NiDB User Guide

March 2015

Greg Book - [gregory.book@hhchealth.org](mailto:gregory.book@hhchealth.org)

Table of Contents

[2 Overview & Database Access 4](#_Toc415563733)

[3 Projects, Instances, and Sites 4](#_Toc415563734)

[4 Subject Management 4](#_Toc415563735)

[4.1 Hierarchy 4](#_Toc415563736)

[4.2 Subject Creation & Enrollment 5](#_Toc415563737)

[4.2.1 Import via DICOM Receiver 5](#_Toc415563738)

[4.2.2 Subject Creation via Imported MRI data 6](#_Toc415563739)

[4.2.3 Manual Subject Creation 7](#_Toc415563740)

[4.3 Modifying Existing Subjects 10](#_Toc415563741)

[4.4 Modifying Subject Enrollment 10](#_Toc415563742)

[4.4.1 Subject group 10](#_Toc415563743)

[4.4.2 Change Project Enrollment 11](#_Toc415563744)

[4.5 Deleting Subjects 11](#_Toc415563745)

[5 Imaging Studies 11](#_Toc415563746)

[5.1 Creating Imaging Studies 11](#_Toc415563747)

[5.2 Modifying Imaging Studies 12](#_Toc415563748)

[6 Imaging Series 13](#_Toc415563749)

[6.1 Importing Series 13](#_Toc415563750)

[6.2 Modality Specific Series 13](#_Toc415563751)

[6.2.1 DICOM originating series 13](#_Toc415563752)

[6.2.2 All other series 15](#_Toc415563753)

[7 Importing Data 15](#_Toc415563754)

[7.1 MRI 15](#_Toc415563755)

[7.1.1 Uploading MRI Data through the NiDBUploader 16](#_Toc415563756)

[7.2 EEG 17](#_Toc415563757)

[7.2.1 Uploading manually 17](#_Toc415563758)

[7.2.2 Uploading via the NiDBUploader 17](#_Toc415563759)

[7.3 Other Data 18](#_Toc415563760)

[7.4 NiDBUploader 18](#_Toc415563761)

[7.4.1 Downloading and starting the uploader 18](#_Toc415563762)

[7.4.2 Using the NiDBUploader 18](#_Toc415563763)

[7.5 Checking Imported Data 20](#_Toc415563764)

[7.5.1 Checking archiving status 20](#_Toc415563765)

[7.5.2 Matching IDs 20](#_Toc415563766)

[8 Searching for Data 20](#_Toc415563767)

[8.1 Verifying uploaded data 21](#_Toc415563768)

[8.2 Searching and Exporting Data Using Search 22](#_Toc415563769)

[9 Groups 27](#_Toc415563770)

[9.1 Managing Groups 27](#_Toc415563771)

[10 Pipelines 28](#_Toc415563772)

[10.1 Definitions 29](#_Toc415563773)

[10.2 Viewing Pipelines 29](#_Toc415563774)

[10.3 Create a Pipeline 30](#_Toc415563775)

[10.3.1 New Pipeline 30](#_Toc415563776)

[10.3.2 Copy Existing pipeline 35](#_Toc415563777)

[10.4 Modify a pipeline 35](#_Toc415563778)

[10.5 Analyses 36](#_Toc415563779)

[10.5.1 Understanding the Analysis Process 36](#_Toc415563780)

[10.5.2 Data Selection 37](#_Toc415563781)

[10.5.3 Analysis Directory Structure 37](#_Toc415563782)

[10.5.4 Analysis Listing 37](#_Toc415563783)

[10.6 Example Pipelines 39](#_Toc415563784)

[10.6.1 Freesurfer 39](#_Toc415563785)

[10.6.2 FSL fMRI pipelines 39](#_Toc415563786)

[10.7 Checking Pipeline Logs 40](#_Toc415563787)

[10.8 Troubleshooting 40](#_Toc415563788)

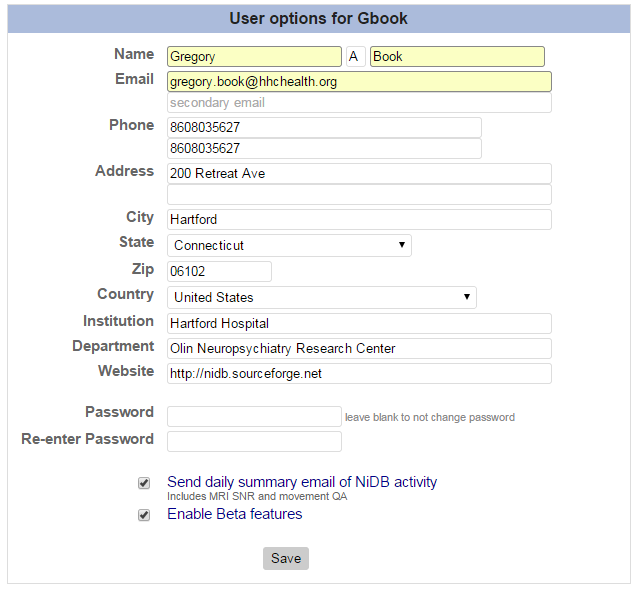
[10.8.1 Why Didn’t the Pipeline Do Anything? 40](#_Toc415563789)

[10.8.2 Why did only some studies start processing? 41](#_Toc415563790)

# Overview & Database Access

To access this database, you must create an account on the server. Visit the website and click the Signup link on the main page. After filling out your information, you will be asked to verify your email address and then your account will be created. Your username is the email address you specified during signup. The first time you sign in, you will be asked to join an instance. To join an instance, mouseover your username/email address on the menu at the top of the page, click **My Account**. At the bottom of the page is a list of available instances. Click the Join Request link for your instance(s). The system administrator will then add you to the requested instance(s).

If you forgot your password, click the reset password link on the login page and it will send you a new password. You can then change your password by clicking your username and the **My Account** link. If you are an administrator, do not create user accounts yourself. Instead, allow the user to create their own account.



# Projects, Instances, and Sites

A project is usually considered to be a collection of subjects that enrolled in a single IRB approved research project. Most projects are collected at a single site, but a single project may collect subjects from multiple sites. An instance will contain multiple related projects. For multi-site collaborations, there is often one instance which contains several projects, with one project assigned to each site. Each of these projects, instances, and sites are assigned a UID (ex. P1234ABCD) and an ID (ex. 1). These IDs and UIDs will be used when importing data to ensure the data is imported into the correct project and instance.

# Subject Management

## Hierarchy

NiDB follows a hierarchy in which subjects are the unique objects in the database. Subjects can be enrolled in multiple projects, and within those enrollments can have multiple imaging sessions (studies). Each of those studies can have multiple series (runs, images, etc). This means that all data that is collected from a particular subject should always be associated with that subject. Permissions on who can view or edit the subject data are controlled on the project/enrollment level. This means that it is very important that original subject IDs, dates of birth, study dates should be intact prior to importing of data.

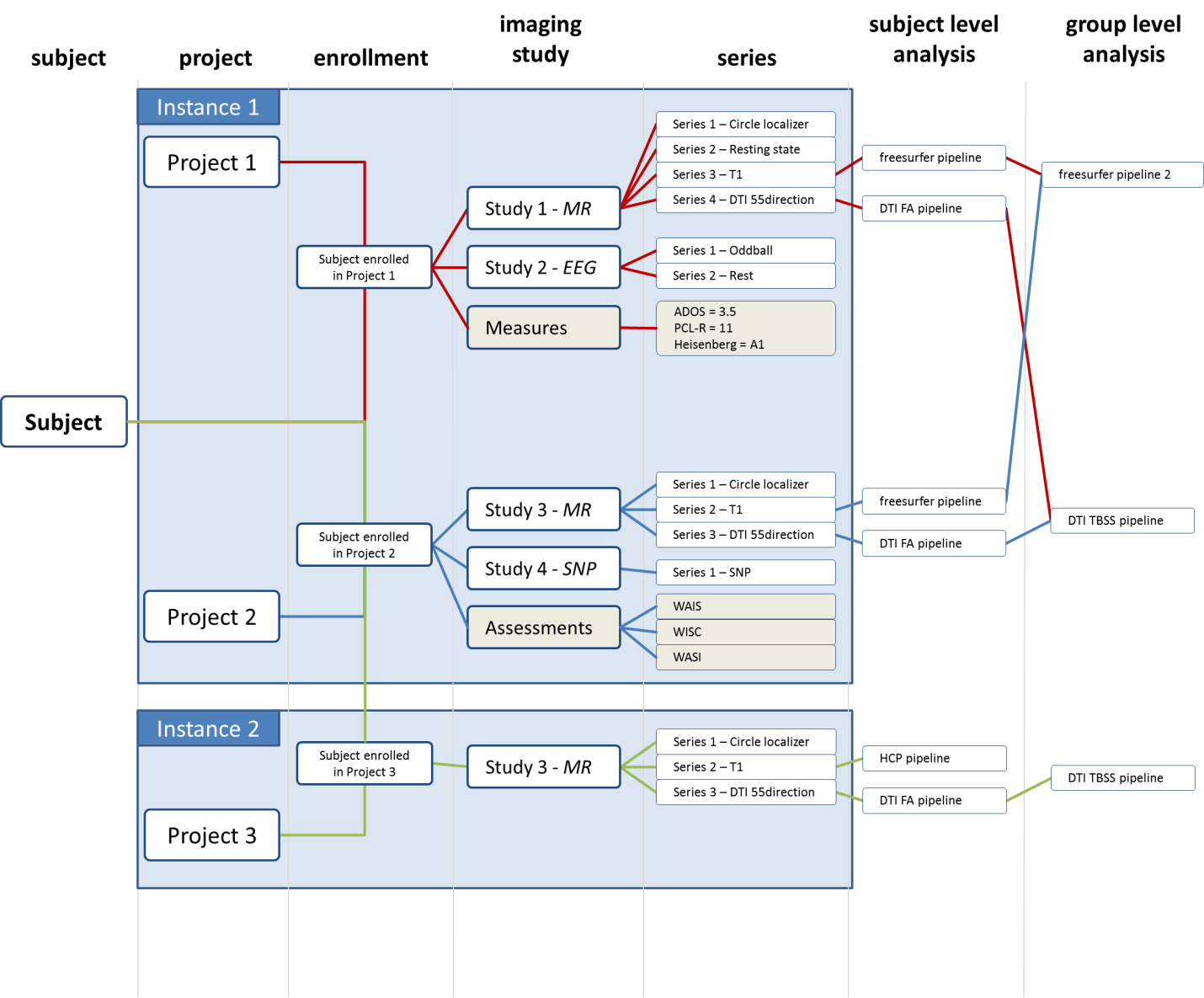


Figure - NiDB hierarchy

## Subject Creation & Enrollment

There are three ways to create subjects in NiDB:

1. DICOM receiver
2. Create the subject manually
3. Let the NiDBUploader create the subject (DICOM only)

### Import via DICOM Receiver

Once your MR (or other) scanner is setup to send DICOM images to NiDB automatically, these images will be received and archived automatically. Series, imaging studies, subjects, and enrollment will be matched to existing objects in the database. Matching is done in the following order:

**Subject**

1. Match DICOM PatientID (0010,0020) to one of the following:
   1. UID or Alternate UID
   2. SHA1(UID) or SHA1(Alternate UID)
2. Match DICOM PatientName (0010,0010), PatientBirthDate (0010,0030), and PatientSex (0010,0040) to existing subjects in NiDB
3. If no Subject match, subject is created

**Project Enrollment**

1. Match the *string in parentheses* in the DICOM StudyDescription (0008,1030) to an existing project number. For example “Project XYZ (12345)” will match to project “Project xyz” with project number 12345
2. If no Project match, subject enrolled in Generic Project (999999).

**Imaging Study**

1. Match DICOM fields StudyDate (0008,0020), StudyTime (0008,0030) +/- 30 seconds an NiDB Study datetime. Also matches with subject UID and project enrollment.
2. Match by DICOM AccessionNumber (0008,0050) to NiDB study number.
3. If no imaging study match, a study is created

**Imaging Series**

1. Match by DICOM field Modality (0008,0060) and SeriesNumber (0020,0011) to NiDB modality and series number.
2. If match, series information (file counts, sizes) is updated. Existing QC is removed
3. If no imaging series match, series is created

### Subject Creation via Imported MRI data

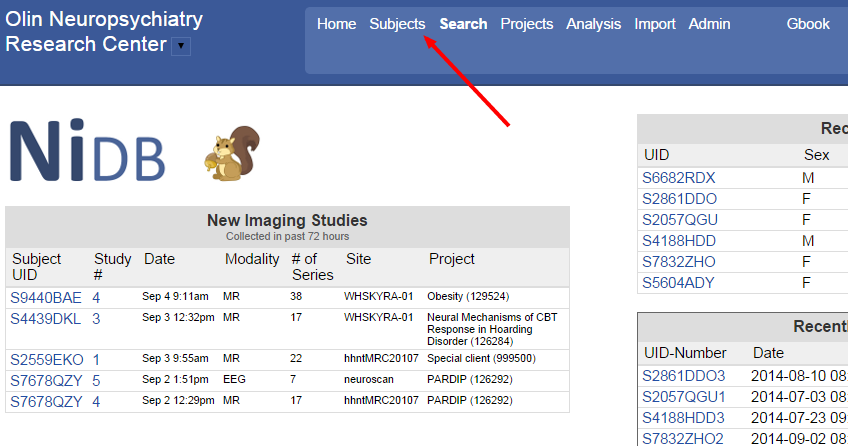
Follow the instructions in the MRI Data 🡪 Importing section of this manual to import your MRI data. Once imported, subject UIDs will be automatically generated for the data, and the imaging sessions and series will be automatically separated into their associated studies. The subject will also be automatically enrolled in the appropriate project. When a subject is imported into NiDB and encrypted, the following changes occur to data fields:

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Original data | Change | New data (encrypted) |
| Name | Bob Smith | SHA1(“smith^bob”) | d6ceae47f32be9fe8f4bcec1fc2e5cf36191a38d |
| DOB | 1973-01-03 | Leave only year | 1973-00-00 |
| Subject ID | 20402030 | SHA1(“20402030”) | ecf68dc6d32a6b206c05cb1ac71d507c82353609 |
| UID |  |  | S1234ABC |

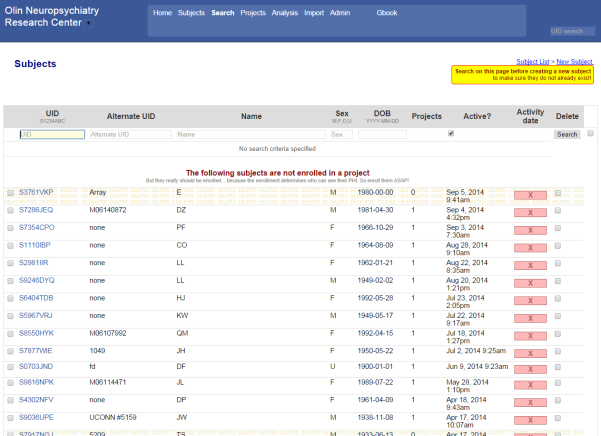
When the subject is created, it will generate a new UID which is unique on NiDB and should be the ID used to identify a subject. This will also imply a mapping between your original ID (20402030) and the new UID (S1234ABC). Subjects can have any number of alternate IDs, which should be encoded like above. This means that because you know the original ID and the mapping, you will be able to identify your subjects on NiDB while other sites will not be able to. Data are archived in the same manner as described in the “Import via DICOM receiver” section.

### Manual Subject Creation

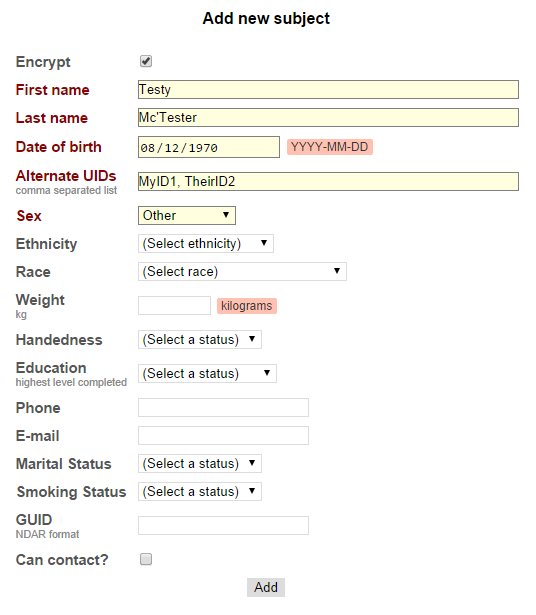
Click on the Subjects menu item.



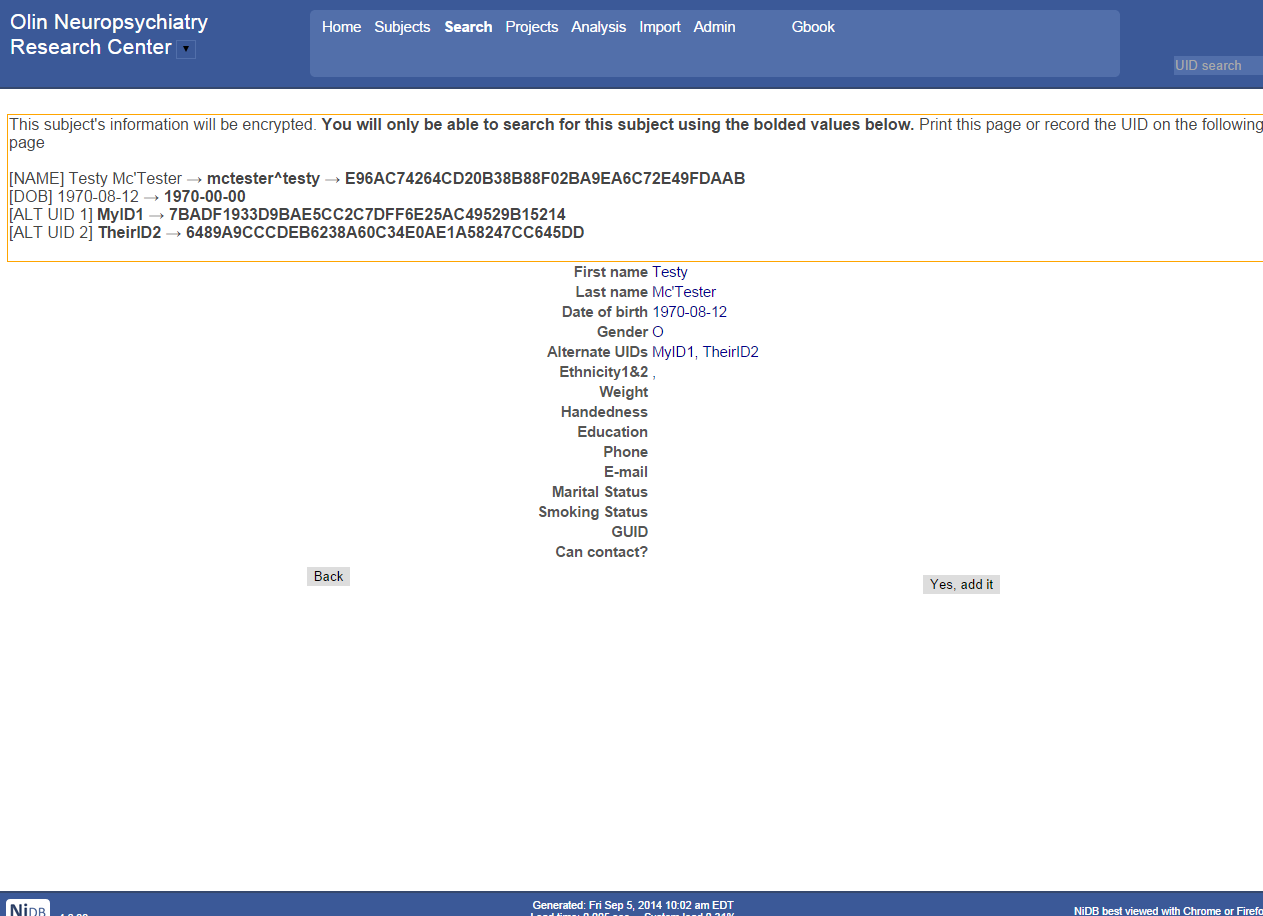
In the upper-right, click the New Subject link. Mousing-over the link will remind you to search for your subject before creating a new one.



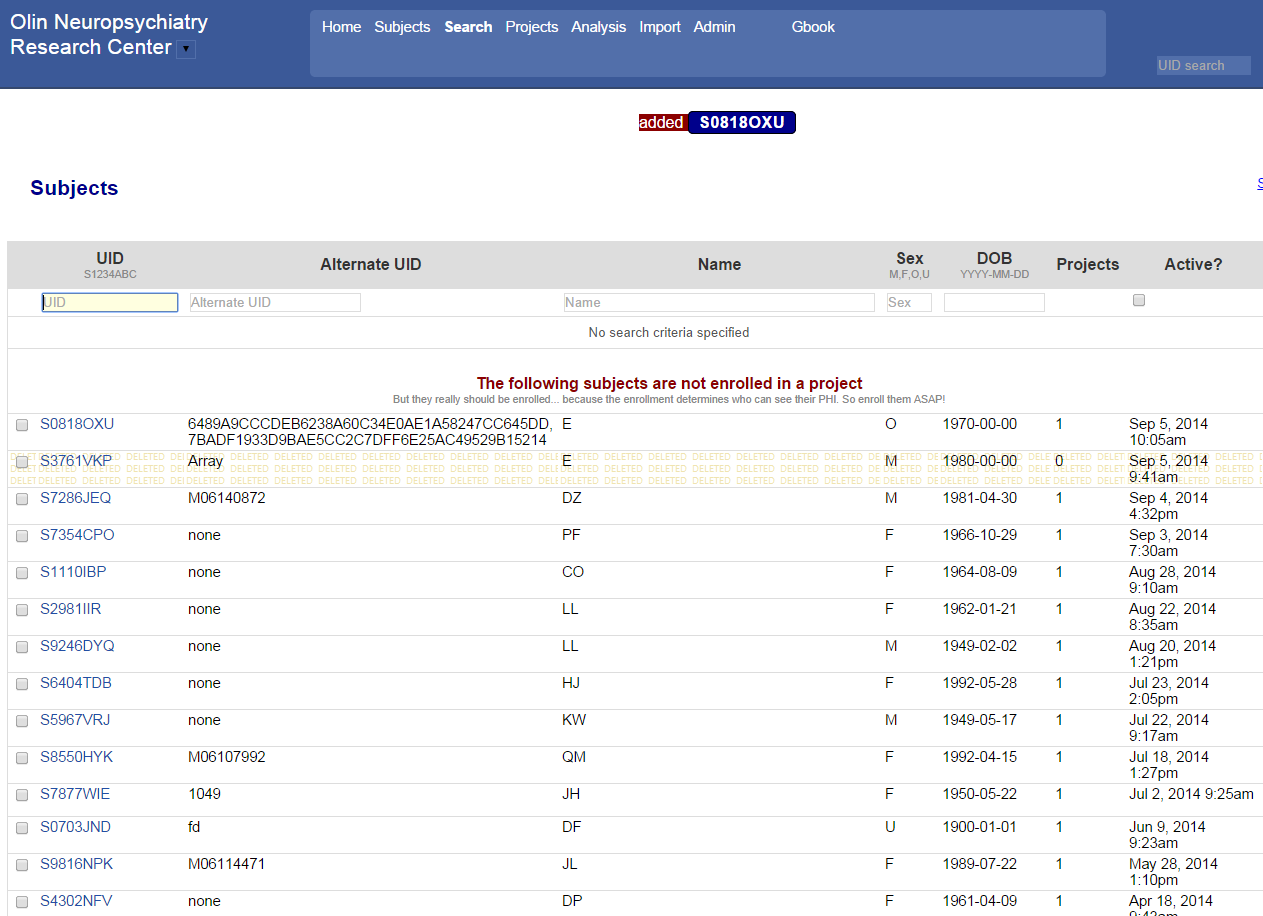
You should then see the new subject form.



If you are importing data into a project into which the subject ID is considered PHI, such as in a retrospective collection of data, then you will want to encrypt the ID. Check the Encrypt box. If you use the subject’s name as a way to identify the subject at your site, then enter their name. First name should contain any middle initial(s) or middle name(s). Enter at least the subject’s birth year so that age-at-scan can be calculated later on. You must enter at least one alternate UID. The alternate UIDs are IDs that you use at your site to identify subjects, and can be used to identify subjects on the NiDB server. The alternate UID should be unique. If your site assigns simple incremental IDs to your subjects like 1,2,3… etc, then add a known string of characters to the beginning of your ID. The server is shared, so it will be easier to find your subjects if you have a less common ID than “1”. Fill out any other information you need, obviously avoiding PHI. Click Add.

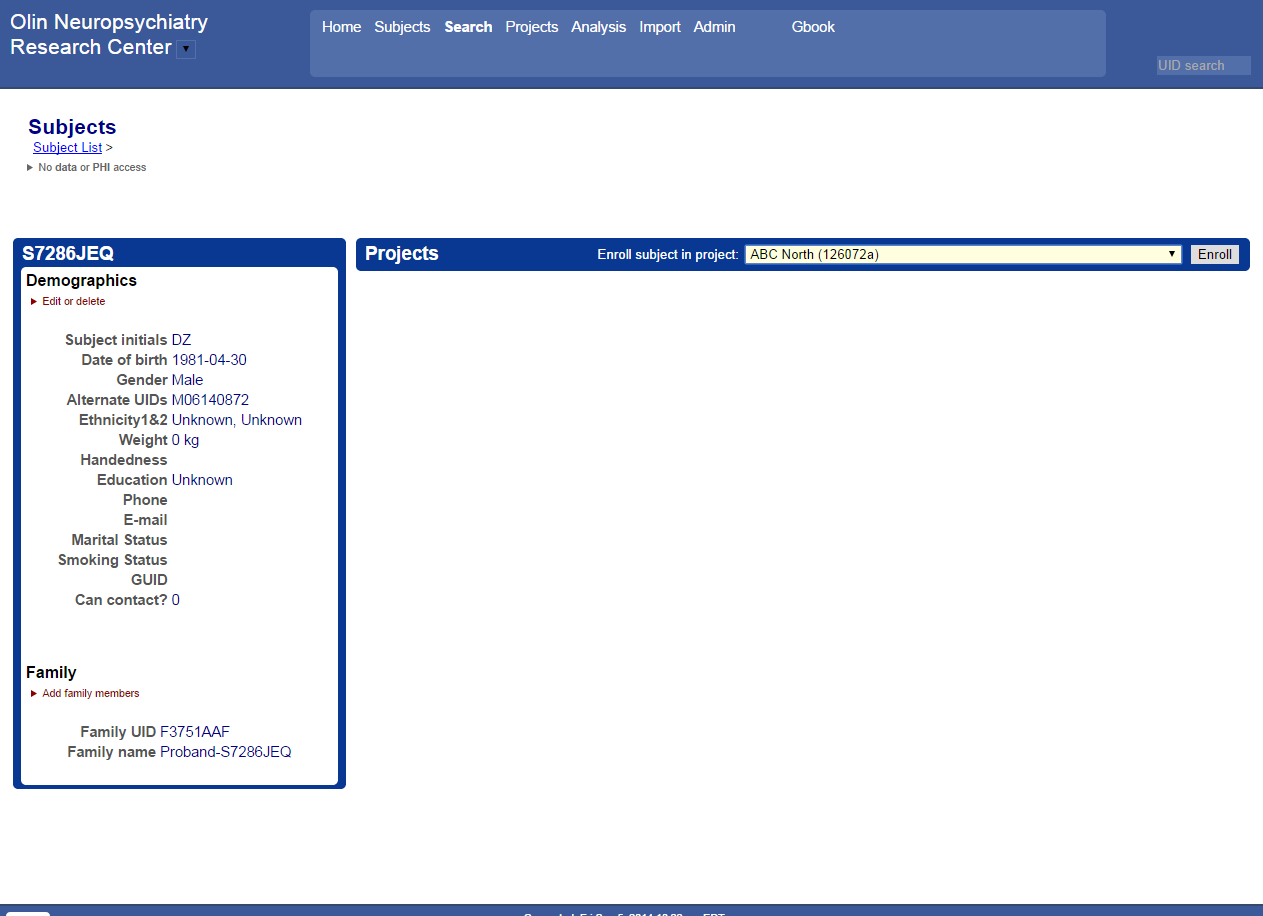


This page isn’t pretty, but it is important because it shows the ID mapping between your site’s ID and the how you would search for that ID on NiDB. It will ask if you are sure the subject is not a duplicate. Once you create the subject, it will give the UID and the subject will be displayed as not being enrolled in a project. **The new UID for this subject is at the top of the page.** Click on the subject ID on the left and enroll the subject in a project. For our example, the new UID is S0818OXU. Remember, UIDs are S + (4 digits) + (3 letters).

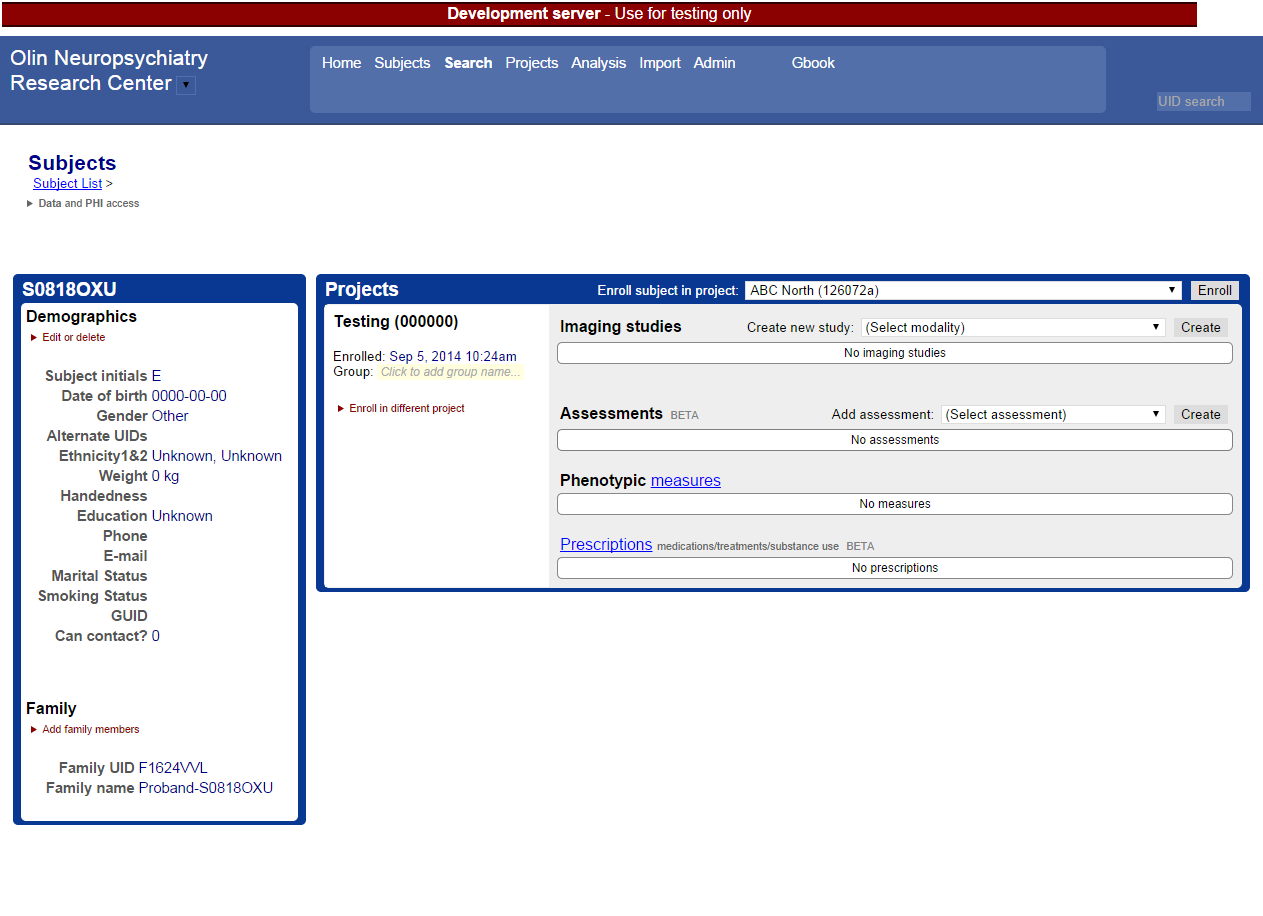


Record this new UID at your site, along with the original ID so you have a proper mapping between the IDs. Any time you reference this subject on NiDB, you will need the UID.

The new subject you created will not be enrolled in a project, and you will need to enroll them. Click the UID on the left side. This will display the subject’s page. On the right side of the page, select a project, and click **Enroll**.



The subject will then be enrolled in a project.



## Modifying Existing Subjects

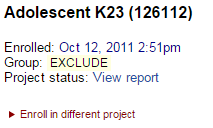
To edit subject demographics, navigate to the subject page. Click Edit or delete, then Edit Demographics. Modify the information you need to change, and click update.

To enroll a subject in a new project, follow the directions in the previous section for project enrollment.

## Modifying Subject Enrollment

### Subject group

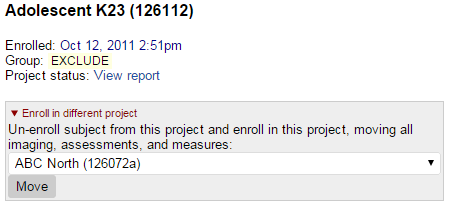
Subjects in research projects are often classified into groups, such as a patient, control, etc. But a subject enrolled in multiple projects may not necessarily be in the same group in all projects. Different inclusion criteria may determine a different group for each project enrollment of the same subject. To specify a group, navigate to the subject page. Under each enrollment, there is a group specified, under the Group: section.



Type in the appropriate group name in the yellow box and hit enter. Group names are not predefined, which means any typos will be carried forward and effectively create new groups within your project, so check group name spelling carefully.

### Change Project Enrollment

Sometimes a subject will be entered into the wrong project. It is possible to un-enroll a subject from a project and enroll them in a different project, while keeping also moving all of the imaging studies to the new enrollment. On the subjects page, click Enroll in different project. Select the new project, and click Move. The subject will be un-enrolled from the old project and enrolled in the new project, and all imaging studies will be moved to the new project. Remember, nothing is deleted in NiDB, so the un-enrolled project will remain visible on the subject’s page.



## Deleting Subjects

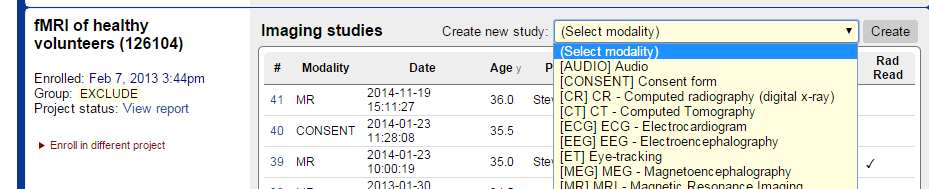
Nothing in NiDB is ever really deleted. When you delete a subject, the data will not be removed, the subject will simply be marked inactive. To delete a subject, navigate to the subject’s page. Click Edit or delete, then click Delete.

# Imaging Studies

All data stored in NiDB, apart from subject demographics, is stored in an imaging study. The imaging study is defined as the period of time during the subject performed a discrete set of experiments on a single piece of equipment. An example is an MRI session: a subject goes into the bore of the MRI, several sequences are run, and they leave the MRI. That would be considered an imaging study. The term ‘study’ comes from the radiology field and while a study may refer to an IRB approved project, or a published analysis, study refers only to an imaging session in the NiDB documentation. A study may also be called an ‘imaging study’, ‘imaging session’, or ‘scan session’.

## Creating Imaging Studies

Most imaging studies from MR, EEG, and other data that is imported either automatically through the DICOM receiver, or manually through the NiDBUploader, will be created and matched automatically when received. Some studies such as tasks, consents, or videos would be uploaded manually.

To create an imaging study, navigate to the subject page. Under the project you want to create to create the study, select the modality and click Create.

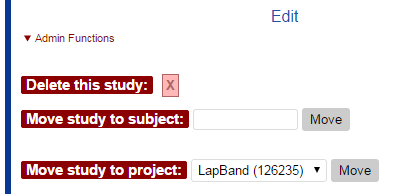
The study will be created and a new study number assigned. Click the View Study link to go to the newly created study, which will be blank.



See the section on creating and modifying imaging series to add data.

## Modifying Imaging Studies

To modify any imaging study, first navigate to the subject page, and click on the imaging study you want to modify. In the “Study Information” box, click the Edit link at the bottom. You will be able to edit key study information on this page. Click **Update** to update the information. If you have administrative privileges, you can more options by clicking Admin Functions. To delete the study, click the **X**. To move the study to another subject (while maintaining the enrollment in the same project for the new subject), enter the full UID and click **Move**. To move the study to an already enrolled project for the current subject, select the project and click **Move**.



# Imaging Series

Each study contains a number of imaging series. Series are numbered in order, but not necessarily sequentially. Some MR studies may have series 1,2,3,4,100,101, etc.

## Importing Series

See the section on [Importing Data](#_Importing_Data) for instructions on how to import series manually or automatically.

## Modality Specific Series

Each modality will contain slightly different data, but there are two main types: 1) series originating from DICOM (CT, MR, etc) 2) all others (EEG, ET, VIDEO, etc)

### DICOM originating series

MR is the most common format, and so we will use that as an example.

#### Uploading behavioral data

Next to each series in a DICOM originating series is a red Upload box. You can click this box to bring up an upload dialog, in which you would select the files you want to add as behavioral data for that series. You can also drag&drop files within Windows explorer. They will be automatically uploaded. Refresh the webpage to see the number of files and total size of the attached behavioral data.

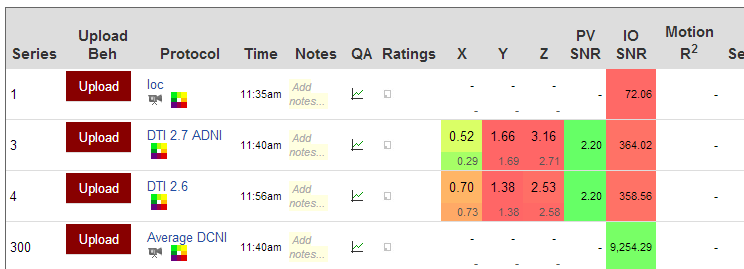




Behavioral data is almost always only associated with a functional MRI run, so if a functional run does not contain behavioral data, it may be marked as red. Click the number in the beh column to view the files, or click the download icon to download the beh files. When you click the beh files, you will see a page listing the files. Here you can view or delete the files.

#### MR Quality Control Measures

Several quality control measures are calculated for MRI data. These are visible on the study page under the X, Y, Z, PV SNR, and IO SNR columns. You can also click the icon under the QA column to get more information and images.



**Timeseries Motion Estimation -** If the MR series is 4D, timeseries motion estimation will be performed. Using FSL, this will align all volumes to the middle volume. The x, y, z, and Pitch, roll, yaw values reported on the webpage are the total displacement of the volume at that timepoint. For example, if a patient moves -0.5mm in the X direction and 1.5mm in the positive direction, the maximum displacement reported will be 2.0mm. The values are color coded according to the voxel size in the displacement dimension: the scale is from green to yellow to red, where a displacement of 0 is green and a displacement equal to or greater than the voxel size is red. Velocity is also calculated, which is the derivative of the displacement, and is color coded the same way. Velocity may be more useful to detect motion spikes rather than slow drift.

**Timeseries per-voxel (PV) SNR -** For 4D volumes, PV SNR is also calculated. For voxels only within the brain (from a brain extracted mask), the mean value for each voxel is divided by the same voxel’s standard deviation, and then average ratio of all in-brain voxels is calculated to produce the PV SNR.

**Volume inside/outside (IO) SNR -** IO SNR, performed on 3D volumes or the first volume of a 4D timeseries, compares the corners of the volume to the center of a brain extracted region of the volume, producing the signal (brain) to noise (corners) ratio.

**Volume Motion Detection -** For 3D volumes, this algorithm will convert each slice of the volume to a 2d FFT, then compute the radial average of the 2D FFT. It will then find the linear regressor (R2) for each slice and average them together. A more negative R value indicates less high frequency signal and therefor more motion. Like SNR, this value is only valid when compared against scans from the same subject and (generally) the same scan session.

#### Other DICOM specific options

* Click the Protocol Name to view a subset of the DICOM header. Click the π symbol at the bottom to view the full DICOM header.
* Mouseover the http://ado2:8080/images/preview.gif icon to view a thumbail of the series
* Mouseover the http://ado2:8080/images/movie.png icon to view an animated gif of the series
* Click the http://ado2:8080/images/colors.png icon to view a 3D rendering of the series in the browser
* Click the Add notes… input box to add notes to the series.
* Click the http://ado2:8080/images/chart.gif icon to view QC information. Much more information may be available in there
* Click the http://ado2:8080/images/rating.png icon to view or edit subjective ratings
* Click the http://ado2:8080/images/download16.png icon to download a series as a .zip file

### All other series

Follow the [instructions](#_EEG) for importing data to create the non-DICOM series. To edit existing series, click on the Protocol, Date, or Notes and edit the item in place. Click the # of files to view the files and delete as necessary. Click the download icon http://ado2:8080/images/download16.png to download the series as a .zip file.

# Importing Data

## MRI

There are three methods available to import MRI data into NiDB:

1. Via the uploader (recommended)
2. Via the website
3. DICOM receiver (not available on the public server)

If your IRB prevents you from sharing PHI (which is the most likely scenario), then you will want to use the uploader. If you import via the website, a copy of the original un-anonymized data may be left on the server, which kind of freaks out most IRBs. The uploader is preferable because it runs on your server and anonymizes data before sending it to NiDB, so no PHI ever reaches the server. There is also a 1GB total size limit when uploading via the website, vs an unlimited size via the uploader.

The uploader is available for Linux and Windows, which can be downloaded from the Import page within NiDB. The uploader is only available for CentOS 6, CentOS 7, Fedora 16, and Windows 7.

### Uploading MRI Data through the NiDBUploader

DICOM is the only format which contains identifiers and meta-data in the header, and is preferred because more information is available for analysis upon data export. However, it is possible to import DICOM and any other data format through the NiDBUploader.

#### DICOM and .par/.rec format

If you are uploading DICOM or .par/.rec, you do not need to rename your files. However, the following DICOM fields must be populated at a minimum for proper archiving:

* PatientID (0010:0020)
* StudyDate (0008:0020) and StudyTime (0008:0030)
* SeriesDate (0008:0021) and SeriesTime (0008:0031)
* StationName (0008:1010)

The following .par/.rec fields must also be populated, at a minimum

* Patient name
* Examination date/time

#### All other data formats

If your MR data is not DICOM or .par/.rec, you must rename your files in the following format:

ID\_YYYYMMDDHHMISS\_task\_operator\_seriesnum\_filenum.ext

* ID – the NiDB UID or Alternate UID
* YYYYMMDDHHMISS – Date/time that the imaging session began. All fields zero-padded: 4-digit year, 2-digit month, 2-digit day, 2-digit hour, 2-digit minute, and 2-digit seconds
* task – the task or protocol name
* operator – the person who administered the imaging session
* seriesnum – the order in which the tasks were administered
* filenum – optional, and is the number of the file within the series
* .ext – the file extension

Table - Example non-DICOM, non-.par/.rec MR filename formats for uploading via the NiDBUploader

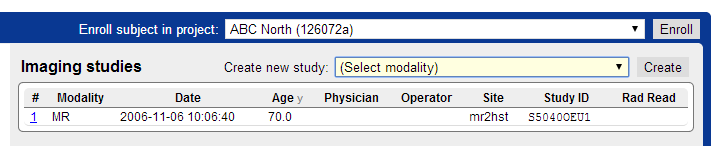
|  |  |
| --- | --- |
|  | Example filenames |
| Structural MR (Nifti) | S1234ABC\_20150205123456\_MPRAGE\_Greg\_2\_0.nii.gz |
| Functional MR (Nifti) | S1234ABC\_20150205123456\_rest\_Greg\_3\_0.nii.gz  S1234ABC\_20150205123456\_rest\_Greg\_3\_1.nii.gz  S1234ABC\_20150205123456\_rest\_Greg\_3\_2.nii.gz  ...  S1234ABC\_20150205123456\_rest\_Greg\_3\_153.nii.gz  S1234ABC\_20150205123456\_rest\_Greg\_3\_154.nii.gz  S1234ABC\_20150205123456\_rest\_Greg\_3\_155.nii.gz |
| Functional MR (Analyze) | S1234ABC\_20150205123456\_rest\_Greg\_3\_0.img  S1234ABC\_20150205123456\_rest\_Greg\_3\_0.hdr  S1234ABC\_20150205123456\_rest\_Greg\_3\_1.img  S1234ABC\_20150205123456\_rest\_Greg\_3\_1.hdr  ...  S1234ABC\_20150205123456\_rest\_Greg\_3\_154.img  S1234ABC\_20150205123456\_rest\_Greg\_3\_154.hdr  S1234ABC\_20150205123456\_rest\_Greg\_3\_155.img  S1234ABC\_20150205123456\_rest\_Greg\_3\_155.hdr |

## EEG

EEG files do not generally contain any metadata in their headers that helps to identify patient or imaging study association, and so need to be uploaded manually or the file names must be formatted.

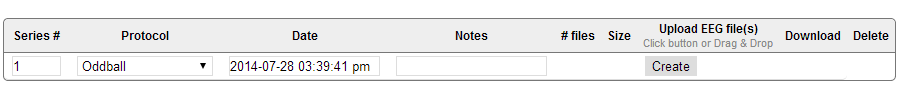
### Uploading manually

To upload EEG data manually, you will need to create a subject (see the create subject section) and a new study. Follow the subject creation section above to create a new subject. To create an imaging study, select the modality, in this case EEG, from the modality list. Then click **Create**.



This will create a study. Click View Study to view and edit it. Click Edit under the Study Information box to edit the study details.

To create series, enter information in the series boxes, and click Create.



You can then upload your files by dragging and dropping onto the series. You can also click on the number of files to edit the files that are part of the series.

### Uploading via the NiDBUploader

Files can uploaded via the uploader, and can save time if you have a lot of data. Subjects must be created and enrolled prior to uploading EEG data. EEG files with the extensions (.cnt .dat .3dd) must be named in the following convention

NiDBUID\_YYYYMMDDHHMISS\_task\_operator\_seriesnum\_filenum.cnt

* ID – the NiDB UID or Alternate UID
* YYYYMMDDHHMISS – Date/time that the imaging session began. All fields zero-padded: 4-digit year, 2-digit month, 2-digit day, 2-digit hour, 2-digit minute, and 2-digit seconds
* task – the task or protocol name
* operator – the person who administered the imaging session
* seriesnum – the order in which the tasks were administered
* filenum – optional, and is the number of the file within the series
* .ext – the file extension

Table - Example file format of EEG data

|  |  |
| --- | --- |
|  | Example filenames |
| Neuroscan | S1234ABC\_20150205123456\_oddball\_Bob\_1\_0.cnt |
| Polhemus | S1234ABC\_20150205123456\_digitize\_Bob\_2\_0.dat  S1234ABC\_20150205123456\_digitize\_Bob\_2\_1.3dd |

## Other Data

The process of importing any data that is not MRI or EEG data is identical to the process for importing EEG data, the only difference is the modality selected and file extensions.

## NiDBUploader

### Downloading and starting the uploader

Download the NiDBUploader from the Import page on NiDB. Choose your operating system and download the file. It may need to be unzipped.

On Linux use

tar –xzvf NiDBUploader.tar.gz

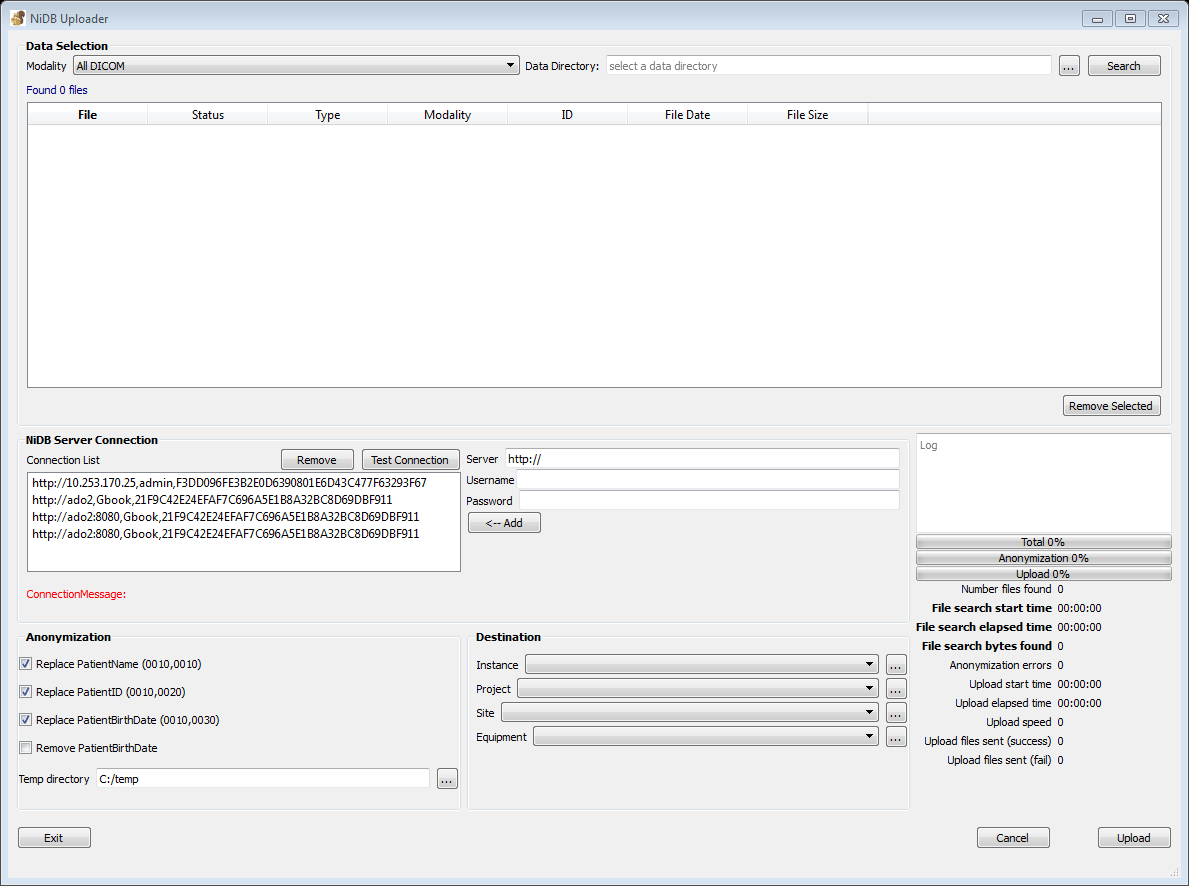
Start the program by typing

> cd NiDBUploader-*directory*  
> ./NiDBUploader.sh

On Windows, unzip using 7-zip or another zip program. Start the program by double clicking the NiDBUploader.exe

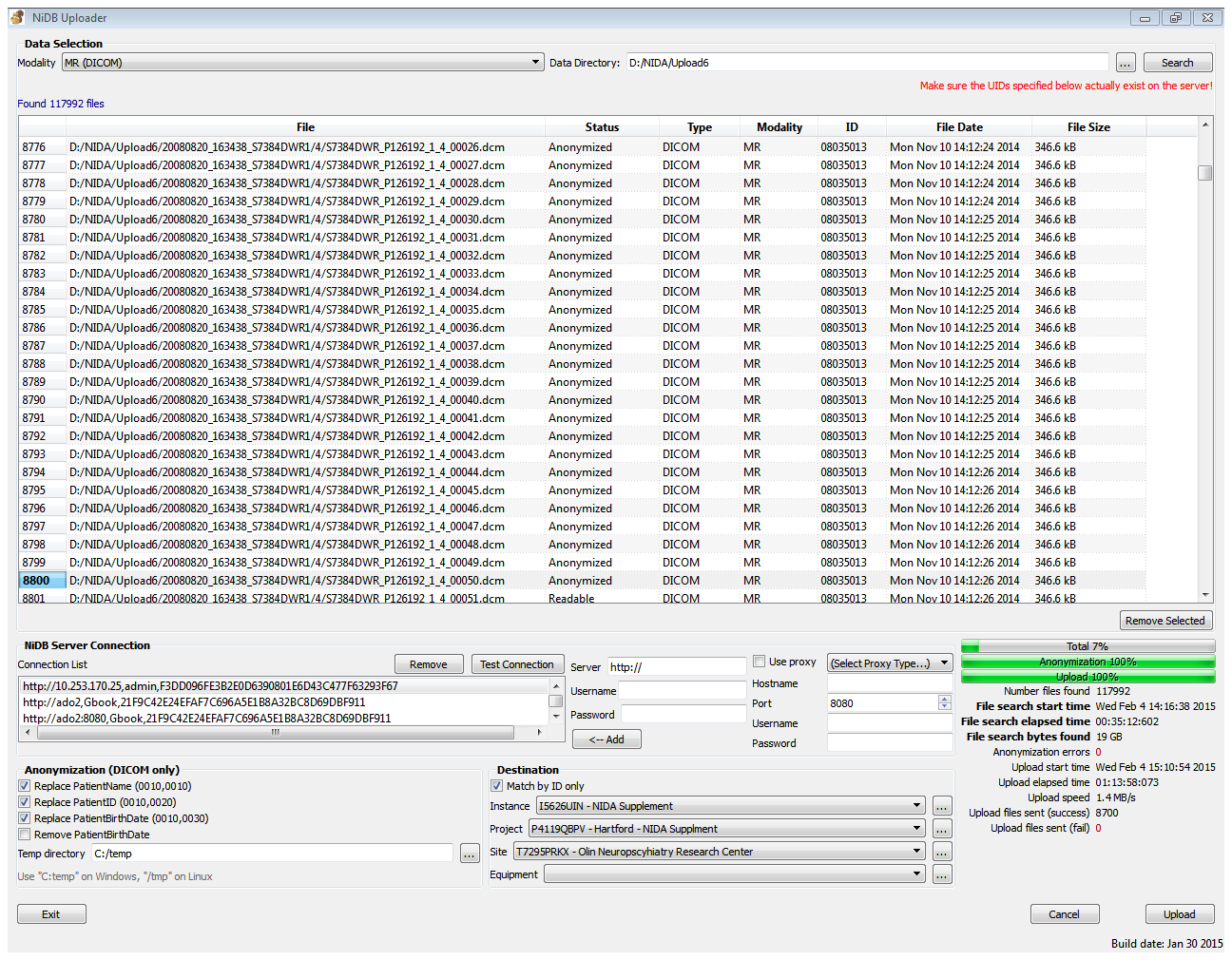
### Using the NiDBUploader

When you first open the uploader, you will see a dialog like this



You can upload data by doing the following, in order (see next screenshot)

1. Select the modality of data (MR, EEG, etc)
2. Select the directory containing all of the data. Don’t worry about sub directories. DICOM data is identified by the header, and EEG (and all other modalities) by the information in the filename.
3. Click Search. This may take a long time if you have a lot of data. MR/DICOM data will take even longer because it must open and read each file’s header. The uploader works ok when uploading up to 100,000 files. If you have more than 100,000 files in a single parent directory, it may be wise to put the files in more than one parent directory so you are not attempting to upload them all at the same time.
4. Enter the NiDB login information (and proxy if necessary), and click **🡨 Add**.
5. You can click on the new connection in the connection list and click **Test Connection**
6. If you have DICOM data, you can select the type of anonymization you want to perform before the data is uploaded. Make sure you set a valid temp directory!
7. Select the Instance, Project, Site, and Equipment from the Destination by clicking the … button
8. When ready, click **Upload**
9. The progress will be displayed. The uploader will upload files in blocks of 100. Any errors will be reported, but if an error occurs, it will not retry the upload.



**9**

**4**

**8**

**7**

**6**

**5**

**3**

**2**

**1**

## Checking Imported Data

### Checking archiving status

Click the **Import** tab, then click **View import logs**. This will display a list of imports from the previous 14 days. Data is often imported in blocks of varying size. Click the **summary** link to view an overall status of each import block. This will display a list of the blocks received, their status toward archiving. Click the **detail** link to view details about the data that was imported. This will display things like UID, information about the data, and whether subjects, studies, etc were created for the first time or simply updated.

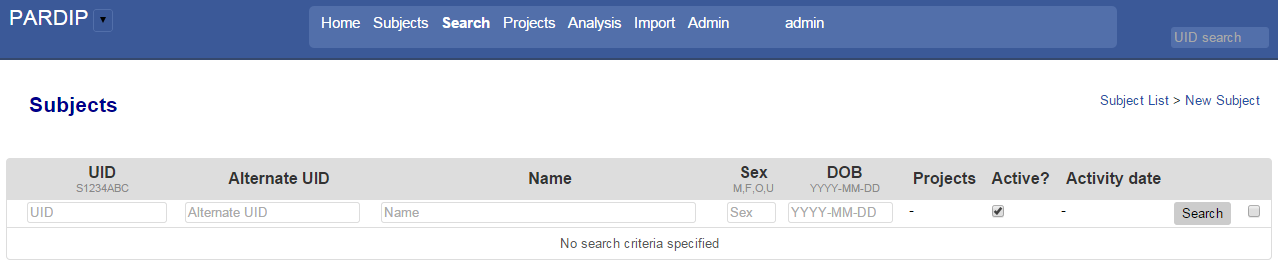
### Matching IDs

Click the **Import** tab, then click **ID mapper**. You can enter any size list of IDs. Click **Map IDs**. You’ll see a list of the foreign IDs, and the new UIDs. If the foreign ID matched to the study ID, it will be marked in red.

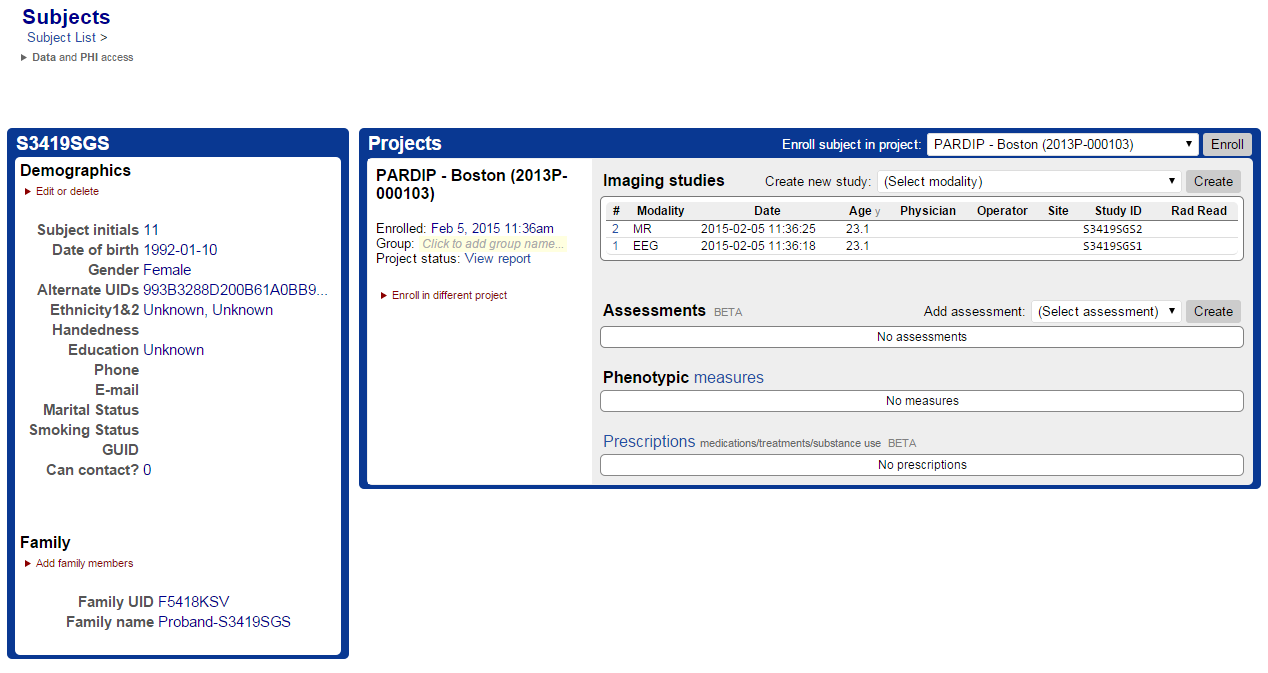
# Searching for Data

You can search for data that anyone has uploaded, and this is useful when it comes time for analysis, but you may also just want to verify that your data uploaded correctly.

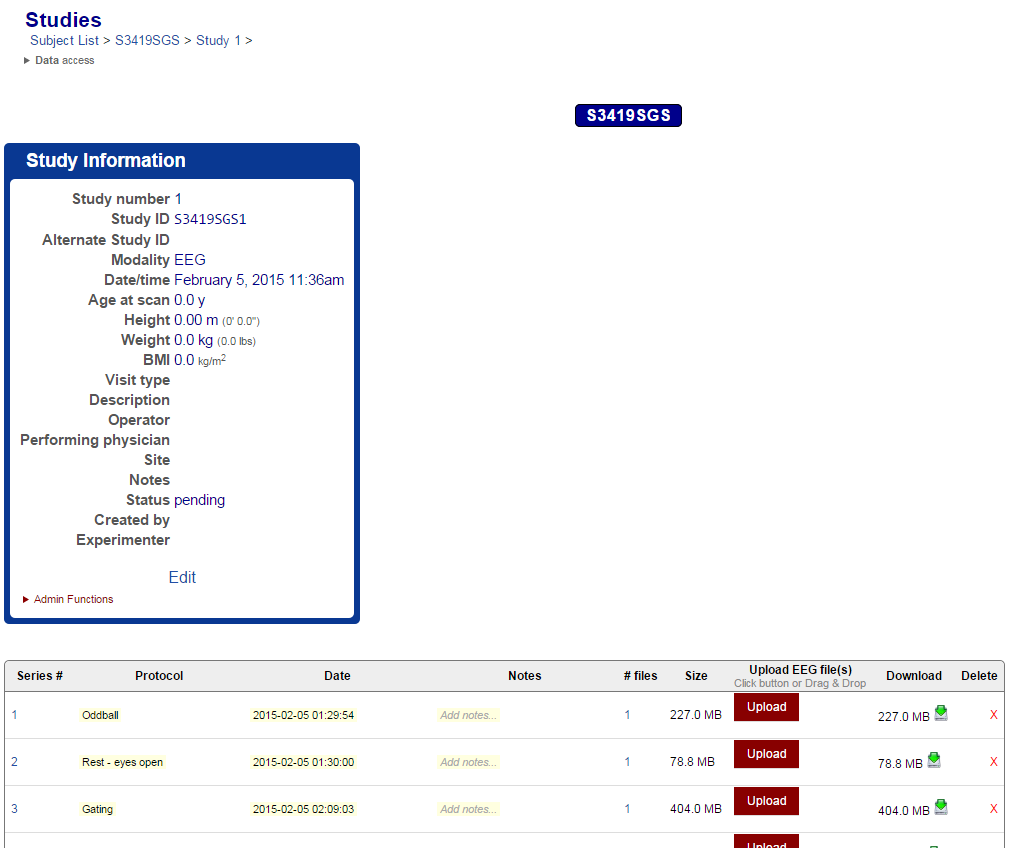
## Verifying uploaded data

On the menu, click the Subjects tab, you will see the page below. 

To find a subject by ID, enter their NiDB UiD in the UID field, or enter the alternate ID in the Alternate UID field and click Search. You may also search for an alternate ID in the Name field as some file formats store ID in the name field in their meta-data. Alternate (original) UIDs and Name may be encrypted, so while you may search for 1234ABC, it may return an alternate UID of 5bec252100af185283441c71606e1d1dde7e1b1e. If your ID is found, click on it to view the imaging studies associated with the subject.



Clicking on imaging study 1 (EEG) in the above screenshot will show



The same method applies to verifying and checking any data uploaded. You can also check the meta-data about the imaging study or subject. **The field ‘age-at-scan’ may be incorrect because the birthdate is removed during anonymization.** Click the Edit link under the study information to edit the age-at-scan.

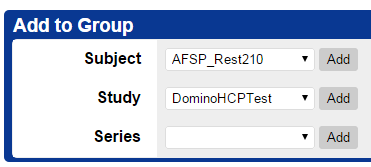
## Searching and Exporting Data Using Search

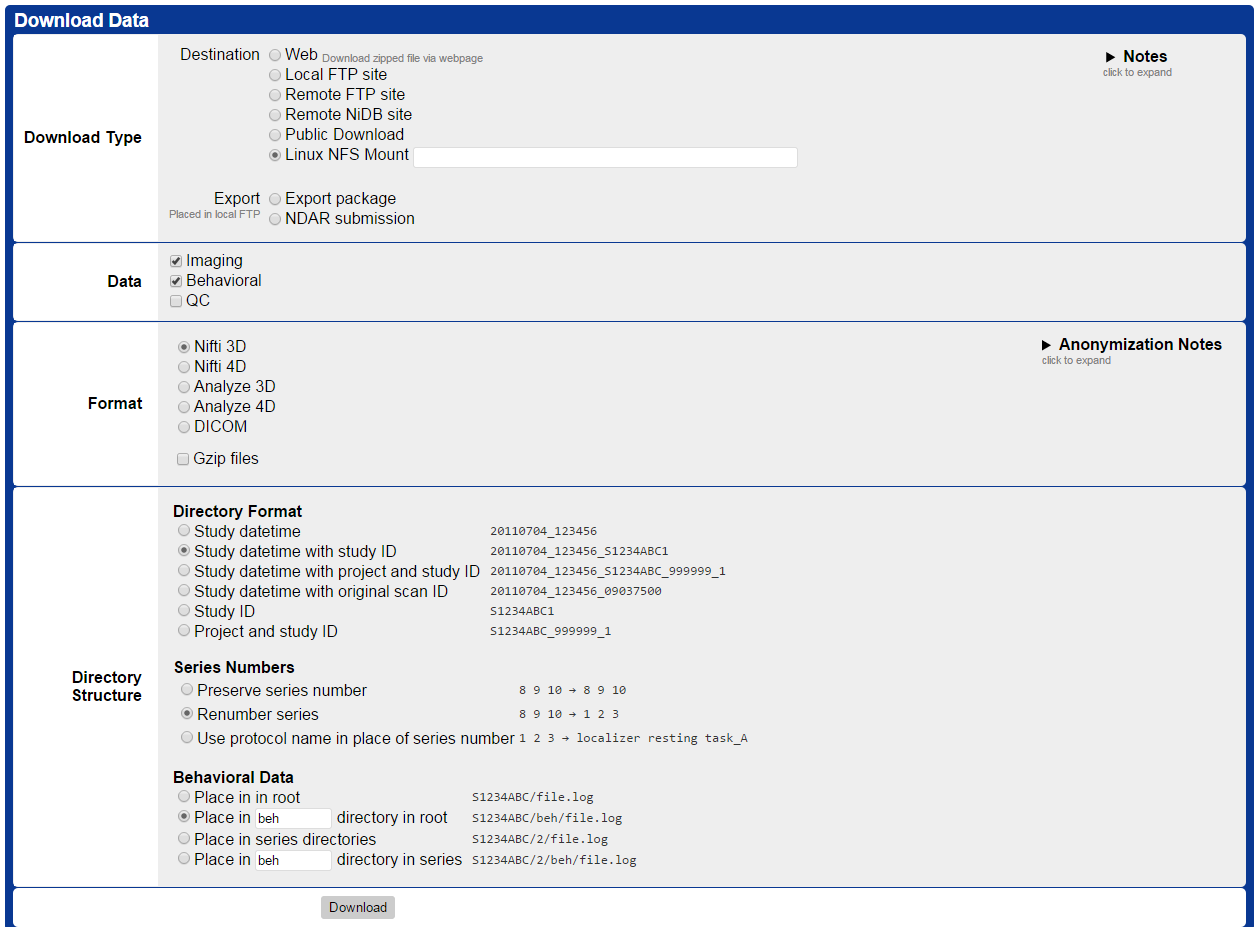
The **Search** tab is a powerful search tool that allows you to search for and download/export data from NiDB. Below is the search page, and following it a description of each search item.



|  |  |  |
| --- | --- | --- |
| Group | Item | Description |
| Subject | UID(s) | The NiDB unique ID (UID). Enter a list of UIDs or single UID. Searching exact UIDs only |
| Alternate UID(s) | Any of the alternate IDs. Enter a list or single ID. Searches exact IDs only. |
| Name |  |
| DOB | Birthdate |
| Gender | Sex |
| Group | The subject level group, specified in the Groups section of NiDB |
| Enrollment | Project | The project |
| Enrollment sub-group | Searches for an exact string matching the enrollment sub-group (things like Patient, Control, etc) |
| Phenotype | Measure Search | Enter a list of measures to search by |
| Measure Columns | Separate from the list of search measures, this is the list that is actually displayed |
| Study | Institution | Originally pulled from the DICOM Institution tag |
| Equipment | Often the hostname of the DICOM sending entity. But can be any other user-defined equipment (Neuroscan1, TMS2, etc) |
| Alternate Scan ID(s) | Each study is assigned an ID. Some institutions assign a new ID for every scan, even if it is the same subject |
| Date | The beginning date/time of the imaging study |
| Modality | The modality |
| Description | Pulled from the DICOM field StudyDescription. Often the field contains the PI or project name |
| Performing Physician | Pulled from the DICOM field of the same name |
| Operator | Pulled from the DICOM field of the same name |
| Visit type | User-defined visit type, defined on the study page. Could be something like visit1, pre, post, etc |
| Group | The study group |
| Series | Protocol | The series name, series description, or protocol name. can be a comma separated list, or a single protocol |
| Use alternate protocol name | Some protocols have common names to consolidate multiple scan names that are actually the same thing. For example, T1w, MPRAGE, Sag T1, etc would all be labeled with alternate protocol name of T1 |
| Sequence | From the DICOM field of the same name |
| Image Type | From the DICOM field of the same name |
| Image Comments | From the DICOM field of the same name |
| TR | Search for the exact TR in milliseconds (MR modality only) |
| Series number | From the DICOM field of the same name |
| Number of files | Total number of files for the series |
| Group | Series group |
| Pipeline | Pipeline | Chose an existing pipeline to view results from |
| Result name | A result name |
| Result unit | Result unit |
| Result type | Value, image, file, or html |
| Result value | Specifies the criteria for the values returned |
| Colorize | For a table of values, colorize each column based on the number in each cell |
| Display correlation matrix | Displays a correlation matrix for a table of values |
| Display result statistics | Displays statistics for each column for a table of values |
| Results | Enrollment List | Displays a list of subjects and all their enrollments |
| Subject List | Displays a list of subjects, their IDs, and their study IDs |
| Group by study | Displays results in a format to download data, grouped by imaging study |
| Series List | Displays the same results the Group by study option, but as one large table |
| Table | Displays the same results as Series List, but with option to download data |
| Spreadsheet | Same result as Table option, but downloaded as a .csv file |
| Pipeline results | Displays pipeline results. Must be selected if any pipeline criteria are specified |
| Pipeline results (.csv) | Same as pipeline results, but downloads them as a .csv |
| Longitudinal | Displays only results where a protocol name was done by a subject more than once. Allows downloading of the data into time slots (time1,time2, etc) |
| Debug | Displays only the SQL used for the query |
| File operations | Admin only. This displays a list of data and a box to change DICOM header values. Really not a good idea to use |
| Audit | Among the search results, it will check each result against the actual files on NiDB and report any discrepancies. Not a good idea to use. Extremely slow. |

For any results that allow you to download data, you’ll see the following Add to Group box at the bottom of the page. Here you can add the selected items to an existing group. Select your group and click **Add**.



You will also see the Download Data box.

Several Download Types are available:

|  |  |
| --- | --- |
| Type | Description |
| Web | Creates a .zip file with a link to download the data via the NiDB website |
| Local FTP site | Places the data in the NiDB FTP site |
| Remote FTP site | Sends the data directly to a remote FTP site |
| Remote NiDB site | Sends data directly to a remote NiDB site. New IDs will be created on the new server, and it is only available for DICOM data |
| Public Download | Make the data available publicly, so that anyone can download the dataset. |
| Linux NFS Mount | A path to the Linux directory where the data should be copied |
| Export package | Not used |
| NDAR submission | Creates an NDAR submission package which can then be uploaded using the NDAR tools |

Several other options are available for the data to be downloaded, image format, and directory structure. When you are all set, click **Download**. You can check the status of a download by mousing over the **Search** tab, and clicking **Download Request Status**. If you select the Web option, the download link will appear here. If you select the public download option, you should mouseover the **Search** tab, and click **Public Downloads**.

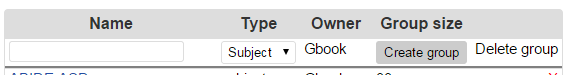
# Groups

Groups are useful to separating subjects and studies into different allocations than are defined by subjects or projects. For example, a sample of subjects may have multiple imaging studies, but you want to analyze or look at only those with particular criteria, like “only those with good imaging”, “only first visits”, etc. Groups are particularly useful when using pipelines so an analysis is only performed on the correct sample of subjects.

Groups can be made up of subjects, studies, or series. Again, these transcend project and other boundaries. **Subject groups** are pointers to the subject and all of its associated imaging studies. **Study groups** are collections of only specified imaging studies. **Series groups** are available, but may be too useful.

## Managing Groups

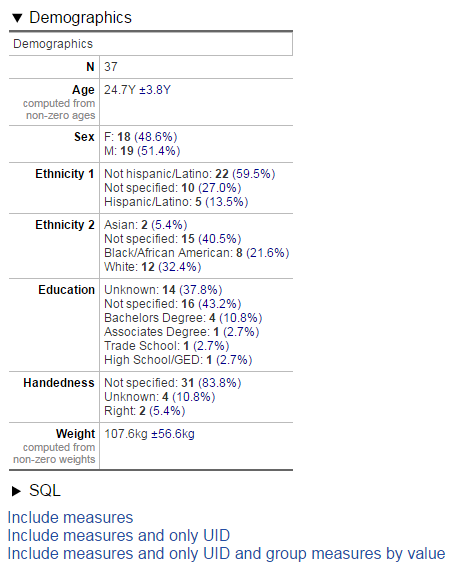
To create a group, mouseover the **Subjects** tab and click **Groups**. Enter a group name (must be unique), the type of group (subject, study, series) and click **Create group**.



Once the group is created, you can add items to it. You can add items in two places:

1. Click on the **Subjects** tab and search for subject(s). On the left, check off the subjects you want to add. At bottom left is a list of groups that you own that you can add the selected subjects to. Select the group and click **Add to group**.
2. Click on the **Search** tab and search for data. Select the items you want by clicking on the checkbox on the left. Select the Subject, Study, or Series and click **Add**. Again, you can only add items to a group you created.

To manage groups, mouseover the **Subjects** tab and click **Groups**. To delete a group, click the red **X**. You will see all groups, but only yours will have a red **X** next to them. Clicking the group will display the entire list of items in that group, along with some specific information for each item. To remove items from the group, select the checkboxes on the right side, and click **Remove**. You can only modify a group that you created. At the top of the page, you will see a **Demographics** section, click the arrow to view the group demographics. The N may not match the number of items in the group if your group includes multiple studies from the same subject. Click the “Include measures…” links to also view any associated measures, and click the SQL link to view the SQL statement.



# Pipelines

The NiDB pipeline is designed to allow you to process data automatically, and with little behind the scenes intervention. It is best used when you have an established command-line based analysis that you want to run on a large batch of data, or you want to run it automatically each time new data arrives on NiDB.

## Definitions

* **Pipeline** – A script that is run on a set of data. Also includes data specifications, dependencies, groups, and compute cluster options.
* **Pipeline Group** – A set of related pipelines. They may all be separate steps of a larger process.
* **Dependency** – A pipeline may depend on another pipeline for results that it requires for its analysis. An example is an fMRI stats pipeline depending on an fMRI preprocessing pipeline.
* **Descendent pipeline** – A pipeline that depends on another pipeline
* **Level 0** – a pipeline that does not operate on any data
* **Level 1** – study/subject level pipeline. This pipeline will download data as part of the analysis
* **Level 2** – second/stats level pipeline that depends on one or more first level pipelines. No data is downloaded as part of a level 2 pipeline
* **Ignored studies** – As a pipeline scans the database, looking for matching data, it will mark non-matching studies so they are not checked again. If you do something like change the data criteria, then some of the ignored studies may get picked up to run, but you must clear the ignore flag.
* **Analysis** – The result of a pipeline running against an imaging study
* **Primary data** – The data criteria on which studies are search first.
* **Associated data** – will only be download if the primary data is found.

## Viewing Pipelines

To view the pipeline page, click the **Analysis** tab. This will display a list of the currently enabled pipelines. By default it will only display the visible pipelines, if you want to view the hidden pipelines, click the **View**: **Normal** link at the top left of the page.

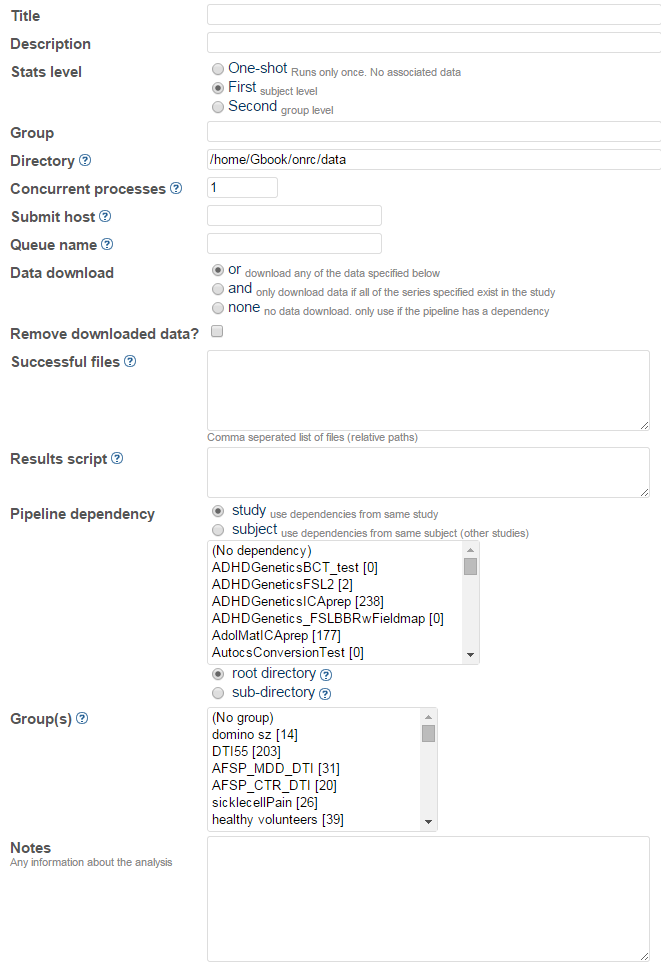


Mousing over the rows will give more information about the item. The first column contains the pipeline group, and when mousing over the label, you will see a graph describing the relationships between the pipeline, its dependencies and the study groups going into them. The second column displays the pipeline, and any descendent pipelines, along with their version numbers. Click the pipeline name to view the pipeline details. The next columns display the pipeline level (0,1,2) and the owner. The **Status** column shows the status and a checkbox to enable/disable the pipeline. If the pipeline is enabled, it will run and process analyses, otherwise it will not run. The **reset** link can be clicked if you absolutely know that a pipeline has stopped processing. The Analysis column displays the number processing and number complete. Click the http://ado2:8080/images/preview.gif icon to view the analysis list for that pipeline. The Disk size column lists the total space used by the analyses. This total includes any data from dependent (parent) pipelines, which are either hard- or soft-linked, and may be higher than the actual disk usage. Finally, the last two columns show the data path and the SGE queue name.

## Create a Pipeline

### New Pipeline

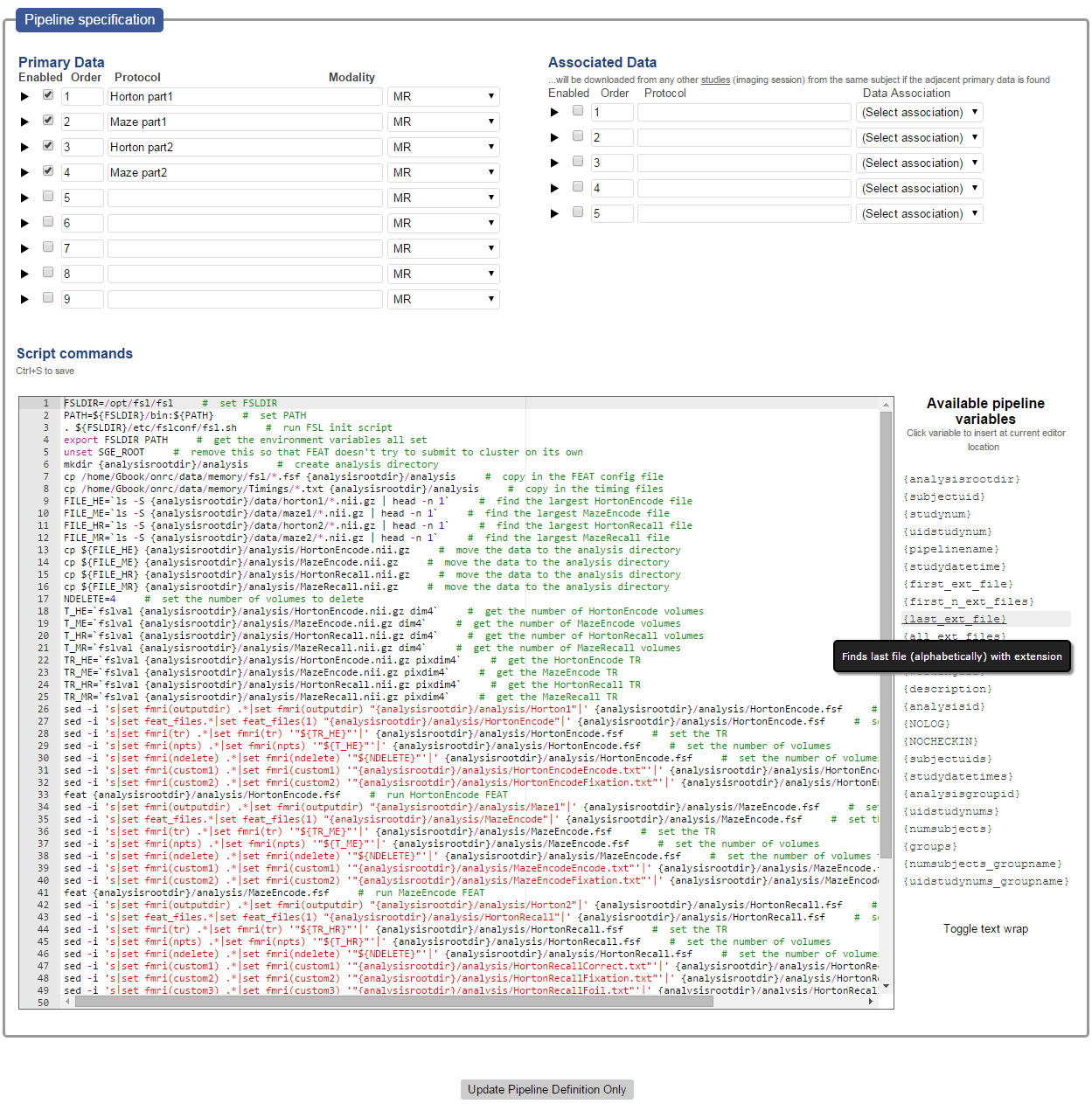
At the top of the pipeline page, click the **New Pipeline** link. You’ll see a pipeline page with many options. Some options have descriptions and help bubbles.



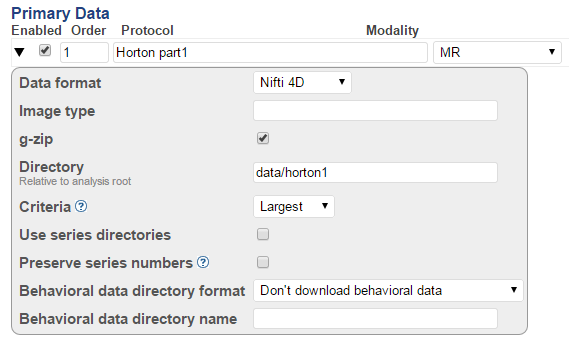
Some descriptions of the pipeline options

* **Title** - an alphanumeric string, with no spaces or special characters. This will become the directory in which an analysis is done. Once the pipeline is created, it cannot be changed.
* **Description** – longer description of the pipeline
* **Stats level** – Corresponds to the levels 0,1,2
* **Group** – (optional) to associate several pipelines together
* **Directory** – Where you want the data to go. If blank, it will be placed the default pipeline analysis directory
* **Concurrent processed** – The total number of analysis jobs allowed to running at the same time. This is a different number than the queue limits in SGE
* **Submit host** – If you have access to more than one cluster, there may be different submit hosts
* **Queue name** – The SGE queue name
* **Data download** – specifies how to interpret the data specification
* **Remove downloaded data** – doesn’t really work, don’t use it
* **Successful files** – a list of files that should exist at the end of the analysis. If they all exist, the analysis is marked complete.
* **Results script** – a script that is run at the end of the pipeline that ports results back into the database
* **Pipeline dependency**
  + **Study** – dependencies will only be used from the same study. For example an fMRI preprocessing pipeline is run on S1234ABC1, so this option will only check the same study for a dependency when running this pipeline
  + **Subject** – this will check the same subject, regardless of study for a dependent pipeline
  + **Dependency list** – the pipeline on which this pipeline depends. Can be multiple. Ctrl+click to select multiple
  + **Root directory** – this will place the dependency(s) in the root directory
  + **Sub**-directory – this will place the dependency(s) in subdirectories of the same name as the dependency
* **Group(s)** – the groups which will be part of this pipeline. This is ONLY study level groups
* **Notes** – extra notes about the pipeline

When you’ve filled out all the information, click **Add Pipeline Info**. This will create the basic pipeline, and the information that generally does not change. On the pipeline list, click your newly made pipeline. Click the pipeline to view it. At the bottom you’ll see a **Pipeline specification** section.



On the upper left is the primary data specification, the upper right the associated data specification, and on the bottom the pipeline script. Clicking the arrows next to each row of the data specifications will display more data options.



Description of the options

* **Enabled** – whether the data criteria will be used or not
* **Order** – change these numbers to reorder the download list
* **Protocol** – This is an exact protocol name. If you list a single protocol name, it will look for only that name. If you use a list like [“Task1” “Task 1” “task one”] it will search for any of those protocol names.
* **Directory** – relative to the analysis root directory
* **Criteria**
  + **All** – all matching series will be downloaded
  + **First** – only the lowest numbered matching series will be downloaded
  + **Last** – only the highest numbered matching series will be downloaded
  + **Largest** – Only one series with the most number of volumes or slices will be downloaded
  + **Smallest** – Only one series, with the least number of volumes or slices will be downloaded
* **Use series directories** – Will place the data in multiple numbered directories, with the series number
* **Preserve series numbers** – if the series directories are used, it will retain the original series number, otherwise it will start the numbering at 1

Below that are the script commands. This is a simple bash script with embedded pipeline variables. These variables (list on the left) are converted to real values when the script is run on each analysis. The bash script is a modified interpretation of bash, it will not run exactly as listed in the script window. Variables are replaced, check-in points are added, and some output is redirected to log files. Trailing semi-colons are also removed from each line.

If you use a loop, each line of the loop should contain the {NOLOG}{NOCHECKIN} variables in the comments, which will prevent the insertion of check-in code or logging, which may interrupt the loop. Each line should have a comment, even an empty comment. If you do any kind of output redirection using > or >>, make sure to include the {NOLOG} variable. If you use a command such as find, which requires a trailing semi-colon, add a second semi-colon to the end of it.

#### Script Variables

Variables are listed to the right of the script area, and clicking the variable will insert it into the script at the current cursor location. Here’s a description and example of the script variables.

|  |  |  |
| --- | --- | --- |
| Variable | Description | Example |
| {analysisrootdir} | Analysis root directory. The absolute path to the analysis directory | /path/to/the/pipelines |
| {subjectuid} | The UID of the subject being analyzed | S1234ABC |
| {studynum} | The study number of the being analyzed | 2 |
| {uidstudynum} | UID and study number | S1234ABC2 |
| {pipelinename} | The name of the pipeline | fMRIPreProc |
| {studydatetime} | Date/time of the start of the study. YYYYMMDDHHMISS | 20150323164523 |
| {first\_ext\_file} | Tries to find the first file with the extension “ext”, replace ext with the extension you want to find. You would want this to be part of a path, and that path and files must exist before the analysis is run | /path/to/{first\_nii\_file}  becomes …  /path/to/S1234ABC1\_001.nii |
| {first\_n\_ext\_files} | Same as above, but finds the first n files with extension | /path/to/{first\_2\_nii\_files}  becomes …  /path/to/S1234ABC1\_001.nii  /path/to/S1234ABC1\_002.nii |
| {last\_ext\_file} | Save as above, but finds last file with extension | /path/to/{last\_nii\_file}  becomes …  /path/to/S1234ABC1\_098.nii |
| {all\_ext\_files} | Save as above, but finds all files with extension | /path/to/{all\_nii\_files}  becomes …  /path/to/S1234ABC1\_001.nii  /path/to/S1234ABC1\_002.nii  …  /path/to/S1234ABC1\_097.nii  /path/to/S1234ABC1\_098.nii |
| {command} | The script command for that line, excluding the comment | feat fmritask.fsf |
| {workingdir} | No idea what this does |  |
| {description} | The comment. The text after the hash (#) character | This is a comment |
| {analysisid} | A unique, but not random number identifying the analysis. Necessary for re-inserting results back into the database | 2354199 |
| {NOLOG} | Will not append a > stepXX.log to the end of the line |  |
| {NOCHECKIN} | Will not prepend a line to checkin |  |
| {subjectuids} | Creates a space separated list of the UIDs involved in the analysis. Will only be used if there are groups | S1223ABC S2349GJH S3589FKW … |
| {studydatetimes} | Creates a space separated list of the study datetimes | 20150304184523 20130304124523 … |
| {analysisgroupid} | The database ID of the analysis group | 23 |
| {uidstudynums} | List of the UID/study numbers | S1359DJJ1 S3952VNS2 S7842WIT1 … |
| {numsubjects} | Number of subjects in the group(s) in total | 345 |
| {groups} | List of groups that the pipeline depends on | fMRIPreProc fMRIStats1 fMRIStats2 |
| {numsubjects\_groupname} | Number of subjects for the specified group | {numsubjects\_fMRIPreProc}  32 |
| {uidstudynums\_groupname} | List of UID/study numbers for the specified group | {uidstudynums\_fMRIPreProc}  Becomes…  S1359DJJ1 S3952VNS2 S7842WIT1 |

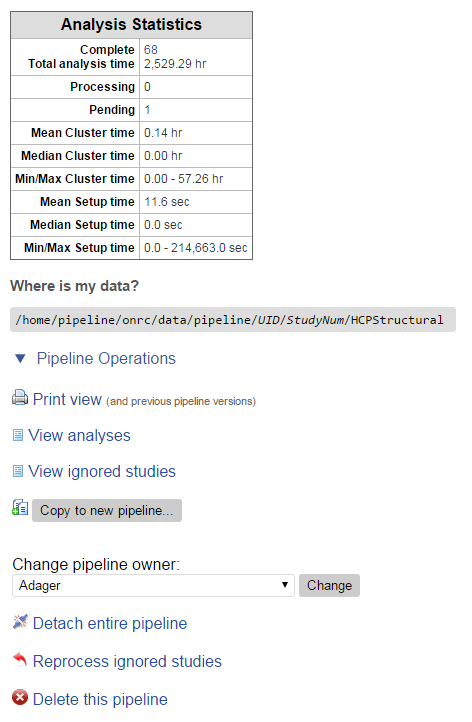
### Copy Existing pipeline

To copy an existing pipeline into a new pipeline, go to the pipeline’s page, and click the **Pipeline Operations** arrow, then click the http://ado2:8080/images/copy16.gif **Copy to new pipeline…** button. It will prompt you for a new pipeline name.

## Modify a pipeline

When clicking on the pipeline name on the main pipeline page, you will be brought to the pipeline’s page. You will be able to modify the pipeline (if you are the owner or administrator) on this page. Items that cannot be changed are gray out. The pipeline is separated into two section 1) the main pipeline details 2) the data criteria and pipeline script. Only the data criteria and script are versioned; so when you save the script or data criteria, a new version will be saved, and old version can be viewed. This is useful when you make a change and suddenly it’s not working anymore. The versioning allows you to check a recently working copy of the script.

At the top of the pipeline page, you’ll see a box with the Analysis Statistics and below that a box for Pipeline Operations.



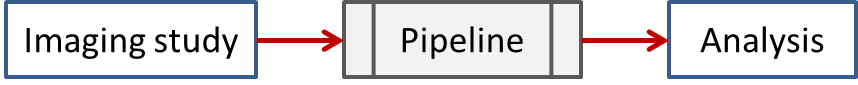
Description of the options:

* **Print view** – View the current (and previous versions) of the pipeline in a print form
* **View analyses** – Displays a list of the current analyses
* **View ignored studies** – Displays a list of imaging studies which were included in the search for data, but were subsequently not run because the data criteria did not match
* **Copy to new pipeline** – Copies this pipeline to a new one with a newly specified name
* **Change pipeline owner** – Changes the owner of the pipeline to another user
* **Detach entire pipeline** – removes the pipeline from the control of the NiDB pipeline system and moves the data to a specified directory (Not Implemented).
* **Reprocess ignored studies** – For any studies which were ignored, this option will try them again to see if they match the data criteria.
* **Delete this pipeline** – This will delete the pipeline and any data associated with it. Dependent or descendent pipelines will not be deleted.

## Analyses

### Understanding the Analysis Process

NiDB pipelines operate on single imaging studies, so that a pipeline is applied to an imaging study and the outputs are associated with that imaging study. In other words the analysis and the imaging study are on the same granular level.



### Data Selection

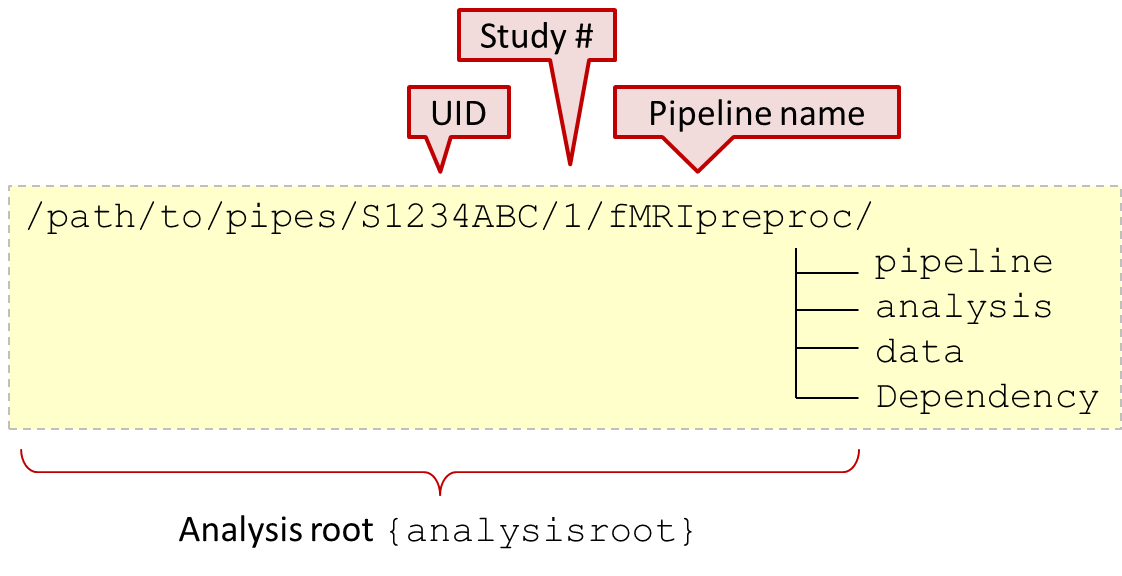
The first step to check which imaging studies will be run through the pipeline starts with the dependencies and groups. Only studies which are part of the specified group (or all studies in the database if there is no group specification) and which have the specified dependency(s) (or all studies is there are no dependency specifications) will be checked. After a list of possible studies is selected, each one is checked against the data criteria. Within each imaging study are several series, each with a protocol name, and these are the data that checked for inclusion in the analysis.

In the data specification, there are two parts 1) primary data 2) associated data. When the pipelines are searching for matching data to determine whether to run the pipeline on that study, they are only looking at the primary data. So if the primary data criteria are met, then the study is analyzed. If there is an associated data spec, then the associated data is also searched for (from studies belonging to the same subject), but if it is not found, the analysis will still run.

For primary data, there is an option to require all of spec, or any of it. In other words if you select “and”, the pipeline will only run on the study if the study contains all of the data criteria specified. If you select “or”, the pipeline will run on the study if the study contains any of the data specified.

### Analysis Directory Structure

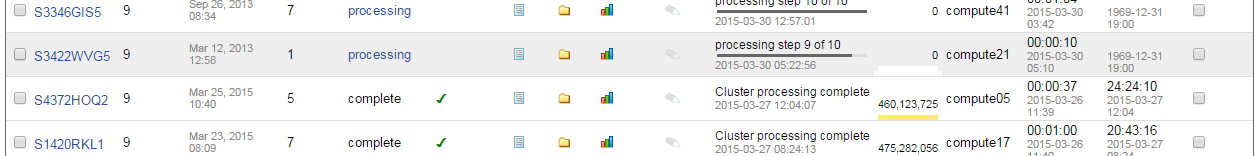
The following figure describes the analysis directory structure.



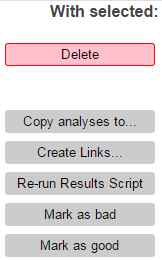
The {analysisroot} variable is considered the root the analysis and should be used in scripts to specify absolute paths. Your pipeline should be assumed to have access to all directories including and below the analysisroot directory. The pipeline subdirectory is always created and stores log files, SGE job files. The analysis directory is optional, but is very useful to contain an analysis. The data directory is optional as well, but is often created when a data directory is specified in the data specification. The Dependency directory is created if the pipeline has a dependency. Any copied (or linked) data will be placed in the Dependency directory.

### Analysis Listing

When viewing the analysis listing page, you’ll be able to see the currently processing and completed analyses. If there a lot of analyses (more than 5000) it will be divided onto multiple pages. You can navigate those pages using the navigation links on the top left.  The analysis listing provides a lot information right on the page, through mouseover messages, or through dialog boxes when clicking icons.



The first column contains a link to the imaging study. Second column contains the pipeline version that was run on that analysis. Next columns contain the imaging study date, and the number of series downloaded. Then the analysis status: processing, pending, or complete. Clicking on a processing link will show the current SGE status for the running job. A green checkmark  indicates the analysis is complete because the file(s) specified in the pipeline for a successful analysis exist. The next three icons allow viewing of analysis details in dialog boxes, without leaving the page. The http://ado2:8080/images/preview.gif icon will display the logs associated with the analysis. The  icon will display a file/directory listing of everything in the analysis root directory. The  icon will display any results that were associated with the analysis. The pencil icon allows you to add notes to each analysis. Click the icon to add or edit notes. The most recent status message and date are next, followed by the disk size of the completed analyses. The hostname of the server the analysis was run is next. This can be helpful in diagnosing random failed analyses: ie, it might be related to one particular incorrectly configured server. Then the analysis setup time, which is basically the time it took to copy in the data. Last is the compute time, this the span of time between when the analysis was first started by the SGE, and when it ended. The right-most column contains checkboxes. Select specific analyses to perform the following options (buttons at bottom of page).



* **Delete** – Completely deletes the analysis.
* **Copy analyses to**… - Will copy the analyses to a specified directory, leaving the original copy
* **Create links** – Will create a directory containing links to the selected data
* **Re-run Results Script** – Will rerun the results script specified in the pipeline
* **Mark as bad** – Will mark the analyses as bad, so **descendent pipelines will not use the analyses**
* **Mark as good** – will unmark bad analyses

## Example Pipelines

Below is a sampling of pipelines that demonstrate the nuances of the pipeline bash format and also the generic quality of the programming necessary to accommodate any data that passes through the pipeline.

### Freesurfer

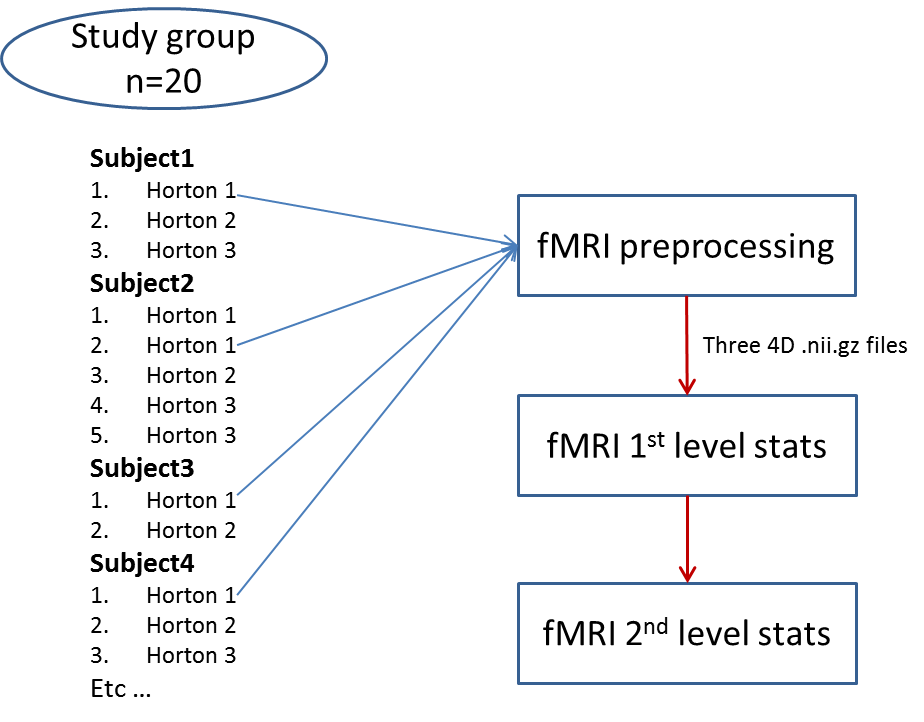
This simple pipeline will take any type and number of T1 images, and pass them through freesurfer’s recon-all process.

|  |  |
| --- | --- |
| **Data** | Multiple T1 images, placed in {analysisroot}/data |
| **Script** | 1 export FREESURFER\_HOME=/opt/freesurfer # Freesurfer home directory  2 export FSFAST\_HOME=/opt/freesurfer/fsfast # ENV variables  3 export MNI\_DIR=/opt/freesurfer/mni # ENV variables  4 source $FREESURFER\_HOME/SetUpFreeSurfer.sh # setup Freesurfer to run  5 export SUBJECTS\_DIR={analysisrootdir} # all FS data will go here  6 freesurfer > {analysisrootdir}/version.txt # {NOLOG} get the freesurfer version  7 perl /opt/autoCS/ImportFreesurferData.pl {analysisrootdir}/data analysis # import data  8 recon-all -clean-tal -notal-check -mprage -no-isrunning -autorecon1 -subjid analysis #  9 recon-all -mprage -no-isrunning -autorecon2 -subjid analysis # Autorecon 2  10 recon-all -mprage -no-isrunning -autorecon3 -subjid analysis # Autorecon 3 |

This script first sets up the environment variables required to find freesurfer and those variables required by freesurfer. Freesurfer requires a SUBJECTS\_DIR which contains the subjects to be processed. This directory is expected to contain a bunch of other directories which would contain their respective subject data. Freesurfer also expects data to be in a specific format prior to running it. Since this analysis only operates on one subject, the SUBJECTS\_DIR directory will have only one directory, and for simplicity, it’s called ‘analysis’. This implies the Subject ID is analysis, and so that ID is passed to the freesurfer calls. The perl script called on line 7 will take the raw data in the data directory and put it into the format expected by freesurfer in the analysis directory.

### FSL fMRI pipelines

This is an example of a dependent pipeline containing a preprocessing step, first level stats, and second level stats. Limited code will not be provided, so we can focus on the concepts of how dependent pipelines work. Let’s start with the simplest case, you have a group of 20 subjects who did an fMRI task called Horton. Each subject should have 3 runs, labeled Horton 1, Horton 2, Horton 3. In this example, the data criteria will be to select the largest (most number of BOLD reps) for each protocol name. Subject 1 is normal, subject 2 has repeated runs, and subject 3 is missing the third run. All of these go into the preprocessing pipeline. The output of the preprocessing pipeline is a 4D preprocessed .nii.gz file, or in the case of this whole pipeline setup: 3 .nii.gz files, one for each run. Each subject will have 3 .nii.gz files.



## Checking Pipeline Logs

This is only available to system administrators of NiDB. Under the Admin tab, go to Modules. Click the pipeline module, and you’ll see a list of the recent log files. If you suspect the pipeline module has failed, you may see an error message at the end of any of the log files. Multiple instances of the pipeline module normally run at the same time, so while one instance may have failed, others may continue to run. You can also see the progress and other debugging information in the log files.

## Troubleshooting

### Why Didn’t the Pipeline Do Anything?

If you created the pipeline, it has data, it has a script, and its enabled, you expect it to run… but it’s not, nothing at all is happening. It’s just sitting there, for an hour, taunting you. It could be many things. If nothing is happening, check these things one by one.

* Are the SGE submit node and queue name valid
* Does the pipeline directory exist, and is it writeable. Especially if you are not using the default pipeline directory
* Is the # of concurrent processes at least 1
* Does the group you selected have any studies in it
* Does the dependency you selected have any analyses in it
* Has it been 6 hours since the data was collected (the pipeline will only process studies older than 6 hours)

Most problems are related to the data criteria, so check the data criteria carefully against the imaging studies you are trying to process.

* Is the criteria step enabled (they are not enabled by default)
* Do your analyses contain the exact protocol name and modality specified
* Make sure the Image type is also valid for your data
* If your data selection criteria is an ‘and’, make sure every study has all the data

### Why did only some studies start processing?

It may, or may not, be more frustrating when only some of the studies you expect to process actually process. Not that some of the analyses started and failed, but that some didn’t start at all and were apparently ignored. For example, you specified a group of 20 studies, but only 18 of them were picked. What happened to the other 2 studies? Check the same things here that you would check if nothing processed, but the problem is most likely

* A dependent pipeline analysis failed
* The data criteria do not match all of the studies in your group

To check the ignored studies, go to the pipeline’s page, and click the **View ignored studies** link. This will display a list of studies that were checked. Click the **view log** link to check the reason they were not included in the analysis. The message may be cryptic, but the last line often describes what was not found.

