

CHAPTER 2

CREATING A NEW STUDY

This [chaptersection](#) describes the processes for creating and managing studies in calnegrator2.

Topics in this [chaptersection](#) include:

- [Creating a Study – Overview on this page](#)
- [Managing a Study](#) on page 22

Creating a Study – Overview

You can create a calIntegrator2 study by importing clinical study data, genomics data and imaging data, using a combination of spreadsheet/files and existing caGrid applications as source data. As the manager creating the study, it is important that you understand the study well and that the data you wish to aggregate has been submitted to the applications whose data can be integrated in calIntegrator2.

- The clinical data should be available in CSV files, with a unique patient identifier in one column, one patient per row. Other relevant data can be supplied in other columns, to be identified as annotations in the file from within calIntegrator2. You, as the study creator, must have access to the clinical data file, as the file does not come from a caBIG[®] repository.
- The array data should be deposited in caArray, either locally or the CBIIT installation. You must also have a mapping file in CSV format. This file consists of two columns: one with the patient ID, and one with the sample ID.
- Imaging data should have been submitted to NCIA grid node as public data, either locally or as part of the CBIIT installation. Image annotations must be in CSV format, with unique image series IDs in one column and annotation IDs in the second column. You must also have a image mapping file in CSV format. This file consists of two columns: one with the patient ID, and one with the NCIA study instance UID.

As you create the study, you define its structure in the process, identifying the data sources and mapping the data between different source data. caIntegrator2 has KM-plot as well as genePattern integration for performing analyses.

Configuring and Deploying Study

Note: Only a user with a Manager role can create a study.

When you create a study, you must specify different data-types (clinical, array, image, tissue etc), data sources (caGrid applications – caArray and NCIA) and map the data, (patient to sample, image series, etc.).

To create a new study, follow these steps:

1. In the Study Management section of the left sidebar, click **Create New Study**.
2. In the Create New Study dialog box that opens, provide a name and description for the study you are creating.
3. Click **Save**.

This opens an Edit Study page where you can add identify data files for your study. [See Creating/Editing a Study.](#)

Creating/Editing a Study

The Edit Study page displays the Name and Description that you entered for a new study, or for an existing study that you are editing.

To continue creating a study or to modify a study, on the Edit Study page complete

these steps:

1. Change the name and or description, if you so choose. Click **Save**.

Note: You can save the study at any point in the process of creating it. You can resume the definition and deployment process later.

- If you choose to add a logo for the study, click the **Browse** button corresponding to **Logo File**. Navigate for the file, then click **Add Study Logo**. Once you save the study (or its edit), the logo displays above the left sidebar.

To continue, you can add clinical data sources, genomic data or imaging data.

Adding Clinical Data

The Edit Study page opens after you save a new study or click to edit an existing study.

Note: To edit information for an existing study, follow the same basic directions in this [topic section](#). Instead of entering new information, you will modify existing information.

To add or edit clinical metadata in this page, follow these steps:

- To add or edit clinical data, on the Edit Study page click the **Browse** button. Navigate to locate the file. Files must be in CSV file format.

In the Clinical Data Sources section, if a file has already been selected, its information displays in the varying fields.

- Click **Add Clinical Data Source**. This opens the Define Fields for Clinical Data page.

Define Fields for Clinical Data

| Study Name: Rembrandt with NCRI Study | |
|---|------------------------|
| Field Definition | Field Header from File |
| Identifier Change Assignment | Subject |
| Age Change Assignment | Age |
| GENDER Change Assignment | Gender |
| Survival Change Assignment | Survival |
| Disease Change Assignment | Disease |
| Grade Change Assignment | Grade |
| RACE Change Assignment | Race |

Clinical data is always added in spreadsheet format (CSV file). The page displays column headers and data values in the file you have designated. You

must map each column name to an existing column name in the calIntegrator2 database or, if it doesn't yet exist, create a custom column name.

| | A | B * | C | D | E | F | G | H |
|----|--------|-------|--------|----------|-------------|-------|-----------|---|
| 1 | Pa | Age | Gender | Survival | Disease | Grade | Race | |
| 2 | ASP221 | 50-54 | M | | ASTROCYTOMA | | WHITE | |
| 3 | ASP308 | 50-54 | M | | GBM | | WHITE | |
| 4 | FPH113 | 20-24 | M | | UNKNOWN | | WHITE | |
| 5 | FPH114 | 40-44 | M | | UNKNOWN | | WHITE | |
| 6 | FPH118 | 55-59 | M | | GBM | | WHITE | |
| 7 | FPH309 | 50-54 | M | | GBM | | WHITE | |
| 8 | E09238 | 45-49 | M | 18-24M | GBM | | WHITE | |
| 9 | E09239 | 25-29 | M | | UNKNOWN | | WHITE | |
| 10 | E09262 | 35-39 | M | | ASTROCYTOMA | | WHITE | |
| 11 | E09278 | 30-34 | M | | UNKNOWN | | WHITE | |
| 12 | E09331 | 35-39 | M | | UNKNOWN | | ASIAN NOS | |
| 13 | E09332 | 55-59 | M | | GBM | | WHITE | |
| 14 | E09336 | 30-34 | M | | GBM | | WHITE | |
| 15 | E09348 | 60-64 | M | | GBM | | WHITE | |
| 16 | E09378 | 45-49 | M | | UNKNOWN | | WHITE | |
| 17 | E09449 | 50-54 | M | | UNKNOWN | | OTHER | |
| 18 | E09454 | 0-4 | M | | UNKNOWN | | WHITE | |
| 19 | E09489 | 55-59 | M | | GBM | | WHITE | |
| 20 | E09515 | 35-39 | M | | UNKNOWN | | WHITE | |
| 21 | E09569 | 45-49 | M | | UNKNOWN | | WHITE | |
| 22 | E09587 | 35-39 | M | | UNKNOWN | | OTHER | |
| 23 | E09601 | 40-44 | M | | GBM | | WHITE | |
| 24 | E09610 | 55-59 | M | | GBM | | WHITE | |
| 25 | E09611 | 60-64 | M | | UNKNOWN | | ASIAN NOS | |
| 26 | E09615 | 45-49 | M | | UNKNOWN | | WHITE | |
| 27 | E09624 | 35-39 | M | | GBM | | WHITE | |
| 28 | E09645 | 45-49 | M | | UNKNOWN | | WHITE | |
| 29 | E09657 | 50-54 | M | | UNKNOWN | | WHITE | |
| 30 | E09730 | 40-44 | M | | UNKNOWN | | WHITE | |

The Field Header from File column on the Define Fields... page displays column headers taken from the source *.csv file.

The MOST important steps in creating a new study in calIntegrator2:

- You MUST designate one column in the file as a unique “identifier” column type.
- You MUST review and define column annotation definitions for each column header in the file.

If calIntegrator “recognizes” the same column header in other files already in the system, a term such as “age” or “survival” which is the current definition appears in the **Field Definition** column above the blue **Change Assignment** link. If the area above the blue **Change Assignment** link is blank, no correlating term exists in the database; you must specify the field type, and then the term will populate the space.

3. To indicate the unique identifier of choice, on the row showing the column header (PatientID in the figure, but other examples are subject identifier, sample identifier, etc), click **Change Assignment** in the **Field Header from File** column.

Assigning An Identifier or Annotation

When you click **Change Assignment** on the Define Fields... page, the Assign Annotation Definition for Column dialog box opens. On this page you can change the column type and the field definition for the specific data field you selected.

Note: When you change an assignment, you must make sure the data types match--numeric, etc.

Assign Annotation Definition for Column: PateintID

Column Type: Annotation

New Save

Search For an Annotation Definition:

Search Search existing studies and caDSR for definitions.

| Matching Annotation Definitions | | | |
|-----------------------------------|---------------|-----------|------------|
| Name | CDE Public ID | Data Type | Definition |
| No matches found for your Search. | | | |

| Matches from caDSR | | | | | |
|--|---------|---------------|---------|--------|------------|
| Name | Actions | CDE Public ID | Context | Status | Definition |
| No matches found in caDSR for your Search. | | | | | |

1. For the column (PatientID) that you choose to be the one and only Identifier column, in the **Column Type** drop-down list, select **Identifier**.

When the definition becomes an identifier, the rest of the page disappears, showing only the Identifier definition.

Assign Annotation Definition for Column: PateintID

Column Type: Identifier

Save

Note: If you select **Annotation** after you have already selected **Identifier**, the rest of the page reappears.

2. Click **Save** to save the identifier. This returns you to the Define Fields for Clinical Data page.
3. After you have defined which field is the Identifier, you must ensure that ALL other fields also have a field definition assignment. For those fields without a Field Definition assignment or for those whose Field Definition you want to review, click **Change Assignment**.
4. In the Assign Annotation Definition for Column: [column header] dialog box, select **Annotation** in the drop-down list.

As you select the column type, you can work with column headers in one of four ways in this dialog box.

- You can accept existing default definitions (those that are inherent in the data file you selected).
 - You can create your own definitions manually.
 - You can search for and use definitions in other caIntegrator2 studies.
 - You can search for and use definitions in caDSR.
5. If there is anything you want to change about the annotation definition of the field such as its name, or if you want to view or edit its definition, click the **New** button under the **Column Type** field. The page expands, now including a Current Annotation Definition section.

Note: If the column header you are working with already has a designated Field Definition, the Current Annotation Definition section of the Assign

Annotation Definition for Column page is already visible when you open this dialog box.

Current Annotation Definition:

Name:
 Definition:
 Keywords:
 Data Type:

Permissible Values:

Non-Permissible
X67129
X56199
X56145
X56123
X89123

6. You can enter a new name annotation, or any other information about the annotation definition:

| Annotation Field | Field Description |
|-------------------------|--|
| Description | Enter any information not covered by other fields. |
| Keywords | Insert keyword(s) that can be used to find the study in a search, separated by commas. |
| Data Type | Enter a string (default), numeric, or date |

Table 2.1

| Annotation Field | Field Description |
|---|--|
| Permissible/Non-permissible Values | <p>Note: The first time you load a file, before you assign annotation definitions (step 3 on page 12), these panels may be blank. If the column header for the data is already “recognizable” by calIntegrator2, the system makes a “guess” about the data type and assigns the values to the data type in the newly uploaded file. They will display in the Non-permissible values sections initially. Use the Add and Remove buttons to move the values shown from one list to the other, as appropriate.</p> <p>When you select or change annotation definitions by selecting matching definitions (described in Searching for Annotation Definitions on page 13), this may add (or change) the list of non-permissible values in this section.</p> <p>If you leave all values for a field in the Non-permissible panel, then when you do a study search, you can enter free text in the query criteria for this field.</p> <p>If there are items in the Permissible values list, then the values for this annotation are restricted to only those values. When you perform a study search, you will select from a list of these values when querying this field. If there are no items in the permissible values list then the field is considered free to contain any value.</p> <p>To edit a field's permissible values, you must change the annotation definition. You can do this even after a study has been deployed.</p> |

Table 2.1

Searching for Annotation Definitions

1. An alternative to creating a new definition is to search for annotation definitions already present in calIntegrator2 studies or in caDSR. Enter search keyword(s)

in the **Search** text box. Click **Search**. After a few moments, the search results display on the page.

Search For an Annotation Definition:

age Search existing studies and caDSR for definitions.

| Matching Annotation Definitions | | | | | |
|---|---|---------------|---------|-----------|--|
| Name | CDE Public ID | | | | Data Type |
| Age | | | | | string |
| Age | | | | | string |
| Age | | | | | string |
| Matches from caDSR | | | | | |
| Name | Actions | CDE Public ID | Context | Status | Definition |
| Pregnancy Age Birth First Child Category | Select View | 2442895 | PS&CC | RELEASED | The age of the participant at the birth of her first child. |
| Weight Person Age 20 Number | Select View | 2443544 | PS&CC | RELEASED | The numeric value to represent a person's weight in pounds at age 20 years. |
| Weight Person Age 50 Number | Select View | 2443527 | PS&CC | RELEASED | The numeric value to represent a person's weight in pounds at age 50 years. |
| Uterus Removed Age Category | Select View | 2442944 | PS&CC | RELEASED | The age the participant had her uterus or womb removed. |
| Edit Over-ride Age Site Morphology Flag Number Code | Select View | 2697798 | PS&CC | DRAFT NEW | Some computer edits identify errors. Others indicate possible errors that require manual review. This code is used to indicate that data in a record (or records) have been reviewed by a clinician. The code is used to indicate that data in a record (or records) have been reviewed by a clinician. The code is used to indicate that data in a record (or records) have been reviewed by a clinician. |
| Demographic Patient Diagnosis Age Value | Select View | 2660065 | PS&CC | DRAFT NEW | The number that represents the age of the patient at diagnosis in complete years. |

- To view the definitions corresponding to any of the “Matching Annotation Definitions”, which are those currently found in other caIntegrator2 studies, click the [term], such as “age”, hypertext link. The definition then appears in the Current Annotation Definition segment of the page just above. [Verify this](#). You can modify an portion of the definition, as described in [step 6](#) on page 12.
- The matches from caDSR display some of the details of the search results. To view more details of a match, such as permissible values, click **View**, which opens caDSR to the term. If you click **Select**, the caDSR definition automatically replaces the annotation definition for this field with which you are working.

Caution: Take care before you add a caDSR definition that it says exactly what you want. caDSR definitions can have minor nuances that require specific and limited applications of their use.

- Once you have settled on an appropriate field definition for the annotation, click **Save**. This returns you to the Define Field for Clinical Data page.

Note: If you have not clicked **Select** for alternate definitions in this dialog box, then click **Save** to return to the Define Field...dialog box without making any definition changes.

- From the Define Fields for Clinical Data screen, be sure and designate the annotations for each field in the file. Click **Save** on each page to save your entries or click **New** to clear the fields and start again. You will not be able to proceed until every Field Definition entry on the Fields for Clinical Data screen has a unique entry.

On the Define Fields for Clinical Data page, the Field Definition column displays the term “Identifier” corresponding to the Column Header you designated as

such. The other cells in the Field Definition column display the column header *values* of the rows you designated as “annotations”.

6. Check the boxes corresponding to each column header/data set whose data you want to display in your browser.
7. Once you have assigned the identifier and assigned all annotations, click **Save**. This saves the study by name and description, but does not deploy the study, which is described in [Deploying the Study](#) on page 21.

Saving the study returns you to the Edit Study page where a “Not Loaded” status now appears for the file whose annotations (column headers) you have defined.

Edit Study

Study Name: Rembrandt with NCRI S

Study Description: Rembrandt/VASARI demo study

Save

Deploy

Logo File (JPEG or GIF): Browse...

Add Study Logo

| Clinical Data Sources | | | |
|-----------------------|---------------------------|------------|----------------------|
| Type | Description | Status | Action |
| DELIMITED_TEXT | clinical_test_for_doc.csv | Not Loaded | Edit Load Clinical |

File: Browse...

8. Click the **Load Clinical** link in the Actions section to load the data file you configured. At this point, the Status changes to “Loaded”.

Note: You can add as many files as are necessary for a study. Patients 1-20 in first file, 21-40 in second file, or many patients in first file and annotations in second file, etc. As long as IDs defined correctly, it works.

9. Once you have assigned data types to every column header in the data file and have loaded the clinical data, click **Save and Deploy**. At that point, caIntegrator2 loads data from the file to the caIntegrator2 database.

Should user wait to deploy until AFTER the genomic and imaging data is identified and uploaded?

When study is deployed, is there anything on the pages that indicates that status?

The Manage Studies page opens when the study is deployed.

Jill, include X ref to deployment section.

Note: You can repeatedly upload additional or updated subject annotations, samples, image data, array data to the study at later intervals. These later imports do not remove any existing data; they instead insert any new subjects or update annotations for existing subjects.

Mapping Genomic Data to Clinical DataJill, add sentence of introduction...

Uploading Control Samples

Deploying the Study

Creating/Editing a Study

Adding Clinical Data

Editing Survival Values

Adding Genomic Data

Adding Imaging Data

Managing Platforms

Editing Survival Values

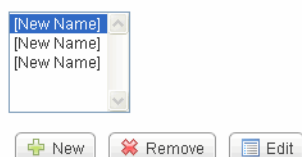
If you plan to analyze your data in caIntegrator2 creating a Kaplan-Meier (KM) Plot, you must define survival values for your data. This is optional, if you do not plan to use the KM plot analysis feature or even if you do not have this kind of data (survival values) in the file.

For some applications, such as REMBRANDT and I-SPY, survival values are pre-defined in the databases when you load the data is loaded. In caIntegrator2, however, you can review and modify survival value ranges in a data set you are uploading to create a study. To be able to do so, you need to understanding what kind of data needs to make up the survival values.

To work with survival values, follow these steps:

1. On the Edit Study page, click **Edit Survival Values**.

Survival Value Definitions for 'jbh test'



2. Click **New** to enter new survival value definitions.

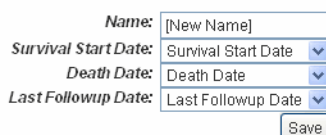
- or -

3. Click **Edit** to edit existing survival value definitions.
4. The dialog box extends, now displaying three drop-down lists that show column headers for date metadata in the spreadsheet you have uploaded.

Survival Value Definitions for 'jbh test'



Survival Value Definition Properties for '[New Name]'



In the drop-down lists, select the appropriate survival value definitions for each field listed. You might want to refer to the column headers in the data file itself. Dates covered by the definitions are already in the data set. You cannot enter a specific date.

- **Name** – Enter a unique name that adequately describes the survival values you are defining here. *Example:* Survival from Enrollment Date or Survival from Treatment Start. The name you enter displays later when you are selecting survivals to create the KM plot.
- **Survival Start Date** – Select the column header for this data
- **Death Date** – Select the column header for this data
- **Last Followup Date** – Select the column header for this data

Note: Dates corresponding to each field in the drop-down lists are already in the data set columns. You cannot enter a specific date.

Adding Genomic Data

Once you have loaded clinical data and identified patient IDs, you can add either array genomic sample data from caArray which calIntegrator2 maps by sample IDs to the patient IDs in the clinical data, covered in this section, or you can load imaging files from NCIA, also mapped by IDs to the patient data. Genomic sample data and imaging data are independent of each other, so neither is required before loading the other.

It is essential that you are well acquainted with the data you are working with--the clinical data, and the corresponding array data in caArray.

calIntegrator2 support a limited number of array platforms. For more information, see [Managing Platforms](#) on page 23.

To add genomic data to your calIntegrator2 study, follow these steps:

1. On the Edit Study page where you have selected and added the clinical data, click the **Add** button under Genomic Data Sources. You can upload genomic data only from caArray.

This opens the Edit Genomic Data Source dialog box. Enter the appropriate information in the fields.

Edit Genomic Data Source
 caArray Server Hostname:
 caArray server JNDI Port:
 caArray Username:
 caArray Password:
 caArray Experiment Id:

- **caArray Host Name** – Enter the URL for your local installation or for the CBIIT installation of caArray
- **caArray JNDI Port** – Enter the appropriate server port. See your administrator for more information.

- **caArray Username** and **caArray Password** – If the data is private, you must enter your caArray account user name and password; you must have been given permissions in caArray for the experiment. If the data is public, you can leave these fields blank.
- **caArray Experiment ID** – Enter the caArray Experiment ID which you know corresponds with the clinical data you uploaded.
- Click **Save**.

caIntegrator2 goes to caArray, validates the information you have entered here, finds the experiment and retrieves all the sample IDs in the experiment. Once this finishes, the experiment information displays on the Edit Study page under the Genomic Data Sources section.



- If you want to redefine the caArray experiment information, you can edit it. Click the **Edit** link corresponding to the Experiment ID. The Edit Genomic Data Source dialog box reopens, allowing you to edit the information.

Mapping Genomic Data to Clinical Data Jill, add sentence of introduction...

To map the samples from the caArray experiment to the patients in the clinical data you uploaded, follow these steps:

- On the Edit Study page, click the **Map Samples** link. This opens the Edit Sample Mappings page.

Edit Sample Mappings

caArray Server Hostname:
caArray server JNDI Port:
caArray Username:
caArray Password:
caArray Experiment Id:
Subject to Sample Mapping File:

Control Samples File:

| Unmapped Samples | |
|------------------|--|
| Sample ID | Sample Name |
| 901 | GeneratedSample.LINKNOWN_DISEASE_L_E10216_U133P2 |
| 902 | GeneratedSample.GBM_L_20070226_15-14-09-358_HF0936_U133P2 |
| 903 | GeneratedSample.OLIGO_L_20070227_12-21-11-104_HF1380_U133P2 |
| 904 | GeneratedSample.GBM_L_20070226_14-05-29-569_HF1262_U133P2 |
| 905 | GeneratedSample.OLIGO_L_20070227_11-49-51-876_HF0899_U133P2 |
| 906 | GeneratedSample.Ast_L_20070226_11-54-59-645_HF0026_U133P2 |
| 907 | GeneratedSample.OLIGO_L_20070227_11-49-51-876_HF1136_U133P2 |
| 908 | GeneratedSample.LINKNOWN_L_E09963_U133P2 |
| 909 | GeneratedSample.LINKNOWN_L_20070227_16-22-37-238_570013_U133P2 |

When you first open this page, all of the samples in the caArray experiment you selected are listed as unmapped, because calIntegrator2 does not know how these sample names correlate to the patient data in the clinical file until you upload the mapping file.

2. Browse for the .csv file that identifies the mapping information. Click the **Upload Mapping File** button.

The mapping file has only two columns—one that shows that subject ID and one that has sample ID with one subject per row. This provides calIntegrator2 with the information for mapping patients to samples.

Note: When you open the mapping file, make sure that the patient ID is used for mapping.

Unmapped samples continue to show at the top of the page. They were loaded from caArray, but they are not in the mapping file. These are not used for integration.

3. Scroll down the page to see samples that are mapped to the patients in the clinical data.

| | | |
|----------------------------|---|--------------------|
| 1336 | GeneratedSample_OUGO_L_20070227_11-49-51-876_HF1356_U133P2 | |
| 1338 | GeneratedSample_UNKNOWN_DISEASE_L_E10029_U133P2 | |
| 1339 | GeneratedSample_GBM_L_20070226_14-05-29-569_HF1356_U133P2 | |
| 1340 | GeneratedSample_OUGOODHROGUOMA_L_HF0599_U133P2 | |
| 1342 | GeneratedSample_GBM_L_20070226_13-30-40-39_HF0142_U133P2 | |
| 1345 | GeneratedSample_GBM_L_20070226_14-31-29-427_HF1409_U133P2 | |
| Samples Mapped to Subjects | | |
| Sample ID | Sample Name | Subject Identifier |
| 901 | GeneratedSample_UNKNOWN_DISEASE_L_E10216_U133P2 | E10216 |
| 911 | GeneratedSample_UNKNOWN_DISEASE_L_E10144_U133P2 | E10144 |
| 914 | GeneratedSample_UNKNOWN_L_20070227_16-22-37-238_E09212_U133P2 | E09212 |
| 918 | GeneratedSample_UNKNOWN_L_20070227_16-22-37-238_E09360_U133P2 | E09360 |
| 922 | GeneratedSample_UNKNOWN_DISEASE_L_E10102_U133P2 | E10102 |
| 925 | GeneratedSample_UNKNOWN_DISEASE_L_E10318_U133P2 | E10318 |
| 930 | GeneratedSample_OUGO_L_20070227_11-27-27-001_E09204_U133P2 | E09204 |
| 940 | GeneratedSample_UNKNOWN_DISEASE_L_E10252B_U133P2 | E10252 |
| 954 | GeneratedSample_GBM_L_20070226_13-14-06-57_E09024_U133P2 | E09024 |
| 957 | GeneratedSample_ASTROCYTOMA_L_E09137_U133P2 | E09137 |
| 958 | GeneratedSample_UNKNOWN_DISEASE_L_E09890_U133P2 | E09890 |
| 968 | GeneratedSample_UNKNOWN_L_20070227_16-57-07-283_E09815_U133P2 | E09815 |
| 1004 | GeneratedSample_UNKNOWN_L_20070227_17-26-09-910_E09722_U133P2 | E09722 |

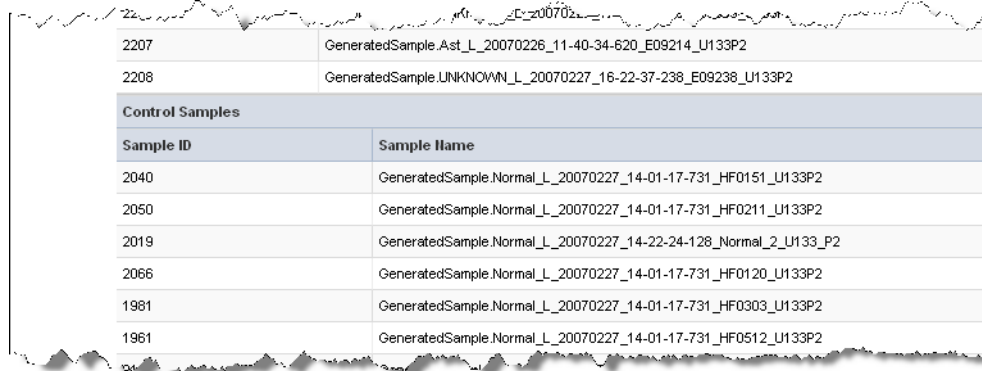
Uploading Control Samples

The Control Samples file is used to calculate fold change data, which compares “tumor” sample gene expression in the caArray experiment to the control samples to identify those that are more highly expressed. Control samples can be the “normal” samples, but that is not necessarily the case.

To upload the control samples, follow these steps:

1. On the Edit Sample Mappings page, click the Map Samples link.

2. Click **Browse** to navigate for the control samples file, and click the **Upload Control Samples** File button. Scroll down the page to view the list of control samples that have been added.



| 2207 | GeneratedSample.Ast_L_20070226_11-40-34-620_E09214_U133P2 |
|------------------------|---|
| 2208 | GeneratedSample.UNKNOWN_L_20070227_16-22-37-238_E09238_U133P2 |
| Control Samples | |
| Sample ID | Sample Name |
| 2040 | GeneratedSample.Normal_L_20070227_14-01-17-731_HF0151_U133P2 |
| 2050 | GeneratedSample.Normal_L_20070227_14-01-17-731_HF0211_U133P2 |
| 2019 | GeneratedSample.Normal_L_20070227_14-22-24-128_Normal_2_U133_P2 |
| 2066 | GeneratedSample.Normal_L_20070227_14-01-17-731_HF0120_U133P2 |
| 1981 | GeneratedSample.Normal_L_20070227_14-01-17-731_HF0303_U133P2 |
| 1961 | GeneratedSample.Normal_L_20070227_14-01-17-731_HF0512_U133P2 |

The control samples now display toward the bottom of the page.

3. This information will be used when performing other tasks in calIntegrator2, to be described in other sections.

Adding Clinical Data

Adding Imaging Data

Managing Platforms

Adding Imaging Data

Once you have loaded clinical data and identified patient IDs, you can add either array genomic sample data from caArray which calIntegrator2 maps by sample IDs to the patient IDs in the clinical data, or you can load imaging files from NCIA, also mapped by IDs to the patient data, covered in this section. Genomic sample data and imaging data are independent of each other, so neither is required before loading the other.

It is essential that you are well acquainted with the data you are working with--the clinical data, and the corresponding imaging data in NCIA.

Any data in NCIA can be uploaded to calIntegrator2. Imaging data consist of images and or annotations for images.

To add imaging data to the study you are creating or are editing, follow these steps:

1. On the Edit Study page, click the **Add** button under Imaging Data Sources section. Imaging data can be NCIA images or image annotations, which are uploaded in spreadsheet format.

This opens the Edit Imaging Data Source dialog box. Enter the appropriate information in the fields..

Edit Imaging Data Source

Enter a NCIA Data Source

Enter Image Annotation Data From file

NCIA Server Grid URL:

NCIA Username:

NCIA Password:

Protocol ID:

Image Series Annotation File:

Clinical/Imaging Mapping File:

- **NCIA Server Grid URL** – Enter the URL for the grid connection to NCIA
 - **NCIA Username and NCIA Password.** This information is not required, as currently all data in the NCIA grid is Public data.
 - **Protocol ID** – Enter the protocol ID assigned to the NCIA file.
 - **Image Series Annotation File** – Click **Browse** to navigate to the appropriate image annotation file in NCIA. This file, which must be in CSV format, is similar to a clinical data file. It contains features and /or descriptions of images and their corresponding series IDs. If you select this option, you must also upload a mapping file, described in the next point.
 - **Clinical/Imaging Mapping File** – Click **Browse** to navigate to the to appropriate clinical/imaging file. It is similar to the genomic data sources mapping file in that the spreadsheet shows only two columns—one that shows that subject ID and one that displays the corresponding image series ID (assigned by NCIA), providing calIntegrator2 with the information for mapping patients to images or annotations.
2. Click **Add** to upload the data to calIntegrator.
 3. Once the data is loaded, you must assign identifiers and annotations to the data in the same way you did with the clinical data. For more information, see [Assigning An Identifier or Annotation](#) on page 10 and [Searching for Annotation Definitions](#) on page 13.
 4. To deploy the study, see [Deploying the Study](#).

[Adding Clinical Data](#)

[Adding Genomic Data](#)

[Managing a Study](#)

[Managing Platforms](#)

Deploying the Study

When you are ready to deploy the study, click the **Save and Deploy** button on the Edit Study page. [See questions about deployment at the bottom of page 16.](#) calIntegrator2 retrieves the selected data from the data service(s) you defined and makes the study available for use to a study manager or to anyone else who may want to analyze the

study's data. The study is then listed by name under Manage Studies on the left sidebar. You can then configure and share data queries and data lists to be available for all investigators who access the study.

When you have deployed the study, you have either uploaded data (clinical and genomic sources) or told calIntegrator2 where all the data you are interested in resides (NCIA), and you have uploaded files that provide the mapping between the systems.

See also *Managing a Study*.

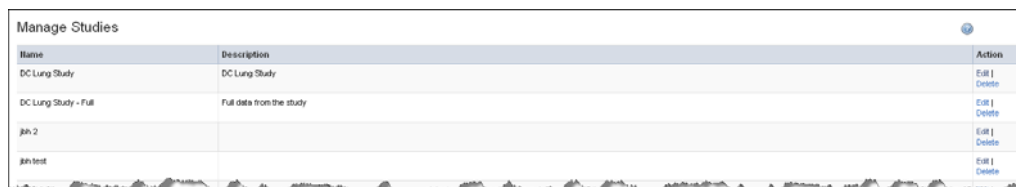
Managing a Study

Once you have started to create a study or have deployed it, you can update an existing study in the following ways:

- Add new attributes (annotations) and upload relevant data to an existing study. Delete a study
- Reload subset of study data (clinical only?) and re-deploy the study and perform new analyses
- Re-deploy the entire study with new set of data and mappings.

To update, edit or delete a study, follow these steps:

1. On the left sidebar, click **Manage Studies**. The Manage Studies page appears.



| Name | Description | Action |
|----------------------|--------------------------|---|
| DC Lung Study | DC Lung Study | Edit Delete |
| DC Lung Study - Full | Full data from the study | Edit Delete |
| jsh-2 | | Edit Delete |
| jsh-test | | Edit Delete |

All of the in process or completed studies display on this page.

2. Click the **Edit** link corresponding to your study of choice to open the Edit Studies page. On this page you can edit any details such as adding or deleting files, survival values, and so forth. For information about working in the Edit Study, see [Creating/Editing a Study](#) on page 8.
3. Click the **Delete** link to delete the corresponding study.

Don't see any of these following features described in development documents:

- Add/remove array platforms via an "Edit Array Platforms" link on the Edit Study page. looks to me like this occurs in Manage Platforms.
- ditto: The "Edit Array Platforms" link will take the user to a new screen for viewing the list of currently supported platforms and uploading new array annotation files.
- The existing "Edit Genomic Data Source" screen will have an additional pulldown for selecting an array name.

Managing Platforms

calIntegrator2 supports a limited number of array platforms, all of which originate from Affymetrix. While they do not represent all of the platforms supported by caArray, calIntegrator2 must have array definitions loaded for the platforms it supports, and be able properly load the data from caArray and parse it.

The supported platforms are identified and managed on the Manage Platforms page.

To manage platforms in calIntegrator2, follow these steps:

1. Click **Manage Platforms** on the left sidebar. The Manage Platforms page that opens lists the platforms calIntegrator2 currently supports, those that the system can pull from caArray.

Manage Platforms

| Affymetrix Annotation File: | |
|------------------------------------|--|
| <input type="text"/> | <input type="button" value="Browse..."/> |
| <input type="button" value="Add"/> | |
| Name | Vendor |
| HG-U133A | Affymetrix |
| HG-U133B | Affymetrix |
| HG-U133_Plus_2 | Affymetrix |
| HG_U95Av2 | Affymetrix |
| HG_U95B | Affymetrix |
| HG_U95C | Affymetrix |
| HG_U95D | Affymetrix |

2. You cannot create a calIntegrator2 study without the genomic platform in the system, so you can add additional Affymetrix array platform annotation files (use chips instead of arrays?) not represented in this list. To do so, click **Browse** to navigate for the Affymetrix Annotation File you want to add.
3. Click the **Add** button.

Will these files that a user adds be able to integrate and parse data uploaded from caArray? In other words, any caveats if user decides to do this?

Anything else to add in this section?

