ID autogenerated by caArray; click the hypertext link to open experiment details
- **Experiment Title**--The name designated for the experiment by the principle investigator
- **Assay Type**--The category of array assay type for this experiment; either **Gene Expression** or **SNP**
- **Samples**--The number of samples used in the experiment. Click the hypertext link to open the Annotations tab to the samples details page.
- **Status**--The current status of the experiment: **Draft** or **In Progress**
- **Permissions**--Click the icon to assign or modify the experiment permissions. See *Managing Experiment Visiblity* on page 49.
- **Edit**--Click the icon to edit experiments with the appropriate permissions. See *Editing an Experiment* on page 49.

Each of these experiment elements is described in separate chapters in this user's guide.

Online help and additional resource links remain fixed through all the user interface pages.

CHAPTER

3

NAVIGATING AND SEARCHING CAARRAY

This chapter describes the processes for browsing and/or conducting searches of the caArray repository.

Topics in this chapter include:

Browsing the caArray Repository

In caArray, you can browse the repository, a feature that can be launched either before login by any user or after login by a registered user.

Note: A non-logged in user can view only public data. A logged in user can view public data and non-public data to which (s)he has assigned permissions.

You can browse experiments organized by organisms, providers, array designs from the home page of a single installation.

Once you are on the caArray Portal Welcome login page, the Browse caArray dialog box on the left center of the page lists the number of experiments saved in the database in each of the listed categories. You can click text categories shown in blue hypertext format to open additional pages that display information about those experiments (*Figure 3.1*).

Note: At any point in using the Browse or Search features, you can return to the home page by clicking the caArray logo at the top of the user interface or the **Browse** link in the left sidebar, if you are logged in.

From this page, you can also launch a search of the caArray repository. For more information, see *Searching the caArray Repository* on page 22.

.

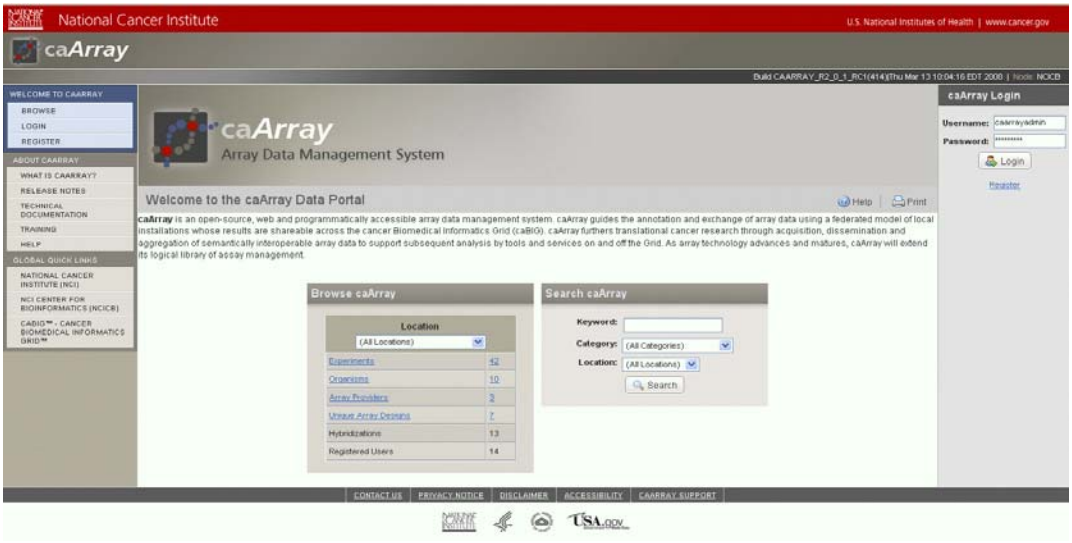


Figure 3.1 Prior to login, you can use these dialog boxes to browse or search the caArray database

To use the browse feature, follow these steps:

1. Click any of the blue hypertext links in the experiment properties category list of the Browse dialog box. The *Table 3.1* lists the categories that display in the Browse dialog box.

   The page that opens depends on the category you selected. The descriptions in *Table 3.1*outlinethe detail that opens for each category that you click.

Browse Dialog Box Category	***Description***
Experiments	Both the experiments and corresponding number links open the Browse by Experiments page.
Organisms	Both the organisms and corresponding number links open the Browse by Organisms page. The page is organized alphabetically by tabs that correspond to unique organism names found among the installation's available experiments.
Array Providers	Both the array providers and corresponding number links open the Browse by Array Providers page. An array provider is generally the company or group that manufactured the array design used in the experiment. The page is organized alphabetically by tabs that correspond to different array manufacturer's names. The count of experiments available within each named group also displays on the tab. **Note:** Only Affymetrix, Illumina and GenePix formats are fully supported with validation and parsers in caArray 2.0. For more information, see the Note about File Types in *Managing Data* on page 65.

Table 3.1 Browse dialog box categories

Browse Dialog Box Category	*Description*
Array Designs	An array design is a file that explains the design of an array. This includes such information as the array layout and design, its substrate, surface type, attachment type, the array strand type and the coordinates of each gene on the array.
Hybridizations	The number of hybridizations in the installation is visible, for information only. You cannot open hybridizations from this page.
Registered Users	The number of registered users in the repository is visible, for information only. You cannot open registered users from this page.

Table 3.1 Browse dialog box categories

Note: **Location** refers to the caArray instance, either at your institution or at NCICB.

2. Once the tab or page opens when you click any of these categories (*Table 3.1*), the same metadata displays on all pages for the list of experiments located for that category (*Table 3.2*).

Note: Only public experiments or non-public experiments which have not been explicitly removed from visibility display in the browse results. You can open only public experiments and non-public experiments with which you are associated.

Experiment Category	*Description*
Experiment ID	The autogenerated identification assigned by caArray. Click the hypertext link to open the corresponding experiment tabs which contain all current experiment information. Only the public data can be opened or private data to which you have been given access.
Experiment Title	The experiment title defined manually, naming and/or briefly describing the experiment
Assay Type	The type of array assay represented by the experiment; for example, Gene Expression, SNP, Exon, etc.
Primary Contact	The person named as the point of contact for the experiment. **Note:** The PI and POC can be the same person, but do not have to be so. Click the hypertext link or the envelope icon ( ) to open an email form where you can draft an email to this contact, if named.
Organism	The organism that is the source of the sample biomaterials used in the experiment
Condition/Disease State	The disease state of the source materials used in the experiment

Table 3.2 Experiment metadata categories

Experiment Category	*Description*
Samples	The number of samples identified in the experiment. Click the hypertext link to open the experiment to the samples details page.
Updated	The date of the most recent update of the experiment

Table 3.2 Experiment metadata categories

Note: Columns with underlined headings are sortable by clicking on the heading. caArray paginates the result sets in groups of 20.

Searching the caArray Repository

In caArray, you can search the repository, a feature that is available before you login or to a non-registered "Anonymous User".

The caArray search feature allows you to locate caArray content based on user-defined search criteria. Once you find the information you seek, you can open the experiment to review or edit details, if you have proper permissions. Additionally, you can extract the data, follow hyperlinks to additional data, or you can return to the search feature to refine the query parameters.

To launch a search for a caArray experiment, follow these steps:

1. *Before login*, from the caArray Portal Welcome page , locate the Search dialog box on the right center of the page.

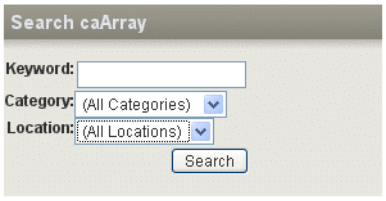


Figure 3.2 Search dialog box

OR

2. *After login*, locate the Search area of the page, in the upper right-hand corner.

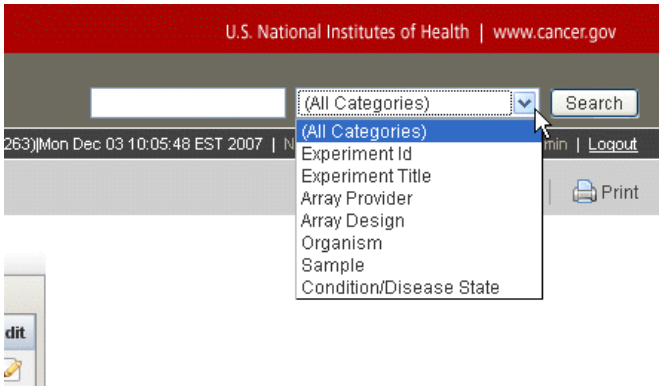


Figure 3.3 Section of the caArray page for launching a search

3. Define the search criteria by using the search options described *Table 3.3*.:

Search Option	***Description***
Category	Select one of the Experiment properties categories listed: • **All Categories** • **Experiment ID** • **Experiment Title** • **Assay Provider** • **Array Design** • **Organism** • **Sample** • **Disease State** Only experiments in the category selected will be searched. If you do not select a category, **All Categories** (default) remains selected, and caArray will search all experiments.
Keyword	In the text box, enter one or more words, separated by spaces. *Example*: ***breast cancer*** **Note:** Queries are case insensitive; wild cards are implied on both sides of the query string. No logic statements, such as AND or OR or SQL statements are supported in these search features.
Location	The list displays only the current caArray instance you are using, either your local institution or NCICB.

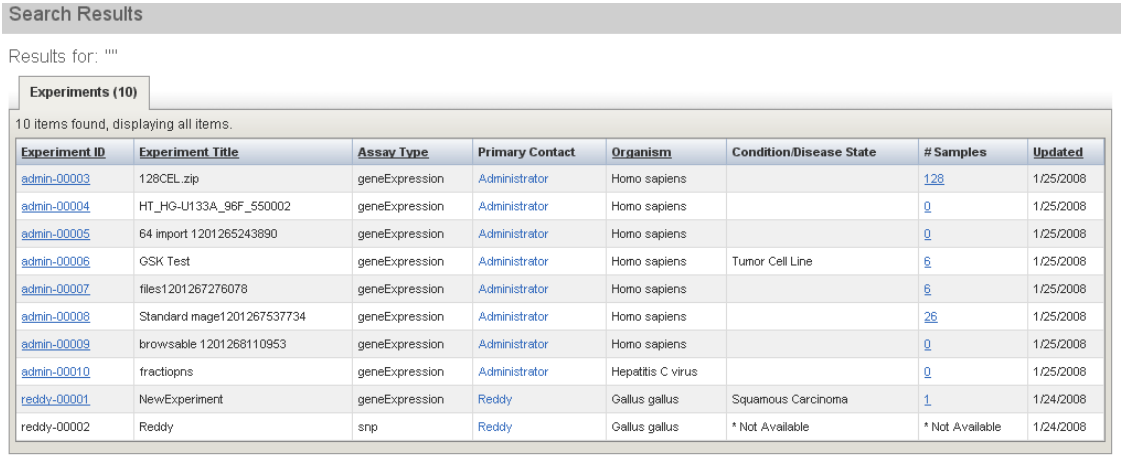
Table 3.3 Search criteria options

4. Click **Search** to execute the search.

Note: If you click **Search** without defining query parameters, the search is unrestricted, and all experiments in caArray that have not been explicitly removed from visibility display on the Search Results page.

Experiment Search Results

Experiment search results display on a new page, Search Results. If no results are found, a message informing you of that fact displays on the Search Results page (*Figure 3.4*).



Experiment ID	Experiment Title	Assay Type	Primary Contact	Organism	Condition/Disease State	# Samples	Updated
admin-00003	128CEL.zip	geneExpression	Administrator	Homo sapiens		128	1/25/2008
admin-00004	HT_HG-U133A_96F_550002	geneExpression	Administrator	Homo sapiens		0	1/25/2008
admin-00005	64 import 1201265243890	geneExpression	Administrator	Homo sapiens		0	1/25/2008
admin-00006	GSK Test	geneExpression	Administrator	Homo sapiens	Tumor Cell Line	6	1/25/2008
admin-00007	files1201267276078	geneExpression	Administrator	Homo sapiens		6	1/25/2008
admin-00008	Standard mage1201267537734	geneExpression	Administrator	Homo sapiens		26	1/25/2008
admin-00009	browsable 1201268110953	geneExpression	Administrator	Homo sapiens		0	1/25/2008
admin-00010	fractiopns	geneExpression	Administrator	Hepatitis C virus		0	1/25/2008
reddy-00001	NewExperiment	geneExpression	Reddy	Gallus gallus	Squamous Carcinoma	1	1/24/2008
reddy-00002	Reddy	snp	Reddy	Gallus gallus	* Not Available	* Not Available	1/24/2008

Figure 3.4 caArray Search Results page

Note: Only public experiments or non-public experiments which have not been explicitly removed from visibility can be found via the search mechanism. You can open only public experiments and non-public experiments with which you are associated.

Search results are listed in table format, with columns displaying properties for each experiment; fields are described in *Table 3.4*. Most of these properties were identified when the experiment was created or edited.

Note: Columns with underlined headings are sortable by clicking on the heading. caArray paginates the result sets in groups of 20.

Search Results Properties	***Search Results Fields Descriptions***
Experiment ID	The autogenerated identification assigned by caArray. Click the hypertext link to open the corresponding experiment tabs which contain all current experiment information.
Experiment Title	The experiment title defined manually, naming and/or briefly describing the experiment
Assay Type	The type of array assay represented by the experiment; for example, Gene Expression, SNP, Exon, etc.
Primary Contact	The person named as the Point of Contact for the experiment. Click the hypertext link or the envelope icon ( ) to open an email form where you can draft an email to this contact, if named.
Organism	The organism that is the source of the sample biomaterials used in the experiment
Disease State	The disease state of the source materials used in the experiment

Search Results Properties	***Search Results Fields Descriptions***
Samples	The number of samples identified in the experiment. If public or if you are a data owner, click the hypertext link to open the Experiment Samples tab.
Updated	The date of the most recent update of the Experiment draft

Table 3.4 Experiment metadata categories

You can open any experiment to which your assigned permissions grant you access. For private experiments to which you have not been assigned permission, only the ability to contact POC is available.

To open the experiment details, click any **Experiment ID** or click the **Samples** number to open the experiment to the samples details page. You can review the experiment or contact the POC for the experiment, or with appropriate permissions, edit it or extract the experiment.

- For information about editing an experiment, see *Editing an Experiment* on page 49.
- For information about contacting the experiment POC, see *Primary Contact* in *Table 3.4*.
- For information about extracting data from an experiment, see *Downloading Data from caArray* on page 75.

Note: At any point in using the Browse or Search features, you can return to the home page by clicking the caArray logo at the top of the page.

CHAPTER

4

CREATING AND MANAGING EXPERIMENTS

This chapter describes the processes for proposing/creating caArray experiments, including all components.

Topics in this chapter include the following:

Overview of an Experiment

A caArray experiment captures all relevant information. This can include general information about the experiment, such as the experimental design and experimental factors; associated publications; biological samples; protocols; array design files; quality control and data processing steps; and so forth. Data files related to the samples used and hybridizations performed in the experiment, and supplementary documents can also be uploaded, validated and imported into the caArray experiment.

Once you log into caArray, the My Experiments Workspace displays by default. The basic elements of a caArray Experiment, shown on the Experiments page, are described in *Table 4.1*:

Term	*Definition*
Experiment Overview	Basic information about an experiment such as IDs, service and assay types, provider of array and array designs, source of biomaterials, and disease state

Table 4.1 Elements of a caArray Experiment

Term	*Definition*
Contacts	Principal Investigator and/or point of contact for the experiment
Annotations	Experimental factors and design, sources, samples, extractions, labeled extracts, and hybridizations
Data	Experimental data files uploaded, validated and imported into caArray; supplemental data
Publications	Publications associated with the experiment, primarily journal articles

Table 4.1 Elements of a caArray Experiment

Managing an experiment in caArray involves two primary features:

1. Creating an experiment with appropriate characteristics and annotations
2. Uploading the experimental research data files into caArray and associating them with the appropriate samples

With the appropriate permissions, you can create ("propose") an experiment, save the draft, edit it, and finally submit an experiment with its corresponding annotations to the caArray repository. A minimum set of information must be entered for an experiment before you can save it, but almost as soon as an experiment is begun, you can save it as a draft to be retrieved and completed at a later time. Alternatively, a principal investigator can have another designee with appropriate permissions complete the draft and submit the experiment.

Creating an Experiment

When you create an experiment in caArray, you begin entering information on the Overview tab. Once you have saved the information on the Overview tab, the experiment becomes a draft and additional tabs for entering experiment information become available.

To create an experiment in caArray, follow these steps:

1. If you plan to use the NCICB instance of caArray, go to the NCICB caArray login page https://array.nci.nih.gov and log in. If you plan you use the local installation of caArray at your center, see your local System Administrator for the URL.

   Once you are on the caArray Portal Welcome login page, the browser displays the experiment workspace.

2. On the left sidebar, click **Create/Propose Experiment** (*Figure 4.1*). This opens the Overview tab for entering overall characteristics for the experiment..

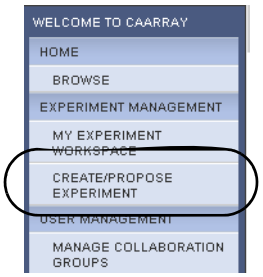


Figure 4.1 Create/Propose Experiment on left sidebar

3. Proceed to the *Overview Tab*.

Overview Tab

When you create a new experiment in caArray, the Overview tab initially displays alone in the user interface. Once you enter the appropriate information on this tab and save it, other tabs, also necessary for adding Experiment information, display as well.

To complete the Overview tab, follow these steps:

1. On the Overview tab, enter the appropriate information for Overall Experiment Characteristics[2] as described in the *Table 4.2*.

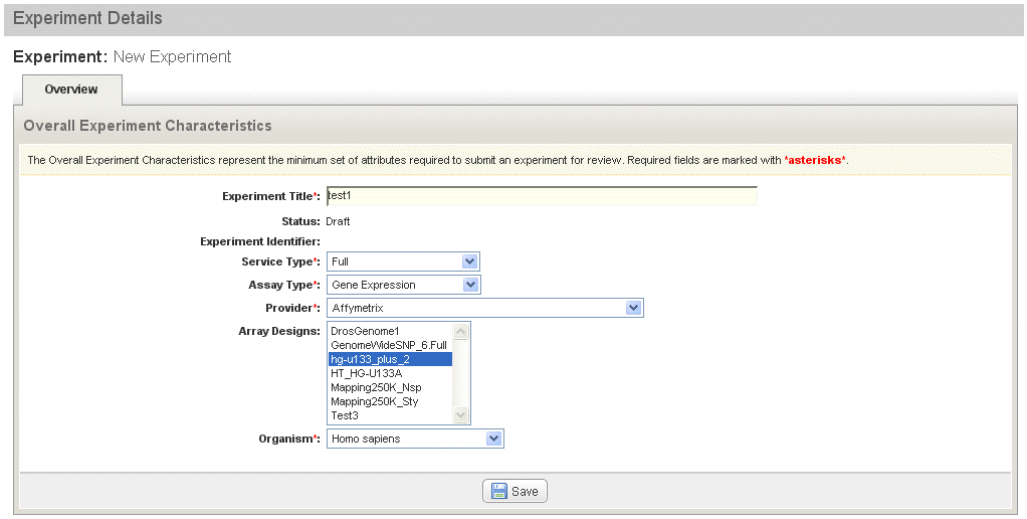


Figure 4.2 : Overview tab for an Experiment

Overview Tab Fields	***Description***
Experiment Title	Title designate by the PI or other user creating the experiment
Status	**Draft** displays by default before the experiment is formally submitted.
Experiment Identifier	This project identifier is autogenerated by caArray upon the initial save of the experiment. The experiment identifier is not editable. The ID is generated using the PIs last name followed by a 5 character number. *Example*: jdoe-90765. After the experiment has been saved or submitted, the experiment ID displays as a hypertext link that opens the experiment.

Table 4.2 Fields for Overall Experiment Characteristics

2. Fields with a red asterisk * are required.

Overview Tab Fields	*Description*
Service Type	Service Type is description of relationship between PI and lab. Select from the drop-down menu the appropriate service type. Options are: • **Full-**--typical relationship between PI and lab; with this selection, the experiment is created in caArray and "submitted to a lab" using the standard workflow. • **Publish**--This option exists for getting a significant amount of pre-existing experiment data into caArray, thereby providing a vehicle whereby the data can be annotated and making the data available to the caArray community. When you select this option, the data validation feature is turned off.
Assay Type	Select from the drop-down menu the appropriate assay type. Options are the following: • **Gene Expression**--experiment using microarrays intended to measure levels of transcribed genes • **SNP**--experiment using microarrays intended to detect nucleotide changes in chromosomal DNA • **aCGH**--array comparative genomic hybridization; a method for the analysis of chromosome copy number changes (gains/losses). • **Exon**--Exon arrays are designed to study which exons are present in an expressed gene. • **microRNA**--Experiment that measures activity among the 217 genes encoding miRNA. Patterns of gene activity that can distinguish types of cancers can be discerned. • **Methylation**--experiment that attempts to establish patterns of methylation genome-wide or within targeted promoters or CpG islands
Provider	Select from the drop-down menu the provider of the array. **Note:** Only Affymetrix, Illumina and GenePix formats are fully supported with validation and parsers in caArray 2.0. For more information, see the Note about File Types in *Managing Data* on page 65. Once selected, caArray automatically loads a corresponding list of array designs (next field).
Array Designs	Select one or multiple array designs, (using CTRL + click or SHIFT + click), from the automatically-generated list of array designs corresponding to the provider you selected. The array design of interest may have already been imported into caArray, or you can choose to import those of your choice. For more information, see *Managing Array Designs* on page 54.
Organism	Select from the drop-down menu the organism that is the source of the sample biomaterial used in the experiment.

Table 4.2 Fields for Overall Experiment Characteristics

2. After entering the information, click the **Save** button at the bottom of the page. Upon saving, caArray validates required fields and saves the experiment as a draft. A confirmation messages displays, verifying that the proposal is saved. If the validation fails, caArray display a message indicating which field(s) need correction.

   When you save the draft successfully, other tabs used for adding additional information for the experiment display.

3. Proceed to the **Contacts** tab *Contacts Tab*.

Contacts Tab

A caArray contact is a person associated with an experiment, either as a principal investigator (PI) or the point of contact (POC). The contact does **not** have to be a registered user of caArray.

To enter contact information for the experiment, follow these steps:

1. On the Contacts tab (*Figure 4.3*), enter information for the fields described in *Table 4.3*.

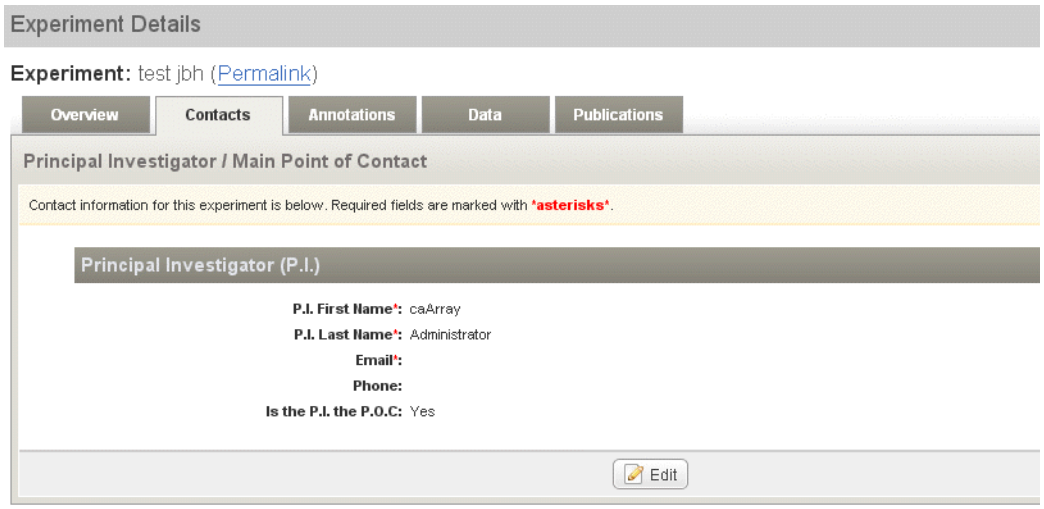


Figure 4.3 Contacts tab

Contact Fields	***Description***
P.I First Name	First and last names of the Principle Investigator (P.I.). These fields automatically populate with the name of the person creating the experiment, but this can be edited. Note that the actual P.I. does not have to be a registered user of caArray.
P.I Last Name	
Email Address	Email address of the P.I.
Phone	Phone number of the P.I.

Table 4.3 Contact fields

Contact Fields	Description
Is the P.I. the P.O.C. (point of contact for the experiment)?	Select **Yes** or **No**. If answer is **No**, caArray opens additional text boxes where you can add contact information for the project point of contact (POC). **Note:** The POC does not have to be registered user of caArray.

Table 4.3 Contact fields

2. Click **Save**.
3. Proceed to the *Annotations Tab*Annotations tab.

Annotations Tab

The Annotations tab opens with seven subtabs for entering annotation data for the experiment you are creating. The browser displays the Experimental Design subtab by default (*Figure 4.4*). Enter the appropriate information as described below in the following topics for each of the seven subtabs.

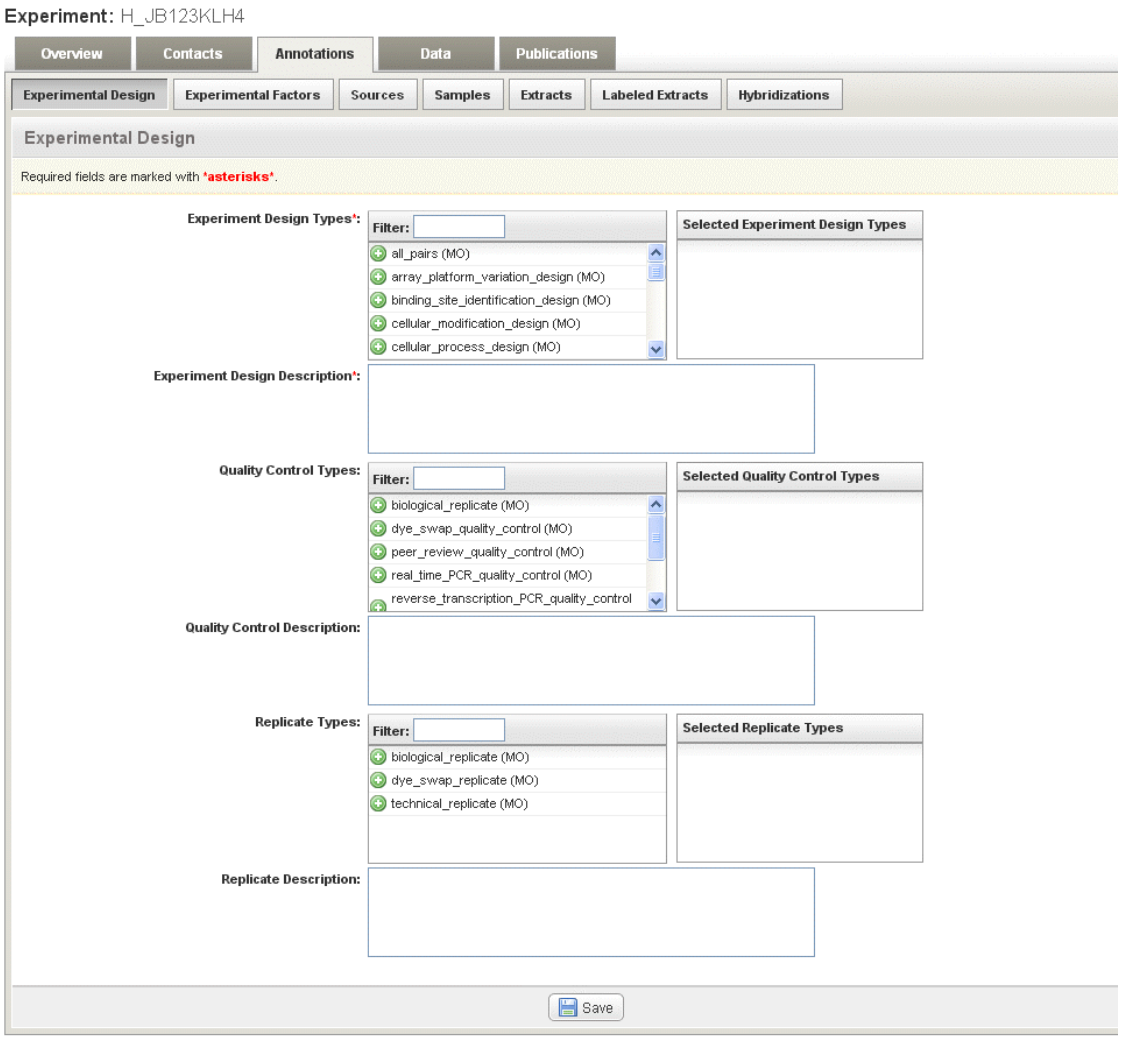


Figure 4.4 Upon opening, the Annotations tab displays the Experimental Design subtab and 6 other subtabs for entering experiment annotation data.

Experimental Design

The experimental design, (*Figure 4.4*), describes the intent of the research and a description that is common to all hybridizations performed in the experiment.

1. Enter on the Experimental Design subtab (under the Annotations tab) the appropriate information as described in *Table 4.4*[3].

Experimental Design Fields	*Description*
Experimental Design Type*	Select the experimental design type among the options in the drop-down menu.
Experimental Design Description	Enter a description for the experimental design used for the experiment.
Quality Control Types	Select the QC type in the displayed list.
Quality Control Description	Enter a description for the quality control used for the experiment.
Replicate Types	Select one or more replicate types kinds from the displayed list. Replicates can be either technical (arrays) or biological (laboratory animals or samples, etc.)
Replicate Description	If there are replicates used in the experiment, describe the number of replications and how the replicates were generated.

Table 4.4 Experimental Design fields

Note: **Experiment Design Type**, **Quality Control Type**, and **Replicate Type** terms are all from the MGED Ontology, http://mged.sourceforge.net/ontologies/MGEDontology.php.

2. Click **Save** to save the draft. Click **Cancel** to return to the subtab without adding the design.
3. Proceed to the *Experimental Factors***Experimental Factors** subtab.

Experimental Factors

Experimental factors are the intended sources of variation in the experiment. The Experimental Factors subtab (under the Annotations tab) displays any previous Experimental factors that may have been added previously to the experiment you are creating.

1. On the Experimental Factors subtab, click the **Add a New Experimental Factor** button.

3. Fields with a red asterisk * are required.

2. In the form that opens, enter the information as described in *Table 4.5*[4].

Experimental Factors Fields	***Description***
Factor Name*	Enter a name for the experimental factor.
Description	Enter a description for the experimental factor.
Category	Select the appropriate category for the experimental factor in the displayed list. **Note:** Terms are from the MGED Ontology, http://mged.sourceforge.net/ontologies/MGEDontology.php

Table 4.5 Experimental Factor fields

3. Click **Save** to save the draft. Click **Cancel** to return to the subtab without adding the factor.
4. Repeat steps 1 - 3 as often as needed to enter all the experimental factors for this experiment.
5. Click **Save** or **Cancel** to abort the action.
6. Proceed to the Sources subtab. See also *Biological Source Material* on this page.

Biological Source Material

Materials of biological origin are used in array experiments, and the state and characteristics of those biomaterials can be objectively documented and described. Biological materials can be treated, extracted, labeled and hybridized on arrays to study relative expression levels of the genes represent on the arrays. All categories of biological materials are created in caArray to be included as essential components in a caArray experiment.

Many biological materials' characteristics are defined in caArray by terms found in the caArray Controlled Vocabulary Terms. For more information, see *Managing [Controlled] Vocabulary [Terms]* on page 61.

In caArray, biological materials are divided into four different categories based on the treatment status of the material. The categories are based on MAGE-TAB specifications, as described in this paper: http://www.biomedcentral.com/1471-2105/7/489. The biological materials consist of **Sources**, **Samples**, **Extracts** and **Labeled Extracts**, defined below and illustrated in *Figure 4.5* in their hierarchical relationship.

- **Source** is any biological site from which the tissue for the array is derived before any preparation of the tissue for the array takes place. *Example:* human brain tumor tissue that has been treated with an anti-cancer drug.
- **Samples** are the original source biomaterials after initial treatment events. *Example:* A tissue or biopsy material treated to create a cell lysate.
- **Extracts** are samples after a treatment event in which DNA or RNA is extracted. *Example:* A sample tissue treated with RNA extraction method yields the extract, RNA.

4. Fields with a red asterisk * are required.

- **Labeled Extracts** are extracts that have been labeled for detection of the nucleic acids on the array. *Example*: The extract RNA is labeled with a fluourescent dye, yielding the labeled extract RNA.

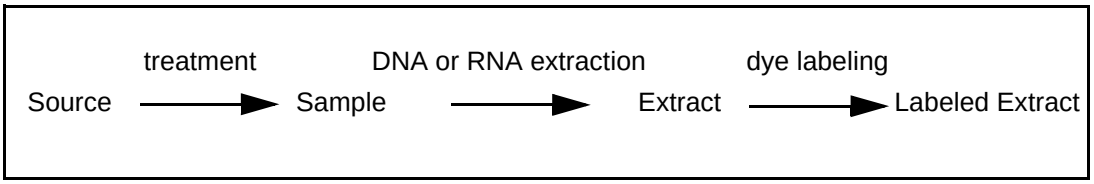


Figure 4.5 Biomaterials components and their relationship in caArray

The nature of each type of biomaterial is defined in varying dialog boxes in caArray.

Biomaterials can be created manually as described in the following sections. Alternatively, they can be generated automatically when data files are imported into caArray. For more information, see *Importing Data* on page 70. When data files are thus associated with biomaterials, the files can be downloaded from each of the corresponding biomaterial tabs. For more information, see the [biomaterials] tab sections and *Downloading Data from caArray* on page 75.

Sources Tab

A source is any biological site from which the tissue for the array is derived before any preparation of the tissue for the array takes place. *Example:* human brain tumor tissue that has been treated with an anti-cancer drug.

The Sources subtab (under the Annotations tab) displays any sources that may have been added previously to the experiment you are creating.

To add a Source, follow these steps:

1. On the Sources tab, click the **Add a new Source** button in the upper right-hand corner of the tab. This opens the Sources page for adding source information *Figure 4.6*).

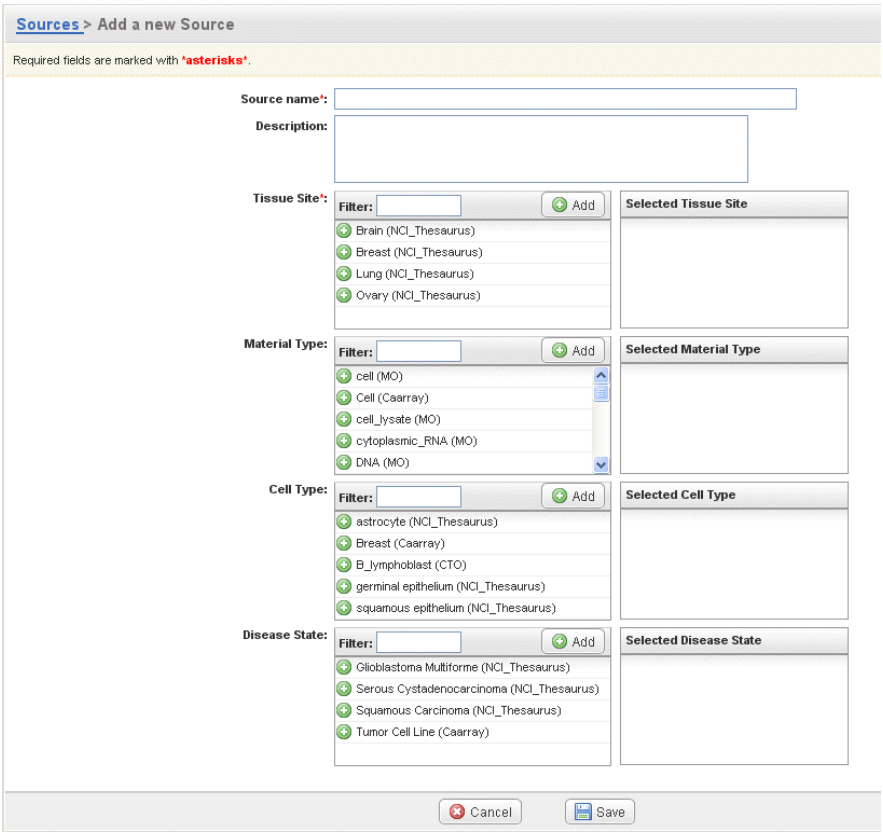


Figure 4.6 Sources subtab

2. In the Sources form, enter the information as described in *Table 4.6*.[5]

Source Fields	***Description***
Source Name*	Name assigned to the source
Description	Description of the source
Tissue Site*	Tissue site is the site from which the source material was obtained. You can choose from available terms or add a new term. For more information about adding a new term to annotate this attribute, see *Adding Vocabulary for Experiments* on page 44.
Material Type	Material type is the descriptor for the type of material gleaned from the tissue site. You can choose from available terms or add one or more new terms. For more information about adding a new term to annotate this attribute, see *Adding Vocabulary for Experiments* on page 44.

Table 4.6 Fields for documenting a source

5. Fields with a red asterisk * are required.

Source Fields	Description
Cell Type	Cell type is the descriptor for the category of cells used for the source material. You can choose from available terms or add one or more new terms. For more information about adding a new term to annotate this attribute, see *Adding Vocabulary for Experiments* on page 44.
Disease State	Disease state is the descriptor used to describe the disease condition when the source was obtained. You can choose from available terms or add one or more new terms. For more information about adding a new term to annotate this attribute, see *Adding Vocabulary for Experiments* on page 44.

Table 4.6 Fields for documenting a source

3. Click **Save** to save the source to the experiment. Click **Cancel** to return to the subtab without adding the source.

   Note: When you create samples for this experiment, you must specify the source(s) for the samples. The Samples column on this Sources tab will then be populated.

4. Repeat steps 1-3 as often as necessary to add all appropriate sources to the experiment.
5. To copy a source, click the corresponding **Copy** icon ( ). caArray copies the source attributes, renames it using the existing source name and adding an incremental number. The copied source now displays under the original.
6. To edit the new source, click the corresponding **Edit** button (Edit) and edit the data. Click **Save** to save the edits.
7. To delete a source, click the **Delete** icon ( ) in the corresponding row.
8. To download data files associated with the source, click the **Download** button ( ).

   Note: Clicking the **Download** link downloads as a .zip file the data files (e.g. .CEL, .CHP, etc.) associated with this source during the import process. For more information, see *Importing Data* on page 70 and *Downloading Data from caArray* on page 75.

9. Proceed to the **Samples** Tab.

Samples Tab

A sample is the original source biomaterial after initial treatment events. *Example:* A tissue or biopsy material treated to create a cell lysate. The Samples subtab (under the Annotations Tab) displays any previous samples that may have been added previously to the experiment you are creating.

The Samples subtab (under the Annotations Tab) displays any samples that may have been added previously to the experiment you are creating.

1. On the Samples tab, click the **Add a New Sample** button to add a new sample. This opens the Samples page when you can add sample information (*Figure 4.7*). The Extracts and Labeled Extracts pages are similar to this page.

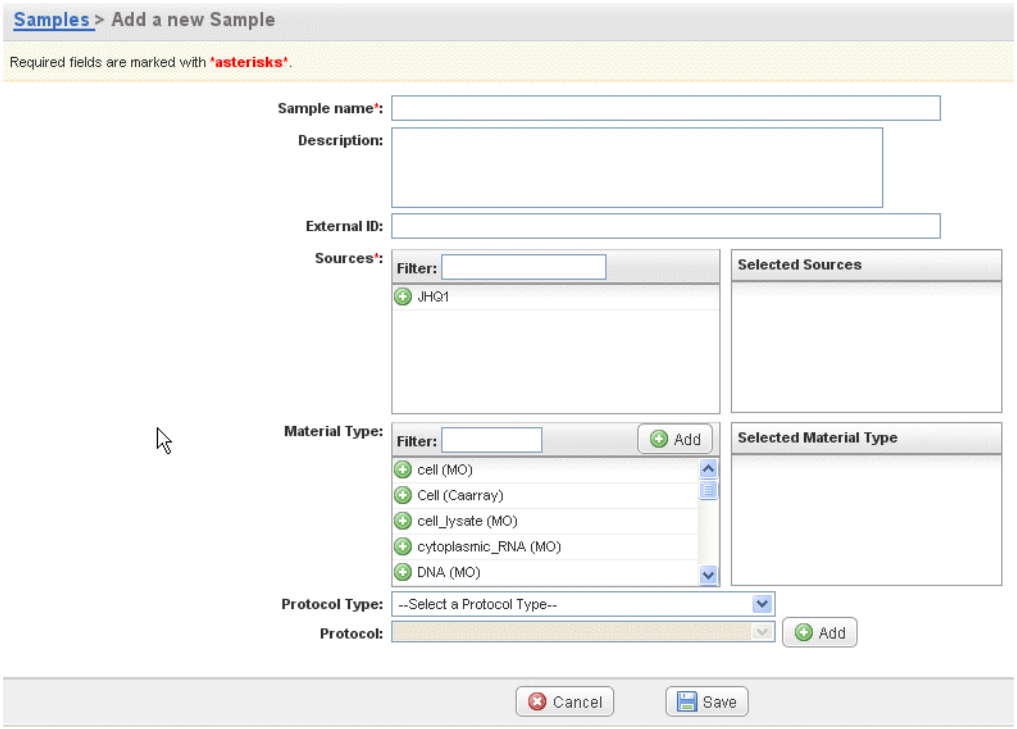


Figure 4.7 Samples page

2. In the Sample form, enter the information described in *Table 4.7*.[6]

Samples Fields	*Description*
Sample Name*	Enter a name for the sample.
Description	Enter a description of the sample.
External ID	Enter an identification value given to the sample outside of caArray, for example, in the lab that did the work
Source(s)*	Sources must already have been saved to caArray. Select one or more sources from which the sample was derived. As you do so, the selected source(s) move into the **Selected Sources** panel. If the appropriate value is not displayed, begin typing a term in the Filter text box. The available terms that display in the panel below are limited according to the text you enter. A message displays if a corresponding term cannot be found. See *Sources Tab* on page 35 for more information.

Table 4.7 Fields for documenting samples

6. Fields with a red asterisk * are required.

Samples Fields	Description
Material Type	Material type is the descriptor for the type of material gleaned from the tissue site. There are three ways you can enter terms for annotating this attribute. See *Adding Vocabulary for Experiments* on page 44 for more information about using this feature.
Protocol Type	Select from the drop-down menu the descriptor for the category of protocol used to prepare the sample.
Protocol	Select from the drop-down list the protocol used to prepare the sample. If the appropriate protocol has not been entered into the system, click **Add** to open the page where you can add a new protocol. For more information, see *Creating a Protocol* on page 59.

Table 4.7 Fields for documenting samples

3. Click **Save**. Click **Cancel** to return to the subtab without adding the sample.

   Note: When you create extracts for this experiment, you must specify the samples for the extracts. The Extracts column on this Samples tab will then be populated.

4. Repeat steps 1 - 3 as often as needed to enter all the samples used in this experiment.

5. To copy a sample, click the corresponding **Copy** icon ( ). caArray copies the sample attributes, renames it using the existing sample name and adding an incremental number. The copied sample now displays under the original.

6. To edit the new sample, click the corresponding **Edit** button (Edit) and edit the data. Click **Save** to save the edits.

7. To delete a sample, click the **Delete** icon ( ) in the corresponding row.

8. To download data files associated with the sample, click the **Download** button ( ).

   Note: Clicking the **Download** link downloads as a .zip file the data files (e.g. .CEL, .CHP, etc.) associated with this sample during the import process. For more information, see *Importing Data* on page 70 and *Downloading Data from caArray* on page 75.

9. Proceed to the **Extracts** tab.

Extracts Tab

An extract is a sample after a treatment event in which DNA or RNA is extracted for the array.

The Extracts subtab (under the Annotations tab) displays any extracts that may have been added previously to the experiment you are creating.

1. On the Extracts subtab, click the **Add a New Extract** button to add a new extract. This opens the Labeled Extracts page that is similar to the Samples page (*Figure 4.7*).

2. In the Extract form, enter the information described in *Table 4.8*.[7]

Extracts Fields	*Description*
Extract Name*	Name assigned to the extract
Description	Description of the extract
Samples*	Samples must already have been saved to caArray. Select one or more samples from which the extract was derived. As you do so, the selected sample(s) move into the Selected Samples panel. If the appropriate value is not displayed, begin typing a term in the Filter text box. The available terms that display in the panel below are limited according to the text you enter. A message displays if a corresponding term cannot be found. See *Samples Tab* on page 37 for more information.
Material Type	Material type is the descriptor for the type of material gleaned from the tissue site. There are three ways you can enter terms for annotating this attribute. See *Adding Vocabulary for Experiments* on page 44 for more information about using this feature.
Protocol Type	Select from the drop-down menu the descriptor for the category of protocol used to prepare the extract..
Protocol	Select from the drop-down list the protocol used to prepare the extract. **Note:** The available selections are limited based on the protocol type selected above. If the appropriate protocol has not been entered into the system, click **Add** to open the page where you can add a new protocol. For more information, see *Creating a Protocol* on page 59.

Table 4.8 Fields for documenting an extract

3. Click **Save**. Click **Cancel** to return to the subtab without adding the Extract.

   Note: When you create labeled extracts for this experiment, you must specify the extracts for the labeled extracts. The labeled extracts column on this Extracts tab will then be populated.

4. Repeat steps 1 - 3 as often as needed to enter all the extracts used in this experiment.

5. To copy an extract, click the corresponding **Copy** icon ( ). caArray copies the extract attributes, renames it using the existing extract name and adding an incremental number. The copied extract now displays under the original.

6. To edit the new extract, click the corresponding **Edit** button (Edit) and edit the data. Click **Save** to save the edits.

7. To delete an extract, click the **Delete** icon ( ) in the corresponding row.

8. To download data files associated with the extract, click the **Download** button ( ).

7. Fields with a red asterisk * are required.

Note: Clicking the **Download** link downloads as a .zip file the data files (e.g. .CEL, .CHP, etc.) associated with this extract during the import process. For more information, see *Importing Data* on page 70 and *Downloading Data from caArray* on page 75.

9. Proceed to the Labeled Extracts tab.

Labeled Extracts Tab

A labeled extract is an extract that has been labeled for detection of the nucleic acids on the array. *Example*: The extract RNA is labeled with a fluourescent dye, yielding the labeled extract RNA.

The Labeled Extracts subtab (under the Annotations tab) displays any labeled extracts that may have been added previously to the experiment you are creating.

1. Click the **Add a New Labeled Extract** button to add a new labeled extract. This opens the Labeled Extracts page that is similar to the Samples page (*Figure 4.7*).
2. In the Labeled Extract form that opens, enter the information described in *Table 4.9*[8].

Labeled Extracts Fields	***Description***
Labeled Extract Name*	Name assigned to the extract
Description	Description of the extract
Extracts*	Extract(s) from which the labeled extract was derived. Extracts must already have been saved to caArray. Select one or more extracts from which the labeled extract was derived. As you do so, the selected extract(s) move into the Selected Extracts panel. If the appropriate value is not displayed, begin typing a term in the Filter text box. The available terms that display in the panel below are limited according to the text you enter. A message displays if a corresponding term cannot be found. See *Extracts Tab* on page 39 for more information.
Material Type	Material type is the descriptor for the type of material gleaned from the tissue site. There are three ways you can enter terms for annotating this attribute. See *Adding Vocabulary for Experiments* on page 44 for more information about using this feature.
Protocol Type	Select from the drop-down menu the descriptor for the category of protocol used to prepare the labeled extract..

Table 4.9 Fields for documenting a labeled extract

8. Fields with a red asterisk * are required.

Labeled Extracts Fields	*Description*
Protocol	Select from the drop-down list the protocol used to prepare the labeled extract. **Note:** The available selections are limited based on the protocol type selected above. If the appropriate protocol has not been entered into the system, click **Add** to open the page where you can add a new protocol. For more information, see *Creating a Protocol* on page 59.

Table 4.9 Fields for documenting a labeled extract

3. Click **Save**. Click **Cancel** to return to the subtab without adding the labeled extract.

   Note: When you create hybridizations for this experiment, you must specify the labeled extracts for the Hybridizations. The Hybridizations column on this Labeled Extracts tab will then be populated.

4. Repeat steps 1 - 3 as often as needed to enter all the labeled extracts used in this experiment.

5. To copy the labeled extract, click the **Copy** icon ( ) in the **Copy** column. caArray copies the labeled extract attributes, renames it using the existing labeled extract name and adding an incremental number. The copied labeled extract now displays under the original.

6. To edit the new labeled extract, click the corresponding **Edit** button (Edit) and edit the data. Click **Save** to save the edits.

7. To delete a labeled extract, click the **Delete** icon ( ) in the corresponding row.

8. To download data files associated with the labeled extract, click the **Download** button ( ).

   Note: Clicking the **Download** link downloads as a .zip file the data files (e.g. .CEL, .CHP, etc.) associated with this labeled extract during the import process. For more information, see *Importing Data* on page 70 and *Downloading Data from caArray* on page 75.

9. Proceed to the *Hybridizations Tab*.

Hybridizations Tab

A hybridization is the process of incubating one or more labeled extracts with an array.

The Hybridizations subtab (under the Annotations tab) displays any hybridization information that may have been added previously to the experiment you are creating.

1. Click the **Add a New Hybridization** button to add a new hybridization. This opens the Hybridizations page where you can add hybridization information (*Figure 4.8*).

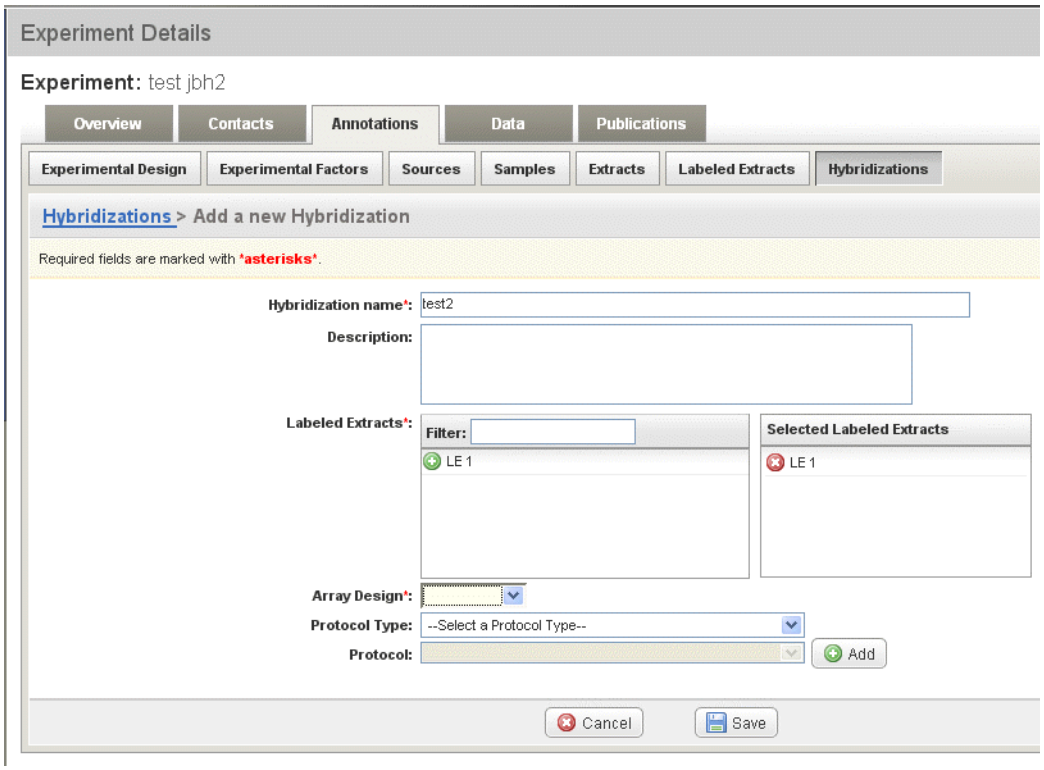


Figure 4.8 Hybridizations page

2. In the Hybridizations form, enter the information described in *Table 4.10*[9].

Hybridizations Fields	*Description*
Hybridization Name*	Name assigned to the extract
Description	Description of the hybridization
Labeled Extracts*	Labeled extract(s) used in the hybridization protocol. Select one or more listed labeled extracts. As you do so, the selected labeled extract(s) move into the Selected Labeled Extracts panel. If the appropriate value is not displayed, begin typing a term in the Filter text box. The available terms that display in the panel below are limited according to the text you enter. A message displays if a corresponding term cannot be found. See *Labeled Extracts Tab* on page 41 for more information.
Array Designs	This field displays only if you associated more than one array design on the *Overview Tab*, described on page 29. Select the array design appropriate for this hybridization.

Table 4.10 Fields for documenting a hybridization

9. Fields with a red asterisk * are required.

Hybridizations Fields	*Description*
Protocol Type	Select from the drop-down menu the descriptor for the category of protocol used to prepare the hybridization.
Protocol	Select from the drop-down list the protocol used to perform the hybridization. If the appropriate protocol has not been entered into the system, click **Add** to open the page where you can add a new protocol. For more information, see *Creating a Protocol* on page 59.

Table 4.10 Fields for documenting a hybridization

3. Click **Save**. The new hybridization object displays on the Hybridization tab. Click **Cancel** to return to the subtab without adding the hybridization.

   Note: On the saved draft, the uncompressed size of the hybridization file is defined by caArray when it becomes available.

4. Repeat steps 1 - 3 as often as needed to enter all the hybridization data used in this Experiment.
5. To edit the new hybridization entity, click the corresponding **Edit** button () and edit the data. Click **Save** to save the edits.
6. To delete a Labeled Extract, click the **Delete** icon () in the corresponding row.
7. To download data files associated with the hybridization, click the **Download** button ().

   Note: Clicking the **Download** link downloads as a .zip file the data files (e.g. .CEL, .CHP, etc.) associated with this hybridization during the import process. For more information, see *Importing Data* on page 70 and *Downloading Data from caArray* on page 75.

8. Click **Save**. The name of the hybridization now displays on the Hybridizations subtab.
9. Proceed to the **Data** tab.

Adding Vocabulary for Experiments

For an experiment you are creating, a number of the annotations subtabs (Source, Samples, etc.) include panels where you can select or add new attribute voculabulary

terms to annotate experiment components. *Figure 4.9* The following figure displays an example page for performing this task.

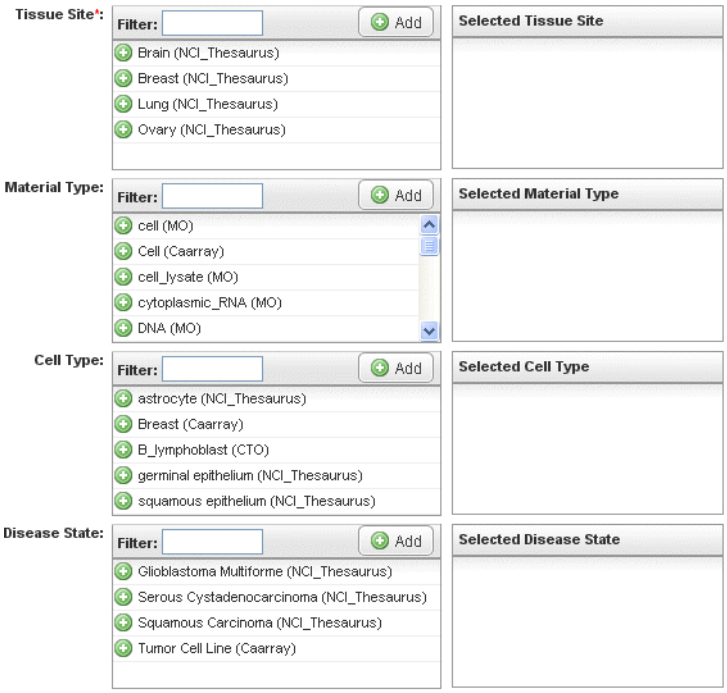


Figure 4.9 Annotation panels for selecting or adding new vocabulary terms to experiments

You can enter terms for any of these attributes in three ways:

1. If the value for the attribute or condition displays in the site list, click the adjoining icon (⊕) to move it into the **Selected {attribute}** panel.
2. If the appropriate value is not displayed, to find a term of interest that might already be in the caArray dictionary, begin typing a term in the **Filter** text box. The available terms that display in the panel below are limited according to the text you enter. A message displays if a corresponding term cannot be found.
3. To add a new term, click the **Add** button (⊕ Add). This takes you to the Manage {Attribute or Condition} page where you can add a new term. *Table 4.11* describes fields for defining the vocabulary term.

Vocabulary Term Category	***Description of Fields***
Term	
Value*	Enter the new term. *Example:* DNA
Description	Enter the description of the term, as appropriate. *Example:* deoxyribonucleic acid
Source	

Table 4.11 Fields for entering a new vocabulary term

Vocabulary Term Category	*Description of Fields*
Create a New Source [for the Term you are adding]	Select **Yes** or **No** • If **No**, select from the drop-down list in the next field, the source for the term. In many cases, the source will be an existing controlled vocabulary such as the NCI Thesarus, or the MGED Ontology (MO). • If **Yes**, the dialog box expands with new fields where you can add the name, URL and version for the new source.
Source*	Select from the drop-down menu the source for the new term you are adding. This field disappears if you select **Yes** in the previous field.
Accession	
Accession URL	Enter the exact URL for accessing the new term. *Example:* http://mged.sourceforge.net/ontologies/MGEDontology.php#DNA
Accession Value	Enter the value given the term in the source vocabulary. *Example:* MO_945

Table 4.11 Fields for entering a new vocabulary term

4. Once you have entered the appropriate information, click **Save**.

   This returns you to the original tab, where you can continue defining the experiment attribute.

You can also work with vocabulary terms in caArray using the curation tool, Manage Vocabularies. For more information see *Managing [Controlled] Vocabulary [Terms]* on page 61.

Data Tab

The Data tab is the vehicle for uploading, validating, importing and downloading data relating to caArray Experiments. When you click on the **Data** tab, four subtabs where you initiate data-related tasks display (*Figure 4.10*). They are described in *Table 4.12*.

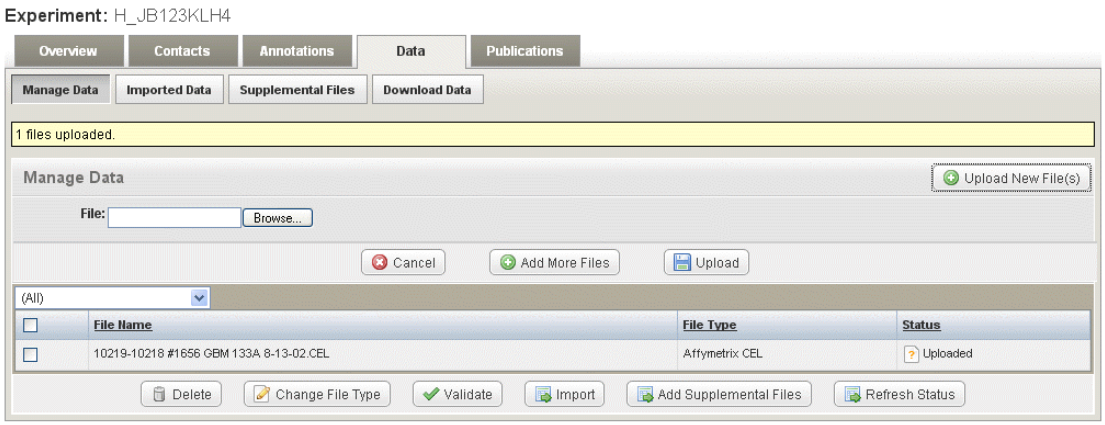


Figure 4.10 caArray experiment data tab, extended to display the browse/upload feature. One file has already been uploaded.

Data Tabs	*Description*
Manage Data	From this tab, you can perform data-related tasks such as uploading, validating and importing data into caArray. Additional tasks such as changing data file types and designating supplemental files also takes place here.
Imported Data	This subtab list all files that have been imported into caArray.
Supplemental Files	This tab lists files and documents that have been uploaded to caArrayand have been designated supplemental on the Manage Data subtab.
Download Data	From this tab, you can download data that has been imported into caArray.

Table 4.12 Tabs for performing data-related tasks

Note: To import data, you must have Write access to the experiment.

All of these data-related tasks are described in detail in *Chapter 6 Submitting Data to an Experiment* and *Chapter 7 Extracting Data from caArray*.

Publications Tab

The Publications tab allows you to create, edit or delete associations of publications with the experiment you are creating or editing. This feature allows you to add journal articles, reviews, and books.

To add publication associations, follow these steps:

1. Locate and open the experiment for which you want to associate publications.
2. Click the **Publications** tab.
3. Click the **Add a New Publication** button.
4. On the form that opens, enter the appropriate information for the article or publication in the fields provided (and described in *Table 4.13*).[10].

Publications Fields	*Description*
Title*	Title of the publication
Authors	Author(s) of the publication
URL	URL for locating the publication
Editor	Editor of the publication, where appropriate
Pages	The page number(s) of the article your are referencing, where appropriate,
Publisher	Publisher of the publication

Table 4.13 Fields for documenting Publications

10. Fields with a red asterisk * are required.

Publications Fields	*Description*
PubMedID	ID for locating the publication in PubMed
Volume	Volume where article is found
Year	Year of publication
Publication	Name of the publication where article is found
Type	Select in the drop-down menu the publication type.
Status	Select in the drop-down menu the publish status: **Published**, **In Preparation**, **Submitted**

Table 4.13 Fields for documenting Publications

5. After entering the appropriate information, click **Save**. A message displays verifying that the publication was successfully saved with the experiment.

The System saves the experiment with the associated publications and returns you to your experiment workspace, which is now updated with the state of the project.

Experiment Status Settings

As you enter details of the experiment on the Overview tab, you must click the **Save** button to open more tabs for recording experiment information. As you step through the successive tabs, you should continue save the experiment information by clicking the **Save** button. This saves the experiment in draft status, which means that the experiment is visible to other users, but they cannot view the details.

Summary of experiment status and buttons to achieve that status:

***Draft*—**Click **Save** on experiment tabs. Details of the experiment are available only to experiment owner (creator); existence of the experiment is visible to all.

***in Progress*—**-Click **Submit Experiment Proposal** of an experiment in "draft" status. All details of the experiment or only selected segments of an experiment can be made available to all users or only to collaboration group(s) as "read-only".

***Public*—**Click the **Make Experiment Public** of an experiment in "In Progress" status. All details of an experiment are then availabe to all users,including anonymous users.

***Retract Public Experiment*—**Click this button to cancel the public status for an experiment. At that point, data associated with the experiment will no longer be viewable or downloadable by users other than the Experiment owner and designated collaborators.

For information about setting the visibility of an experiment, see *Managing Experiment Visiblity* on page 49.

Updating An Experiment Proposal

At any point, after you have saved an experiment draft, you, as its creator, can open the draft and edit it, add or delete data.

Note: After data has been imported into an experiment, the array design associated with that data cannot be removed from the experiment's list of array designs.

Editing an Experiment

At any point after an experiment has been saved, you can edit it.

Note: A good rule of thumb is if you can see an **Edit** button or **Delete** option, then you can modify or delete an item. If you cannot see those options, then you cannot edit or delete it.

To edit an experiment, follow these steps:

1. In the My Experiment Workspace, locate the experiment you want to edit.
2. On the row corresponding to the experiment you want to edit, click the **Edit** button () and edit the data. All information is editable except the automatically generated experiment ID and the status.
3. Click **Save** to save the edits to the draft.

Note: To import data, the Experiment must be in Edit mode. See *Managing Data* on page 65.

Editing Experiment Annotations

Once an experiment is in draft or in progress status (having been submitted), as the experiment creator or having Write access as a collaborator, you can add data such as annotations, files and publications to the experiment. To do so, follow these steps:

1. In the My Experiment Workspace, locate the experiment of interest on the Work Queue tab.
2. Click the **Edit** button () that corresponds to the selected experiment.
3. Navigate to the tab for the attribute or item you want to add. The tabs where you can add data have an **Add {attribute}** button in the upper right corner. To mention a few, you could add samples, labeled extracts, upload data or associate publications with your experiment.
4. Click the **Add {attribute}** button in the upper right hand corner of the tab.
5. Enter the appropriate information.you wish to add. For more specific information, see *Creating an Experiment* on page 28.
6. Click **Save** to complete adding the data.

For details about each tab where you can edit annotations, see topics under *Annotations Tab* on page 32.

Managing Experiment Visiblity

Once you create an experiment draft, it is listed on the **Work Queue** tab of the My Experiment Workspace. At that point, you can review and/or modify the access of other users to the project, using the permissions feature of caArray. Public access to the experiment can be configured, and well as collaboration group access.

Setting Public Visibility

To assign or modify experiment visibility, follow these steps:

1. Go to My Experiment Workspace and locate the experiment of interest on the Work Queue tab.
2. Click the **Permissions** icon ( ) in the row listing the experiment.
3. The Experiment Permissions page that opens displays a panel that describes the visibility policies of caArray and allows you to set visibilty for your experiment (*Figure 4.11*).

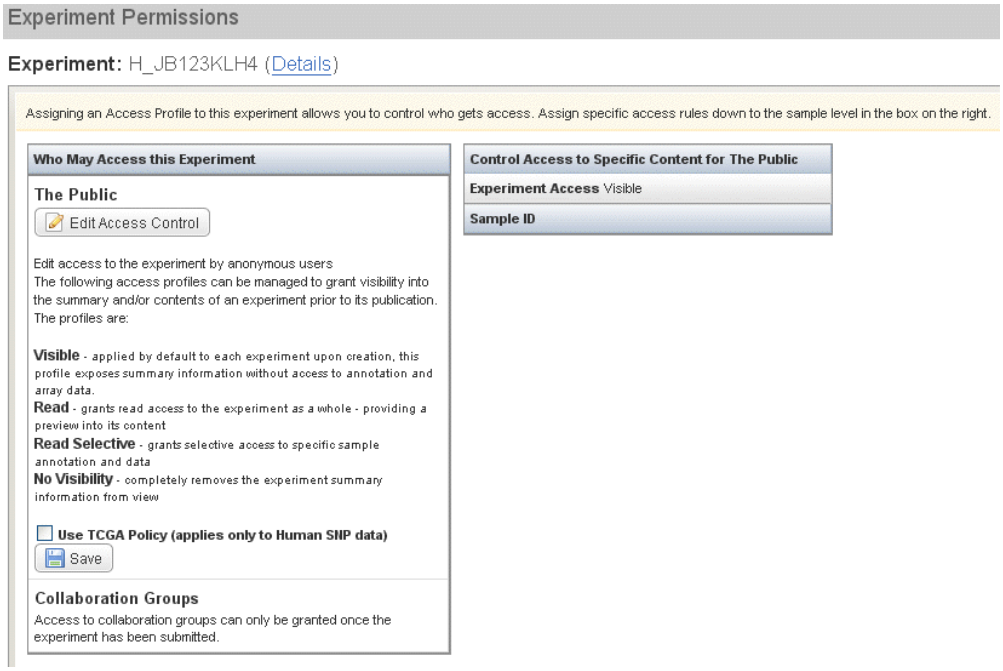


Figure 4.11 Experiment Permissions page

The visibility options are:

- **Visible**--applied by default to each experiment upon creation, this profile exposes summary information without access to annotation and array data.
- **Read**--grants read access to the experiment as a whole, providing a preview into its content
- **Read Selective**--grants selective access to specific sample annotation and data
- **No Visibility**--completely removes the experiment summary information from view

Note: Your ability to set visibility options varies according to the draft/ submission status of the experiment. See below.

***For an Experiment in Draft status*:**

Note: Only **Visible** and **No Visibility** options are available for an experiment in draft status.

a. To control public access, under **The Public**, click the **Edit Access Control** button.

b. In the Control Access to Specific Content panel that displays on the right, select the visibility option from the drop-down list. The only available options for and experiment draft) are: **Visible** and **No Visiblity**.

c. The option at the bottom of the left hand panel relates to The Cancer Genome Atlas (TCGA) policy (for human SNP data), which is that only four fields of annotation can be provided to the open public:

— Clinical Diagnosis

— Histologic Diagnosis

— Tissue Anatomic Site

— Pathologic Status

Click this option to assign TGCA policy limits to your experiment.

Note: The Collaboration Groups visibilty option is not available to an experiment in Draft status.

d. Click **Save** under the TCGA option to execute the Public access choices.

For a "submitted" Experiment:

Note: Four visibility options plus collaboration group visibility are available for experiments that have been submitted to the caArray repository.

a. To control public access, under **The Public**, click the **Edit Access Control** button.

b. In the Control Access to Specific Content panel that displays on the right, select the visibility option from the drop-down list. The four available options are those described above: **Visible**, **Read**, **Read Selective** and **No Visibility**.

c. The option at the bottom of the left hand panel relates to The Cancer Genome Atlas (TCGA) policy (for human SNP data), which is that only four fields of annotation can be provided to the open public:

— Clinical Diagnosis

— Histologic Diagnosis

— Tissue Anatomic Site

— Pathologic Status

Click this option to assign TGCA policy limits to your experiment.

d. Click **Save** under the TCGA option to execute the Public access choices.

Setting Collaboration Group Visibility

A section at the bottom of the left panel of the Experiment Permissions page (*Figure 4.11*) allows you to set experiment visibility for collaboration groups. You can use one of the existing groups, or you can create a group from this page. For information about how to create collaboration groups through the user management features, see *Managing Collaboration Groups* on page 81.

To configure experiment visibility for a collaboration group, follow these steps:

1. Go to My Experiment Workspace and locate the experiment of interest on the Work Queue tab.
2. Click the **Permissions** icon ( ) in the row listing the experiment for which you want to alter permissions.

   The Experiment Permissions page that opens displays a panel the describes the visibility policies of caArray and allows you to set visibilty for your experiment. For information about setting public visibility, see *Managing Experiment Visiblity* on page 49.

If the Collaboration Group already exists:

1. At the bottom of the left panel, under **Collaboration Groups**, select in the drop-down list the group to which you want to assign visibility.
2. Click the **Edit Access Control** button.
3. In the Control Access to Specific Content to [Group] panel that displays on the right, select the visibility option for the group from the Experiment Access drop-down list. The four available options are: **None**, **Read**, **Read/Write** and **Read/Write/Selective**.
4. Click **Save** in that panel to execute the collaboration group visibility choices.

If the Collaboration Group must be created:

1. If the group to which you want to assign experiment visibility does not yet exist, click the **Add New Group** button.
2. In the Add a New Collaboration Group form, enter the name of the group and click **Save**.
3. The Manage Collaboration Groups page that opens display all groups in the system, including the one you just created. Click the **Edit** icon ( ) to specify members for the group. For more information about creating and working with a collaboration group, see *Managing Collaboration Groups* on page 81.
4. Return to the Experiment Permissions page to continue assigning visibility for the experiment. You may need to return to the My Experiment Workspace and re-click the **Permissions** icon ( ) in the row listing the experiment to get back to that page.
5. Continue from step 1 in this section of this topic: *If the Collaboration Group already exists*.

See also *Setting Public Visibility* on page 49.

CHAPTER 5
CURATION TOOLS

This chapter describes the processes for completing curation tasks in caArray.

Topics in this chapter include the following:

Curation Tasks

Curation tasks in caArray 2.0 are available to all logged in users. These tasks govern global, and reusable data elements, namely array designs, protocols and vocabulary terms. Curation tasks are available under the Curation options on the left sidebar (*Figure 5.1*):

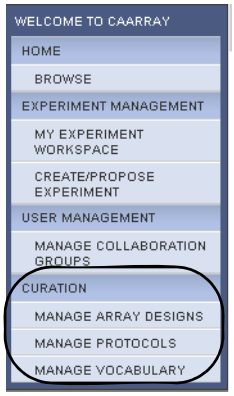


Figure 5.1 Curation options display in the left sidebar

Managing Array Designs

Any caArray user can upload, validate and import array designs for supported providers (Affymetrix, Illumina, and Genepix), and anyone can view, edit or replace the files. An array design only needs to be loaded once and is available to all users.

Note: After data has been imported into an experiment, the array design associated with that data cannot be removed from the experiment's list of array designs.

You can also view array designs, or edit those for which you have permissions.

Note: Manage Array Designs does not allow for the upload of array designs from providers for which caArray does not have parsers. If you attempt to upload such an array design, an intercept message will inform you of that fact.

Viewing Array Designs

To view array designs in the system, follow these steps:

1. After logging into caArray, on the left sidebar under **Curation**, click **Manage Array Designs**.

   The Manage Array Designs page that opens displays all array designs that have been imported into caArray (*Figure 5.2*). See *Importing Data* on page 70. Properties corresponding to those array designs are described in Table *5.1* .

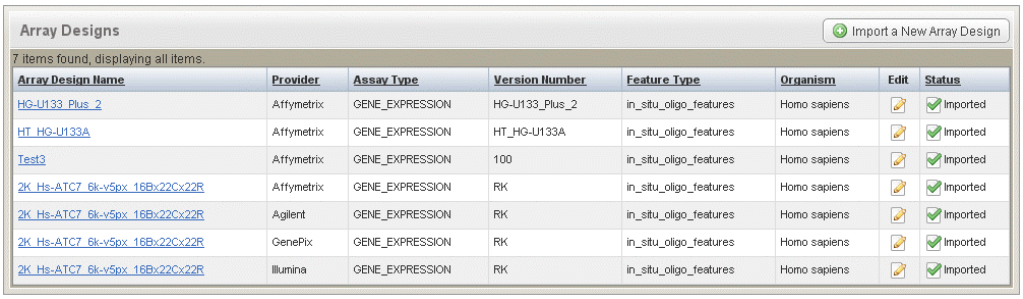


Figure 5.2 Array Designs imported into caArray

Note: Columns with underlined headings are sortable by clicking on the heading.

Array Designs Properties	***Description***
Array Design Name	Name assigned to the array design

Table 5.1 Array Designs properties

Array Designs Properties	*Description*
Assay Type	The assay type used for the Array Design. • **Gene Expression**--experiment using microarrays intended to measure levels of transcribed genes • **SNP**--experiment using microarrays intended to detect nucleotide changes in chromosomal DNA • **aCGH**--array Comparative Genomic Hybridization; a method for the analysis of chromosome copy number changes (gains/losses). • **Exon**--Exon arrays are designed to study which exons are present in an expressed gene. • **microRNA**--Experiment that measures activity among the 217 genes encoding miRNA. Patterns of gene activity that can distinguish types of cancers can be discerned. • **Methylation**--experiment that attempts to establish patterns of methylation genome-wide or within targeted promoters or CpG islands
Provider	Select from the drop-down menu the provider of the array. **Note:** Only Affymetrix, Illumina and GenePix formats are fully supported with validation and parsers in caArray 2.0. For more information, see the Note about File Types in *Managing Data* on page 65.
Version Number	The version number of the array design
Feature Type	The technology type or platform of the reporters on the array. Note that these terms are from the MGED Ontology.
Organism	The organism the array was designed to assay.
Edit	If you do not have permissions to edit this Array Design, this icon is not visible. If it is, click the **Edit** icon ( ) to open the Array Designs details page where you can edit the data. For more information, see the following section.
Status	**Imported**

Table 5.1 Array Designs properties

If you click the **Array Design Name** in the Import Array Designs page, the details page that opens displays the name including file type extension of the uploaded/imported array design file.

Adding an Array Design

Note: Array design files must be in one of the following formats to be successfully imported into caArray. Unsupported array designs can be imported, but will end up in an "Imported, not parsed" state.

--Affymetrix .cdf
--Illumina Design .csv

--Genepix .gal
--Agilent .csv or .xml
--UCSF Spot .spt
--ImaGene .tpl
--Nimblegen .ndf

To add an array design to caArray, follow these steps:

1. After logging into caArray, on the left sidebar under **Curation**, click **Manage Array Designs**.

   The Import Array Designs page that opens displays all array designs that have been imported into caArray (see previous topic *Importing Data*).

2. To import a new array design, click **Import a New Array Design** in the upper right corner.

3. On the form that opens, enter the appropriate information in the Array Design Details fields provided (and described in Table *5.2*).[11]

Array Designs Details Properties	***Description***
[Array Design Name]	**Note:** This field does not appear in the array design form; the name is automatically generated by caArray based on the file you import.
Description	Enter an appropriate description for the array design you are adding.
Assay Type	The assay type used for the array design. • **Gene Expression**--experiment using microarrays intended to measure levels of transcribed genes • **SNP**--experiment using microarrays intended to detect nucleotide changes in chromosomal DNA • **aCGH**--array Comparative Genomic Hybridization; a method for the analysis of chromosome copy number changes (gains/losses). • **Exon**--Exon arrays are designed to study which exons are present in an expressed gene. • **microRNA**--Experiment that measures activity among the 217 genes encoding miRNA. Patterns of gene activity that can distinguish types of cancers can be discerned. • **Methylation**--experiment that attempts to establish patterns of methylation genome-wide or within targeted promoters or CpG islands
Provider	The provider of the array design. This is generally the company or group that manufactured the array design.
Version Number	The version number of the array design

Table 5.2 Array Designs properties

11. Fields with a red asterisk * are required.

Array Designs Details Properties	*Description*
Feature Type	The technology type or platform of the reporters on the array.
Organism	The organism used for the Array described by the array design.

Table 5.2 Array Designs properties

4. In the **Upload Array Design File** section, click the **Browse** button to navigate to the file.
5. Select the **File Format** in the drop-down list. Compatible file types are listed at the beginning of this topic.

   Note: If you select **Automatic,** caArray tries automatically to infer the array design type.
6. Click **Save** to launch the array design import process.

   The process includes uploading the file, validating it and importing it into the system. You should not leave the user interface once this process is underway, or you may have to start all over.

Note: After data has been imported into an experiment, the array design associated with that data cannot be removed from the experiment's list of array designs.

Editing an Array Design

In caArray 2.0, any logged in user can edit an array design. An array design already in the system can be replaced by a new array design by following the edit steps.

Note: After data has been imported into an experiment, the array design associated with that data cannot be removed from the experiment's list of array designs.

To edit an array design, follow these steps:

1. On the row corresponding to the array design, click the **Edit** icon ( ).

OR

1. Open the array design by clicking on its name, and click the **Edit** button (Edit) at the bottom of the details page.
2. All required fields become editable; enter any edits.
3. From this page, you can initiate uploading of a new array design file to replace the existing file.
4. Save any edit by clicking the **Save** button.

Managing Protocols

In caArray, you can create and manage protocol(s) for referencing in an experiment. A protocol provides detailed documentation about the precise actions taken in any procedure that might be part of an experiment. For example, a protocol could describe

the steps a laboratory used for any kind of process used in an experiment, such as the way a source material or sample is derived, the method used for labeling an extract or the methods used for running a hyrbridization or creating an image file of array results.

A protocol can be created independently of a specific experiment, or added during the course of creating biomaterials or a hybridization for an experiment. See the topics under *Annotations Tab* on page 32 for more information. A protocol can be used by any caArray user but it can only modified by the owner of the protocol or another user with assigned permissions.

Viewing Protocols

To view existing protocols in caArray, follow these steps:

1. After logging into caArray, on the left sidebar under **Curation**, click **Manage Protocols**. The **Protocols** tab displays by default (*Figure 5.3*).

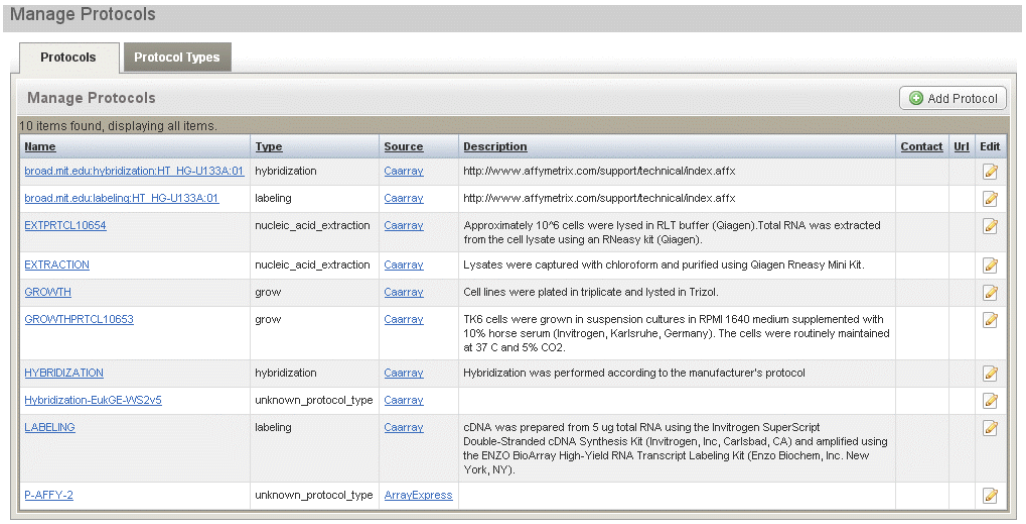


Figure 5.3 Protocols page

All protocols that have been created in caArray display on this tab. Properties corresponding to those protocols are described in Table *5.3*.

Protocol Properties	***Description***
Name	Name assigned the protocol
Type	Descriptor of the protocol type, such as labeling or hybridization.
Description	Description of the protocol procedure. Include any and all appropriate details, such as the detailed steps taken in a laboratory procedure. *Example*: Enter a description of a procedure for labeling RNA with fluorescent tags to be used in a hybridization procedure.
Contact	The name of the person to contact for information about the protocol.

Table 5.3 Protocol properties

Protocol Properties	Description
URL	Link to a source of external documentation related to the protocol
Edit	Click the **Edit** icon ( ) to open the protocol details page where you can edit the data. For more information, see *Editing a Protocol* on page 60.

Table 5.3 Protocol properties

Note: Columns with underlined headings are sortable by clicking on the heading.

2. To view details of a protocol, click its name.

   Note: Any protocols you did not create are in read-only mode.

Viewing Protocol Types

To view existing protocol types in caArray, follow these steps:

1. After logging into caArray, on the left sidebar under **Curation**, click **Manage Protocols**.
2. By default, the Manage Protocols page displays. To view protocol types, click the **Protocol Types** tab.

   The **Protocol Types** tab displays all protocol types that have been created in caArray. Properties corresponding to those protocol types are described in Table *5.3*.

Protocol Type Properties	Description
Value	The descriptor of the protocol type, such as labeling or hybridization.
Description	The description of the protocol type.
Source	The controlled vocabulary that is the source for the descriptor term value for the protocol type. The source name is a hypertext link that takes you to the website for the source.
Edit	Click the **Edit** icon ( ) to open the protocol type details page where you can edit the data. For more information, see *Editing a Protocol Type* on page 61.

Table 5.4 Protocol properties

3. To view details of a protocol type, click its **Value**.

Creating a Protocol

To create a protocol, follow these steps:

1. After logging into caArray, on the left sidebar under **Curation**, click **Manage Protocols**.

The Manage Protocols page that opens displays all protocols that have been created in caArray (*Figure 5.3*).

2. To create a new protocol, click the **Add Protocols** button in the upper right hand corner of the page.
3. In the Manage Protocols form that opens, enter the appropriate information for the new protocol. Fields are described in Table *5.5*.[12]

Protocol Properties	***Description***
Name*	Name assigned the array design
Description	Description of the protocol procedure. Include any and all appropriate details, such as the detailed steps taken in a laboratory procedure. *Example*: Description of a procedure for labeling RNA with fluorescent tabs to be used in a hybridization procedure.
Type*	Descriptor of the protocol type such as "labeling" or "hybridization" from a controlled vocabulary, fpr example MGED. Select a listed type or if the appropriate value is not displayed, begin typing a term in the **Filter** text box. The available terms that display in the panel below are limited according to the text you enter. A message displays if a corresponding protocol type cannot be found.
Contact	The name of the person to contact for information about the protocol.
Software	Name of software used in the protocol. *Example*: GenePix Pro 3.0.1.22
Hardware	Name of hardware used in the protocol. *Example*: GeneChip(R) Fluidics Station 450®
URL	Link to a source of external documentation related to the protocol

Table 5.5 Protocol fields

4. Click **Save** to save the protocol. Click **Cancel** to halt the action. In both cases, you are returned to the Manage Protocols page. The protocol you just added is listed first in the list of protocols.

Editing a Protocol

Note: A protocol can be edited by anyone, not just the owner of the protocol. All experiments that reference the protocol will be updated to reflect the changes made.

To edit a protocol, follow these steps:

1. After logging into caArray, on the left sidebar under **Curation**, click **Manage Protocols.**
2. On the row corresponding to the protocol, click the **Edit** icon ( ).

12. Items with an asterisk are required.

OR

1. Open the protocol by clicking on its name, and click the **Edit** button (Edit) at the bottom of the details page that opens.
2. All information for a protocol is editable. Make the appropriate edits on the form that opens.
3. Save any edits by clicking the **Save** button. To abort the edit, click the **Cancel** button. This returns you to the Manage Protocols page.

Editing a Protocol Type

1. After logging into caArray, on the left sidebar under **Curation**, click **Manage Protocols.**
2. On the Manage Protocols page, click the **Protocol Types** tab.
3. On the row corresponding to the listed protocol type, click the **Edit** icon ( ).

OR

1. Open the protocol type by clicking on its name, and click the **Edit** button (Edit) at the bottom of the details page.
2. All information for a protocol type is editable. Make the appropriate edits on the form that opens.
3. Save any edits by clicking the **Save** button. To abort the edit, click the **Cancel** button. This returns you to the Manage Protocols page.

Managing [Controlled] Vocabulary [Terms]

In caArray, when you are creating or editing experiments, many experiment attributes are available for entering descriptive terms or annotations. These attributes are:

- **Tissue Site**
- **Cell Type**
- **Disease State**
- **Material Type**

Viewing Vocabulary Terms

To view existing vocabulary terms in caArray, follow these steps:

1. After logging into caArray, on the left sidebar under **Curation**, click **Manage Vocabulary**. The **Manage Vocabularies** page opens, displaying four tabs that correspond to the attribute vocabulary categories in caArray (*Figure 5.4*).

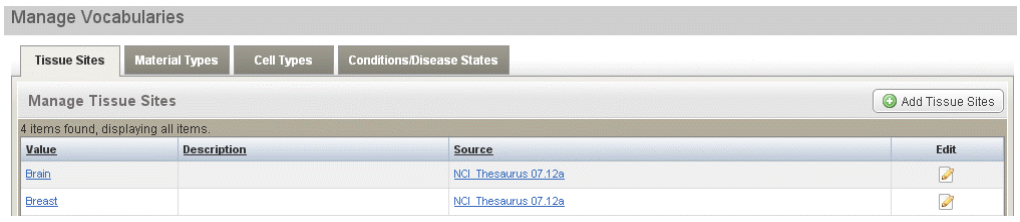


Figure 5.4 Manage Vocabularies page

All vocabulary terms that have been imported into caArray display on the tabs. Properties corresponding to those protocols are described in Table *5.3*.

Protocol Properties	***Description***
Value	Name assigned the term
Description	Description of the vocabulary term
Source	The external source or standardized vocabulary that is the source of the term.
Edit	Click the **Edit** icon ( ) to open the vocabulary details page where you can edit the information for the term. For more information, see *Editing a Vocabulary Term* on page 63.

Table 5.6 Protocol properties

Note: Columns with underlined headings are sortable by clicking on the heading.

2. To view details of a vocabulary term, click its value.

Adding Vocabulary Terms

In caArray, you can enter a new vocabulary term while you are adding annotations to an experiment (see *Adding Vocabulary for Experiments* on page 44) or you can work with vocabulary terms using on of the curation tools of the application, described in this section.

To enter a vocabulary term, follow these steps:

1. After logging into caArray, on the left sidebar under **Curation**, click **Manage Vocabulary.**
2. Select the attribute category that corresponds to the the term you want to add.
3. Click the Add {attribute} button on the upper right of the page.

   This takes you to the Manage {Attribute or Condition} page where you can add a new term. Table *5.7* describes fields for defining the vocabulary term.

Vocabulary Term Category	***Description of Fields***
Term	
Value*	Enter the new term. *Example:* DNA
Description	Enter the description of the term, as appropriate. *Example:* deoxyribonucleic acid
Source	

Table 5.7 Fields for entering a new vocabulary term

Vocabulary Term Category	*Description of Fields*
Create a New Source [for the Term you are adding]	Select **Yes** or **No** • If **No**, select from the drop-down list in the next field, the source for the term. In many cases, the source will be an existing controlled vocabulary such as the NCI Thesarus, or the MGED Ontology (MO). • If **Yes**, the dialog box expands with new fields where you can add the name, URL and version for the new source.
Source*	Select from the drop-down menu the source for the new term you are adding. This field disappears if you select **Yes** in the previous field.
Accession	
Accession URL	Enter the exact URL for accessing the new term. *Example:* http://mged.sourceforge.net/ontologies/MGEDontology.php#DNA
Accession Value	Enter the value given the term in the source vocabulary. *Example:* MO_945

Table 5.7 Fields for entering a new vocabulary term

4. Once you have entered the appropriate information, click **Save**. Click **Cancel** to halt the action.

   This returns you to the original Manage Vocabularies {attribute} tab.

Editing a Vocabulary Term

Note: A vocabulary term can be edited by anyone.

To edit a vocabulary term, follow these steps:

1. After logging into caArray, on the left sidebar under **Curation**, click **Manage Vocabulary.**
2. Select the attribute tab where the term you want to edit is listed.
3. On the row corresponding to the protocol, click the **Edit** icon ( ).

OR

4. Open the term details page by clicking on its value, and click the **Edit** button (Edit) at the bottom of the page.
5. Make the appropriate edits on the page.

   You can edit details of the term itself, its source, and the accession number for the term in the source database. All information is editable. For more information about the fields for defining the terms, see *Adding Vocabulary Terms* on page 62.
6. Save any edits by clicking the **Save** button. To abort the edit, click the **Cancel** button. This returns you to the Manage Vocabularies page.

CHAPTER

6

SUBMITTING DATA TO AN EXPERIMENT

This chapter describes the processes for submitting data such as annotation and array content into caArray experiments.

The following topics are part of this chapterchapter:

Managing Data

Note: Array designs are not imported through the tasks described in this section. For more information about working with array designs, see *Managing Array Designs* on page 54.

In caArray, you can submit data to an experiment by performing the following tasks on the Experiment Data tab:

1. **Uploading data**: moving data into caArray from a local or networked location
2. **Validating data**: determining by caArray if the data uploaded is valid according to pre-determined rules

3. **Importing data**: making validated or unvalidated data available through the user interface and or an API for download from caArray.

Note about file types in caArray:

caArray supports the ability to upload, validate, parse and import many data file types for the following providers: Affymetrix, GenePix and Illumina. The list of file types shown in caArray indicates those that caArray currently supports with full validation and parsing.

For unrecognized files, there is an option to set the file type to **Supplemental** which turns off validation and parsing and allows for the file to be downloaded. caArray also supports the ability to upload files for providers for which it does not have a parser: Agilent, Imagene, Nimblegen and UCSF Spot. For those files, validation and parsing is turned off and the end state of those files will be "imported not parsed". This will allow for the system to recognize that those files need to be parsed as new parsers are developed.

The Data tab is the vehicle for uploading, validating, importing and downloading data relating to caArray experiments. When you click on the **Data** tab, four subtabs where you initiate data-related tasks display. They are described in *Table 6.1*.

Data Tabs	*Description*
Manage Data	From this tab, you can perform data-related tasks such as uploading, validating and importing data into caArray. Additional tasks such as changing data file types and designating supplemental files also takes place here.
Imported Data	This subtab list all files that have been imported into caArray.
Supplemental Files	This tab lists files and documents that have been uploaded to caArray, and identified as supplementary (reference) files.
Download Data	From this tab, you can download data that has been imported into caArray.

Table 6.1 Tabs for performing data-related tasks

All kinds of data can be uploaded and otherwise managed in caArray, although the majority of data will likely be annotation and array content files.

Notes:

- To perform tasks on the Data tab, your experiment must be in Edit mode.
- Importing MAGE-TAB is the only mechanism for entering annotations that are not displayed as generically available and editable fields in the annotation user interface. The unique data will be visible but uneditable.
- Importing Array Design files is performed through the Curation tool in caArray, not on the Data tab. For more information, see *Managing Array Designs* on page 54.

Note: It is not possible to import source or sample annotations directly into caArray 2.0 from their respective tabs in the user interface. You can, however, import MAGE-TAB files that contain source and sample information. See *Importing Data* on page 70.

Uploading Data Files

Through the process of uploading annotation and array data, the content becomes available for validation and import into caArray. Imported files can be shared for download or deleted.

Notes:

- In caArray 2.0, the ability to upload files is restricted to files (raw or compressed in .zip format) that are 2GB or less. It is highly encouraged that upload jobs are no larger than 1GB at any one time. In addition, uncompressed files can only be half as large as the amount of memory on your server in order for validation and import processing to occur. caArray 2.0 has been tested using a 2GB allocation of memory and therefore the maximum size for any individual file is 1GB.
- caArray supports the upload of .zip compressed files only. NO other compression formats are supported in v.2.0 for extraction. caArray automatically extracts the files from a .zip file, discards the original and displays each of the files in the .zip to the user, indicating that each has been uploaded.

Steps for Uploading

Note: In caArray 2.0, the ability to upload files is restricted to files (raw or compressed in .zip format) that are 2GB or less. It is highly encouraged that upload jobs are no larger than 1GB at any one time. In addition, uncompressed files can only be half as large as the amount of memory on your server in order for validation and import processing to occur. caArray 2.0 has been tested using a 2GB allocation of memory and therefore the maximum size for any individual file is 1GB.

To upload data into caArray, follow these steps:

1. Go to My Experiment Workspace and locate the experiment of interest on the Work Queue tab.
2. In the My Experiment workspace or the search results page, click the **Edit** button (

) corresponding to the experiment to which you want to upload data. This opens the experiment to the Overview tab.
3. Select the **Data** tab, and **Manage Data** subtab.
4. Click the **Upload New Files** button.
5. In the form that opens, click **Browse** to navigate to the file you want to upload. To select several files for upload at once, click the **Add More Files** button.This opens more File/Browse options where you can locate several files.
6. Click the **Upload** button. Click **Cancel** if you decide to halt the task.

Note: caArray launches the upload process, which occurs in the background, allowing you to navigate through and use the application while the upload is in progress. The Upload window, which must remain open during the process, monitors the percentage of the upload completed as well as its status. caArray will inform you when the upload process is complete.

When the upload has finished, the list of files displays on the page, as well as their status (uploaded) and file type. As you continue to work with the data, their status

updates (**Uploading**, **Uploaded**, **Validating**, **Validated**, **Importing** and **Imported**) (*Figure 6.1*).

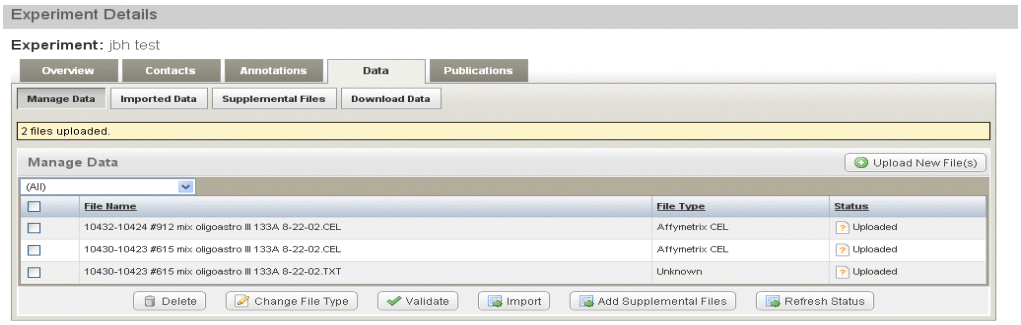


Figure 6.1 Manage Data subtab displays files in caArray and their status

Note: caArray supports the upload of .zip compressed files only. NO other compression formats are currently supported for extraction. caArray automatically extracts the files from a .zip file, discards the original and displays each of the files in the .zip to the user, indicating that each has been uploaded.

Check boxes corresponding to each file allow you to select one or more at a time individually for further management.

Once files are uploaded, the files workflow should be continued by validating the file(s) and importing the file(s). From the Manage Data tab, you can also change file types, designate files as supplementary, and delete files.

Deleting a File

On the Data tab, Manage Files subtab, check the box corresponding to one or more files you wish to discard and click the **Delete** button. The file(s) are physically deleted from caArray.

Validating Data Files

Once data have been uploaded into caArray, anyone associated with writing an experiment can validate and import annotation and array content files into the project. File validation verifies that data content adheres to a certain format; it does not evaluate the accuracy of the data from scientific viewpoint.

Note: The Validate feature for annotation data files whose Service Type is "Publish" is turned OFF. These files can still be imported, however. For more information, see *Service Type* in *Table 4.2*.

The following file types support information sharing and can be uploaded, but because they are not array data files, they are not validated and no validation routines are available. These files should be identified as Supplemental Files, as described in *Supplemental Files* on page 73, and then imported.

- Word documents
- Excel spreadsheets
- PowerPoint files

- PDFs

These files are not associated with any samples, but are associated at an experiment level.

Note: In caArray, many data file types can be uploaded, but not all file types can be validated (see definition above). All file types, even those that cannot be validated, can be imported. If you choose to import data that cannot be validated, validation is turned off, and a message indicating the data cannot be validated displays. The data gets imported, and its final state is "imported but not parsed".

The following array content file types can be validated and parsed:

- **Affymetrix**: .cel, .chp, .cdf
- **GenePix:** .gpr, .gal
- **Illumina**: .csv, .some .txt

Steps for Validating Data

To validate uploaded data files in caArray, follow these steps:

1. Go to your My Experiment Workspace and locate the experiment of interest on the Work Queue tab.
2. Click the **Edit** button (

). This opens the experiment to the Overview tab; you are in Edit mode.
3. Select the **Data** tab, and the **Manage Data** subtab.
4. Check the box corresponding to the file(s) you want to validate, and click the **Validate** button.

   Note: If you choose to import a file before validating it, validation is launched automatically prior to import.

   For files where the type cannot be inferred, their status is marked **Unknown**. You must change the file type to a known format before validation can proceed.

caArray performs structural and then content validation against each file you have selected, updating the status of each file, in the yellow message box, periodically (10 seconds) until all files display the validation status in the **Status** column: **Validated** or **Validation Failed**.

Validation Errors

If validation fails, the file cannot be imported and a **Validation Failed** message displays in the row corresponding to the file.

A validation error can be structural or content-based. Validation can fail for the following reasons:

- Format unknown (based on file extension and array type)
- Reference file not found
- File incomplete
- Vocabulary failure--annotation terms not found in supported ontology

To view a validation error description, click the hypertext **Failed Validation** link in the **Status** column

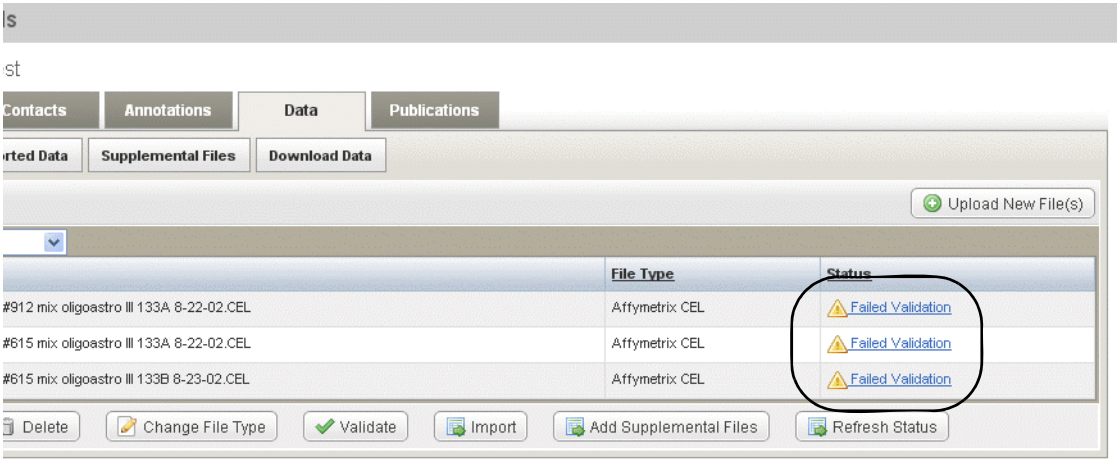


Figure 6.2 Validation failures display on the Manage Data table.

Importing Data

Once files are uploaded into caArray, anyone associated with an experiment can validate and import annotation and array content files of the appropriate formats into the project. The import feature allows the array data values to be parsed into the database such that the data is available for download when the Experiment is Public. In addition, if a parser is available for this file type, discrete data values are will be available through the API.

If you import just data files (for example, .cel, .chp, etc.), caArray automatically creates a source, sample, extract, labeled extract and hybridization for each data file. If multiple files with the same name (but different extensions) are imported, only one annotation chain of source > sample > extract > labeled extract > hybridization will be created, and all of the files will be associated with that single linked chain.

If you import a MAGE-TAB set (idf and sdrf) along with the data files, where the sdrf refers to each of the data files, the sdrf tells the system what sources, samples, extracts, labeled extracts and hybridizations to create.

Files can be downloaded from the Data tab or from the Annotations tabs. For more information, see *Downloading Data from caArray* on page 75.

Note: While data is being imported into an experiment, all attributes and annotations of that experiment become read-only, so that user interface changes do not conflict with annotations being created as part of the import.

Note: Required fields are turned off for data entry when the Service Type is **Publish**.

Importing MAGE-TAB Data

Files Types that can be Imported into caArray

The file types listed in *Table 6.2* can be imported into caArray:

File Types	*Acceptable File Formats*
Raw/processed data files	• Affymetrix .CEL • Affymetrix .CHP • Affymetrix.DAT (can be imported even though they cannot be validated in caArray) • Affymetrix.EXP (can be imported even though they cannot be validated in caArray) • GenePix .GPR • Illumina .CSV
Array Design files	• Affymetrix .cdf • Illumina Design .csv • Genepix .gal • Agilent .csv or .xml • UCSF Spot .spt • ImaGene .tpl • Nimblegen .ndf **Note:** These can be uploaded, validated and imported only through the Manage Array Design feature described in *Managing Array Designs* on page 54. • MAGE-TAB ADF (Array Design Format) **Note:** Any array design reference in a MAGE-TAB SDRF must refer to the LSID of an array design that has already been imported into caArray.
MAGE-TAB files	• MAGE-TAB with single SDRF (Sample and Data Relationship Format) • MAGE-TAB with multiple SDRFs • IDF (Investigation Description Format) only, no referenced SDRFs **Note:** Only one IDF is allowed per import, since the import is in the context of a single experiment. • SDRF with only source > Hybridizations • ADF (Array Design Format) • MAGE-TAB with existing samples in caArray.

Table 6.2 File types that cam be imported into caArray

Steps for Importing Data

To import data, follow these steps:

1. Go to your My Experiment Workspace and locate the experiment of interest on the Work Queue tab.

2. On the row that corresponds to the experiment, click the **Edit** button (

). This opens the experiment to the Overview tab; you are in Edit mode.
3. Select the **Data** tab, and the **Manage Data** subtab. Select the **Data** tab, and **Manage Data** subtab.
4. Check the box corresponding to the file(s) you want to import, and click the **Import** button.

   Note: If the file has not been previously validated, caArray performs the validation.

   — If validation fails, import does not proceed. Validation errors can be viewed as described in *Validating Data Files* on page 68.

   — If the validation completes successfully, caArray continues with the import and automatically auto-refreshes the status of the file set until the Import is complete and the Import Status of the file displays.
5. After a successful import, the files automatically move to the Imported Data subtab and the **Status** of the file set is **Imported**.
6. After a successful import, you can delete the file set. See *Deleting a File* on page 68.

Importing MAGE-TAB Data

Array data can be uploaded, validated and imported as MAGE-TAB files. The common flows are represented as follows:

- MAGE-TAB IDF, with no referenced SDRFs

Note: SDRF is a data format that represents Sample and Data Relationships.

- MAGE-TAB SDRF, with only source hybridizations.

In the MAGE-TAB SDRF being imported, the sdrf tells the system what biomaterials (sources, samples, extracts, labeled extracts) and hybridizations to create. If one or more biomaterial nodes is missing (no samples, extracts and/or labeled extracts are explicitly mentioned) in the Source > Sample > Extract > Labeled Extract > Hybridization chain, caArray generates appropriate intermediate nodes to complete the chain. (Protocols and characteristics can be present between the source and hybridization data.) The number of nodes generated will depend on the left side of the "flow".

Examples:

- If the SDRF describes one source connected to three extracts, one sample will be auto-generated and inserted in the chain.
- If the SDRF describes 3 sources combined to generate 1 extract, 3 samples will be auto-generated and inserted into the chain.

If biomaterials missing in the SDRF are auto-generated, caArray "intelligently" associates protocol applications with the most appropriate auto-generated node. Note that this applies only to protocol applications in the SDRF whose association to

biomaterial nodes is ambiguous. For example, if the SDRF contains a sample followed by a labeling protocol followed by a hybridization, the labeling protocol will be associated with the auto-generated extract > labeled extract portion of the chain. Similar logic is used for extraction and hybridization protocols.

Note: From the biomaterial (source, sample, extract, and/or labeled extract) or hybridization annotation pages (described in the sections beginning with *Biological Source Material* on page 34), you can download data files (e.g. .CEL, .CHP, etc.) that have been associated with these biomaterials and hybridizations. For more information, see *Downloading Data from caArray* on page 75.

Supplemental Files

Many file types can be uploaded into caArray, but only validated array content files can be imported into the application and parsed, therefore making the file content extractable through the API. Other files types can be designated on the Manage Data page as “supplemental files”.

To identify uploaded files as supplemental files, follow these steps:

1. Go to your My Experiment Workspace and locate the experiment of interest on the Work Queue tab.
1. Open the experiment and select the **Data** tab, and the **Manage Data** subtab.
2. Check one or more boxes for the file(s) you want to identify as supplemental files.
3. Click the **Change File Type** button.
4. On the drop-down list, for each appropriate file, scroll down and select **Supplemental File.** The file type then changes to Supplemental File.
5. Back on the Manage Data subtab, make sure these files are still checked, and click the **Add Supplemental Files** button.

   As you do so, the selected files are moved to the Supplemental Files tab.

Downloading Files

With appropriate permissions, you can extract files that have been imported into caArray. For more information, see *Chapter 7 Extracting Data from caArray*.

CHAPTER 7

EXTRACTING DATA FROM CAARRAY

This chapter describes the processes for extracting data from the caArray repository.

Topics in this chapter include:

Downloading Data from caArray

Any user can download public data from any publicly available instance of caArray. No permission restrictions are required to acquire the data, either at an experiment level or sample level within or across projects.

From the Download Data page, you can download any data files that have been uploaded in caArray.

From the biomaterial annotation pages associated with an experiment, you can download imported data files (e.g. .CEL, .CHP, etc.) that have been associated with those biomaterials or hybridizations. The files download as a .zip file. The MAGE-TAB files themselves must be downloaded from the Data tab described in this section. For more information about MAGE-TAB data, see *Importing MAGE-TAB Data* on page 72. For information about downloading data files from the Annotation tabs, see the sections beginning with *Biological Source Material* on page 34.

Note: Only files that have been imported into caArray can be downloaded on the biomaterial annotation pages. For more information, see *Importing Data* on page 70.

To download data of interest from the caArray Data tab, follow these steps:

1. Go to My Experiment Workspace and locate the experiment of interest on the Work Queue tab.
2. Open the experiment, and select the **Data** tab, and the **Download Data** subtab.

Note: You can sort the columns of the list by clicking on the column headers. You can also filter the list of files by choosing the file type on the Filter drop-down list.

All files that are part of this experiment display on this tab (*Figure 7.1*).

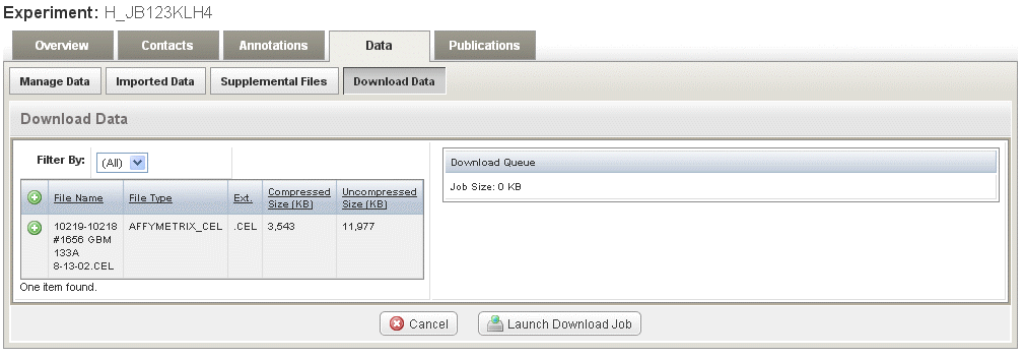


Figure 7.1 Download data subtab

3. Identify one or more file(s) to be downloaded. Click the plus icon ( ) to the left of the **File Name** column on the row of the file(s) you select. That places the data set in the download queue, visible in the right panel on the tab. The identifying information for the file, including the total file size displays there.

   Note: If you select multiple or all files to download, and there is a large amount of data, caArray calculates the total size of the download. If the size is greater than 1.5 GB (after compression), the system breaks up the download into batches of files, with each batch limited to a size nor more than 1.5 GB (after compression). A download button is provided for each batch. When you have downloaded a batch, it is marked as having been downloaded.

4. To remove selected files from the queue, click the **Remove** icon(s) ( ) corresponding to the data file or click the **Cancel** button.

5. Click the **Launch Download Job** button to intitiate the download process. On launching the process, caArray displays the following message: *The Download job is being assembled*. The job proceeds until all designated files are downloaded.

   - The length of time for the download is dependent upon the file size.
   - You can continue to work in caArray during the download process.

   In the dialog box that opens, indicate whether you want to open or save the file to be downloaded. To save, navigate to the destination where the file will be saved, always with the title `caArray.zip`.

When the download is complete, your local system displays an on-screen message telling you that the download is finished or that it failed.

Grid availability

For those Institutions that register their caArray instance on caGrid, the public data is available to the integrated tools that use the caGRID service. For more information, see https://cabig.nci.nih.gov/workspaces/Architecture/caGrid.

Note: The ability to browse and search from the caArray user interface features across the Grid is not available in caArray 2.0.

Extracting Data Programmatically by API

Data in caArray can be extracted programmatically using an API. caArray also provides a grid service which allows software engineers to acquire data from caArray.

For more information about using the remote Java API and grid service to extract data, see the *caArray 2.0 Technical Guide* which can be downloaded from this site: https://gforge.nci.nih.gov/frs/?group_id=305https://gforge.nci.nih.gov/frs/?group_id=305.

CHAPTER

8

USER ACCOUNT MANAGEMENT

This chapter describes the process for creating and managing accounts for users and groups of collaborators in caArray.

Note: The options for performing user management tasks are visible in caArray on the left sidebar of the browser only if you have these Admin privileges.

Topics in this chapter include:

Roles in caArray

When a new user account is created within caArray, the user can be assigned one or more roles that then determines that user's access to specified private data.

Role	*Description*	*Permissible Actions*
Anonymous User	User without a caArray account or a non-logged in user **Note:** Because of the use of CSM, the Anonymous Group is crafted, but it does not need to be assigned by the System Administrator. caArray does it automatically upon Login.	Browse and Search tasks, downloading Public data
System Administrator	Person responsible for the effective operation of caArray	Manages users

Role	Description	Permissible Actions
Principal Investigator [PI]	Owns experiments and studies and/or projects	Manages experiments Manages experiment visibility
Lab Administrator	Responsible for managing lab operations. They typically interact with submitting investigators, assign work, and run reports on the operations of the lab.	Same as PI in caArray 2.0
Lab Scientist	The primary handler of samples in the lab. They run the experiments, collaborate with the statisticians and document their activities step by step.	Same as PI in caArray 2.0
Biostatistician	A special form of submitter who is responsible for statistical analysis of project data. The key actions to be performed are review of experiment designs, submission of quality control metadata, and uploading of normalized data and the annotation of the parameters used.	Same as PI in caArray 2.0
External System **Note:** Not listed under Roles	Systems other than caArray from which caArray data can be extracted programmatically using an API.	For more informations, see *Extracting Data Programmatically by API* on page 77.

When the account is registered and roles are assigned, the user can fully access caArray according to the roles provisioned.

Managing User Accounts

In caArray 2.0, all tasks related to creating and managing user accounts can be performed only by a System Administrator. The System Administrator must use the NCICB User Provisioning Tool (UPT) v. 3.2 for performing these tasks. All instructions for managing user accounts are described in the UPT 3.2 User's Guide which can be accessed from this website link to NCICB GForge: http://gforge.nci.nih.gov/frs/download.php/2634/UPT_User_Guide.pdf

Note: If you are a person who wants to be caArray user, you can launch a request for a user account from the Welcome/Login page. For more information, see *Requesting a User Account* on page 10

Managing Collaboration Groups

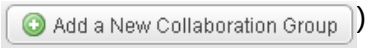
Any registered user in caArray can create, edit and delete collaboration groups and the users associated within them. This set of users ("collaborators") can then be given access to an experiment for which the user is the Data Owner (usually the creator of the experiment) or to particular samples and their underlying array data. Only registered users are available to be a part of a collaboration group.

Creating a Collaboration Group

A group can be created by any registered user of caArray.

For more information, see *Viewing Group Details* on page 82.

To create a group, as a logged in user**,** follow these steps:

1. Click the **Manage Collaboration Groups** option on the left sidebar.
2. On the Manage Collaboration Groups page, click the **Add a New Collaboration Group** button () in the upper right corner of the page.
3. On the New Collaboration Group page, enter the **Group Name** in the appropriate text box.
4. Click **Save**. This returns you to the Manage Collaboration Groups page that lists groups that you have created in the system. When a group you create first displays here, it is empty (circled in *Figure 8.1*).

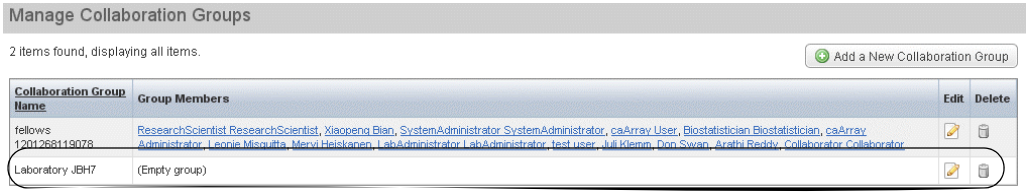


Figure 8.1 Collaboration groups are listed on the Manage Collaboration Groups page

5. To add members to a Collaboration Group, click the **Edit** icon ( ) corresponding to the group.
6. The group details page that opens lists any and all current members of the group. Click the **Add a New Group Member** button ().
7. In the form that opens, as appropriate, enter the registered users last name, first name, and organization (just the first letters), and click the **Filter** button () to find the user.

   Note: To return all registered users, enter nothing and click the **Filter** button (not recommended due to time).

   The user displays in the Member Name column.

8. For each member to be added to the group, click the **Add** icon ( in the far right column of the screen.

   The System automatically saves the user as a member and removes the name from the filter results.

9. Click on the **Collaboration Groups** bread crumb link or the left hand menu item at the top of the page to return to the main collaboration group page .

Viewing Group Details

From the list of groups on the Collaboration Groups page, you can view all groups you have created, the first 20 members and the ability to edit or delete them. The group details lists all users in the group, along with their corresponding Institution and email address.

Editing Collaboration Group Details

To edit collaboration group details, follow these steps:

1. From the Manage Collaboration Groups page, click the **Edit** icon () corresponding to the group you select.
2. The page that opens lists group members, their institution and email address. On this page, you can perform the following edits:
   - Edit the Group Name. (A Group Name must be unique within the system.)
   - Add or delete group members.
   - View user details using the hypertext link corresponding to a user in the group.
   - Delete the group.
3. To perform these edits, do the following:.

Edit Function	***Description***
Edit the group name	Enter new name in the Group Name text box.
Add a new group member	**Note:** The new member must already have a valid caArray user account. Click the **Add a New Group Member** button at the top right of the page. In the new section of the page that opens, you can search for the group member using one or more criteria. Enter the last name, select the Role category, the Institution, and the Status to be searched. Click the Filter button.
Remove a group member	On the Collaboration Group page, in the **Remove** column, click the icon () that corresponds to the group member.
Review group member details	Click the name of the group member. The page that opens displays contact information about the member in Read-only format.

Edit Function	***Description***
Delete the group	To delete the entire Collaboration Group, click the **Delete** button () at the bottom center of the page.

APPENDIX A

caARRAY REFERENCES

- **NCICB Production Site**: https://array.nci.nih.gov
- **Product Summary Site**: https://cabig.nci.nih.gov/tools/caArray - the summary of caArray capabilities and direction
- **Public Information Site**: http://caarray.nci.nih.gov/ – a public web site that allows anyone to download the latest version, access documentation, launch the portal and visit sites that provide analysis of the data contained in caArray.
- **caArray Work Group Site**: https://cabig.nci.nih.gov/workspaces/ICR/caArray-wg/ – this public web site provides access to the schedule, monthly meeting notes and links to the listserv for the stakeholder community
- **Microarray Gene Expression Data Society -** http://mged.org/ The providers and curators of microarray standards, software and models.

INDEX