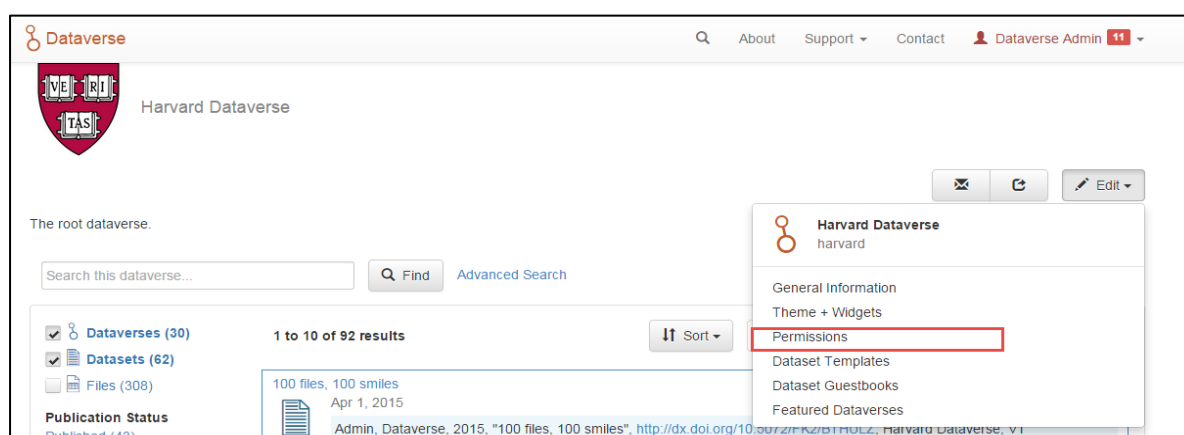# Dataverse, Dataset and File Management

A. Dataverse Management

Edit Dataverse

To edit your dataverse, navigate to your dataverse homepage and select the "Edit Dataverse" button, where you will be presented with the following editing options:

- *General Information*: edit name, identifier, category, contact email, affiliation, description, Metadata Elements, and facets for your dataverse
- *Theme*: upload a logo for your dataverse, add a link to your department or personal website, and select colors for your dataverse in order to brand it
- *Widgets*: get code to add to your website to have your dataverse display on it
- Permissions: give Dataverse users permissions to your dataverse, i.e.-can edit datasets, and see which users already have which permissions for your dataverse
- *Dataset Templates*: these are useful when you have several datasets that have the same information in multiple metadata fields that you would prefer not to have to keep manually typing in
- *Dataset Guestbooks*: allows you to collect data about who is downloading the files from your datasets
- *Featured Dataverses*: if you have one or more dataverses, you can use this option to show them at the top of your dataverse page to help others easily find interesting or important dataverses
- **Delete Dataverse**: you are able to delete your dataverse as long as it is not published and does not have any draft datasets

General Information

The General Information page is how you edit the information you filled in while creating your dataverse. If you need to change or add a contact email address, this is the place to do it. Additionally, you can update the metadata elements used for datasets within the dataverse, change which metadata fields are hidden, required, or optional, and update the facets you would like displayed for browsing the dataverse. If you plan on using templates, you need to select the metadata fields on the General Information page.

Tip: The metadata fields you select as required, will appear on the Create Dataset form when someone goes to add a dataset to the dataverse.

Theme

The Theme feature provides you with a way to customize the look of your dataverse. You can decide either to use the customization from the dataverse above yours or upload your own image file. Supported image types are JPEG, TIFF, or PNG and should be no larger than 500 KB. The maximum display size for an image file in a dataverse's theme is 940 pixels wide by 120 pixels high. Additionally, you can select the colors for the header of your dataverse and the text that appears in your dataverse. You can also add a link to your personal website, the website for your organization or institution, your department, journal, etc.

Widgets

The Widgets feature provides you with code for you to put on your personal website to have your dataverse displayed there. There are two types of Widgets for a dataverse, a Dataverse Search Box widget and a Dataverse Listing widget. From the Widgets tab on the Theme + Widgets page, you can copy and paste the code snippets for the widget you would like to add to your website. If you need to adjust the height of the widget on your website, you may do so by editing the heightPx=500 parameter in the code snippet.
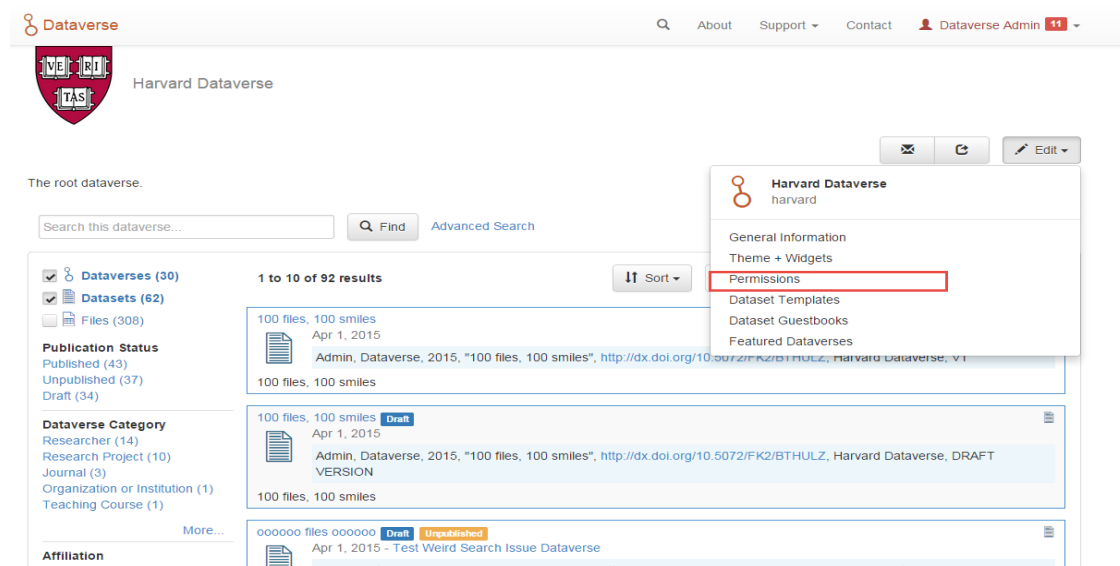
Dataverse Search Box Widget

The Dataverse Search Box Widget will add a search box to your website that is linked to your dataverse. Users are directed to your dataverse in a new browser window, to display the results for search terms entered in the input field.

Dataverse Listing Widget

The Dataverse Listing Widget provides a listing of all your dataverses and datasets for users to browse, sort, filter and search. When someone clicks on a dataverse or dataset in the widget, it displays the content in the widget on your website. They can download data files directly from the datasets within the widget. If a file is restricted, they will be directed to your dataverse to log in, instead of logging in through the widget on your website.

Permissions

When you access a dataverse's permissions page, you will see there are three sections: Permissions, Users/Groups, and Roles.

Clicking on Permissions will bring you to this page:

By clicking on the Edit Access button, you are able to change the settings allowing no one or anyone to add either dataverses or datasets to a dataverse.

The Edit Access pop up allows you to also select if someone adding a dataset to this dataverse should be allowed to publish it (Curator role) or if the dataset will be submitted to the administrator of this dataverse to be reviewed then published (Contributor role). These Access settings can be changed at any time.

Assign Role

You can also give access to a Dataverse user to allow them to access an unpublished dataverse as well as other roles. To do this, click on the Assign Roles to Users/Groups button in the Users/Groups section. You can also give multiple users the same role at one time. This roles can be removed at any time.
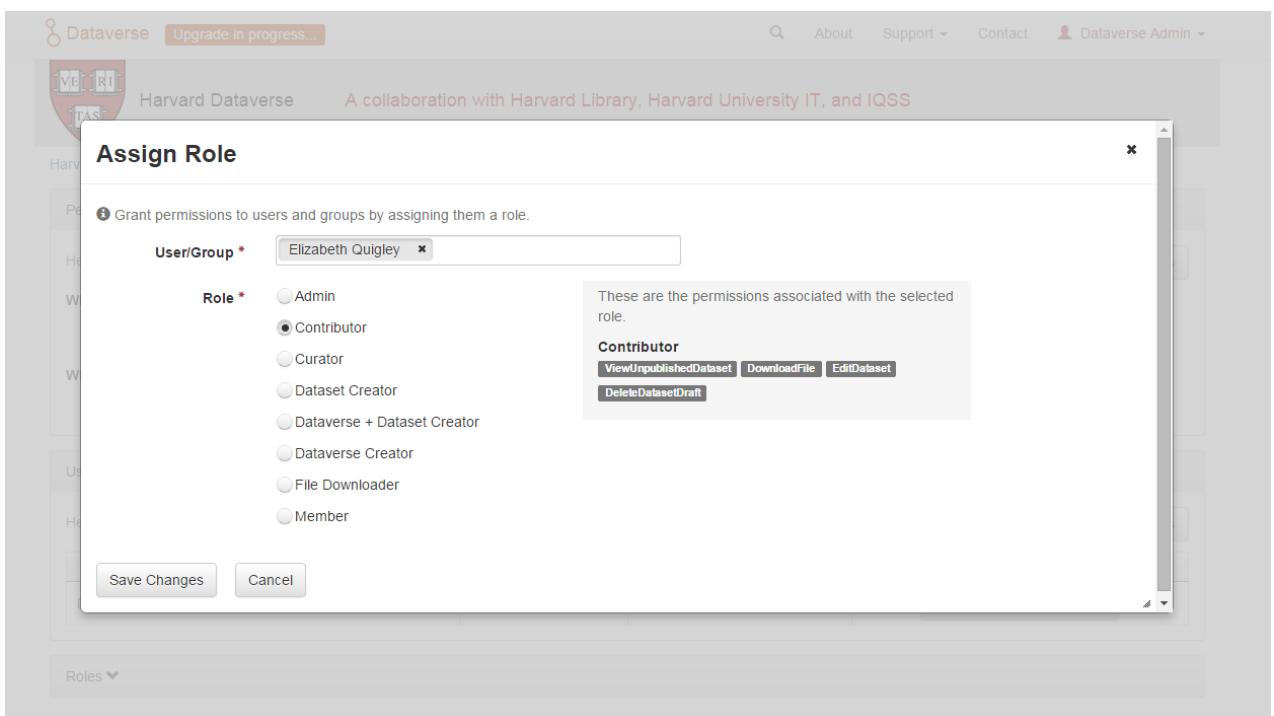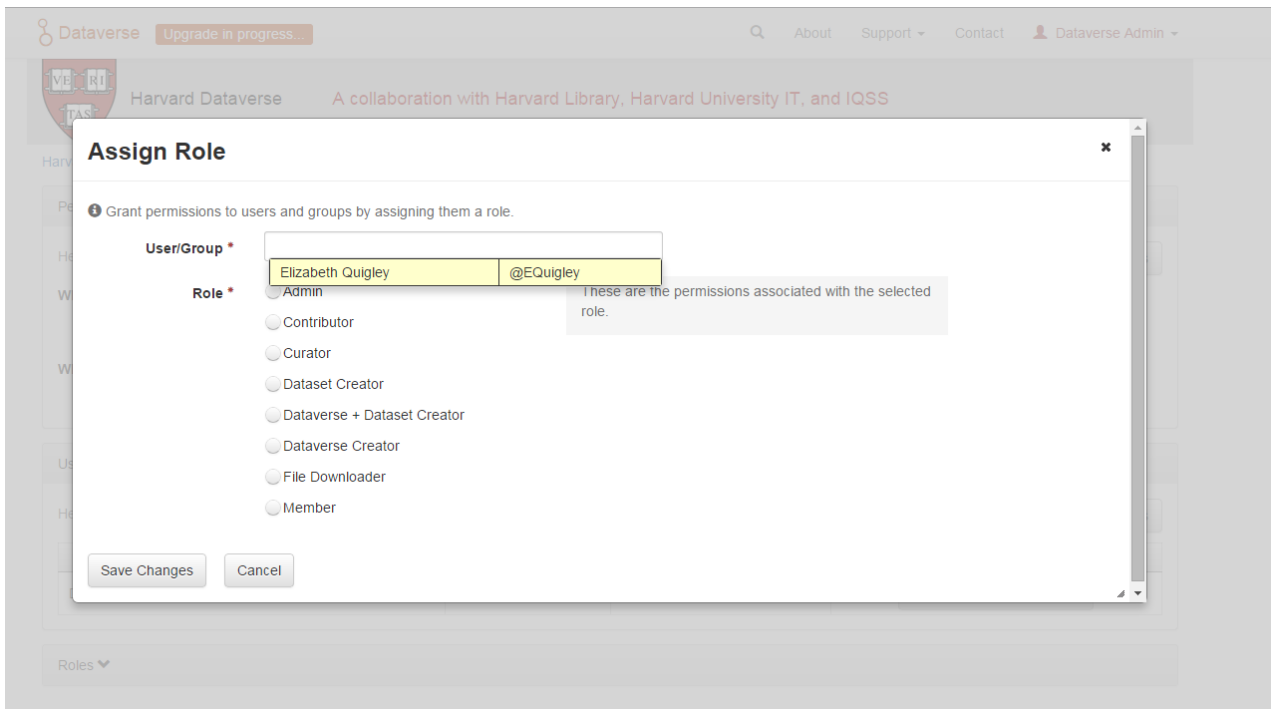
Dataset Templates

Templates are useful when you have several datasets that have the same information in multiple metadata fields that you would prefer not to have to keep manually typing in or want to use a custom set of Terms of Use and Access for multiple datasets in a dataverse. In Dataverse 4.0, templates are created at the dataverse level, can be deleted (so it does not show for future datasets), set to default (not required), or can be copied so you do not have to start over when creating a new template with similiar metadata from another template. When a template is deleted, it does not impact the datasets that have used the template already.

How do you create a template?

1. Navigate to your dataverse, click on the Edit Dataverse button and select Dataset Templates.

2. Once you have clicked on Dataset Templates, you will be brought to the Dataset Templates page. On this page, you can 1) decide to use the dataset templates from your parent dataverse 2) create a new dataset template or 3) do both.
3. Click on the Create Dataset Template to get started. You will see that the template is the same as the create dataset page with an additional field at the top of the page to add a name for the template.
4. After adding information into the metadata fields you have information for and clicking Save and Add Terms, you will be brought to the page where you can add custom Terms of Use and Access. If you do not need custom Terms of Use and Access, click the Save Dataset Template, and only the metadata fields will be saved.
5. After clicking Save Dataset Template, you will be brought back to the Manage Dataset Templates page and should see your template listed there now with the make default, edit, view, or delete options.
6. A dataverse does not have to have a default template and users can select which template they would like to use while on the Create Dataset page.
7. You can also click on the View button on the Manage Dataset Templates page to see what metadata fields have information filled in.

* Please note that the ability to choose which metadata fields are hidden, required, or optional is done on the General Information page for the dataverse.

Dataset Guestbooks

Guestbooks allow you to collect data about who is downloading the files from your datasets. You can decide to collect account information (username, given name & last name, affiliation, etc.) as well as create custom questions (e.g., What do you plan to use this data for?). You are also able to download the data collected from the enabled guestbooks as Excel files to store and use outside of Dataverse.

How do you create a guestbook?

1. After creating a dataverse, click on the Edit Dataverse button and select Dataset Guestbook
2. By default, guestbooks created in the dataverse your dataverse is in, will appear. If you do not want to use or see those guestbooks, uncheck the checkbox that says Include Guestbooks from Root Dataverse.
3. To create a new guestbook, click the Create Dataset Guestbook button on the right side of the page.
4. Name the guestbook, determine the account information that you would like to be required (all account information fields show when someone downloads a file), and then add Custom Questions (can be required or not required).
5. Hit the Create Dataset Guestbook button once you have finished.

What can you do with a guestbook? After creating a guestbook, you will notice there are several options for a guestbook and appear in the list of guestbooks.

- If you want to use a guestbook you have created, you will first need to click the button in the Action column that says Enable. Once a guestbook has been enabled, you can go to the License + Terms for a dataset and select a guestbook for it.
- There are also options to view, copy, edit, or delete a guestbook.
- Once someone has downloaded a file in a dataset where a guestbook has been assigned, an option to download collected data will appear.

Featured Dataverses

Featured Dataverses is a way to display sub dataverses in your dataverse that you want to feature for people to easily see when they visit your dataverse.

Click on Featured Dataverses and a pop up will appear. Select which sub dataverses you would like to have appear.

Note: Featured Dataverses can only be used with published dataverses.

Linked Dataverses + Linked Datasets

Currently, the ability to link a dataverse to another dataverse or a dataset to a dataverse is a super user only feature.

If you need to have a dataverse or dataset linked in the DR-NTU (Data) Dataverse installation, please contact rdm@ntu.edu.sg.

Publish Your Dataverse

Once your dataverse is ready to go public, go to your dataverse page, click on the "Publish" button on the right hand side of the page. A pop-up will appear to confirm that you are ready to actually Publish, since once a dataverse is made public, it can no longer be unpublished.

B. Dataset Management

Edit Dataset
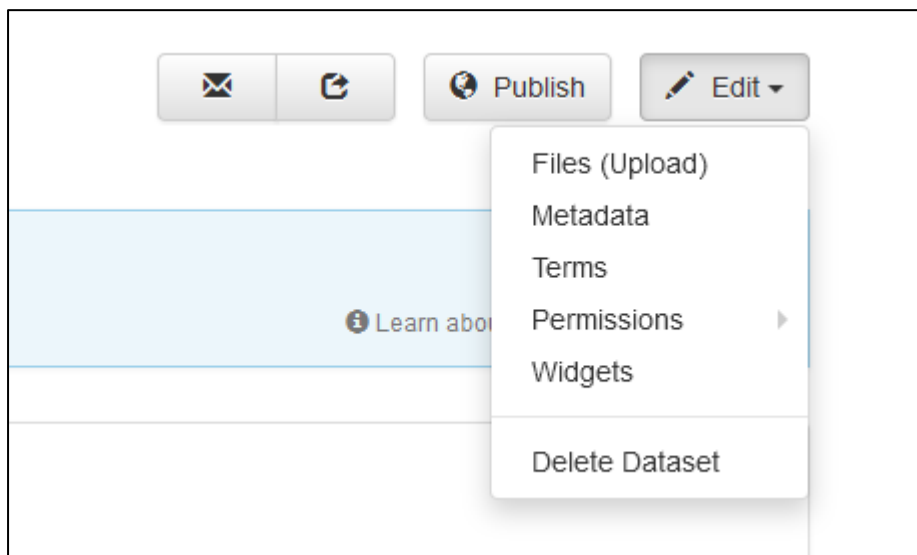
Files (Upload)

To upload new files to a dataset, go to the dataset you want to update and click on the Upload Files button in the files tab. From there you will be brought to the Upload Files page for the dataset. Once you have uploaded files, you will be able to edit the file metadata, restrict, add tags, or delete them before saving.

Metadata

See: Supported Metadata and References

Terms

In the Terms tab, which can also be found by clicking on the Edit dropdown button of a Dataset, you can setup how users can use your data once they have downloaded it (CC-BY-NC or custom Terms of Use), how they can access your data if you have files that are restricted (terms of access), and enable a Guestbook for your dataset so that you can track who is using your data and for what purposes. These are explained in further detail below:

CC-BY-NC + Dataset Terms of Use

CC-BY-NC facilitates reuse and extensibility of research data. Our Community Norms as well as good scientific practices expect that proper credit is given via citation. If you are unable to give your datasets a CC-BY-NC you may enter your own custom Terms of Use for your Datasets.

* **Legal Disclaimer:** these Community Norms are not a substitute for the CC-BY-NC or custom terms and licenses applicable to each dataset. Please be advised that the Community Norms are not a binding contractual agreement, and that downloading datasets from Dataverse does not create a legal obligation to follow these policies.

Setting up Custom Terms of Use for Datasets

If you are unable to use a CC-BY-NC for your datasets you are able to set your own custom terms of use. To do so, select "No, do not apply CC-By-NC- "Creative Commons license - attribution required and for non-commercial use only" and a Terms of Use textbox will show up allowing you to enter your own custom terms of use for your dataset. To add more information about the Terms of Use, click on "Additional Information [+]".

Here is an example of a Data Usage Agreement for datasets that have de-identified human subject data.

Restricted Files + Terms of Access

If you restrict any files in your dataset, you will be prompted by a pop-up to enter Terms of Access for the data. This can also be edited in the Terms tab or selecting Terms in the "Edit" dropdown button in the dataset. You may also allow users to request access for your restricted files by enabling "Request Access". To add more information about the Terms of Access, click on "Additional Information [+]".

Guestbook

This is where you will enable a particular Guestbook for your dataset, which is setup at the Dataverse-level. For specific instructions please visit the Dataset Guestbooks section in this guide.

Permissions

Dataset-Level

Dataset permissions are located under Permissions in the Edit button on a dataset page. The dataset permissions page has two sections: Users/Groups and Roles.

To give someone access to view your unpublished dataset or edit your published or unpublished dataset, click on the Assign Roles to Users/Groups button in the Users/Groups section.

File-Level

If you have restricted specific files the file-level permissions is where you will need to go to grant users/groups access to specific restricted files. Dataset file permissions are located under Permissions in the Edit button on a dataset page. The file permissions page has two sections: Users/Groups and Files.

To give someone access to your restricted files, click on the Grant Access to Users/Groups button in the Users/Groups section.

Widgets

The Widgets feature provides you with code for your personal website so your dataset can be displayed. There are two types of Widgets for a dataset: the Dataset Widget and the Dataset Citation Widget. The Widgets are found by going to your dataset page, clicking the Edit button (the one with the pencil icon) and selecting Widgets from the dropdown menu.

On the Widgets page, you can copy and paste the code snippets for the widget you would like to add to your website. If you need to adjust the height of the widget on your website, you may do so by editing the heightPx=500 parameter in the code snippet.

Dataset Widget

The Dataset Widget allows the citation, metadata, files and terms of your dataset to be displayed on your website. When someone downloads a data file in the widget, it will download directly from the datasets on your website. If a file is restricted, they will be directed to your dataverse to log in, instead of logging in through the widget on your site.

To edit your dataset, you will need to return to the Dataverse repository where the dataset is stored. You can easily do this by clicking on the link that says "Data Stored in (Name) Dataverse" found in the bottom of the widget.

Dataset Citation Widget

The Dataset Citation Widget will provide a citation for your dataset on your personal or project website. Users can download the citation in various formats by using the Cite Data button. The persistent URL in the citation will direct users to the dataset in your dataverse.

Publish Dataset

When you publish a dataset (available to an Admin, Curator, or any custom role which has this level of permission assigned), you make it available to the public so that other users can browse or search for it. Once your dataset is ready to go public, go to your dataset page and click on the "Publish" button on the right hand side of the page. A pop-up will appear to confirm that you are ready to actually Publish since once a dataset is made public it can no longer be unpublished.

Whenever you edit your dataset, you are able to publish a new version of the dataset. The publish dataset button will reappear whenever you edit the metadata of the dataset or add a file.

Note: Prior to publishing your dataset the Data Citation will indicate that this is a draft but the "DRAFT VERSION" text will be removed as soon as you Publish.

Submit for Review

If you have a Contributor role (can edit metadata, upload files, and edit files, edit Terms, Guestbook, and Submit datasets for review) in a Dataverse you can submit your dataset for review when you have finished uploading your files and filling in all of the relevant metadata fields. To Submit for Review, go to your dataset and click on the "Submit for Review" button, which is located next to the "Edit" button on the upper-right. Once Submitted for Review: the Admin or Curator for this Dataverse will be notified to review this dataset before they decide to either "Publish" the dataset or "Return to Author". If the dataset is published the contributor will be notified that it is now published. If the dataset is returned to the author, the contributor of this dataset will be notified that they need to make modifications before it can be submitted for review again.

Private URL for Reviewing an Unpublished Dataset

To share an unpublished dataset using Private URL

Creating a Private URL for your dataset allows you to share your dataset (for viewing and downloading of files) before it is published to a wide group of individuals who may not have a user account on Dataverse. Anyone you send the Private URL to will not have to log into Dataverse to view the dataset.

1. Go to your unpublished dataset
2. Select the "Edit" button
3. Select "Private URL" in the dropdown menu
4. In the pop-up select "Create Private URL"
5. Copy the Private URL which has been created for this dataset and it can now be shared with anyone you wish to have access to view or download files in your unpublished dataset.

To disable a Private URL

If ever you had shared a Private URL to your dataset and wish to revoke access, follow the same steps as above until step #3 but in the pop-up select "Disable Private URL".

Dataset Versioning

Versioning is important for long term-research data management where metadata and/or files are updated over time.

Once you have published a dataset, any metadata or file changes (e.g, by uploading a new file, changing file metadata, adding or editing metadata) will be tracked in our versioning feature. For example if you were at version 1 of your dataset, and you edit your dataset a new draft version of this dataset will be created. To get to the already published version 1 of your dataset, click on the "View Dataset Versions" button on the top left section of your dataset. To go back to the unpublished version click on the same button. Once you are ready to publish this new version of your dataset, select the "Publish Dataset" button on the top right side of the page. If you were at version 1 of your dataset, and depending on the types of changes you have made, you will be asked to select to publish your draft as either version 1.1 or version 2.0 (**important note**: if you add a file, your dataset will automatically be bumped up to a major version (example: if you were at 1.0 you will go to 2.0).
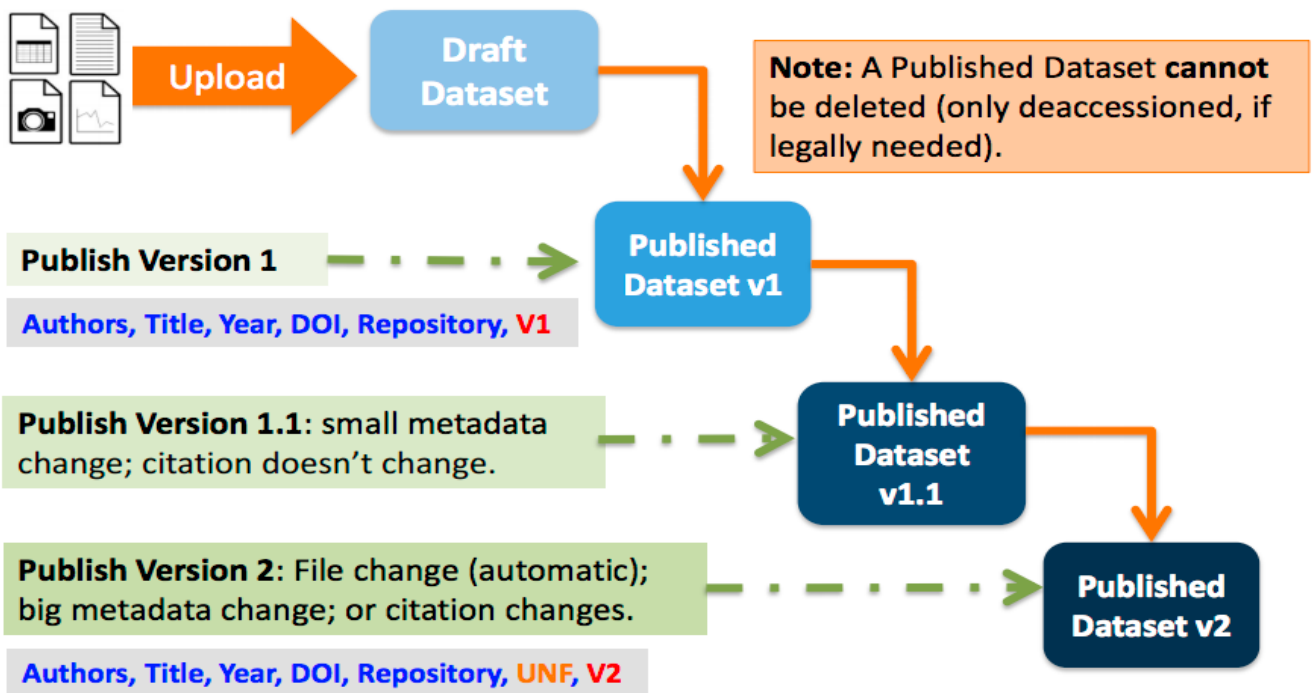
Version Details

To view what has exactly changed starting from the originally published version to any subsequent published versions: click on the Versions tab on the dataset page to see all versions and changes made for that particular dataset. Once you have more than one version (can be version 1 and a draft), you can click the Show Details link in the Versions tab to learn more about the metadata fields and files that were either added or edited.

If you have more than two versions of a dataset, you can select any two versions to compare the differences between them. After selecting two versions, click on the "Show Differences" button to see the version differences details.

Deaccession Your Dataset [not recommended]

Deaccessioning a dataset or a version of a dataset is a very serious action that should only occur if there is a legal or valid reason for the dataset to no longer be accessible to the public. If you absolutely must deaccession, you can deaccession a version of a dataset or an entire dataset. To deaccession, go to a dataset you've already published (or add a new one and publish it), click on Edit Dataset, then Deaccession Dataset. If you have multiple versions of a dataset, you can select here which versions you want to deaccession or choose to deaccession the entire dataset. You must also include a reason as to why this dataset was deaccessioned from a dropdown list of options. There is also a free-text box to add more details as to why this was deaccessioned. If the dataset has moved to a different repository or site you are encouraged to include a URL (preferably persistent) for users to continue to be able to access this dataset in the future.

If you deaccession the most recently published version of the dataset but not all versions of the dataset, you are able to go in and create a new draft for the dataset. For example, you have a version 1 and version 2 of a dataset, both published, and deaccession version 2. You are then able to edit version 1 of the dataset and a new draft will be created.

Important Note: A tombstone landing page with the basic citation metadata will always be accessible to the public if they use the persistent URL (Handle or DOI) provided in the citation for that dataset. Users will not be able to see any of the files or additional metadata that were previously available prior to deaccession.

C. File Management

Edit Files
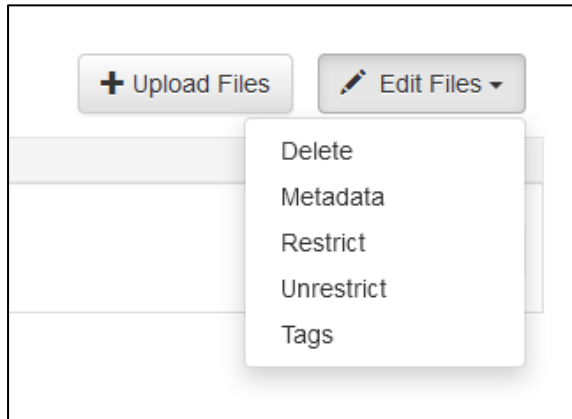
Go to the dataset you would like to edit where you will see the listing of files. Select the files you would like to edit by using either the Select All checkbox or individually selecting files. Next, click on the Edit button above the files and select if you would like to:

- Delete the selected files
- Edit the file metadata (file name, description) for the selected files
- Restrict the selected files
- Unrestrict the selected files (only if the selected files are restricted)
- Add tags to the selected files

All of these actions, besides editing file metadata, will happen within this page and not bring you to another page. If you restrict files, you will also be asked to fill out the Terms of Access for the files. If Terms of Access already exist, you will be asked to confirm them.

File Tags

The File Tags are comprised of custom, category (i.e. Documentation, Data, Code) and tabular data tags (i.e. Event, Genomics, Geospatial, Network, Panel, Survey, Time Series). Use the dropdown select menus as well as the custom file tag input to apply these tags to the selected files. There is also a Delete Tags feature that, if checked, will allow you to delete unused file tags within that dataset.

File Handling + Uploading

The file types listed below are supported by additional functionality, which can include downloading in different formats, subsets, file-level metadata preservation, file-level data citation; and exploration through data visualization and analysis.

Tabular Data Files

Files in certain formats - Stata, SPSS, R, Excel(xlsx) and CSV - may be ingested as tabular data (see "Tabular Data Ingest" section for details). Tabular data files can be further explored and manipulated with TwoRavens - a statistical data exploration application integrated with Dataverse. It allows the user to run statistical models,

view summary statistics, download subsets of variable vectors and more. To start, click on the "Explore" button, found next to each relevant tabular file (the application will be opened in a new window). To download subsets of variables click on the "Download" button found next to a relevant tabular file and select "Data Subset" in the dropdown menu. You will then be able to create your subset using the interface opened in a new window (this functionality is also provided by the TwoRavens project). See the TwoRavens documentation section for more information.

For example, the ingest functionality for tabular files in the DR-NTU (Data), a file can only be up to 2GB in size. To use the ingest functionality for RData files, a file can only be up to 1MB in size, however, to upload a RData file without using ingest, a file can be up to 2GB in size.

Additional download options available for tabular data (found in the same drop-down menu under the "Download" button):

- As tab-delimited data (with the variable names in the first row);
- The original file uploaded by the user;
- Saved as R data (if the original file was not in R format);
- Variable Metadata (as a DDI Codebook XML file);
- Data File Citation (currently in either RIS or EndNote XML format);
- All of the above, as a zipped bundle.
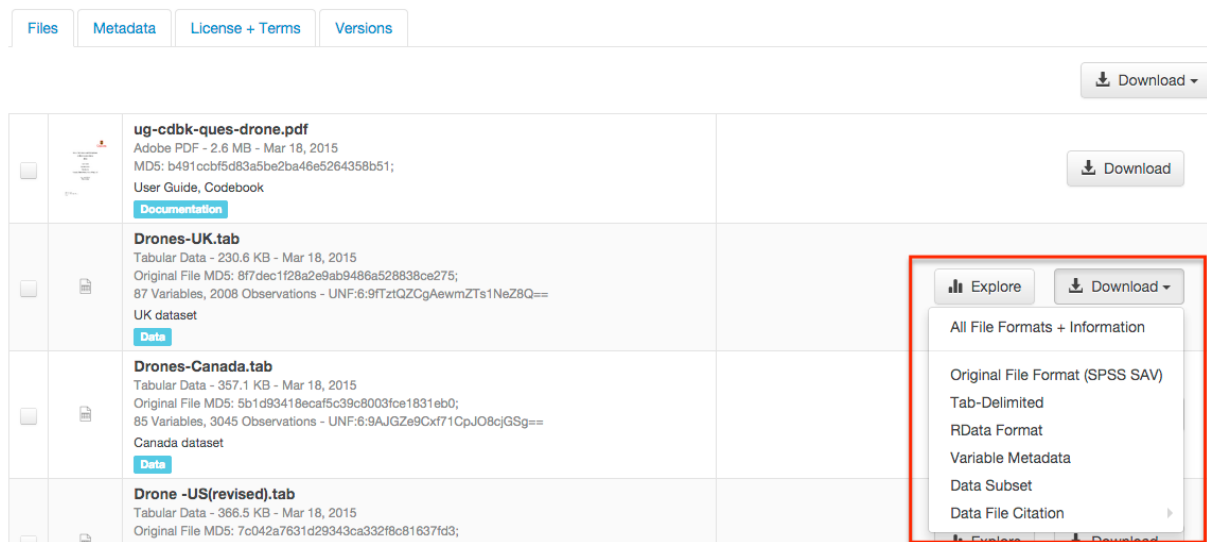
Geospatial

Geospatial shapefiles can be further explored and manipulated through our integration with WorldMap, a geospatial data visualization and analysis tool developed by the Center for Geographic Analysis at Harvard University. A shapefile is a set of files, often uploaded/transferred in .zip format. This set may contain up to 15 files. A minimum of 3 specific files (.shp, .shx, .dbf) are needed to be a valid shapefile and a 4th file (.prj) is required for WorldMap–or any type of meaningful visualization.

For ingest into Dataverse and connecting to WorldMap, these 4 files are the minimum required:

- .shp - shape format; the feature geometry itself
- .shx - shape index format; a positional index of the feature geometry to allow seeking forwards and backwards quickly
- .dbf - attribute format; columnar attributes for each shape, in dBase IV format
- .prj - projection format; the coordinate system and projection information, a plain text file describing the projection using well-known text format

For a zipped shapefile, we require 4 files with these extensions. Other files may be included within the zipped shapefile, but they are not required:

- .shp
- .shx
- .prj
- .dbf

For example, if these files were included within a .zip, the "Map Data" button would appear:

- subway_line.shp
- subway_line.shx
- subway_line.prj
- subway_line.dbf

Once you publish your dataset with your shape files, you will be able to use the "Map Data" button using GeoConnect to visualize and manipulate these files for users to Explore this geospatial data using the WorldMap interface. Please note: In order to map your data file, a copy will be sent to Harvard's WorldMap platform. You have the ability to delete any maps, and associated data, from the Harvard WorldMap platform, at any time.

Astronomy (FITS)

Metadata found in the header section of Flexible Image Transport System (FITS) files are automatically extracted by Dataverse, aggregated and displayed in the Astronomy Domain-Specific Metadata of the Dataset that the file belongs to. This FITS file metadata, is therefore searchable and browsable (facets) at the Dataset-level.

Compressed Files

Compressed files in zip format are unpacked automatically. If it fails to unpack, for whatever reason, it will upload as is. If the number of files inside are more than a set limit (1,000), you will get an error message and the file will uploads as is.

Support for unpacking tar files will be added when this ticket is closed: https://github.com/IQSS/dataverse/issues/2195.

Advanced Options

There are several advanced options available for certain file types.

- Image files: jpgs, pngs, and tiff files are able to be selected as the default thumbnail for a dataset. The selected thumbnail will appear on the search result card for that dataset.
- SPSS files: SPSS files can be tagged with the language they were originally coded in. This is found by clicking on Advanced Options and selecting the language from the list provided.