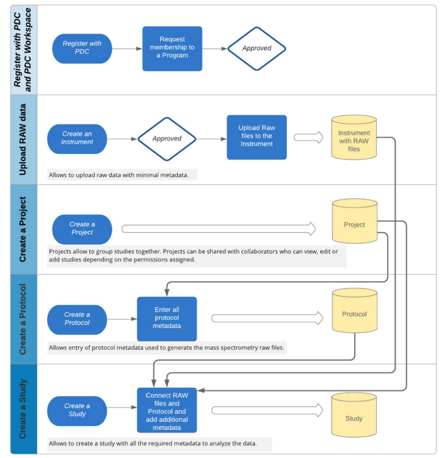



# PDC Workspace Data Submission Overvie


**Quick Links:**
0:00 (https://www.youtube.com/watch?v=aRNXAaWckrg&t=0s) PDC Data Model and Dictionary
1:05 (https://www.youtube.com/watch?v=aRNXAaWckrg&t=65s) Data Submission Overview
2:01 (https://www.youtube.com/watch?v=aRNXAaWckrg&t=121s) Register your program with PDC
2:19 (https://www.youtube.com/watch?v=aRNXAaWckrg&t=139s) Register with PDC as a data submitter
3:21 (https://www.youtube.com/watch?v=aRNXAaWckrg&t=201s) Create instrument and upload raw files
5:05 (https://www.youtube.com/watch?v=aRNXAaWckrg&t=305s) Create a project
5:35 (https://www.youtube.com/watch?v=aRNXAaWckrg&t=335s) Create a protocol
8:03 (https://www.youtube.com/watch?v=aRNXAaWckrg&t=483s) Create a study
11:23 (https://www.youtube.com/watch?v=aRNXAaWckrg&t=683s) Biospecimen data submission overview

Please click here (https://docs.google.com/spreadsheets/d/1wgCaEwLHYlfbjUZDUjc8o-u66vHAaMSs/edit?usp=sharing&ouid=102791987143514485948&rtpof=true&sd=true) to download the PDC submission template in spreadsheet format.

## How to submit data:

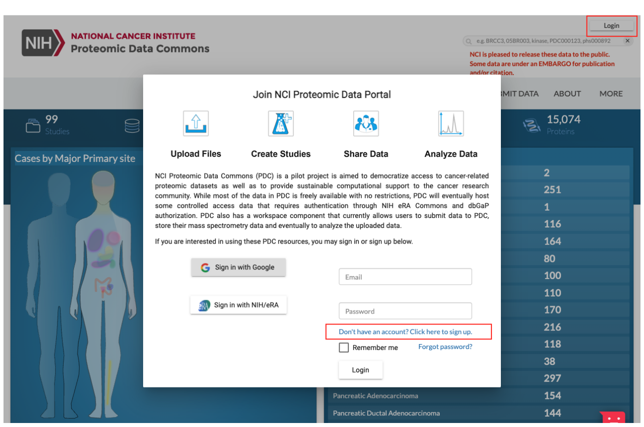
1. Register your Program Name
2. Register with PDC
3. Register on the PDC workspace as a data submitter for a Program
4. Login to the PDC Workspace
5. Submission process
6. Upload Clinical & Biospecimen Metadata
7. Upload additional supplementary files

### Register your Program Name:

To register your program, please send an email to PDCHelpDesk@mail.nih.gov (mailto:PDCHelpDesk@mail.nih.gov?Subject=PDC Help).

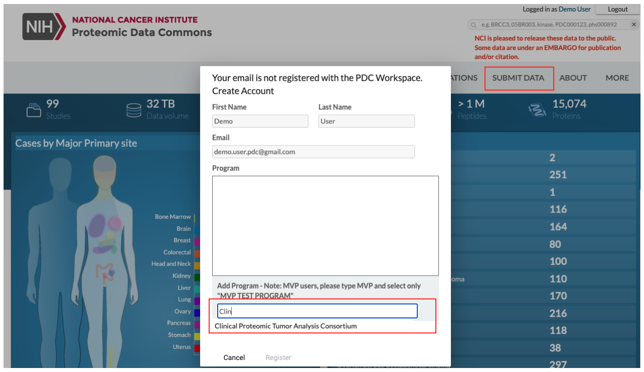
### Register with PDC:

Go to https://pdc.cancer.gov (https://pdc.cancer.gov), click Login at the top right corner. You may use a Google or eRA commons account, or NIH email to register with PDC. If you do not have any of these, you may create an account by clicking 'Click here to sign up' and by following the instructions. You will receive an email; click the provided link to confirm your registration.



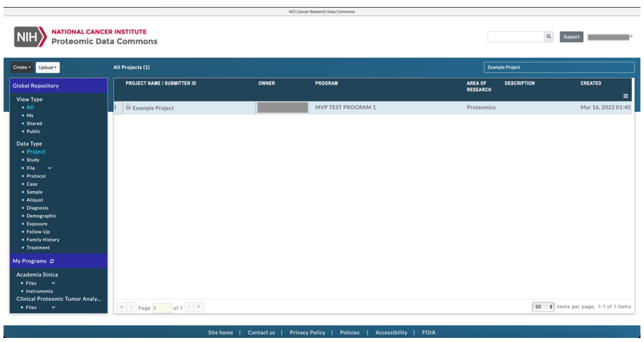
### Register on the PDC workspace as a data submitter for a Program:

Return to https://pdc.cancer.gov (https://pdc.cancer.gov) and click 'SUBMIT DATA' and select 'Register with Workspace'. In the dialog box, start typing the program name and select the relevant program from the drop-down list. Click submit and you will receive an email confirmation for your request. You will receive another email once the request for membership to the program is approved.



### Login to the PDC Workspace:

Return to https://pdc.cancer.gov (https://pdc.cancer.gov) and click Login at the top right corner. A pop up wizard will appear which will allow you to either sign in with your Google or NIH/eRA account or you can login with your registered credentials. To enter the PDC Workspace, Click 'SUBMIT DATA' then click 'Go to Workspace'. You will see a screen displaying something like this:



### Submission process:

**Note: Before you begin the submission, we strongly recommend you use the PDC submission template to collect the information required for a successful submission.**

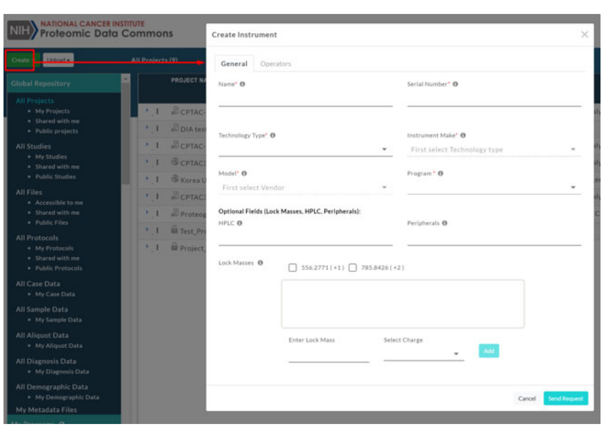
Now you are ready to begin the submission process. **You need to follow the steps in the order given in the Submission Overview (image) at the top of this document to avoid duplicating steps**

The submission process involves the following major activities,

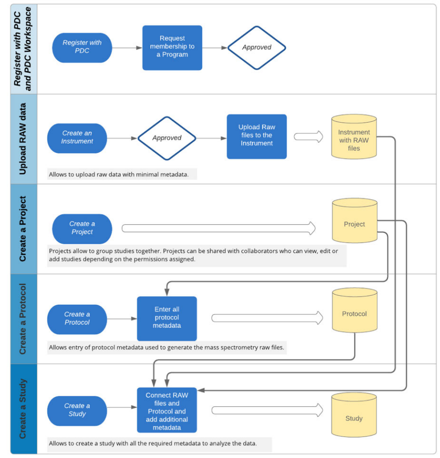
- Creating an **instrument** to upload raw data
- Create a **project** under which to group the raw data into a **Study** that was generated using a specific experimental **protocol**
- Upload biospecimen and clinical metadata for the samples used in the Studies
- Upload any additional supplementary files

1. **Create an instrument:**

   An instrument is a representation of the mass spectrometer from your lab. Go to create, click instrument, and enter the relevant information about the mass spectrometer used to collect the data. If you are a program member, you will receive an email once approved by the PDC admin or the program lead (PI).

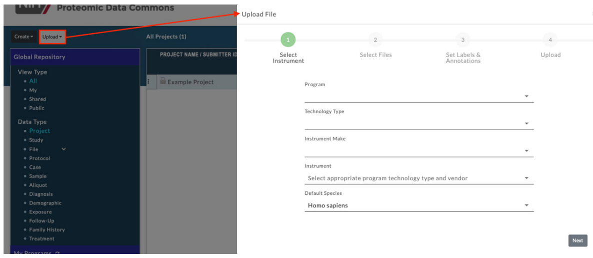
   Note: You must be added as an assigned instrument operator to upload data to an instrument



unless you are the creator of that instrument.

2. **Upload raw files:**

   Click upload, and in the dialog requesting information, select the instrument you previously created. You can upload directly from your local computer or from the cloud (AWS S3 is supported) if your files are on S3 already.



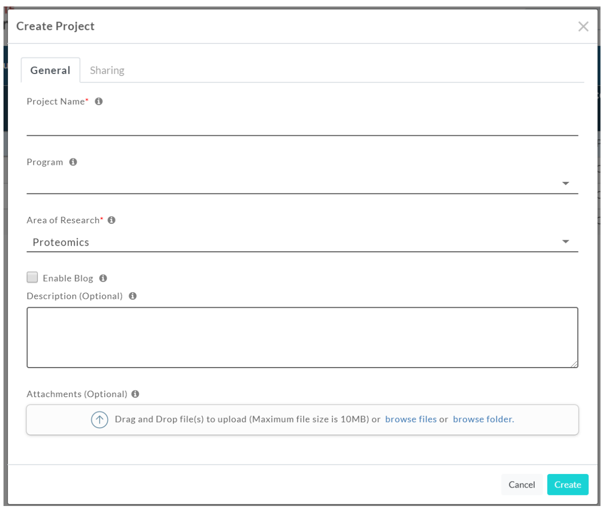

Note: The PDC Workspace accepts files with the ".raw", ".d"; ".wiff"; ".wiff.scan"; ".wiff.mtd"; ".dat"; ".mis". These files are produced directly on the mass spectrometer acquisition computer and are typically in proprietary formats.

Tip: You may create a separate instrument for each study that you would like to create, even if all those files are generated from the same mass spectrometer. This is not required, but may help to organize and separate your files, facilitating study creation.

3. **Create a Project:**

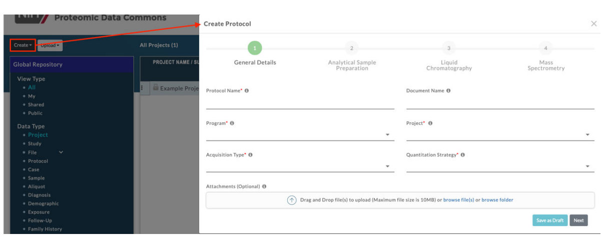
   **Before you can create a Study or add your Protocol, you first need to create a Project.**

   Click Create->Project to open the wizard. Next, give your Project a name, select a Program and area of Research. You may also add collaborators to your project using the Sharing tab.

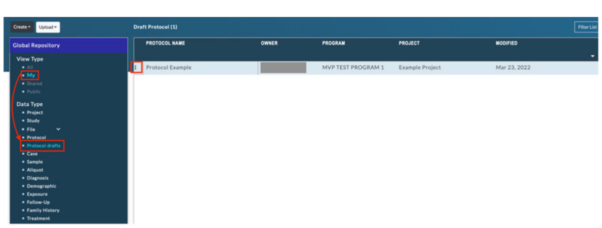



4. **Create a Protocol:**

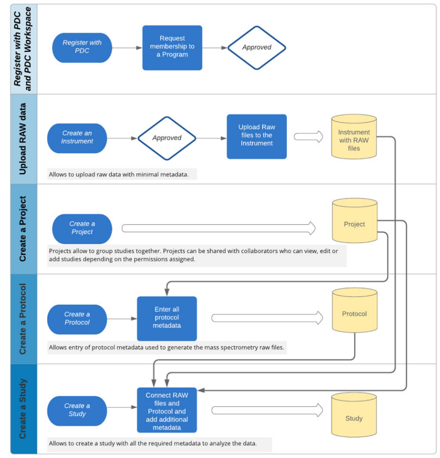
   Next, you will add the protocol metadata used in your Study. Click Create->Protocol to open the wizard. Required fields are marked with *. If you are unable to complete your protocol, you can 'Save as Draft' and resume later.

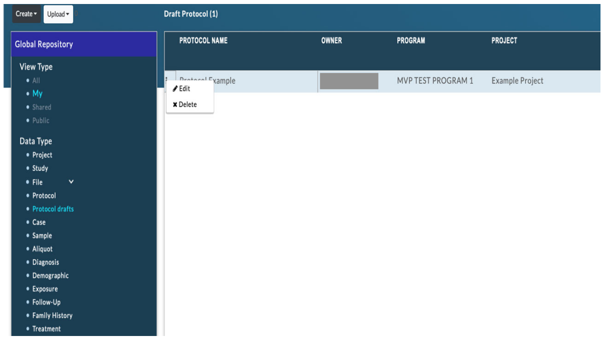



You can locate your protocol draft using the left navigation panel. Under "View Type" Click "My" and under "Data Type" click "Protocol drafts." Locate your protocol name and click the vertical dots in front of your protocol name for a menu to appear.



Click "Edit" and proceed with entering your protocol data.





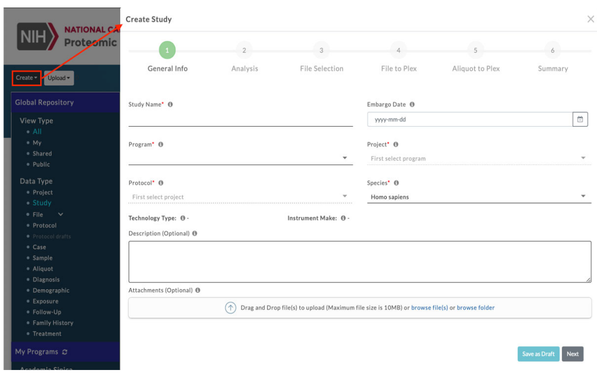


If you feel that the wizard does not adequately meet your annotation needs for your particular Protocol, please contact the PDC by email for help resolving any issues.

5. **Create a Study**:

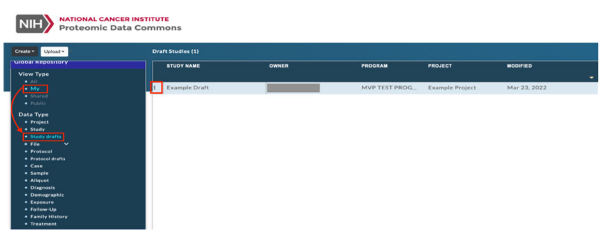
   In the PDC workspace, a study is a group of **instrument** raw files that are generated using a specific experimental **protocol**, with a well-annotated experimental design (sample to file mapping) that can be analyzed through a single bioinformatics pipeline.

   To create the study, click Create->Study to launch the wizard. In the wizard, for steps requiring you to enter data in tabular form for the sample/aliquot to file mapping, you may copy and paste from the Experimental Run Metadata sheet from the PDC submission template you prepared. If you are unable to complete your study, you can click 'Save as Draft' to resume later.



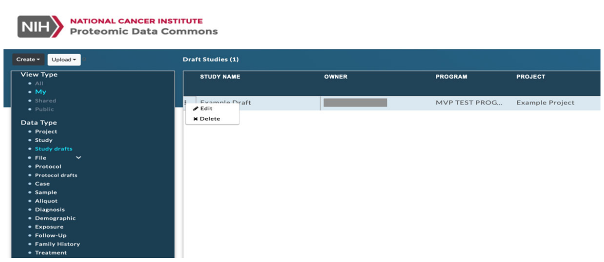


   You can locate your study draft using the left navigation panel. Under "View Type" Click "My" and under "Data Type" click "Study drafts." Locate your study name and click the vertical dots in front of your study name for a menu to appear.





   Click "Edit" and proceed with entering your study data.



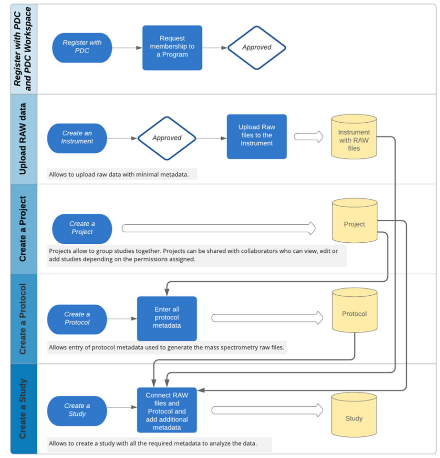


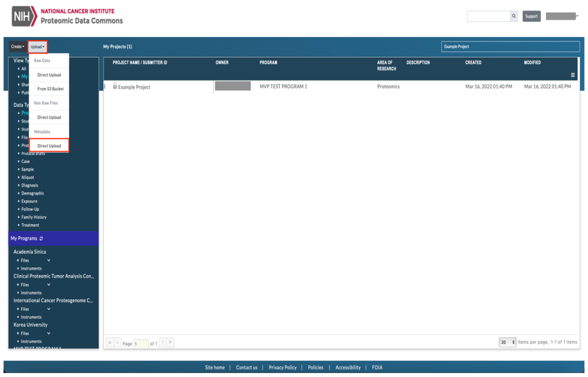
6. **Upload Clinical & Biospecimen Metadata**:

   To load the Clinical and Biospecimen data, save the Case Matrix, Case, Demographics, Diagnosis, Exposure, Family History, Treatment, Follow-Up, Sample, and Aliquot worksheets as TSV files from the PDC Submission template you prepared.

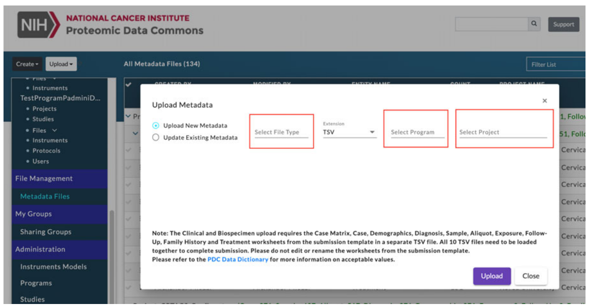
   Note: Clinical and biospecimen metadata should be submitted for all the aliquots (their parent samples and cases) that are entered in the Study creation step.

   All 10 files are required for upload, but if you have no data for the Exposure, Family History, Treatment, and Follow-Up entities you may include empty files containing only the column headers. Place all 10 TSV files in a new folder on your local computer. In the workspace click "Upload" then under Metadata click "Direct Upload."

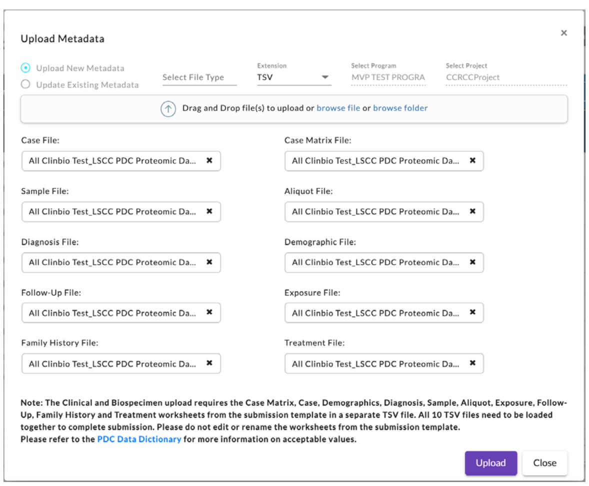




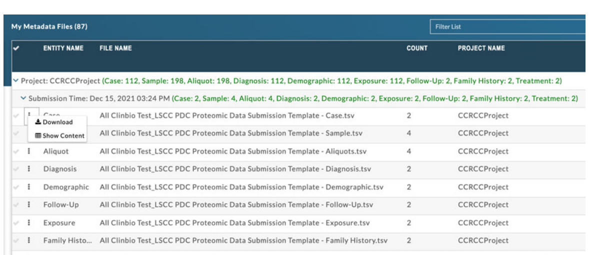
A popup wizard to upload the metadata will appear. Click “Select a File Type” and select “All Clinical & Biospecimen.” Next, click “Select Program” to select the program the metadata belongs to and “Select Project” to choose the project your metadata is associated with. An icon will appear in the wizard giving you the option to upload your files. You can either drag and drop your files or select a folder the files belong to. A notification will appear requesting you to confirm the upload of all files from that folder. Click “Upload” to confirm.



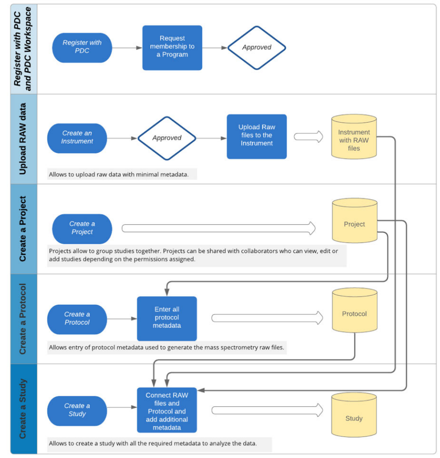
When all 10 TSV files are present, click "Upload" and you will have successfully uploaded your files.

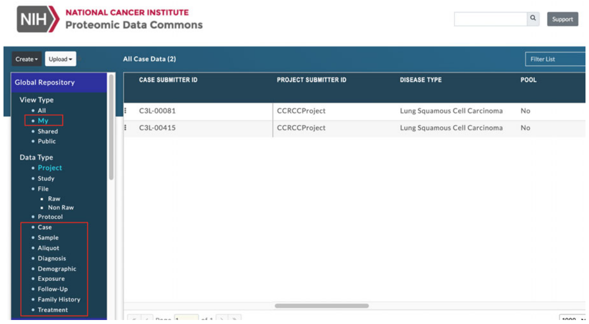



The clinical and biospecimen data you uploaded can be viewed by locating the navigation panel on your left and under the “File Management” section select “Metadata Files.” Select the vertical dots in front of the metadata file name and select “Show Content.”

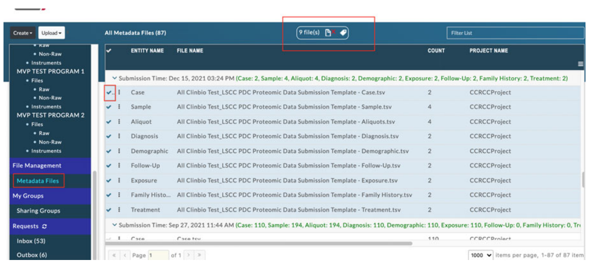


To view all the metadata you have uploaded, select “My” under the “View Type” section and select Case, Sample, Aliquot, Diagnosis, or Demographic under the “Data Type” section. As seen below, by selecting View Type “My” and Data Type “Case” you would see the metadata you uploaded for Case.

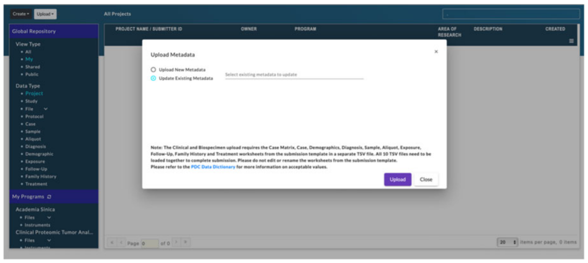




If you need to make changes to your clinical and biospecimen metadata, you will need to either reupload or update your data. To reupload, you first need to delete all your metadata files. Locate the navigation panel on your left and under File Management select “Metadata Files.” Select the files to delete and click the delete icon at the top.



To update your metadata. Click 'Upload', then select 'Direct Upload' under 'Metadata'. Within the 'Upload Metadata' window, select 'Update Existing Metadata'.



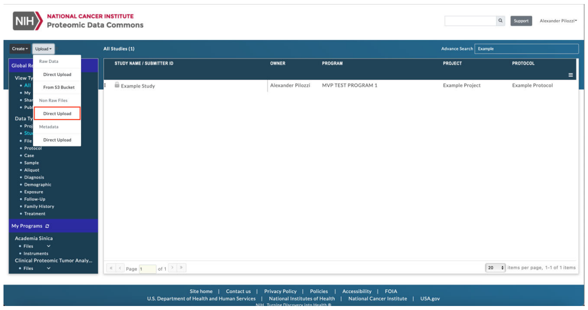
Select the metadata upload set you wish to update; the name of the set contains the prefix or suffix of your TSV files and a timestamp of the original upload. Once you have selected a set to update, upload the set of 10 TSVs for the cases, samples, and aliquots you wish to update. You may choose to include only a subset of the cases, samples, and aliquots included in the original upload, but cannot include those that were not included in the original set.

7. **Upload additional supplementary files:**

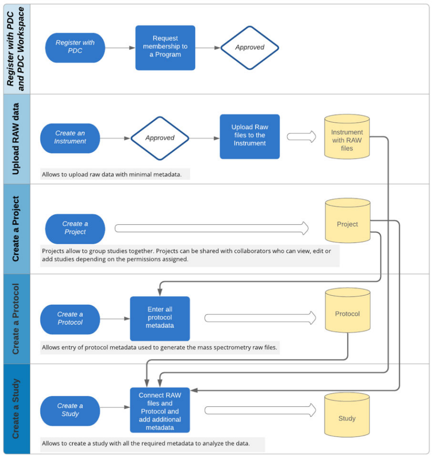
You may also upload additional supplementary files (non-raw files), such as processed outputs from your own data analysis pipeline, SOPs, clinical data, etc. You can select the appropriate data category for the files you wish upload - ‘Alternative Processing Pipeline’, ‘Other Metadata’ and ‘Supplementary Data’.

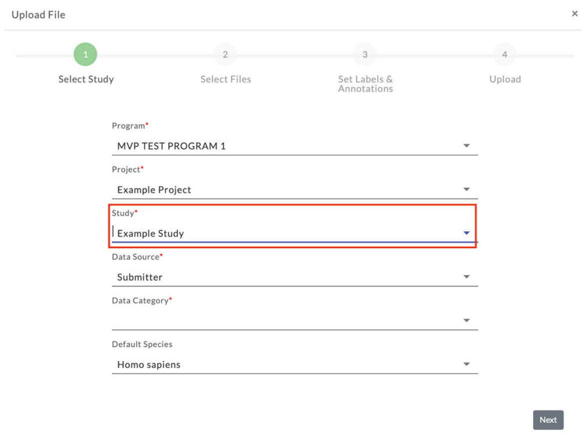
Once you complete the Study creation, you may proceed to upload the supplementary files associated with the study. We recommend you contact the PDC team to validate your study on Workspace before you begin uploading the non-raw files.

To begin uploading any of these types of files, select “Direct Upload” under the ‘Non-Raw Files’ tab in the ‘Upload’ dropdown.



In the upload interface, first you must select the study you wish to associate the files with, and the type of file. You may choose to directly select a study, and the ‘Program’ and ‘Project’ options will be automatically filled.






The Data Source 'Submitter' will be automatically filled in. The other option, CDAP, is used by the PDC team to handle the upload of files produced by the "Common Data Analysis Pipeline." Choose the data category you wish to upload from the 'Data Category' dropdown. Note that you will **only** be able to upload files with extensions belonging to that category. You may choose to upload a directory of files simultaneously, and file extensions that do not match the selected category will be automatically removed. However, note that some data categories have overlapping file extensions. You may wish to organize your files by data category prior to uploading.

**This completes the process of submission of a dataset for release through the PDC data portal. If you have any questions about this process, please contact the PDC by email for help PDCHelpDesk@mail.nih.gov (mailto:PDCHelpDesk@mail.nih.gov?Subject=PDC Help).**