SSHOC

social sciences & humanities open cloud

Dataverse End User Manual

SSHOC, "Social Sciences and Humanities Open Cloud", has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-04-2018, Grant Agreement #823782.

Research and Innovation Action

Social Sciences & Humanities Open Cloud

Project Number: 823782 Start Date of Project: 01/01/2019 Duration: 40 months

Dataverse End User Manual

The information in this document reflects only the author's views and the European Community is not liable for any use that may be made of the information contained therein. The information in this document is provided 'as is' without guarantee or warranty of any kind, express or implied, including but not limited to the fitness of the information for a particular purpose. The user thereof uses the information at his/ her sole risk and liability. This deliverable is licensed under a Creative Commons Attribution 4.0 International License.

Author List

Organisation	Name	Contact Information
AUSSDA	Veronika Heider	veronika.heider@univie.ac.at
DANS-KNAW	Simon Saldner	simon.saldner@dans.knaw.nl
DANS-KNAW	Laura Huis in 't Veld	laura.huisintveld@dans.knaw.nl
DANS-KNAW	Marion Wittenberg	marion.wittenberg@dans.knaw.nl

Summary

Contents and aims: This Dataverse manual aims to provide a practical, user-friendly, and concise guide intended for typical end users of Dataverse. The manual includes instructions for the essential functions of Dataverse such as creating, managing and sharing Dataverse collections and datasets; a technical setup guide; legal questions, including usage agreements; as well as references to additional resources such as technical setup guides. Please refer to the table of contents below for a full overview of the topics covered. This manual is not intended to cover advanced functions such as those for administrators or developers, but throughout the manual, references to further reading that do cover more advanced topics will be provided.

Use: We recommend that users of this manual refer to the table of contents below to find instructions for their topic of interest. Screenshots of the Dataverse interface are provided throughout the manual. An overview of these can be found in the table of figures at the end of this document.

Sources and acknowledgements: This manual is based on the exhaustive User Guide by the Dataverse Project[1], the AUSSDA Self-Deposit Manual[2], as well as other guides cited in this manual. This manual aims to provide a condensed, user-friendly and updated version of these existing guides. This version of the manual was created January 2022, and applies most closely to Dataverse versions 5.6 and 5.7.

[1] https://guides.dataverse.org/en/5.6/user/index.html

[2] https://aussda.at/fileadmin/user_upload/p_aussda/Documents/Self-Deposit_Manual_v1_0.pdf

Abbreviations and Acronyms

API	Application Programming Interface
AUSSDA	Austrian Social Science Data Archive
CMM	CESSDA Metadata Model
DANS	Data Archiving and Networked Services
DDI	Data Documentation Initiative
ELSST	European Language Social Science Thesaurus
EOSC	European Open Science Cloud

Table of Contents

1. Account Creation and Management *3*

1.1. Sign up / Login **3**

1.2. Account Information **4**

2. Finding and using data *5*

2.1. Searching and Browsing **5**

2.1.1. Searching all contents of a Dataverse installation 5

2.1.2. Sorting and viewing search results 5

3. Using Data *7*

3.1. Searching within a dataset **7**

3.2. View Files **7**

3.3. Tree View **8**

3.4. Cite Data **8**

3.5. Download (meta)data **9**

4. Dataverse Management *12*

4.1. Creating a New Dataverse **12**

4.2. Editing a Dataverse **13**

4.2.1. Theme + Widgets 14

4.2.2. Permissions 15

4.2.3. Groups 17

4.2.4. Dataset Templates 17

4.2.5. Dataset Guestbooks 17

4.2.6. Featured Dataverses 19

4.3. Publishing **19**

5. Dataset + File Management *20*

5.1. Creating a Dataset **20**

5.2. Editing a Dataset **21**

5.2.1. Metadata 21

5.2.2. Files 22

Uploading 22

Data File Ingest 23

File metadata 24

File Tags 26

File Restrictions 26

Replace File 27

Provenance and other file features 27

5.2.3. Terms 28

Terms of Use 28

Terms of Access 29

Guestbook 30

5.2.4. Roles & Permissions 30
Access requests 31
5.2.5. Private URL 32
5.2.6. Thumbnails + Widgets 33
5.2.7. Delete Dataset 34
5.2.8. Deaccession 34

5.3. Publish or Submit for Review 35

5.4. Versions 36

6. References / Additional resources 38

7. List of Figures 39

1. Account Creation and Management

1.1. Sign up / Login

The first step to take when using Dataverse is the creation of a user account. There are two forms of registration. You either register a local account under 'Sign up' or use your institution's login data under 'Log in'.

If you choose 'Sign up', you can choose a username and password and create a local account in Dataverse. The next time you visit Dataverse, you can use that username and password to access your account under 'Log in'.

If you choose 'Log in' for institutional login, you will be asked to 'connect with your institution's login'. You can look for your institution, either by scrolling through the list or by typing the name. You are then forwarded to your institution's login page, where you log in using your institutional username and password. If you cannot find your institution in the list, choose 'Sign up' in the meantime and contact the Dataverse support via the button in the upper-right corner.

If at first you used the 'Sign up' option and want to switch to institutional login with the same email address, the system will ask you if you want to merge the two accounts. If you merge the accounts, be aware that the 'reset password' function won't work anymore, as it is your institutional password that you now use to log in.

If you want to find out more about account creation, for example switching from institution login to a local account or how to log in using ORCID, GitHub, Microsoft, or Google (if this feature is activated in the Dataverse), see the Dataverse User Guide.[3]

When creating an account or logging in using an institutional account, you have to agree to the 'General Terms of Use'. These Terms are set by the institution hosting the Dataverse and list the terms and conditions for using the application and services. Users have to mark a box, agreeing to the General Terms of Use of the Dataverse.

[3] https://guides.dataverse.org/en/5.6/user/account.html#id13

1.2. Account Information

After being logged in, you can find information about datasets you created, notifications, your account and API tokens in the upper-right corner if you click on your username. You can also log out there.

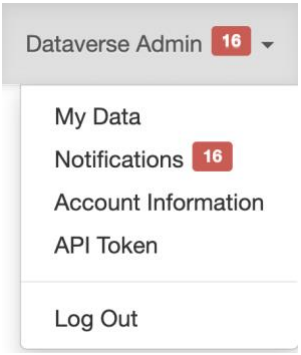

FIGURE 1: ACCOUNT INFORMATION

- **My Data:** Under 'My Data' you can see all Dataverses, datasets and files associated with your account. You can also filter these categories as well as filter for 'Publication Status' and 'Roles' you have.
- **Account Information:** Here you can see and edit your account information and password as well as verify your email address. You can only edit your account information if you are logged in with a local Dataverse account.
- **API Token:** If you are interested in using Dataverse via its Application Program Interfaces (APIs), you can create an 'API Token' that is valid for one year. Please be aware that the usage of the APIs may be bound by API Terms of Use/Service. Your 'API Token' is like a password, anyone who has it can access, edit and delete your data, so take care with your API token. For a whole list of APIs[4] and who is able to use them, as well as on how to get started[5], see the Dataverse User Guide.

[4] https://guides.dataverse.org/en/5.6/api/intro.html#id16

[5] https://guides.dataverse.org/en/5.6/api/getting-started.html

2. Finding and using data

Users can browse or search the contents of a Dataverse installation to find sub-Dataverses (i.e. Dataverses contained within other Dataverse collections), datasets, and files, as well as to view, subset, analyse or visualise data (this may require external tools to be installed, and for you to be logged in to an account that has access to them). This can be done for both restricted and non-restricted data without requiring you to log in, however accessing unpublished or restricted data requires that the administrator grants you access (which does require logging in). Users can also search within a specific dataset using the search bar under the 'Files' tab of the dataset.

2.1. Searching and Browsing

2.1.1. Searching all contents of a Dataverse installation

A 'Basic Search' of the contents of a Dataverse installation can be done using the search bar on the main/root page of the Dataverse installation or via the navigation bar button at the top of the webpage (this may be unavailable on some versions of Dataverse). The Basic Search searches through all available metadata fields of the Dataverse installation, depending on the full-text indexing configuration of the Dataverse installation you may also be able to search in file contents. This search bar accepts terms, queries (including operators such as AND, OR), and exact phrases (in parentheses).

Clicking 'Advanced Search' on the main web page or on the search results page allows you to refine the search further. Here, you can search for specific Dataverses (e.g. by name, affiliation or subject), Datasets (e.g. by title, ID, or keywords), and files (e.g. by file types, tags, and variables).

2.1.2. Sorting and viewing search results

Results: Search results are displayed in results cards under the search box. Each card shows which metadata fields matched the search term or query used in bold.

Facets: using the facets on the left-hand side of the search results (e.g. category, year, name etc.), users can sort and filter the results shown by a number of factors. To select a facet, simply click on the desired item(s) (e.g. year), and only the result that matches that/those criteria will be displayed. To remove a filter, simply click the X symbol next to the item or above the search results. By default, the 5 most common facets are shown under each category. Date facets will be displayed in chronological order. To show more facets, just click more under the category in question.

Sorting: Results can also be sorted using the sort button to the top right of the search results. This lets you sort by name, date, or relevancy of the search results (the default).

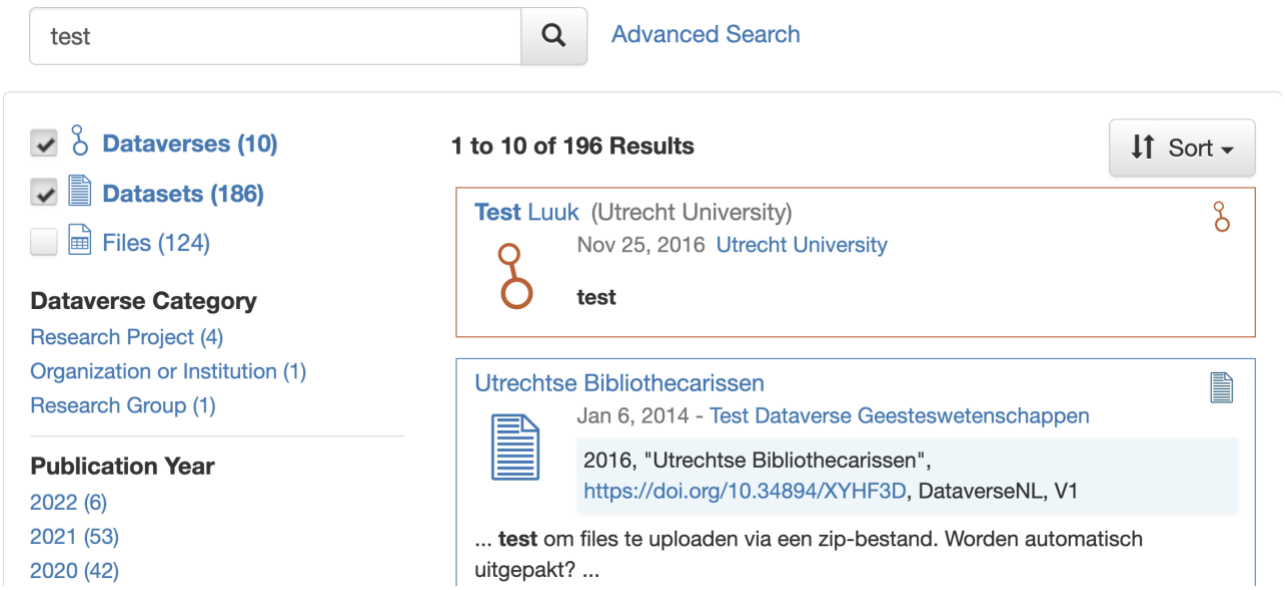

FIGURE 2: SORTING AND VIEWING SEARCH RESULTS

3.Using Data

When you want to access a Dataverse or dataset, simply click its name or icon, and you are taken to its page. The Dataverse or dataset page lets you view all its associated content, metadata, terms, versions, and so on.

3.1. Searching within a dataset

Datasets that contain multiple files can be searched for individual files using the search bar under the 'Files' tab of the dataset. Performing a search in this search bar will filter the file table below the search bar to only show search results (if any).

After performing a search, you can filter the file table by file type, file tag, and access level using the 'Filter by' search facets below the search bar. To return to the full list of files, simply perform another search with the search bar empty.

3.2. View Files

Per default, files in the ‘File’ tab are displayed in ‘Table’ view. You can see how many files are attached to the dataset in the upper left corner. You can filter and sort all files.

Each file can have a name, a path (which is useful in the 'Tree' view), and more information like number of downloads and checksums. It also has additional file metadata, short descriptions you can use to learn more about the file as well as tags (in dark grey below) to categorize data files. You can also see file restricted (red lock next to thumbnail below), this indicates that the file is restricted and cannot be downloaded right away.

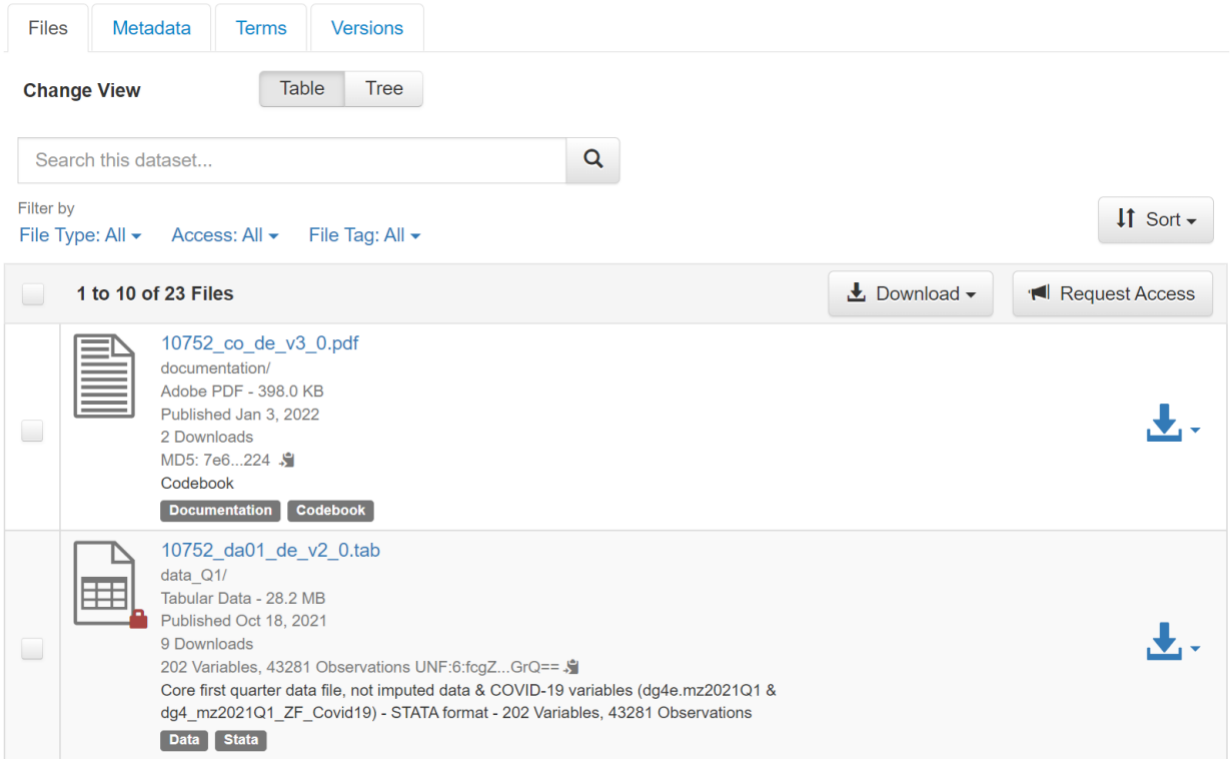

FIGURE 3: VIEW FILES[6]

3.3. Tree View

If you click on 'Tree', files can be displayed in another order, assigned by the depositor.

In the example to the right, data has been split into two separate folders.

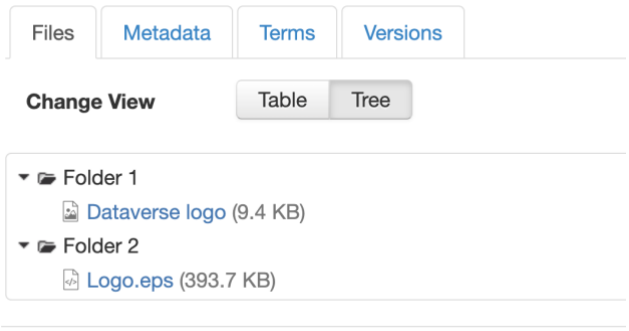

FIGURE 4: TREE VIEW WITH FOLDER STRUCTURE

3.4. Cite Data

Data Citation is important and it is very easy in Dataverse, as it is prominently displayed right below the title. You can export different styles by clicking 'Cite Dataset' to add the dataset to your preferred reference management system.

[6] *Screenshot: Statistics Austria, 2021, 'Microcensus Labour Force Survey / Housing Survey 2021 (SUF edition)', https://doi.org/10.11587/AROIHY, AUSSDA, V3*

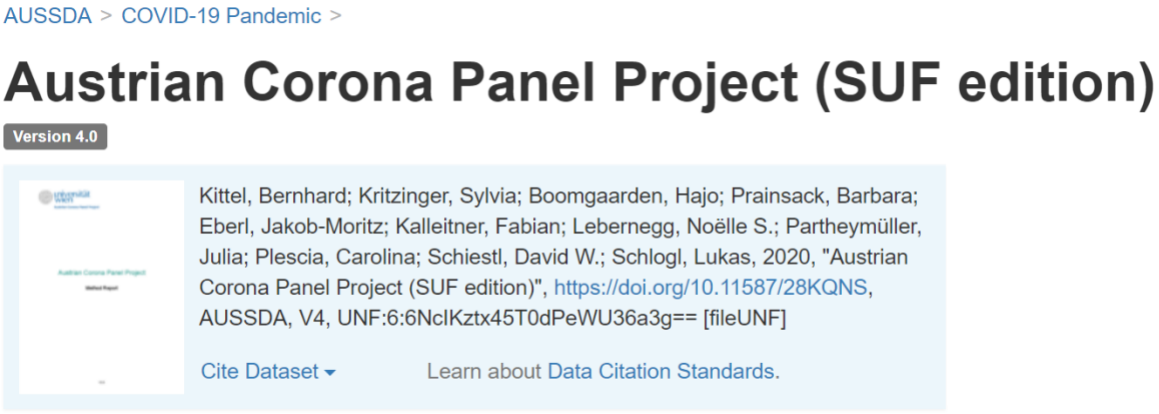

FIGURE 5: CITE DATA[7]

3.5. Download (meta)data

To download all data in a dataset, click the 'Access Dataset' dropdown menu and select one of the options. This will download the dataset's files in a compressed .zip format, maintaining the folder structure that the dataset creator set up.

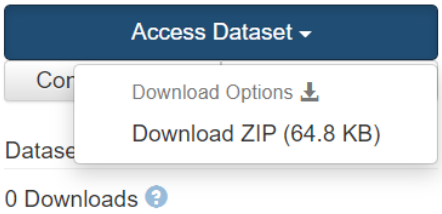

FIGURE 6: DOWNLOAD (META)DATA

To download individual files, go to the files tab in the lower part of the page, then select the files you want to download before clicking the 'Download' button in the top-right corner. This will download the selected files as a .zip file. You can also download individual files in their original file format by clicking the dropdown menu next to the arrow to the right of each file. Through this dropdown menu, you can also download the file metadata in various citation formats. Finally, you can also download specific files directly using its 'Download URL' (this can also be used by APIs). This can be found in the file page under the metadata section, which can be accessed by clicking the file name in the table below, or by searching for a specific using the search functions described previously.

[7] *Screenshot: Kittel, Bernhard; Kritzinger, Sylvia; Boomgaarden, Hajo; Prainsack, Barbara; Eberl, Jakob-Moritz; Kalleitner, Fabian; Lebernegg, Noëlle S.; Partheymüller, Julia; Plescia, Carolina; Schiestl, David W.; Schlogl, Lukas, 2020, 'Austrian Corona Panel Project (SUF edition)', https://doi.org/10.11587/28KQNS, AUSSDA, V4*

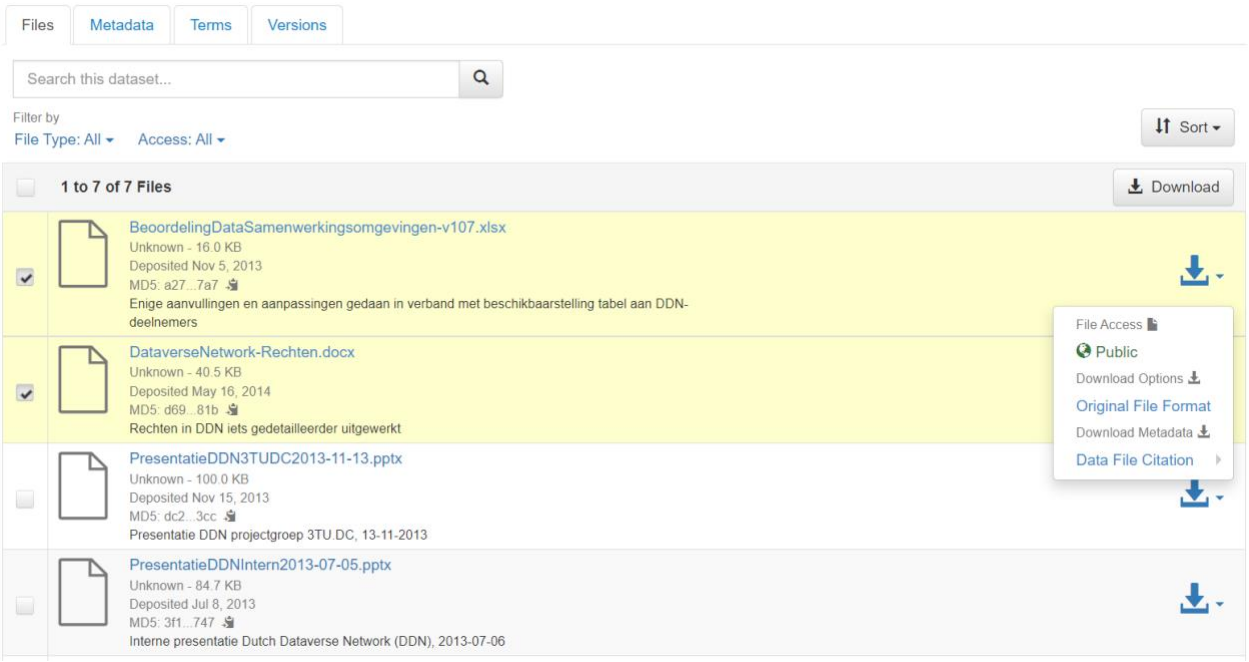

FIGURE 7: SELECTING FILES FOR DOWNLOAD[8]

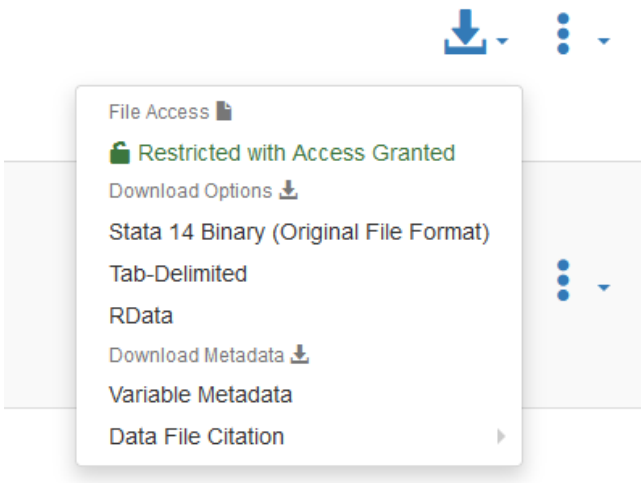

FIGURE 8: DOWNLOADING METADATA AND FILES IN ORIGINAL AND OTHER FORMATS

[8] https://dataverse.nl/dataset.xhtml?persistentId=doi:10.34894/CZGASF

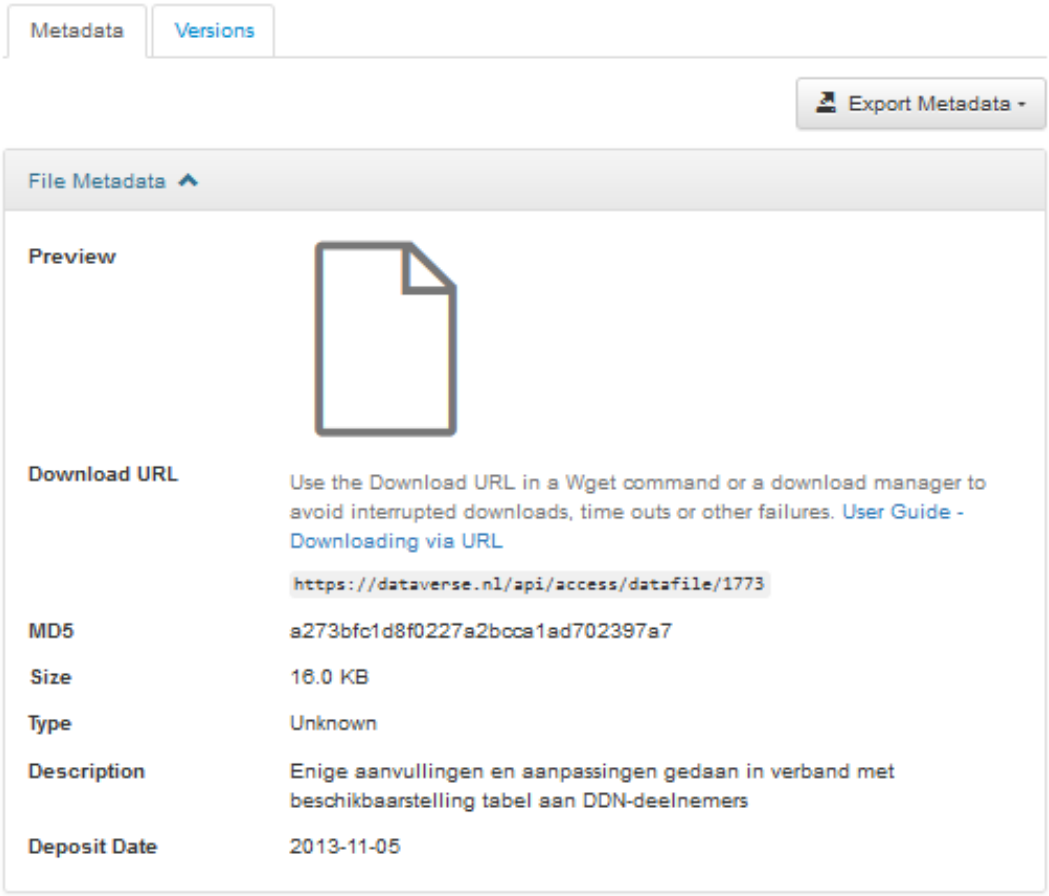

FIGURE 9: FILE TAB 'FILE METADATA' PAGE

For more information about advanced features, such as how to reliably download extremely large files, how to download via command line terminal, and how to download or synchronize data via rsync, please refer to the Dataverse Project user guide, at: guides.Dataverse.org/en/5.6/user/find-use-data.html. This guide also contains guidance for how to explore data in Dataverse using external tools (if installed), as well as the experimental feature Exploratory Data Analysis Using Differentially Private Metadata.

4. Dataverse Management

Dataverse collections are containers for one or more datasets and other Dataverses (in this case called a sub-Dataverse). A Dataverse can be considered as a folder, the name and category of which indicates its contents, even though the Dataverse itself never contains files. For example, a university may have a Dataverse where its researchers can either store their datasets directly, or where different projects can store their own Dataverses and datasets.

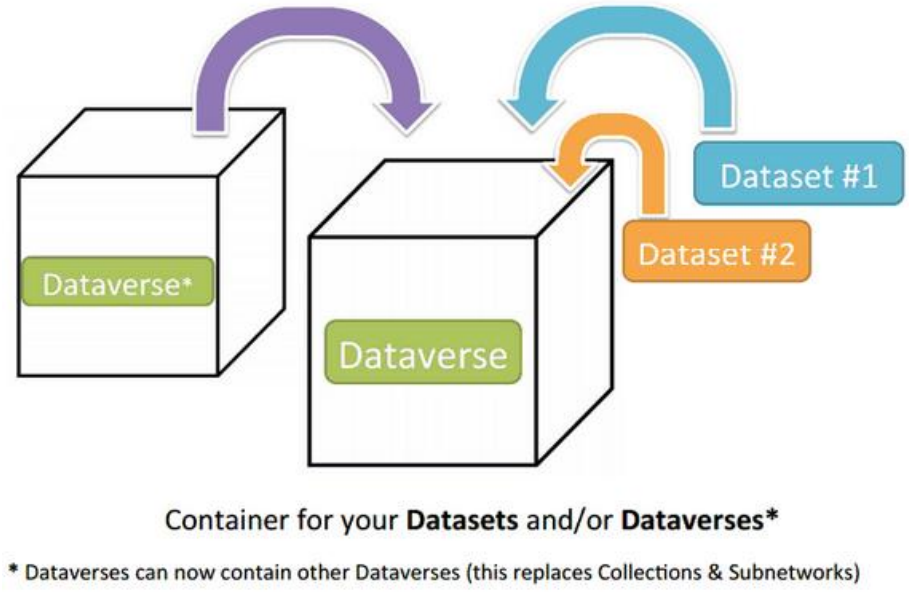

FIGURE 10: SCHEMATIC DIAGRAM DATAVERSE IN DATAVERSE IN DATAVERSE 4.0[9]

By default, the user who creates a Dataverse becomes its administrator. These administrator rights are required to manage the settings that will be discussed in this section. Datasets can be created within a (sub)Dataverse given that the user has the required permissions (see section 5: Dataset + File management below for more details).

4.1. Creating a New Dataverse

Creating a Dataverse requires that you first have a registered user account (see the account creation and management in section 1 for more information). After logging in, click the 'Add Data' button (either in a navigation bar link or in a separate button), and select 'New Dataverse' in the dropdown menu. If you cannot see the 'Add New Dataverse' field in the drop-down list, you don't have permissions to add

[9] https://guides.dataverse.org/en/5.6/user/dataverse-management.html#create-a-new-dataverse-collection

Dataverses. If you think this is an error, contact the Dataverse support. To complete the creation of the Dataverse, add 'General Information' by completing the following steps:

1. Filling in the fields in the 'New Dataverse' page: Beyond simple information such as choosing a Dataverse name, category, and contact details, this includes choosing a **Host Dataverse**, which determines where the new Dataverse will be stored. This could either be the Dataverse you were viewing at the moment you clicked the 'Add data' button or the Root Dataverse if you used the 'Add Data' button in the navigation bar on top. You can choose any Dataverse that you have sufficient permissions to. You also need to enter an **Identifier**, which is an abbreviation that will form part of the URL linking to the new Dataverse (e.g. http://.../[identifier])**.** The abbreviation cannot contain spaces or special characters (e.g. !, ?, %), and should be chosen with care, since changing it will break any links using the old URL.

2. Choosing a set of Metadata Fields: by default, this will be the same as that of the Dataverse collection that the new Dataverse is created in, but there are several available sets for different domains. You can also decide which fields should be required, and which should be optional (note that this requires you to uncheck the 'Use metadata fields from Root' box first). These settings will also determine which fields get used when creating Dataset Templates (more information about metadata and dataset templates can be found in the Metadata section 5.2.1 below.
   For data catalogues associated with CESSDA, Dataverse has a new feature which lets you use the CESSDA Metadata Model (CMM). The CMM is the standardised metadata model used by CESSDA and their service providers, making sure that sharing is possible despite different data documentation across services. The CESSDA metadata block can be selected here, if it was made available by the Dataverse administrator.

3. Deciding which metadata fields will be used for the search and browsing facets of your Dataverse collection: by default, the same settings will be used as for the Dataverse Collection it is created in, but you can also add or remove facets using any field(s) of choice (this requires you to first uncheck the box 'Use browse/search facets from Root').

4. Once all the above settings are set, click the 'Create Dataverse' button to complete the creation of your new Dataverse.

4.2. Editing a Dataverse

You can edit your Dataverse by going to the landing page of a Dataverse and clicking the 'Edit' button, as pictured below. The dropdown menu lets you edit several categories of options, and each of these will be gone through below.

If you go to 'Edit' and click on 'General Information', you can edit information such as the name, category, contact email and so on that was described above. It also allows you to edit the metadata options that

were also discussed in the previous section, such as what fields are hidden, required, or optional. You can also edit the facets used when searching or browsing the Dataverse.

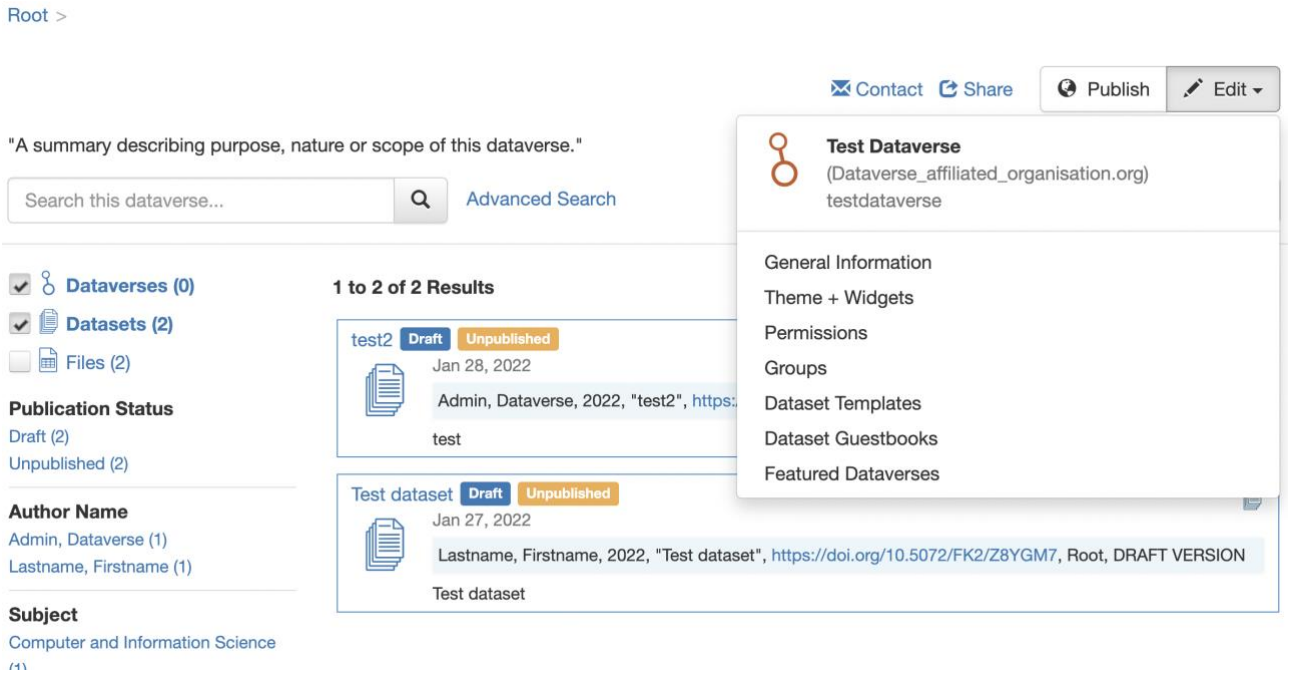

FIGURE 11: EDITING A DATAVERSE

4.2.1. Theme + Widgets

This takes you to a menu with two tabs. The **Theme tab** lets you change visual aspects of the Dataverse, such as changing colours, uploading a logo and footer, as well as adding a link to a website and a catchy tagline for the Dataverse (this tagline will be visible on the Dataverse page and will act as a link to the website). You can also choose to inherit the same theme as the Dataverse which the edited Dataverse is placed in (i.e. using the same theme options as it is).

The **widgets tab** lets you get code that can be used to display/embed a Dataverse on a website (note that the Dataverse must be published before this can be done). Two kinds of widgets can be used for a Dataverse: the **Search Box Widget**, and the **Listing Widget:**

- Search Box Widget: Once added to a website, this lets you search your Dataverse directly from the website. The results will be displayed in a new tab of the browser.
- Listing Widget: this provides a listing of all the Dataverses and datasets that users can search or browse from the website. Files can be downloaded from here as well, provided that doing so does not require permission, in which case the user will be asked to log in to the Dataverse installation.

Widgets are also available on 'Dataset' level, see the Thumbnails + Widgets section below (section 5.2.6). Widgets can be used in different websites, such as OpenScholar websites. In Openscholar, this is done

by navigating to the page where you want the widget to be added, clicking the Settings Cog > Layout > Add New Widget.

4.2.2. Permissions

These settings let you edit or view other users' permissions to your Dataverse, including assigning roles. Roles and permissions define what different users can and cannot do with its associated sub-Dataverses, datasets, and files. These roles and permissions can also be assigned to Groups (see section 4.2.3 for more details. The permissions page contains three sections as pictured below:

- **Permissions:** this section lets you define the conditions that decide which types of users can add sub Dataverses as well as datasets, and what permissions they gain when they do so. Admins of a Dataverse have the right to edit these permissions. To do this, clicking the Edit Access button gives you a number of options about who can add content to your Dataverse.
- **Users/Groups:** This lets you assign roles to certain users or groups, which determines what actions they are permitted on your Dataverse. This section also lets you see and remove roles that you already have assigned to users and groups. To change roles, click the Assign Roles to Users/Groups button, and type in a user or group to change their roles. More information about roles can be found below in this section.
- **Roles:** this section lists all the roles that are available to assign, as well as what permissions each role offers (see figure 13). In the figure below, the role 'Admin' has been assigned to a group (called Ingest).

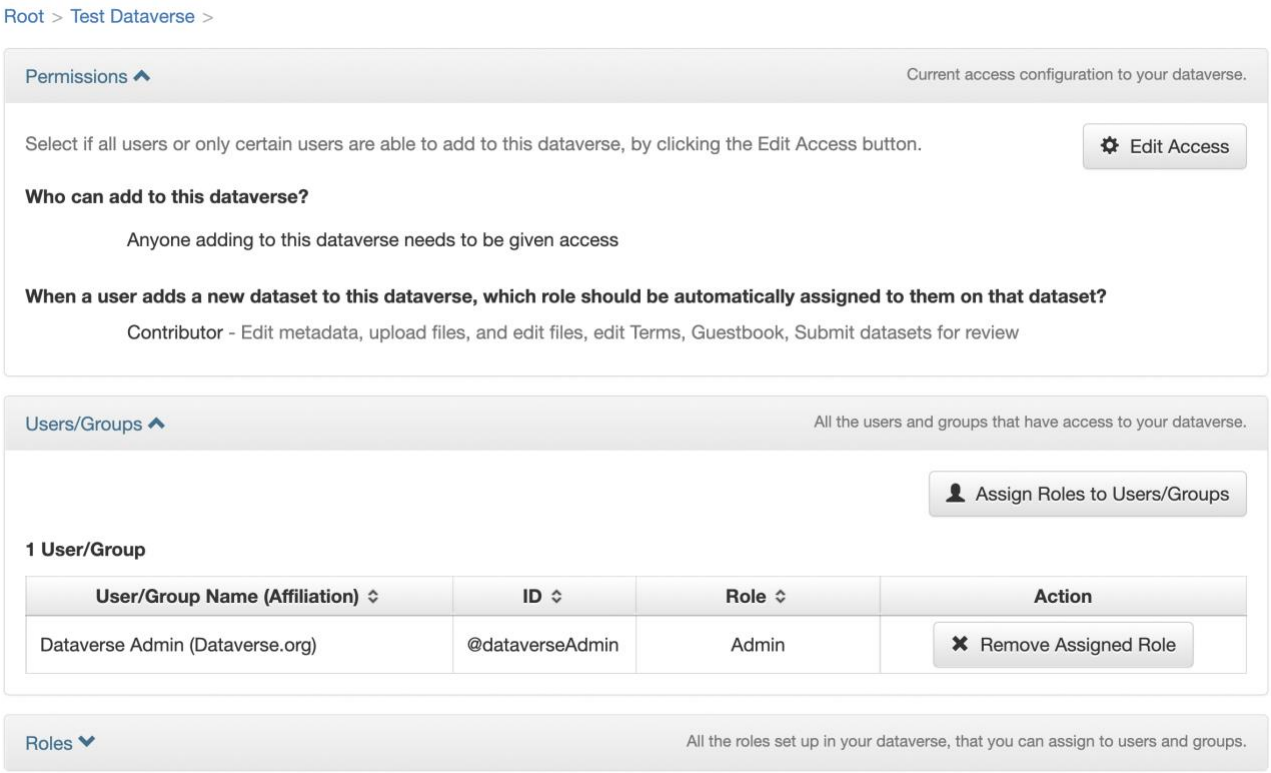

FIGURE 12: DATAVERSE PERMISSIONS

FIGURE 13: DATAVERSE PERMISSIONS MENU

Managing access configurations: In Dataverse, there is a lot of variety and ability to customize permissions and roles. Which settings should be used is highly dependent on the context and use case, and in some cases, users may not be able to change these settings at all. For more information about how to configure advanced access configurations, refer to the Dataverse.org User Guide.[10]

[10] https://guides.dataverse.org/en/5.6/user/dataverse-management.html

4.2.3. Groups

Groups allow you to assign users to groups to then give the whole group permission to use a dataset. This can be very useful if a whole class attending the same course needs access to a dataset, as after the course ends, the permissions can be revoked for the whole group with less effort than 'Remove Access' one person at a time.

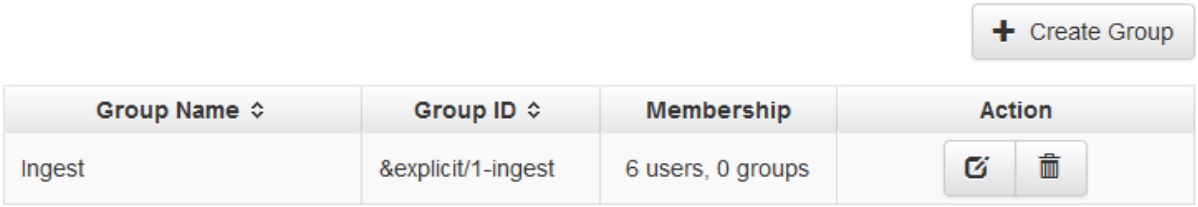

FIGURE 14: DATAVERSE EDIT GROUPS MENU

4.2.4. Dataset Templates

If a dataset creator adds more than one dataset and they are very similar to each other, they might consider using a dataset template. In templates, metadata fields like Author name, and terms can be pre-filled, which saves time if two or more datasets are very similar.

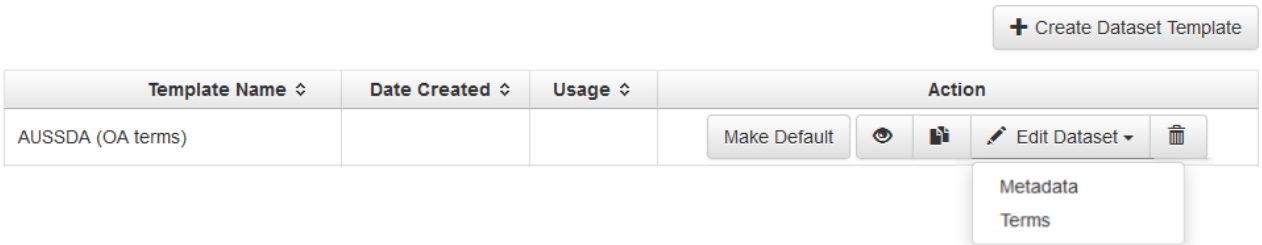

FIGURE 15: EXAMPLE DATASET TEMPLATE

4.2.5. Dataset Guestbooks

Guestbooks is a feature that lets you gather information about who is downloading files from your Dataverse, helping you understand your users better. Once activated, questions will be part of the pop-up when downloading files. Questions can include information such as name and affiliation, as well as more open-ended questions about what the user intends to use the data for (see figure 16 below). Data collected through the Guestbook can also be downloaded for use outside of Dataverse.

To create a guestbook, simply click the 'Edit Dataverse' button shown in figure 11. Then click 'Create Dataset Guestbook', and it will take you to the menu below. This menu lets you choose what information you want to collect, and what information will be required. Questions added to the 'Custom Questions' section can either be in text or multiple-choice format. To add or remove questions or response options, simply click the plus and minus symbols next to them.

Once a 'Guestbook' has been created, you can edit it and others in the 'Guestbooks' menu (see figure 17) via the 'Edit Dataverse' menu (figure 11). For each Guestbook, you can see usage numbers and responses, as well as edit, preview, and view responses using the buttons in the 'Action' column. The 'View Responses' menu lets you see more detailed information about who used the guestbook, and to download all responses collected.

When creating a dataset within the Dataverse, you will be able to select which guestbook you would like to use. See also section about 'Terms', 5.2.3.

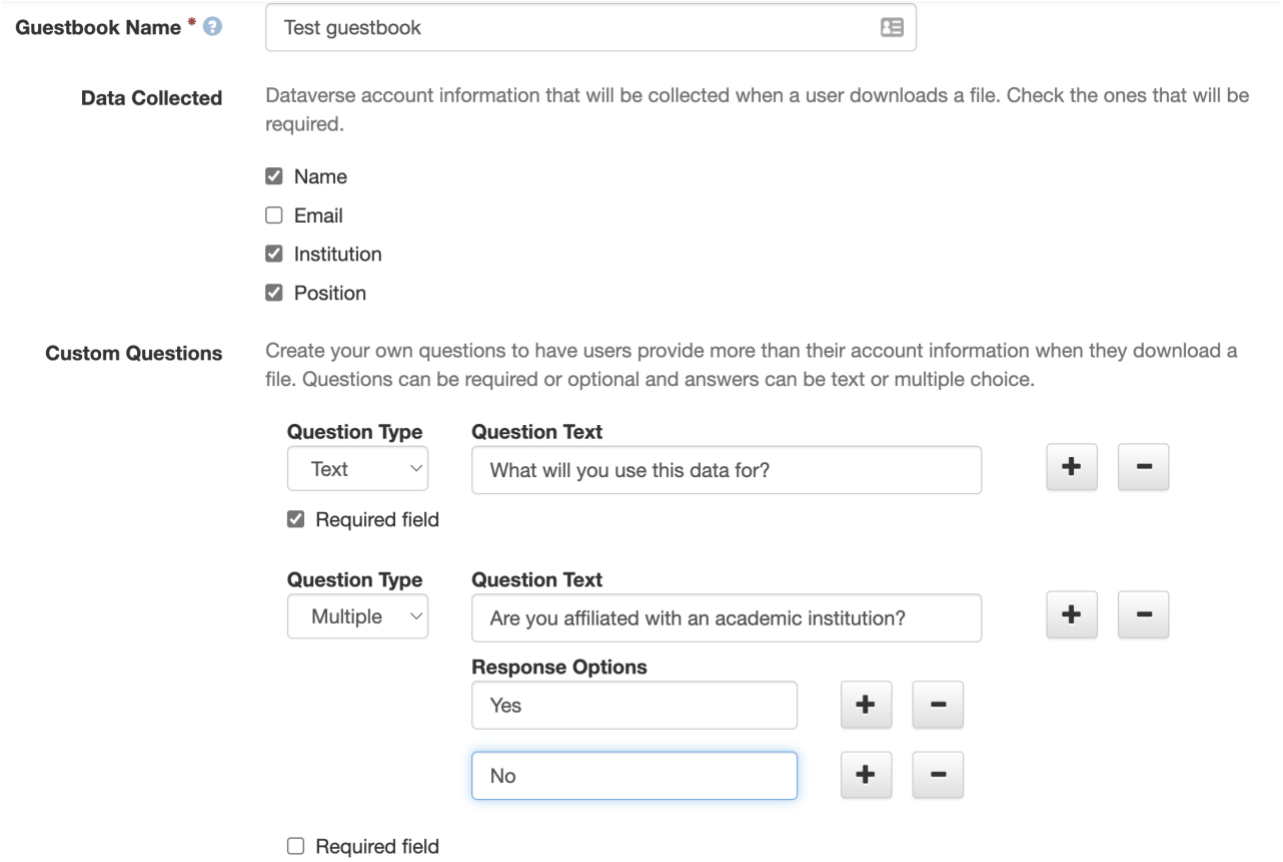

FIGURE 16: GUESTBOOK CREATION/EDITING MENU

FIGURE 17: MANAGING GUESTBOOK(S)

4.2.6. Featured Dataverses

You can highlight specific Dataverses by selecting them from the list of available Dataverses and display them prominently on the landing page of the Dataverse. You can only select published Dataverses.

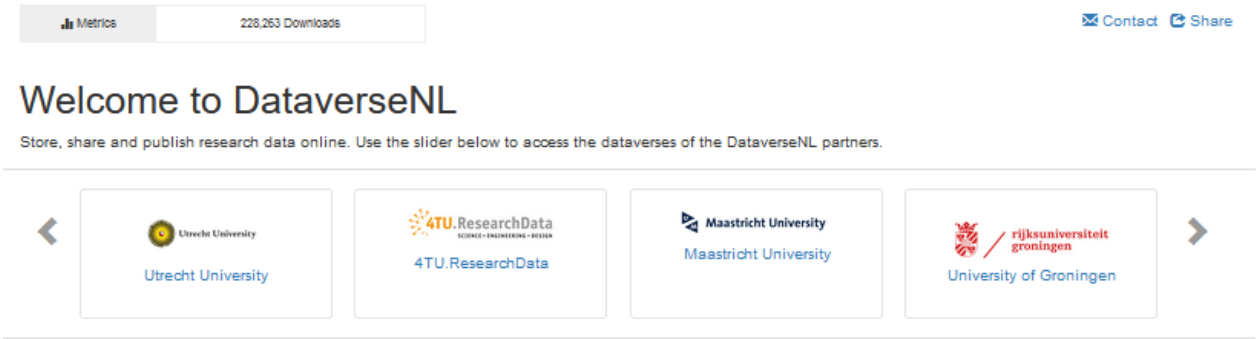

FIGURE 18: FEATURED DATAVERSES IN DATAVERSENL

4.3. Publishing

Once you are ready to make your Dataverse public, simply go to the Dataverse start page (figure 11) and press the 'Publish Dataset' button. You will then receive a pop-up asking you to confirm that you want to publish the Dataverse. **NOTE:** as this menu warns – once you decide to publish a Dataverse, it must remain published. In other words, once a Dataverse is public, you cannot un-publish it.

5. Dataset + File Management

5.1. Creating a Dataset

To create a dataset, start by going to the Dataverse where you want it to be located. In that Dataverse, click on 'Add Dataset', and this will take you to the dataset creation menu. In this menu, fill in the citation metadata (e.g. the dataset name, author, description etc), and ensure that required fields (marked by asterisk) are filled (for more information about metadata, see the Metadata section (section 5.2.1). You can also add files to the new dataset straight away, by uploading them under the 'Files' tab at the bottom (uploading files and providing more details metadata can be done at a later point). Finally, click the Save Dataset button, and the dataset is created.

After clicking 'Save Dataset' at the bottom of the menu, the dataset is created (if clicking save fails, make sure that you filled all required fields, and try again). From here, you can add or edit your dataset's Metadata (5.2.1), Files (5.2.2), Terms (5.2.3) and Versions (5.4). You can also assign Roles and Permissions (5.2.4) to your dataset or Delete a Dataset (5.2.7) under 'Edit Dataset'.

Once your draft is finished and all your settings are in place, you can either Publish your Dataset or Submit it for Review (5.3).

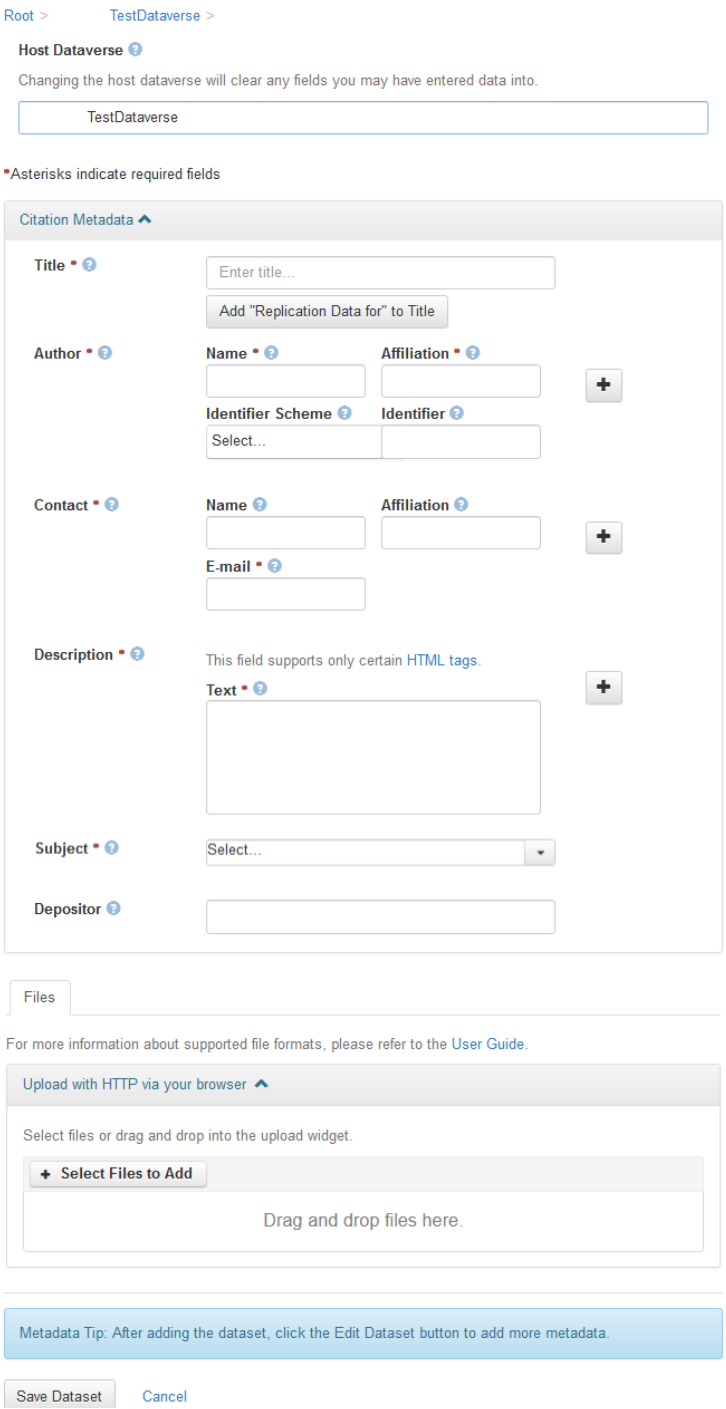

FIGURE 19: ADDING A NEW DATASET TO A DATAVERSE, USING ONLY THE CITATION METADATA BLOCK

5.2. Editing a Dataset

Under 'Edit Dataset' on the right side of the dataset page, you can navigate through all dataset settings. We will go through each of these settings in the following sections.

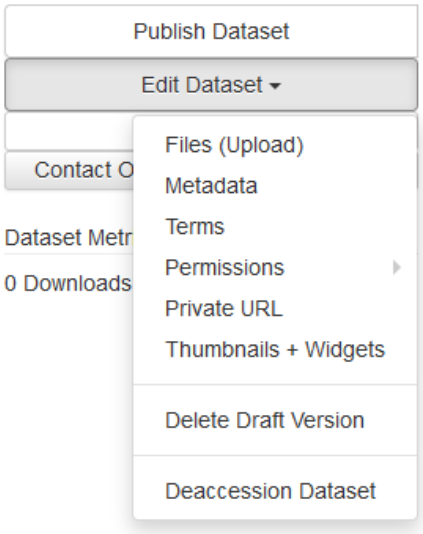

FIGURE 20: EDITING A DATASET

5.2.1. Metadata

Metadata describe your data resources and facilitate searching for data. Dataverse differentiates between three levels of metadata: citation metadata, domain specific metadata and file-level metadata, visit the File Metadata section 5.2.2 to find out more about the latter.

Minimum required metadata in Dataverse that are needed for the dataset citation (citation metadata block) are the following fields: the 'Title', 'Author', 'Contact Email', 'Description' and 'Subject' of a dataset. Depending on the Dataverse you work in, the number of metadata that are required for depositing your data may vary significantly. When you add metadata, required fields are indicated by an orange asterisk next to the name of the metadata field. Depending on the discipline you work in, you may see different metadata blocks (domain specific metadata), and these can be adapted on Dataverse level, see Creating a Dataverse section 4.1, and Editing a Dataverse section 4.2. Some examples are the 'Geospatial Metadata' and 'Social Science and Humanities Metadata' blocks.

Archives may use controlled vocabularies to catalogue their data holdings, like Data Documentation Initiative (DDI)[11] or European Language Social Science Thesaurus (ELSST),[12] to make metadata comparable across the holdings or in international catalogues like the CESSDA Data Catalogue.[13]

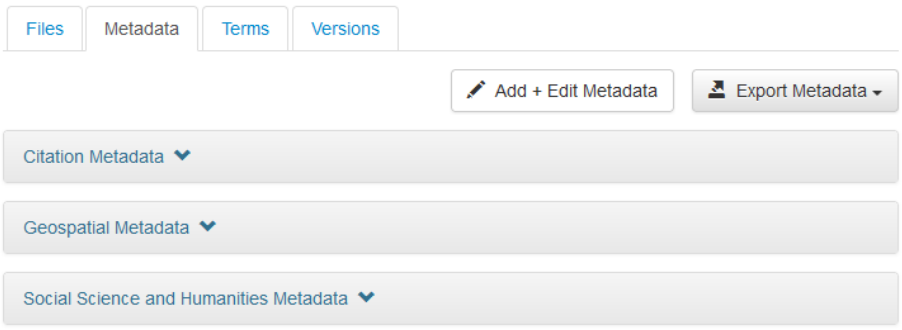

FIGURE 21: A METADATA ENTRY PROVIDING INFORMATION FOR THREE METADATA BLOCKS

[11] https://ddialliance.org/

[12] https://elsst.cessda.eu/

[13] https://datacatalogue.cessda.eu/

If you add more than one dataset and they are very similar, consider using a dataset template. In templates you can prefill metadata fields and terms, which saves you time if you have two or more very similar datasets. Dataset templates can only be created on Dataverse level, see Dataset Templates section 4.2.4.

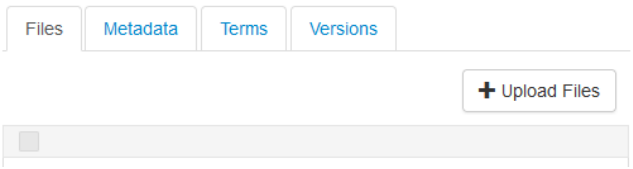

FIGURE 22: FILES TAB

It is also possible to assign Persistent Identifiers (PIDs) to datasets in Dataverse, currently, Digital Object Identifier (DOI) and Handle (hdl) are supported. They are then displayed in the 'Metadata' tab and in the citation box at the top of the dataset page.

Exporting metadata is supported in many formats (e.g. Dublin Core, JSON, DDI, OpenAIRE, ...), see all in the Dataverse User Guide.[14]

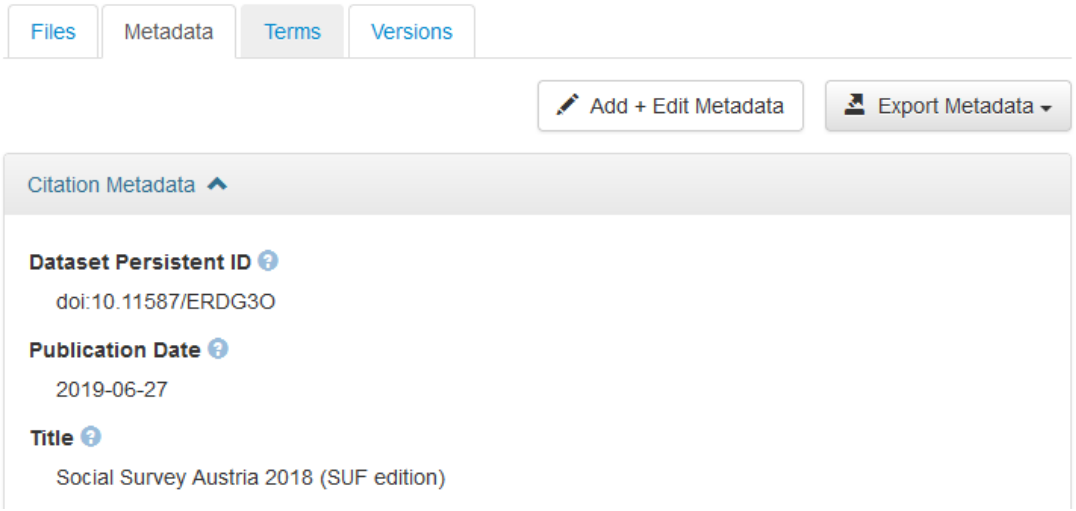

FIGURE 23: DOI AND 'EXPORT METADATA' BUTTON OF DATASET[15]

5.2.2. Files

UPLOADING

You can either upload files while you create your dataset (see bottom of Figure 19), or you can do so at a later time by going to the 'Files' tab in your dataset draft and clicking on '+Upload Files'. Clicking this opens a pop-up where you can either drag and drop files or click on '+ Select Files to Add' to add files from your computer.

[14] https://guides.dataverse.org/en/5.6/user/dataset-management.html#id8

[15] Hadler, Markus; Höllinger, Franz; Muckenhuber, Johanna, 2019, 'Social Survey Austria 2018 (SUF edition)', https://doi.org/10.11587/ERDG3O, AUSSDA, V3.

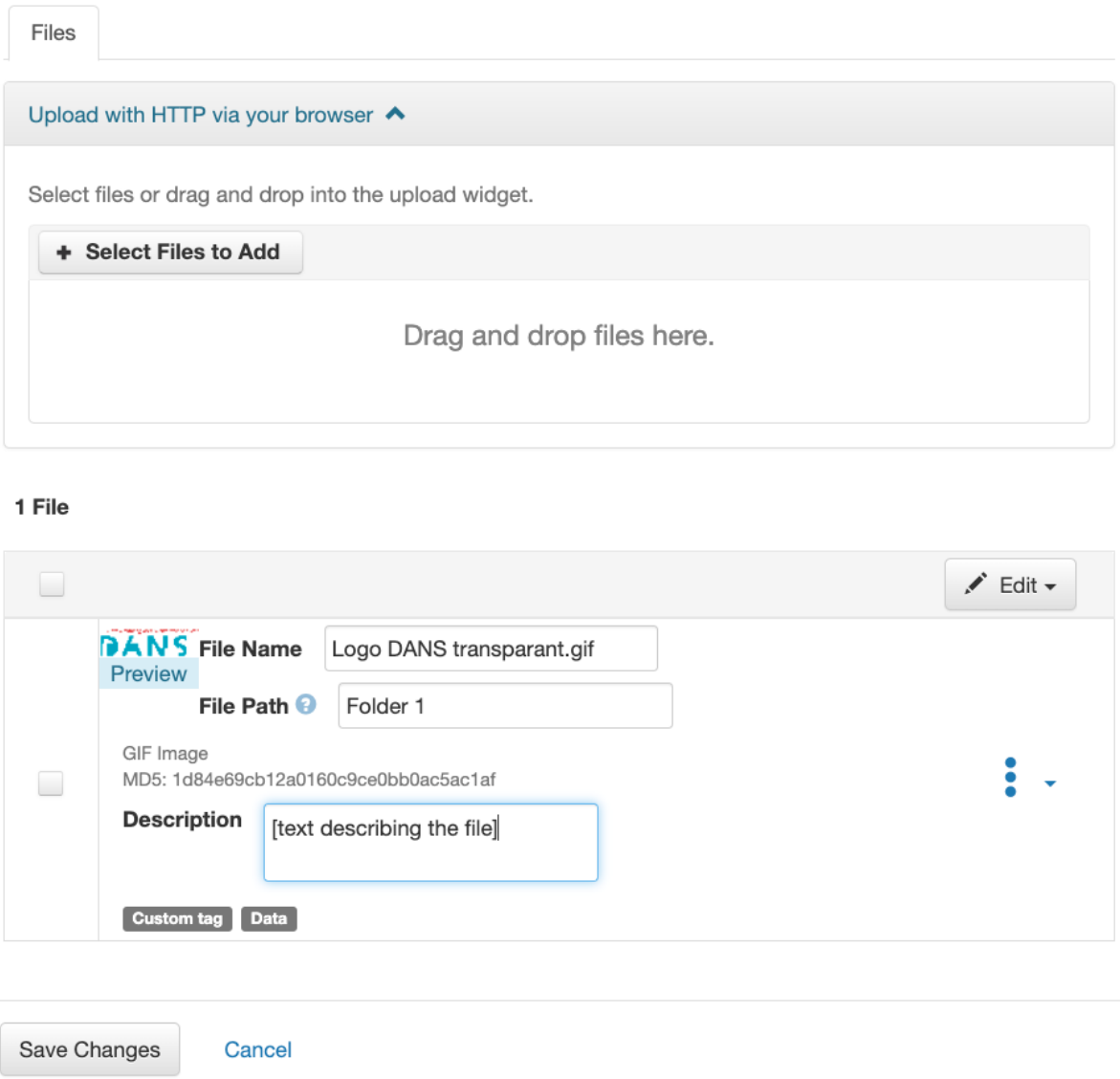

FIGURE 24: UPLOADING FILES

Data File Ingest

During the upload, Dataverse automatically generates a tab-delimited version of many data files, for example of Stata and SPSS-files. This process is called 'ingest' and takes some time. You will receive an e-mail after the ingest is complete. 'The goal of our ingest process is to extract the data content from the user's files and archive it in an application-neutral, easily-readable format.'[16] When clicking the 'Download' button, the original file format is still the first download option for users but in addition, open formats are provided for usage in other software. By conversion to *.tab-format, long-term re-use is secured.

[16] https://guides.dataverse.org/en/5.9/user/tabulardataingest/ingestprocess.html

Also pay attention when uploading compressed files (e.g. .zip). Dataverse will automatically unpack the file(s). If there is a folder structure, it will be displayed in the 'File Path'. Double-compressing a file, or having a very high number of files within a compressed file (1000 per default) will prevent unpacking, it will then be uploaded as is.

For more details on this topic and how Dataverse treats other formats and duplicate files, please see the Dataverse user guide.[17]

File metadata

File metadata include 'File Name', 'File Path', 'Description', 'Tags', 'Provenance' and 'Replace Files'. You can select which files you want to edit by marking the checkbox(es) (see figure below). Clicking Metadata in the Edit Files drop down menu let you edit the following properties:

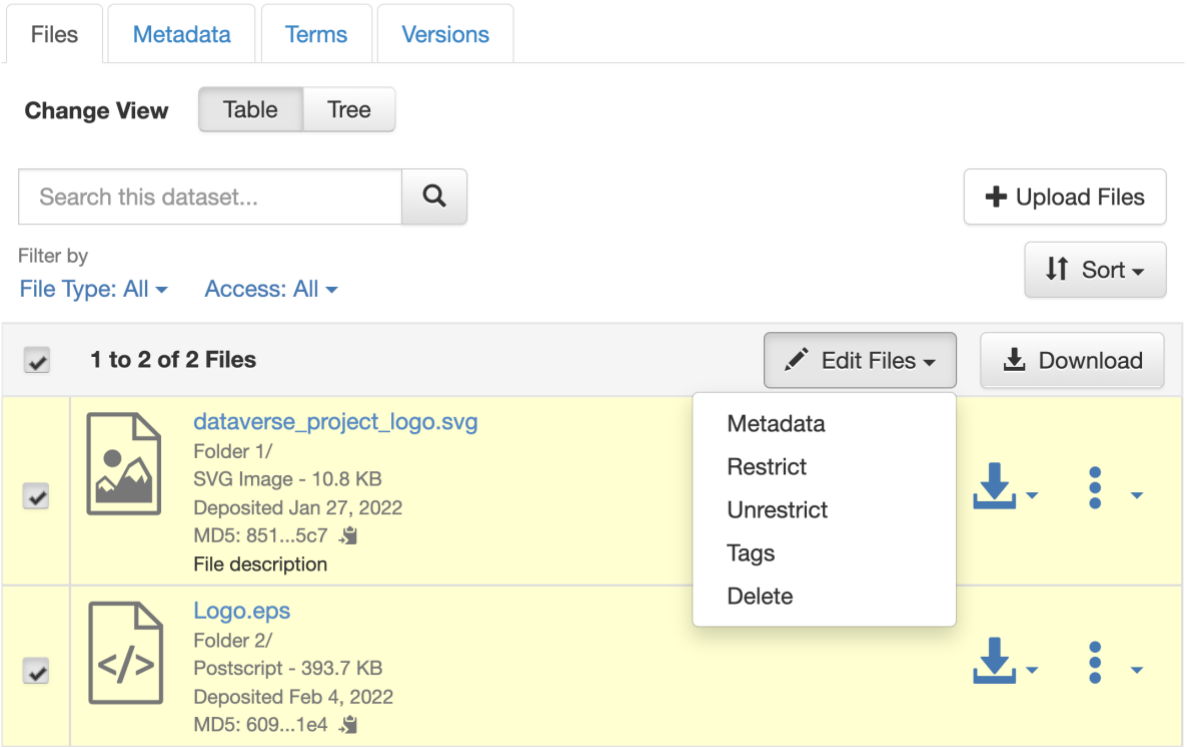

FIGURE 25: SELECTING FILES TO EDIT

[17] http://guides.dataverse.org/en/latest/user/tabulardataingest/index.html

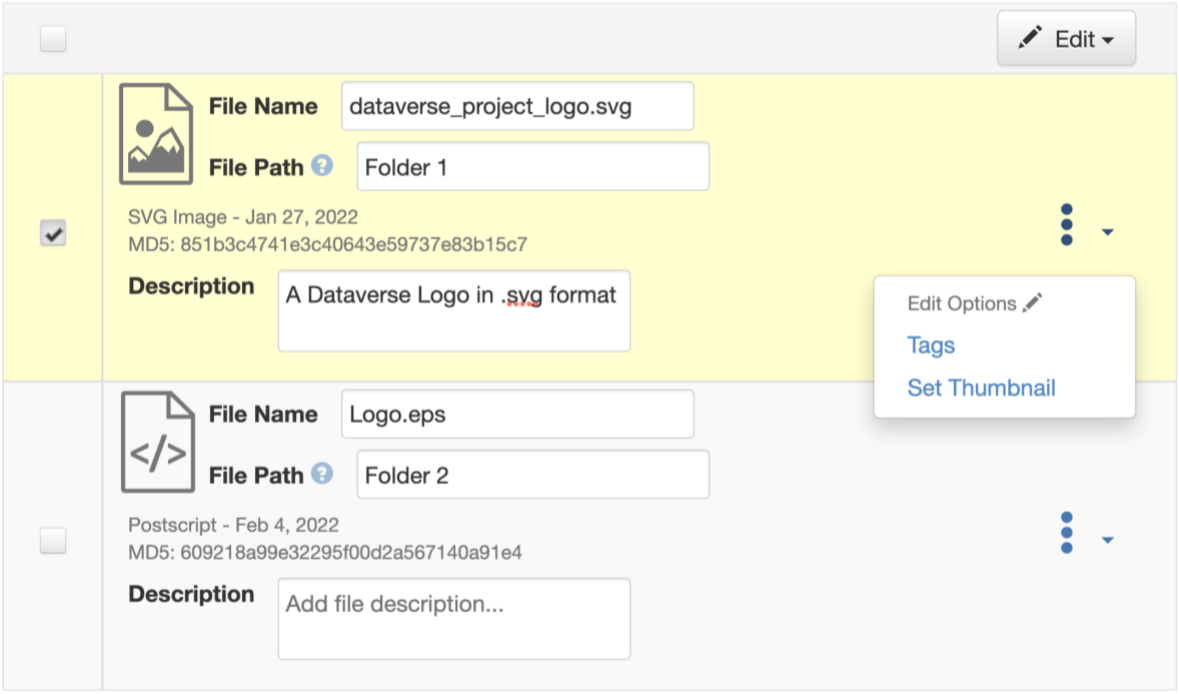

FIGURE 26: EDITING FILE METADATA

- **File Name:** When uploading a file, the current name of the file is displayed here and can be changed when uploading or at a later point in time using the current menu ('Edit Files' - 'Metadata').
- **File Path:** The 'File Path' is on the one hand another way to display your data files in a folder structure and on the other hand it is a nice feature for file downloaders. If you have a lot of files and assign paths to them, for example if you assign all data files per quarter to the same path, file downloaders will see this structure after downloading the .zip file. It can be changed when uploading or at a later point in time under the current menu ('Edit Files' - 'Metadata'). You can switch between the classic 'Table' view and the 'Tree' view by selecting the desired view under 'Change View' (see figure 4).
- **Description**: You can use this free text field to add a short descriptive text to describe the content of your file.
- **Tags and thumbnails**: Clicking the three-dot menu next to a file lets you change the Tags (more details in the following section), and file thumbnails (see section 5.2.6).

FILE TAGS

Tags are another way to describe your files or to categorise them. You can use existing tags, like categories ('Documentation', 'Data', 'Code') and tabular data tags ('Panel', 'Survey', 'Network', ...) or create your own custom tags. You can add tags when uploading files under 'File Options' (Figure 24) or at a later point in time under 'Edit Files' - 'Tags' (Figure 25).

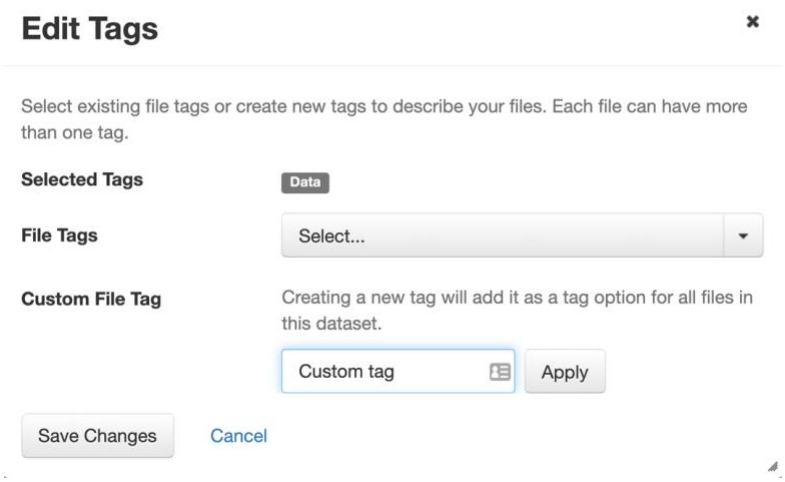

FIGURE 27: EDIT FILE TAGS

When adding custom tags, click on 'Apply' to see the tags under 'Selected Tags'. Add custom tags if you like, 'Apply custom tags' and then 'Save Changes'.

FILE RESTRICTIONS

If this functionality is active in your Dataverse you can 'Restrict' other users' access to your files.

Go to 'Edit Files' and click on 'Restrict' to restrict one or more selected files (see figure 25). Add 'Terms of Access' to the pop-up field and mark the 'Enable Access Request' box as in the figure below. If you do not select the box, users can see your restricted file(s) but cannot request access. By clicking 'Save Changes' the file(s) should be restricted. If you see a green, open lock next to the file thumbnail as in the below figure, the restriction is in place. If you see no lock, try again. The green lock signifies that the data file is restricted for users that do not have been granted a download permission (they see a red, closed lock). If you assign Permissions on either dataset or file level, logged in users with the right permission also see a green, open lock and can download the file(s) immediately. Find out more about Roles and Permissions in section 5.2.4.

If you want to 'Unrestrict' one or more files, mark them, go to 'Edit Files' and click on 'Unrestrict' (see figure 25).

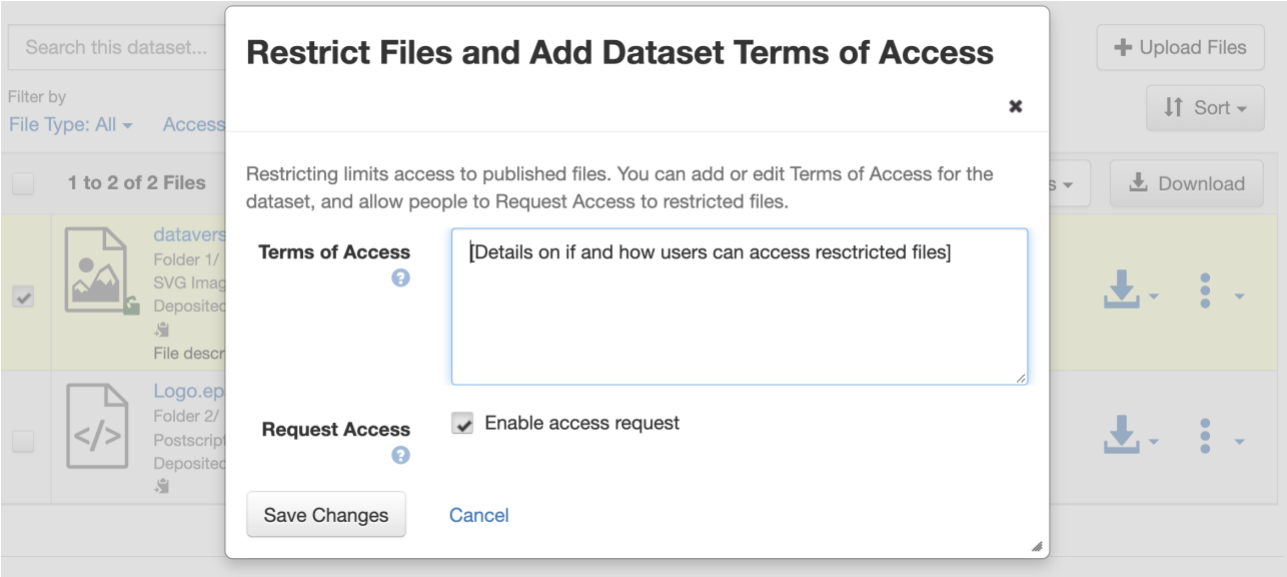

FIGURE 28: RESTRICTING FILES AND ADDING TERMS OF ACCCESS

REPLACE FILE

If you want to replace an uploaded file, you can go to the 'File Options' - 'Replace'. If you provide a second version of a file, instead of deleting the first version, by choosing 'Replace' the history of the file can be tracked across versions.

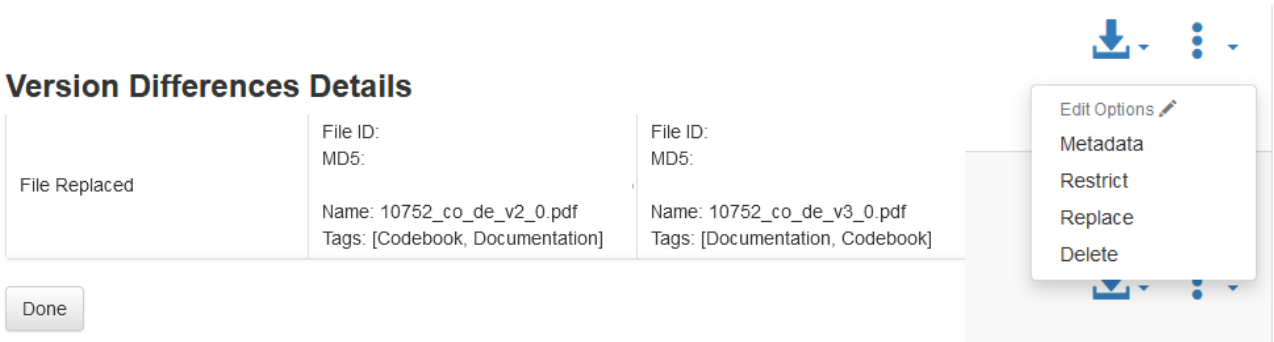

FIGURE 29: REPLACING FILES AND VIEWING VERSION HISTORY

PROVENANCE AND OTHER FILE FEATURES

A feature not everyone is able to see at the moment is the 'Provenance' of your file. By clicking 'Edit Files' - 'Provenance' you can upload a .json file and a 'Provenance Description'.

Learn more about the provenance and other features like 'Edit File Variable Metadata' in the Dataverse User Guide.[18]

[18] https://guides.dataverse.org/en/5.9/user/dataset-management.html#edit-file-metadata

5.2.3. Terms

Under the tab 'Terms', you can define for your dataset how users can use a dataset (Terms of Use), the access conditions (Terms of Access) and you can choose guestbooks. Click the 'Edit Terms Requirements' button to edit each of these options.

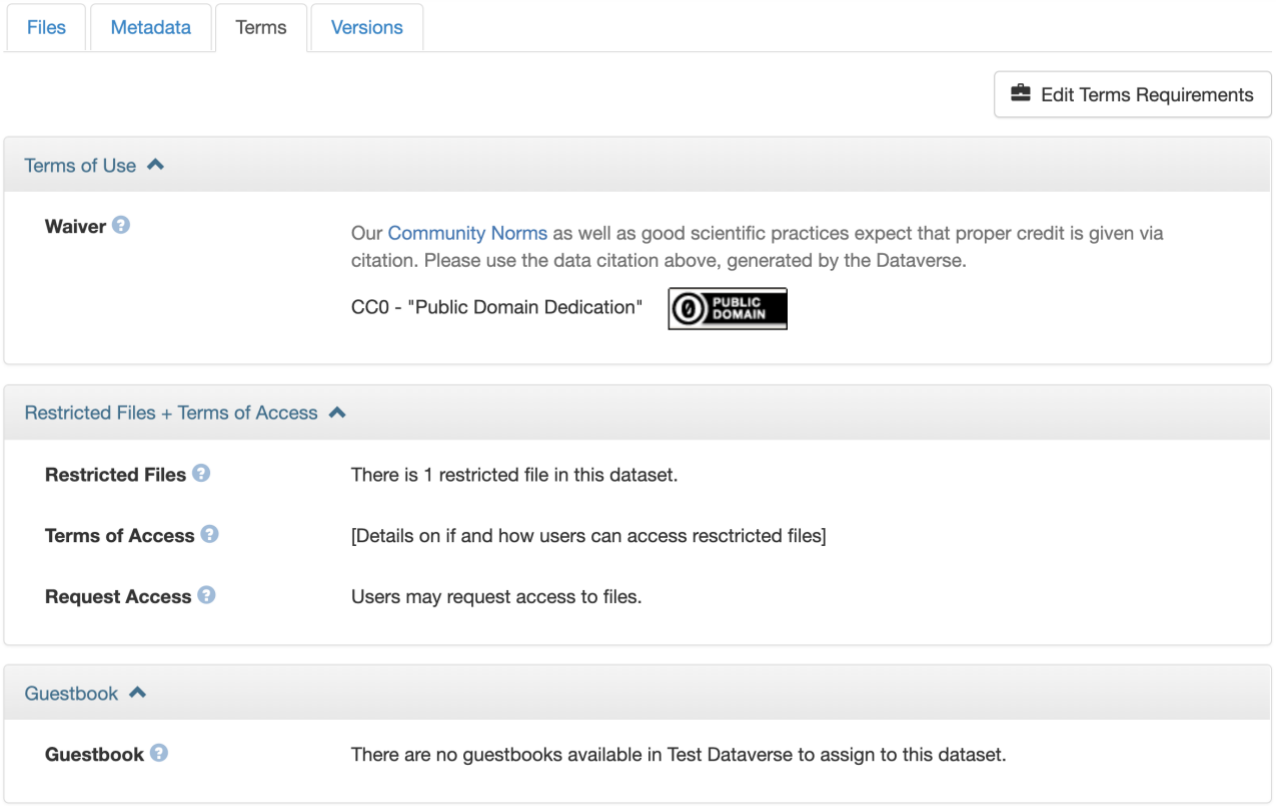

FIGURE 30: TERMS TAB

TERMS OF USE

By default, all datasets created in Dataverse receive a CC0 Public Domain Dedication (except for datasets [created with a dataset template](#) section number 5.1 (that has different terms listed).[19]

Terms of Use can easily be adapted by clicking 'Edit Terms Requirements' and activating the box 'No, do not apply CC0 - 'Public Domain Dedication''.

[19] https://guides.dataverse.org/en/5.6/user/dataset-management.html#id35

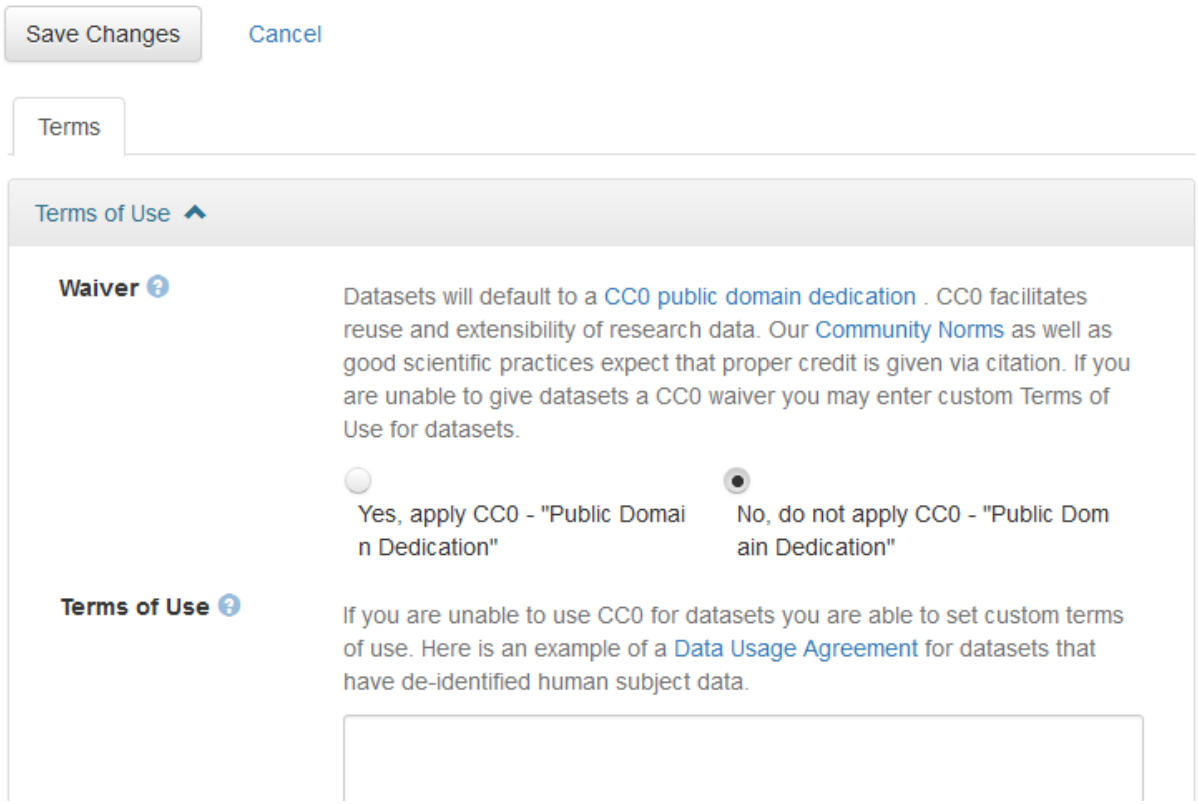

FIGURE 31: DATASET TERMS OF USE

In the free text field 'Terms of Use' you can now add the conditions under which users can use your data. Same as in other free text fields, you can use certain html tags to format your terms of use. In this section of Dataverse, you can also add information on:

- **Confidentiality Declaration**: indicates whether accessing the data will require a confidentiality declaration.
- **Special Permissions:** Indicates whether any special permissions are required to access the data (e.g. special forms).
- **Restrictions:** Details on any restrictions to access or use of the data (e.g. privacy certification or distribution restrictions).
- **Citation Requirements:** Any special citation requirements for the data to be cited properly when used.
- **Depositor Requirements:** Any requirements on users to inform dataset depositors/authors/curators to provide citation to any published work that uses the data.
- **Conditions:** Additional information explaining access conditions.
- **Disclaimer:** Information regarding responsibility for uses of the data.

Terms of Access

The 'Terms of Access' could play a very important role in your dataset. If files in your dataset are restricted, the 'Terms of Access' have to be provided for users to know the access conditions. In this section under

'Request Access', you can also decide if users are allowed to request access to restricted files. Same as in other free text fields, you can use certain html tags to format your terms of access.

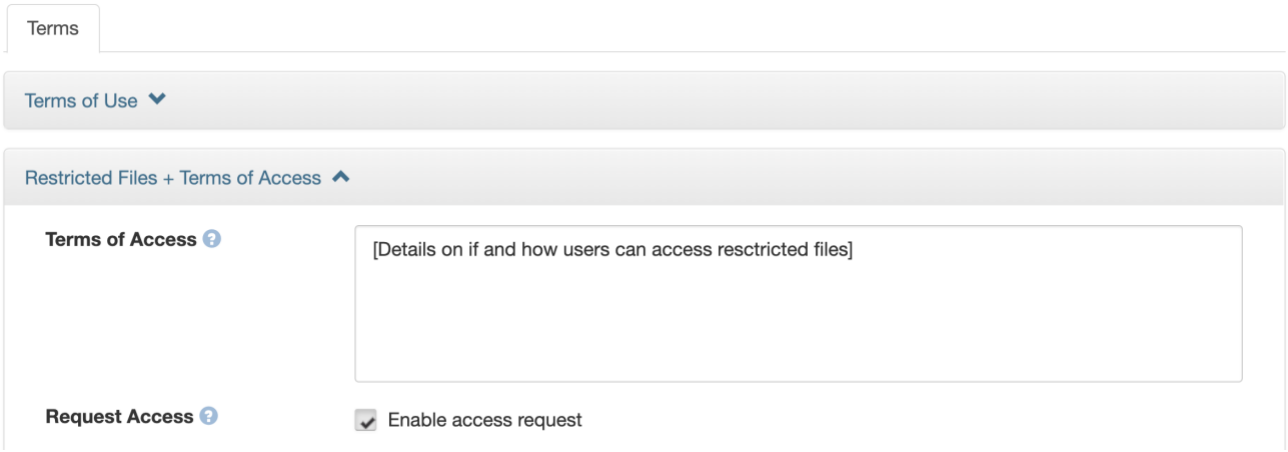

FIGURE 32: DATASET TERMS OF ACCESS

Other information that can be added here is 'Data Access Place', 'Original Archive', 'Availability Status', 'Contact for Access', 'Size of Collection' and 'Study Completion'.

GUESTBOOK

In the tab 'Terms' you can also select a guestbook you want your users to fill out. Guestbooks can help you understand your users better by asking them questions about themselves or their projects. They will be part of the pop-up when downloading files. Guestbooks need to be activated on Dataverse level to be used here. For more information about Guestbooks and how to activate them, see the Dataset Guestbooks section 4.2.5.

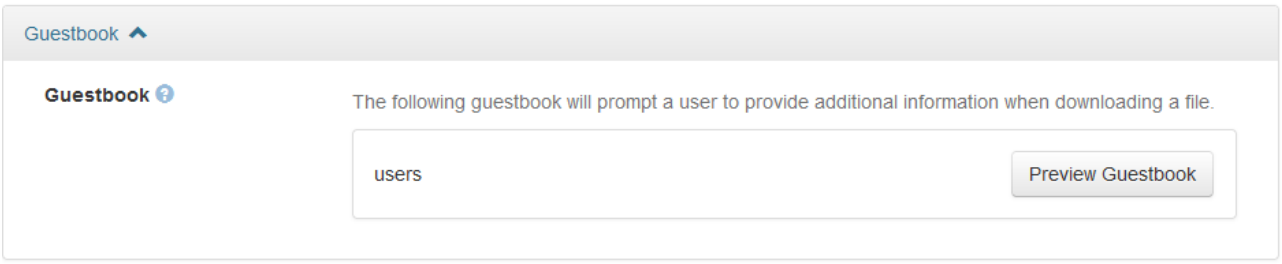

FIGURE 33: A DATASET WITH AN ACTIVE 'GUESTBOOK' CALLED 'USERS'

5.2.4. Roles & Permissions

Similar to Roles & Permissions on Dataverse level (see Permissions section 4.2.2), users can also receive different usage roles on dataset level. Those are either assigned on dataset level (for all files) or on file level (for selected files). Different roles have a certain set of permissions which defines what actions different users may take. If you are admin or curator of a dataset, you can assign users or groups specific

roles (see more in the Groups section 4.2.3). For more details on various permissions settings and considerations, refer to the Dataverse guide in the footnote below.[20]

To edit permissions, click the 'Edit Dataset' button on the dataset page (figure 20), click 'Permissions', and then chose whether you want to edit permission on the dataset or file level. Note that not all Dataverse installations let you modify these settings.

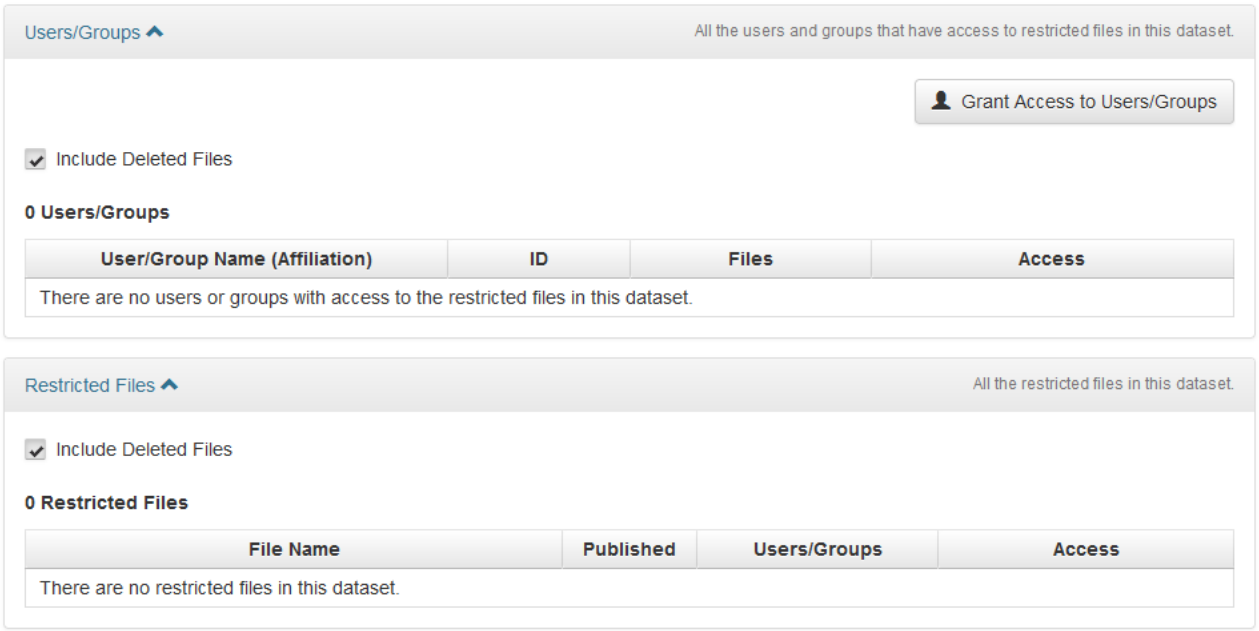

FIGURE 34: PERMISSIONS MENU (FILE LEVEL)

ACCESS REQUESTS

If users are allowed to do an access request for a restricted file, they will see a 'Request Access' button in the User Interface. Users need to be logged in to do such a request. After the button is clicked, an email will be automatically sent to the curators and admins of the dataset. In this message there will be a link to the page where you can reject or grant access to the file(s) (see figure 35). You can also navigate to this page by clicking 'Edit Dataset>Permissions>File'. On this page, you can either click 'Grant' to allow the user to download the file, or click 'Reject' to refuse access to the file.

[20] https://guides.dataverse.org/en/5.6/user/dataset-management.html#id40

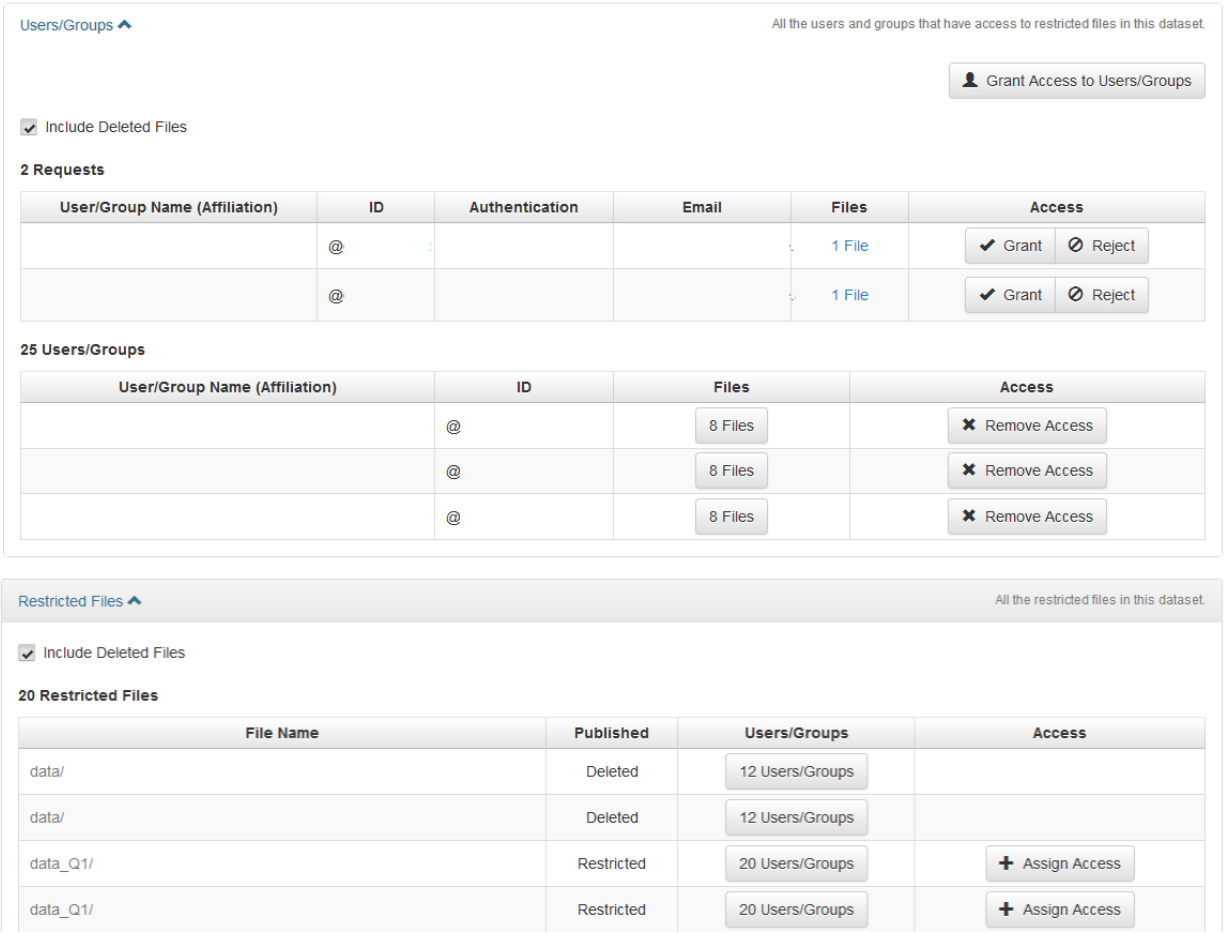

FIGURE 35: MANAGING ACCESS TO RESTRICTED FILES (FILE LEVEL)

5.2.5. Private URL

You can create a preview URL of the dataset to share with trusted third persons while the dataset is not published yet, they do not need to be logged in to do so. Also be aware, that the link also allows for downloading of data files.

The 'Private URL' can be deactivated again by clicking the 'Disable Private URL' button.

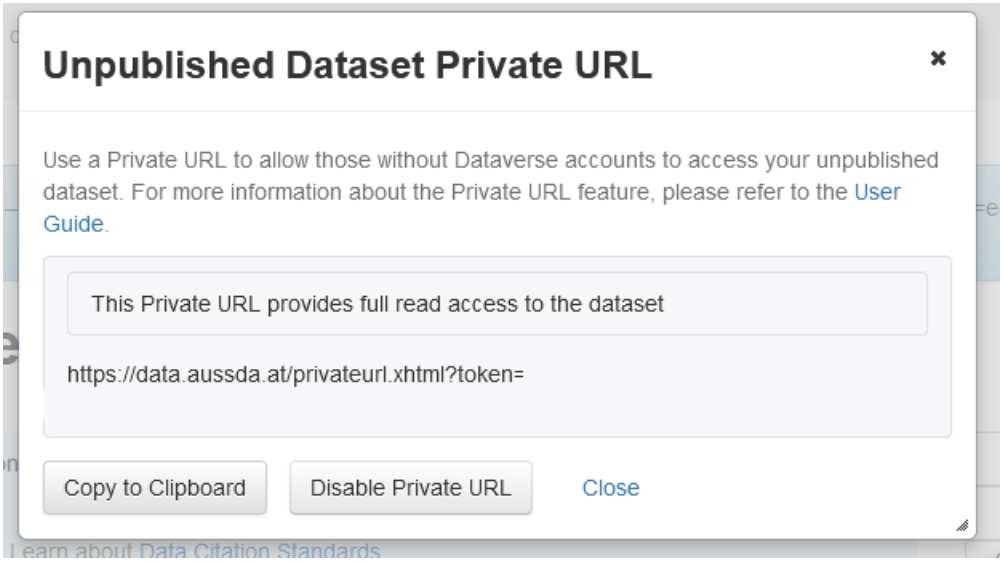

FIGURE 36: DISABLE AN ACTIVE 'PRIVATE URL'

5.2.6. Thumbnails + Widgets

If the feature is activated, Dataverse automatically chooses a 'Thumbnail' as a picture that shows up in the search results and on the dataset page. You can also change that thumbnail manually (see File metadata, figure 26).

FIGURE 37: SEARCH RESULTS WITH AND WITHOUT THUMBNAIL

If you want to link to your Dataverse or dataset from your own or your institution's website, you can implement 'Widgets', either for datasets or for dataset citation.

To learn more about 'Thumbnails' and 'Widgets' visit the Dataverse User Guide[21].

[21] https://guides.dataverse.org/en/5.6/user/dataset-management.html#id44

5.2.7. Delete Dataset

Under dataset settings you can also delete dataset drafts. If you delete the draft, be very sure, it cannot be restored. If you already published a version and created a draft of the next version, you can delete this draft here (see figure 20).

5.2.8. Deaccession

Deaccessioning a dataset should not be taken lightly (see the below warning), especially if you provide persistent identifiers for your datasets. A reason must be given why a deaccession is necessary. If you provide persistent identifiers like Handle or DOI and those are provided as an URL in the citation of the dataset, a tombstone landing page will always be accessible to users. This tombstone page will have basic information but no files. If your dataset has more than one version, you can deaccession selected versions or all versions at once. The deaccession options (figure 38) can be found from the 'Edit Dataset' menu on the dataset page (figure 20).

If the deaccesioned dataset was moved to another location, it is recommended to add a link to this location. This link will then be shown on the tombstone page.

> 'Warning:
>
> It is not recommended that you deaccession a dataset or a version of a dataset. This 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.'[22]

[22] Dataset Deaccession, Dataset + File Management User Guide, https://guides.dataverse.org/en/5.6/user/dataset-management.html#id59

Also be aware, that if you deaccession a version of your dataset, difference notes between versions will not be available.

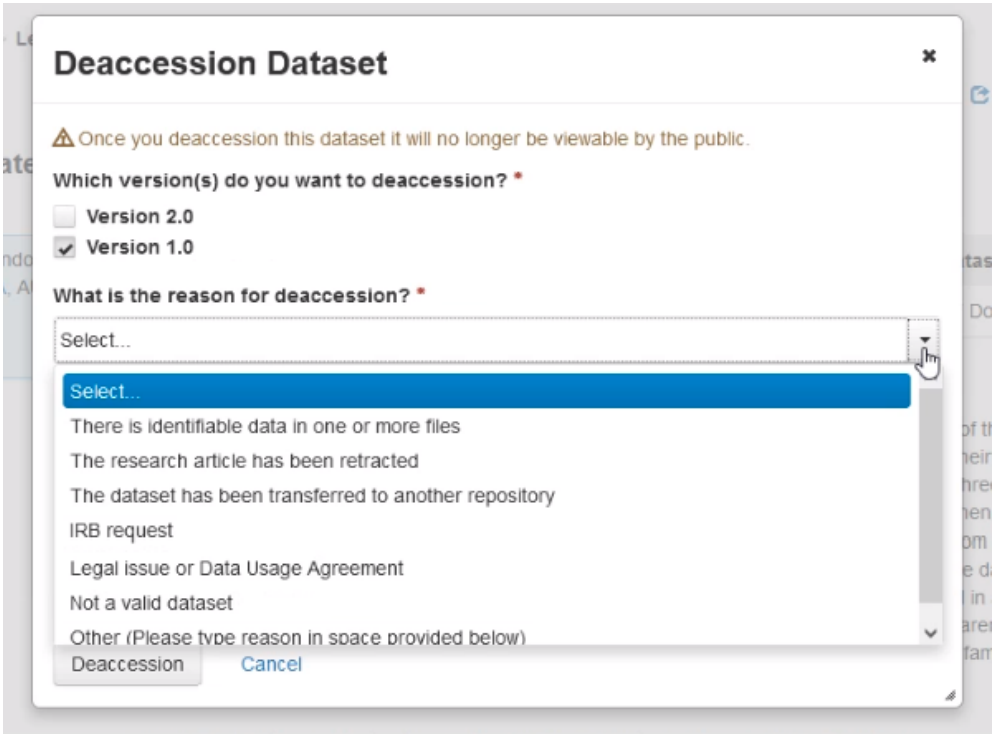

FIGURE 38: DEACCESSION OF A DATASET

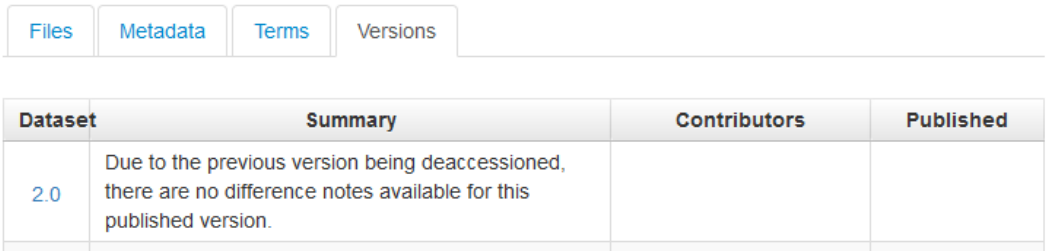

FIGURE 39: VERSION NOTES UNAVAILABLE FOR DEACCESSIONED DATASET

5.3. Publish or Submit for Review

Thoroughly check your dataset in Dataverse before the publication or submission for Review. Your draft is now finished, congratulations!

If you have submit rights: You can go ahead and submit your entry! Go to 'Submit for Review' Button right beside/below the dataset citation box. The Admin or Curator of the Dataverse will be notified and can publish the dataset or contact you by clicking 'Return to Author'. Either way, you will receive a notification.

If you have publishing rights: You can go ahead and publish your entry! '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.'[23] Go to 'Publish Dataset' Button right beside/below the dataset citation box. Confirm by clicking 'Continue' in the pop-up. Be very sure, once a dataset is published, it cannot be unpublished. If you do not get a 'Success' message at the top of the page, try again.

Publishing may take some time as Dataverse validates your files, so please be patient. After publishing, the 'DRAFT VERSION' in the citation box below the title will turn into the current version, for example 'V1'.

5.4. Versions

In the tab 'Versions' you can find links to all published versions of a dataset. You see the version, a summary of what changed between versions, who contributed and the publish date of the version.

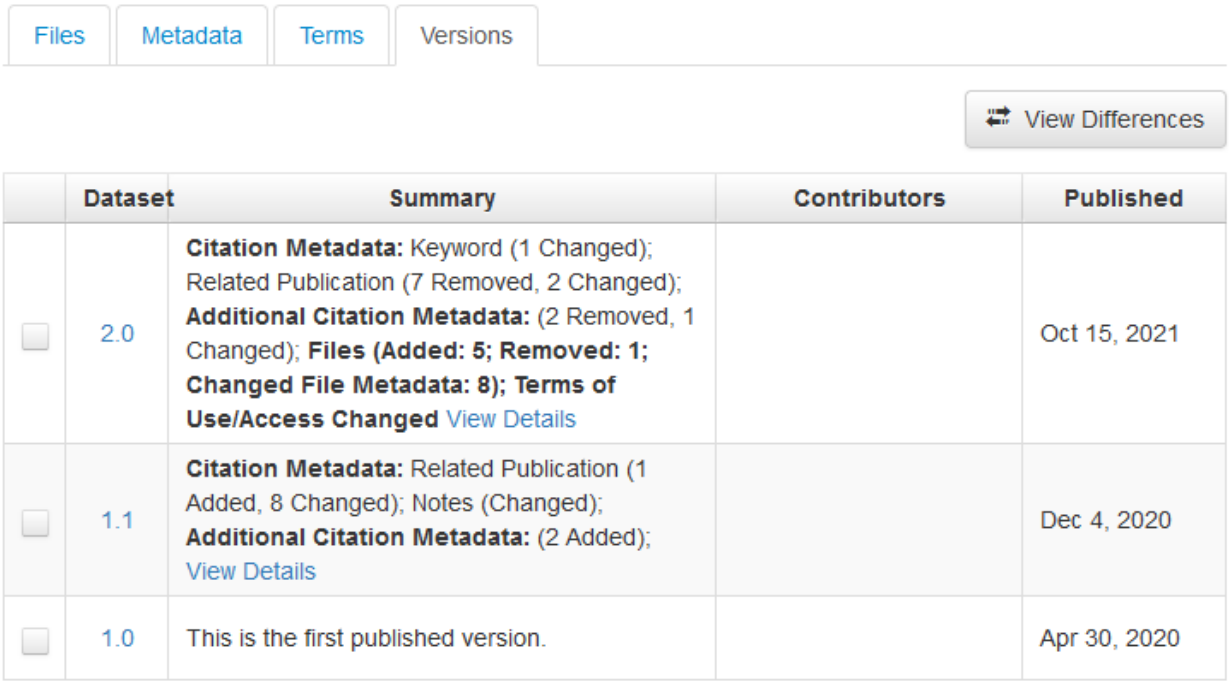

FIGURE 40: DATASET VERSIONS

In this example, the current version is 2.0, by selecting 'View Details' a more detailed report of changes is displayed. You can also compare two versions by selecting 'View Differences'.

[23] https://guides.dataverse.org/en/5.9/user/dataset-management.html#id50

FIGURE 41: VERSION HISTORY UNAVAILABLE FOR DEACCESSIONED DATASET

6. References / Additional resources

- **Dataverse User Manual**[24]

This guide provides an in-depth guide to all aspects of Dataverse, including more advanced features. This guide is hosted and written by the Harvard IQSS Dataverse team.

All information above is more or less also displayed in the 'User Guide' section of the Dataverse Guides. The Dataverse Project also provides the 'Admin Guide'[25], the 'API Guide'[26], the 'Installation Guide'[27], the 'Developer Guide'[28] and the 'Style Guide'.[29]

- **AUSSDA – Self-Deposit Manual**[30]

AUSSDA created a Manual for its Self-Depositors who are users that add their own data to the AUSSDA Dataverse. In addition to covering the topics data curation, it also shows how to put datasets into the AUSSDA Dataverse.

- **University of Manitoba Dataverse Guide**[31]

Comprehensive Dataverse guide for end-users

- **University of Manitoba Quick Guides**[32]

Practical one-page guides for users and admins respectively

[24] https://guides.dataverse.org/en/5.6/user/index.html
[25] https://guides.dataverse.org/en/5.6/admin/index.html
[26] https://guides.dataverse.org/en/5.6/api/index.html
[27] https://guides.dataverse.org/en/5.6/installation/index.html
[28] https://guides.dataverse.org/en/5.6/developers/index.html
[29] https://guides.dataverse.org/en/5.6/style/index.html
[30] https://aussda.at/fileadmin/user_upload/p_aussda/Documents/Self-Deposit_Manual_v1_0.pdf
[31] https://libguides.lib.umanitoba.ca/dataverseguide/
[32] https://libguides.lib.umanitoba.ca/c.php?g=711163&p=5067688%20#dverseQuickGuides

7.List of Figures

Figure 1: Account Information 4
Figure 2: Sorting And Viewing Search Results 6
Figure 3: View Files 8
Figure 4: Tree View With Folder Structure 8
Figure 5: Cite Data 9
Figure 6: Download (Meta)Data 9
Figure 7: Selecting Files For Download 10
Figure 8: Downloading Metadata And Files In Original And Other Formats 10
Figure 9: File Tab 'File Metadata' Page 11
Figure 10: Schematic Diagram Dataverse In Dataverse In Dataverse 4.0 12
Figure 11: Editing A Dataverse 14
Figure 12: Dataverse Permissions 16
Figure 13: Dataverse Permissions Menu 16
Figure 14: Dataverse Edit Groups Menu 17
Figure 15: Example Dataset Template 17
Figure 16: Guestbook Creation/Editing Menu 18
Figure 17: Managing Guestbook(S) 18
Figure 18: Featured Dataverses In Dataversenl 19
Figure 19: Adding A New Dataset To A Dataverse, Using Only The Citation Metadata Block 20
Figure 20: Editing A Dataset 21
Figure 21: A Metadata Entry Providing Information For Three Metadata Blocks 21
Figure 22: Files Tab 22
Figure 23: Doi And 'Export Metadata' Button Of Dataset 22
Figure 24: Uploading Files 23
Figure 25: Selecting Files To Edit 24
Figure 26: Editing File Metadata 25
Figure 27: Edit File Tags 26
Figure 28: Restricting Files And Adding Terms Of Acccess 27
Figure 29: Replacing Files And Viewing Version History 27
Figure 30: Terms Tab 28
Figure 31: Dataset Terms Of Use 29
Figure 32: Dataset Terms Of Access 30
Figure 33: A Dataset With An Active 'Guestbook' Called 'Users' 30
Figure 34: Permissions Menu (File Level) 31
Figure 35: Managing Access To Restricted Files (File Level) 32
Figure 36: Disable An Active 'Private Url' 33
Figure 37: Search Results With And Without Thumbnail 33
Figure 38: Deaccession Of A Dataset 35

Figure 39: Version Notes Unavailable For Deaccessioned Dataset 35
Figure 40: Dataset Versions 36
Figure 41: Version History Unavailable For Deaccessioned Dataset 37

SSHOC
social sciences & humanities open cloud

www.sshopencloud.eu
@SSHOpenCloud
in/company/sshoc
info@sshopencloud.eu