Data Preparation Tool (DPT)

The Data Preparation Tool (DPT) organizes study items (e.g., datasets, documents, and information about biospecimens) associated with your study and allows you to annotate these items prior to uploading them into NICHD DASH, either for a new study submission or to add more items to an existing study. DPT is a standalone application that you can download to your computer and use offline. Once you have begun to organize your study items, you can save your work at any time and return to it later.

DPT automatically categorizes files belonging to the study (e.g., datasets and documents). You will be asked to approve the categorizations or make the appropriate changes. Next, you will be asked if you are going to submit information about biospecimens available for sharing that are stored in the NICHD Contracted Biorepository. If yes, the DPT will also generate a separate template for your biospecimens along with a dataset and document annotation template in Microsoft Excel. You will then be instructed to review and provide key information (annotations/metadata) about each study item, such as dataset type, format, document type, biospecimen type, etc. Once you have saved your completed annotation templates, DPT will encrypt the files in preparation for uploading them into NICHD DASH.

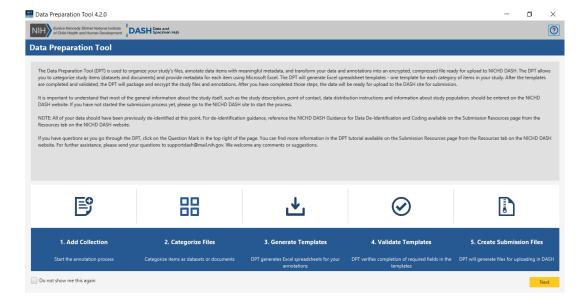
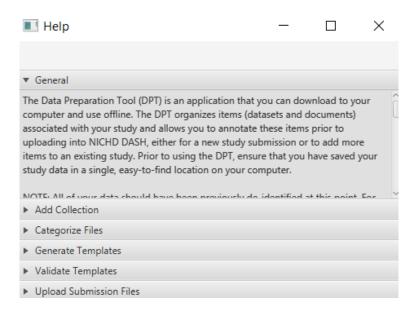


Figure 1: Data Preparation Tool

At any time, if you need help, click on the question mark (?) at the top right of the screen and the following Help screen will appear.

Figure 2: DPT Help Screen



Select any heading in this Help screen by clicking on the arrow head to the left of the heading for step-by-step instructions in DPT.

1. Prepare for Annotation

Note: Ensure that all files belonging to your study have been de-identified prior to beginning the annotation process (for de-identification guidance, please follow the "<u>Data and Biospecimen Catalog De-identification</u> Guidance").

After de-identifying your study items, follow the steps below:

- 1. Ensure that all files belonging to your study (datasets and documents) are de-identified and saved in a single, easy-to-find location on your computer.
- 2. If you are logged out of NICHD DASH at this stage, you can return to the study submission page by logging into NICHD DASH and clicking on the "Submit Study" tab located on the top menu bar of the homepage. You will be directed to the "Submission Dashboard" page displaying your study submissions. Click on "Edit Submission" to the right of the study that you wish to annotate, and navigate to the "Data Preparation" page.
- Download and install DPT from the "Data Preparation" page. The DPT can also be downloaded from the Submission Resources page in the Resources tab of the upper navigation bar.
- 4. Open and run DPT.

Note: If you do not have the latest version of the DPT, you will see a modal that will inform you that a later version of the DPT is available. Please download the current version of DPT from the "Data Preparation" page or from the Submission Resources page in NICHD DASH to ensure that you are using the latest version available.

The landing page for DPT provides an overview of the tool and the steps you will need to complete. Select the yellow "Next" button at the bottom right to proceed to the DPT Dashboard page.

2. Dashboard

The DPT Dashboard displays a list of all the studies you have annotated or are currently annotating via DPT. You can save your work and revisit it later by clicking the yellow "Resume" button to the right of the study in the main DPT Dashboard.

- 1. To start annotating a new collection, click the yellow "+Add Collection" button in the lower right corner.
- 2. To resume a previous annotation, click the yellow "Resume" button to the right of the study name to pick up where you left off.
- To delete a previous annotation, click the yellow "Delete" button to the far right of the study name you would like to remove. A confirmation window will appear allowing you to complete deleting the study or cancel from deleting the study.

Data Preparation Tool 4.2.0 X Eunice Kennedy Shriver National Institute of Child Health and Human Development (?) **Data Preparation Tool** Dashboard The DPT performs automatic categorization of your study items (datasets and documents). You will be asked to approve these categorizations or make the appropriate changes. The DPT will then generate Microsoft Excel templates for each category with file names automatically populated in the first column. Templates will be saved in a folder of your choice on your computer. You will then be instructed to review and provide key information (annotations/metadata) about you data, such as a description, type, and/or format. You will be able to further annotate your data, providing more information to facilitate future data discovery. The DPT dashboard will display a list of all of the collections you have annotated or are currently annotating via the DPT. You can save your work and revisit it later by clicking the "Resume" button from the main dashboard. To start annotating a new study, click the "Add Collection" button in the lower right of the page. For more assistance, click on the Question Mark in the top right of Collection Name Collection Type Resume Delete 31 Study Submission C:\Users\582399\Documents... Upload Submission Resume Delete 32 Delete Study Update C:\Users\582399\Documents... Content Sent > Return To Homepage Add Collection

Figure 3: DPT Dashboard

Note: For step-by-step instructions within the DPT, click on the "?" button at the top right of the screen at any time.

3. Add Collection

To start the annotation process, click on the yellow "+Add Collection" button in the DPT Dashboard, a pop-up modal appears:

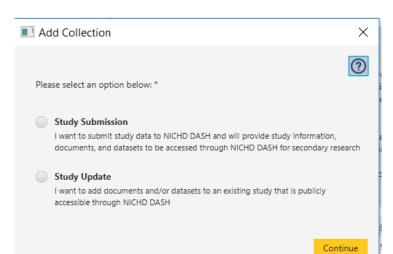
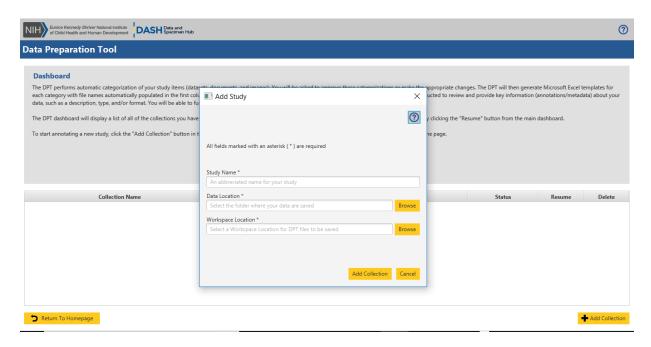


Figure 4: Add Collection

Please select if you are going to submit a new study or add more files to (update) an existing study. After you make your selection, DPT will display another modal:

Figure 5: Specify Collection Location



Follow the steps below to add the study collection:

- Enter the name of your study in the "Study Name" field. You do not need to use the
 complete study name—an abbreviated study name will help you remember what data
 you are submitting. The abbreviated study name is strictly internal to help you manage
 your data submission and will not be accessible to end users or the NICHD DASH Curator.
- 2. Make sure you have saved all study items (datasets and documents) associated with your study in a single, easy-to-find folder on your computer. Do not save any of your study items in a single compressed folder, such as a .zip file, as they will be treated as a single file and the individual study items will not be listed separately for annotation.
- Click the yellow "Browse" button next to the "Data Location" field, search for the folder, and click "Select Folder."
- 4. Create an empty Workspace folder on your computer that is separate from the folder containing your study items. The "Workspace Location" is where DPT will store the templates that it generates for you to use to annotate all your study items. Indicate where you would like these templates to be saved on your computer by clicking "Browse" and choosing the Workspace folder on your computer. DPT will then generate a sub-folder titled "DPT_Annotations_[Study Name]." This sub-folder is where DPT will save your templates.

Note: Your "Workspace Location" folder must be completely separate from your "Data Location" folder. Please make sure that you specify two distinct folders. They may both be contained within the same parent folder (e.g., the "My Documents" folder), but the "Data Location" folder cannot be a sub-folder within the "Workspace Location" folder or vice versa. You will receive an error message if the "Workspace Location" folder is not separate from your "Data Location" folder. To address this error, select a "Workspace Location" folder on your computer that is completely separate from your "Data Location" folder.

5. Click the yellow "Add Collection" button to start the annotation process. You will be taken to the "Categorize Files" page, where you will be able to view a list of your study items.

4. Categorize Files

Each study item is assigned to a category (e.g., dataset or document). Each category has a different set of annotation properties (metadata) that are associated with it. DPT analyzes your study items and automatically assigns a category based on the file extension. You can either approve the category automatically selected by DPT or select a different category from the dropdown list. Follow the steps below to categorize your study items:

- 1. Each folder in your "Data Location" is listed on the left side of the table; the study items in that folder will be listed to the right. DPT will have performed a first pass at categorizing your items based on the file extension.
- 2. Check to see that each study item is appropriately categorized. For example, a Microsoft Excel (.xlsx) file will be categorized as a dataset but may actually be one of your study documents (e.g., codebook). Use your knowledge of the study items to ensure proper categorization. To recategorize an item, click the category dropdown list aligned to that item and select the correct categorization type. If you have a study item that is categorized as "Unknown", you will see an error message prompting you to recategorize accordingly. If during categorization you see a file that should not be included in the submission, you can select the "Ignore" option from the category dropdown list.
- 3. Every categorization must be approved; click the "Approve" checkbox for each item to confirm categorization. You can also click the yellow "Approve All" button located at the top of the table to approve all categories for the study items in the selected folder.

(?) NIH Eunice Kennedy Shriver National Institute of Child Health and Human Development DASH Data and Specimen Hub **Data Preparation Tool** Add Study **Categorize Files** Categorize Files In the DASH system, each item is assigned to a category (dataset, document, or image). Each category has a different set of properties (metadata) that are associated with it. The DPT will analyze your study items and assign a category based on the file extension Generate Templates Validate Templates ...ation: C:\Users\. Create Submission Files GPN LS Docs a FILE CATEGORY APPROVE Demo_LS_original_consented.csv Dataset GPN Protocol V9 10-19-2009.pdf Document

Dataset

Dataset

Dataset

Dataset

Dataset

Next

Document

GPN_LS_Dictionary_10-14-2010.xls

Manual Of Operations 9-19-2011.pdf

PD01_Screen_eligible_LS_original_consented.csv

PD02_Intake_Form1_LS_original_consented.csv

PD03 Intake Form2 LS original consented.csv

PD04_Intake_Form3_LS_original_consented.csv

PD05 CurrentMed LS original consented.csv

Return to Dashboard

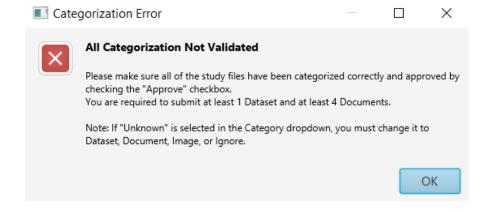
Figure 6: Categorize Files and Approve Categorizations

4. You also have the option to "Unapprove All." If you have clicked "Approve All" by mistake and received an error message regarding issues with your categorizations, clicking the yellow "Unapprove All" button allows you to review your categorizations and make the necessary changes. If you have added or removed files in your "Data Location" folder, use the yellow "Refresh" button to update the categorization list.

Save

5. Click on "Save" to save your work or "Next" to save and continue. If an item has not been categorized and approved, you will see an error message when you attempt to continue. You will not be able to move forward until all of your study items have been categorized and approved.

Figure 7: Categorization Error Message



6. Once all your items are categorized and approved, click the yellow "Next" button to navigate to the "Generate Templates" page

Note: You cannot recategorize your files once you click "Next." To recategorize files, you will need to start a new submission in DPT.

5. Generate Templates

On the "Generate Templates" page, you will have to answer if your study has biospecimens that are stored in the NICHD Contracted Biorepository that will be made available for request through NICHD DASH. If there are biospecimens available for sharing, select "Yes" and the DPT will open a field for you to enter the number of biospecimens associated with your study. You do not need to provide specific number, rounding up a number will suffice. The best practice is to overestimate the number so the template generated will contain enough rows for your available biospecimens.

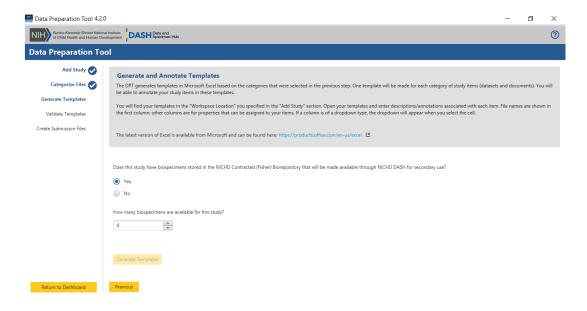


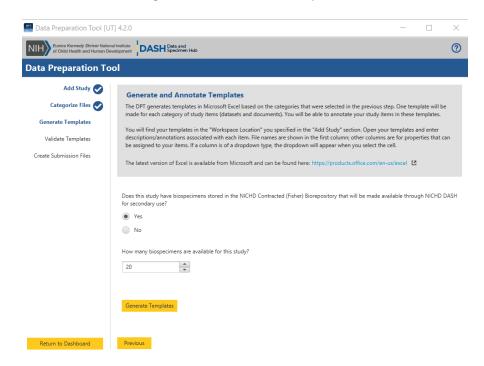
Figure 8: Biospecimen Number Selection

If your study does not have biospecimens available for sharing, select "No".

DPT generates annotation templates in Microsoft Excel based on the categories that were selected in the previous step; one template will be generated for each category of study items (datasets, documents, and biospecimens). For example, all your documents will be referenced in the DocumentTemplate.xlsx spreadsheet. You will be able to annotate your study items in the provided templates. Follow the steps below to generate your templates:

1. Click on the yellow "Generate Templates" button.

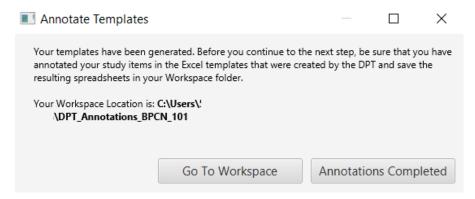
Figure 9: DPT – Generate Templates



2. When you click on "Generate Templates," a pop-up modal will appear with a path to your Workspace folder where DPT saved your templates and a reminder to annotate your study items. Templates are provided in Microsoft Excel format.

Note: A link for downloading Microsoft Excel is provided on the "Generate Templates" page

Figure 10: Workspace Location for Annotating Templates



3. Select the "Go To Workspace" button on the left and you will be directed to your workspace location to annotate your study items. Or, if you have already completed your study annotations, you may click on "Annotations Completed." **Do not click on "Annotations Completed" until you have competed your annotations.**

6. Annotate Templates

Templates are Microsoft Excel spreadsheets. Open your templates and enter descriptions/annotations associated with each study item. Each row in the template represents a study item. Filenames are shown in the first column and are listed in alphabetical order; do not make any changes to this column. If additional sorting is needed, sort rows 4 through the end of the list and leave the first three rows unsorted. Other columns in the templates are for properties that can be assigned to your items. Follow the steps below to annotate each template:

1. Find your templates. They have been saved by DPT in the "Workspace Location" you specified in the "Add Study" section.

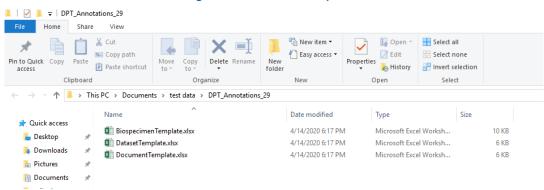


Figure 11: Find Your Templates

2. Open your templates and enter descriptions/annotations associated with each study item. Each row in the template represents a study item. Study item names are shown in the first column; other columns are for properties that can be assigned to your study items. Each column has a header that provides information about the specific property to be captured in this column. Some properties are required (e.g., Required: true) and therefore must be completed for each study item. Follow the instructions in the headers for each column to complete the annotation of each study item in each template.

Figure 12: Annotate Your Study Item Files

File Path	Title	Description	Type of Data	Dataset Format	Dataset Format	Keywords
Type: Text Required: true Instructions: The values have been populated for you. Please do not make any changes in this column.	Instructions: Enter a title for the file	Instructions: Enter a description of the	Required: false Instructions: Select one value from the drop-down list.	Instructions: Select one value from the drop-down list. If there isn't a suitable	Type: Text Required: If 'Other' is selected in previous column Instructions: Enter value for Dataset Format	Type: Text Required, folse Instructions: Enter keywords to describe the study or item. These keywords may be used in search and navigation to the file. Separate keywords by a semi-colon.
Demo_LS_original_consented.csv						
GPN_LS_Dictionary_10-14-2010.xls						
PD01_Screen_eligible_LS_original_co nsented.csv						
PD02_Intake_Form1_LS_original_cons ented.csv						
PD03_Intake_Form2_LS_original_cons ented.csv						

3. Some columns in each spreadsheet will have a dropdown list with annotations you can select to describe your study item. If an appropriate descriptor is not available in the dropdown list, you can choose "Other" and enter a new descriptor in the next column.

Figure 13: Dropdown Menu for Descriptors

File Path	Title	Description	Type of Data	Dataset Format	Dataset Format
Type: Text Required: true Instructions: The values have been populated for you. Please do not make any changes in this column.	Type: Text Required: true Instructions: Enter a title for the file	Type: Text Required: true Instructions: Enter a description of the file	Instructions: Select one value from the drop-down list.	Instructions: Select one value from the drop-down list. If there isn't a suitable	Type: Text Required: If 'Other' is selected in previous column Instructions: Enter value for Dataset Format
Demo_LS_original_consented.csv					•
GPN_LS_Dictionary_10-14-2010.xls				xlsx csv sas sas7bcat sas7bdat xls xpt	
PD01_Screen_eligible_LS_original_co nsented.csv				Other	
PD02_Intake_Form1_LS_original_cons ented.csv					
PD03_Intake_Form2_LS_original_cons ented.csv					

4. You can utilize the drag and drop feature in Microsoft Excel to copy information from row to row. This is especially helpful when you have many similar study items.

5. Be sure to annotate all templates and pay particular attention to the required properties. The required properties are labeled as such in the column headings (e.g., Required: true). The more descriptions/annotations you provide, the easier it will be for others to find your study items in NICHD DASH.

Note: For a new study submission, the "Publicly Available" column in the Document Template is a required property where you will identify which documents will be publicly available on the study overview page in NICHD DASH. Four document types (e.g., Study Protocol, Codebook/Variable Dictionary, Data Collection Instruments, and De-identification Methodology) are required to be publicly available to ensure meaningful use of your study data and to prevent misuse, misinterpretation, and confusion. You are encouraged to upload additional study documents and make them publicly available. Datasets will not be publicly available.

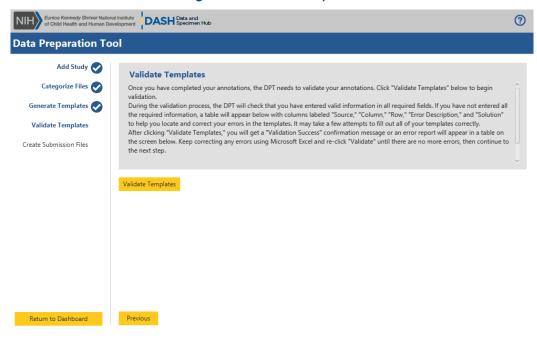
- 6. Save your spreadsheet in the same folder and using the same file name as the original template; do not "save as" a new file name or save your templates in a different location.
- 7. Once you have completed your study annotations, open DPT and click on "Annotations Completed." **Do not click on "Annotations Completed" until you have competed your annotations.**

7. Validate Templates

During the validation process, the DPT will check that you entered information in all required fields in the templates. If you have not entered all of the required information, you will see an error message with information to help you locate and correct your errors in the templates. It may take a few attempts to annotate your templates correctly.

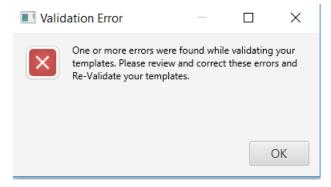
Follow the steps below to validate your templates:

Figure 14: Validate Templates



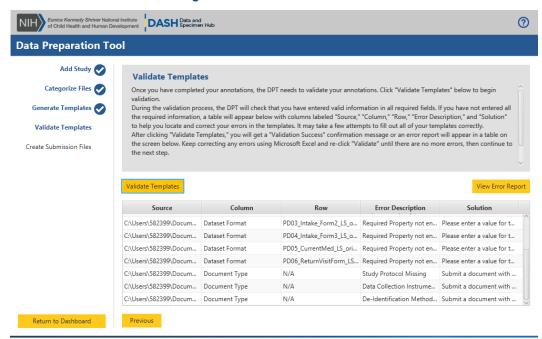
- 1. Click on the yellow "Validate Templates" button.
- 2. During validation, DPT checks that you have entered information in all the required fields of your templates.
- 3. If you have not entered all the required information, you will receive the following error message:

Figure 15: Validation Error



4. DPT will help you locate missing values by generating a table with columns labeled "Source," "Column," "Row," "Error Description," and "Solution."

Figure 16: Validation Error Table

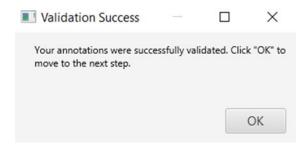


5. Return to your templates in the "Workspace Location" you specified in the "Add Study" section to fix any annotations with validations errors. It may take a few attempts to correct the validation errors in all your templates correctly. Complete the changes and save the templates.

Note: Save your spreadsheet in the same folder and using the same file name as the original template; do not "save as" a new file name or save your templates in a different location.

- 6. You will either receive a "Validation Success" message or a message detailing a list of remaining errors.
- 7. Continue correcting any annotation errors in your templates and click "Validate Templates" until there are no more errors, and then click "OK."
- 8. When you see the "Validation Success" message, you are ready to upload your study items for submission to NICHD DASH.

Figure 17: Validation Success



8. Upload Submission Files

You are now ready to submit your annotations and study items to NICHD DASH. Please click the yellow "Upload Submission" button to encrypt and securely upload your study collection. You will then be required to connect to the internet and enter your NICHD DASH User Credentials to select the study submission that these study items belong to. After your encrypted submission has been uploaded, you can log in to NICHD DASH to complete your study submission.

The following steps will guide you through uploading your submission files.

Data Preparation Tool

Add Study
Categorize Files
Cick on the "Upload Submission" button.

Validate Templates
Upload Submission

Upload Submission

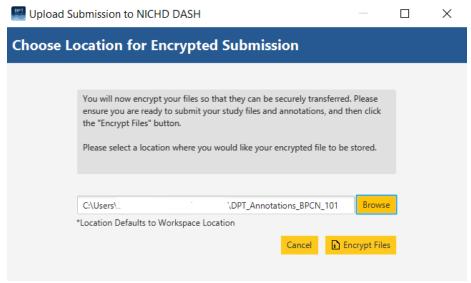
Upload Submission to NICHD DASH.

Click on the "Upload Submission" button.

Figure 18: Upload Submission Files

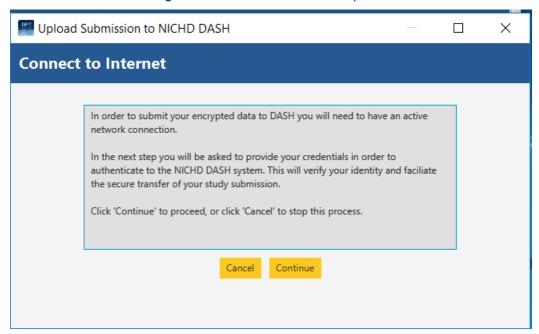
1. Click on the yellow "Upload Submission" button. By default, DPT will auto-populate the Location field for the encrypted study items with the same folder you selected as your workspace location in section 3 "Add Collection". If you would like to change the default location, click "Browse" and select the folder where you have saved all study items (datasets and documents) associated with your study in a single, easy-to-find folder on your computer. Click "Encrypt Files" to create a set of encrypted submission files in the specified location.

Figure 19: Choose Location for Encrypted Submission



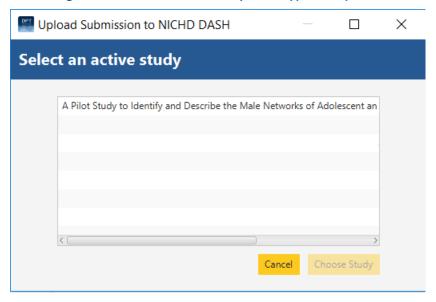
2. A pop-up modal will prompt you to ensure you are connected to the internet for the upload process. Click "Continue" once you have confirmed you are connected to the internet.

Figure 20: Internet Connection Required



3. You will be prompted to log in to the DPT with your NICHD DASH email and password. Enter your credentials and click "Login."

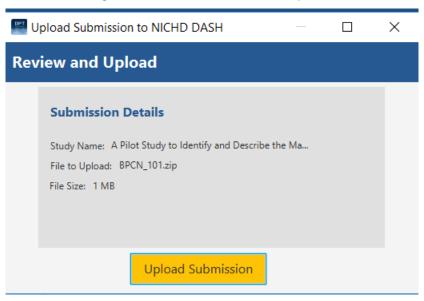
Figure 21: Select an Active Study to Encrypt and Upload



4. A pop-up modal will display the active studies in your "Submission Dashboard" in NICHD DASH. Select the study you want to upload the DPT encrypted study items to and then click "Choose Study."

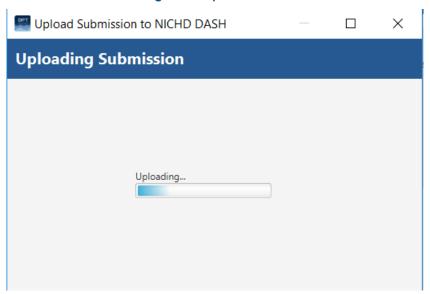
5. A pop-up modal will provide you with one last chance to review your submission. Confirm that you have selected the correct study. Click "Upload Submission." If you have selected the incorrect study, close the modal and go back to the "Upload Submission" page.

Figure 22: Review Submission and Upload



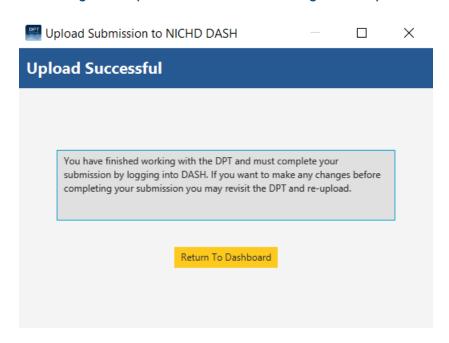
6. The upload status bar will be displayed in a pop-up modal.

Figure 23: Upload Status



7. Once the upload is completed, a message will be displayed informing you that the upload was successful. If you were working on submitting a new study, you can now log in to the NICHD DASH website and complete the submission process as described in the "Submission Tutorial" under the "Submission – Review and Submit" section.

Figure 24: Upload Successful for Submitting New Study



If you were adding items to an existing study, no further action is needed. A NICHD DASH Curator will review your additional items and you will receive an email notification of approval or a denial, along with justification.