

# Curator Guide

[site.uit.no/dataverseno/admin-en/curatorguide/](http://site.uit.no/dataverseno/admin-en/curatorguide/)

Below you will find guidelines on how research data support at DataverseNO partner institutions are meant to curate their collections in DataverseNO. In case you need additional help or have questions about curation, please contact the [support services](#) at your institution.

## Curation of datasets

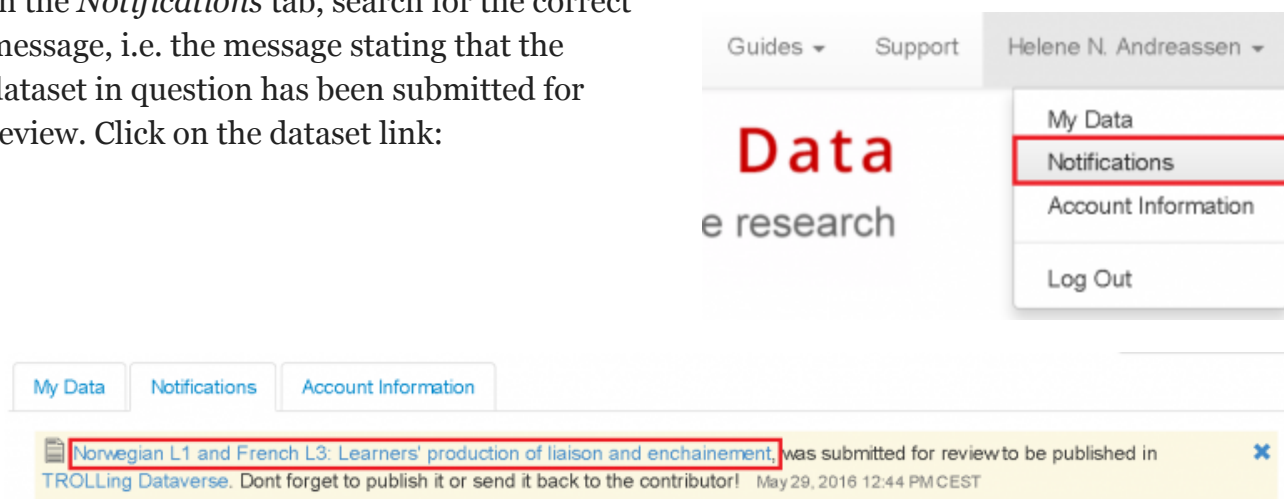


### General



When a user has created a dataset and submitted it for review, the curator(s) of the collection / (sub)dataverse in question are automatically noticed by email. As a curator, you log into DataverseNO and click on your username in the upper right corner, and then click on *Notifications*:

In the *Notifications* tab, search for the correct message, i.e. the message stating that the dataset in question has been submitted for review. Click on the dataset link:



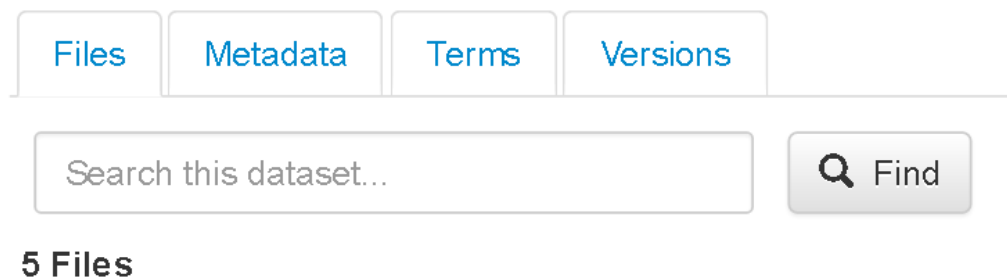
The link takes you to the landing page of the dataset, and you are now ready to curate it.

The first thing you as a curator should check is whether the contents and the author(s) of the dataset meet the requirements in the **DataverseNO Accession Policy**. Here is a short summary of the most important points:

- At least one of the authors of the dataset is or has been affiliated with the partner institution in question. Other rules may apply for special collections.
- The dataset must be suitable for open access publication.

You should also check if the dataset is created in the **right collection** (= (sub)dataverse). For a 'common' user from a DataverseNO partner institution (e.g. UiT), this will be the institutional collection for that institution (e.g. UiT Open Research Data). But for other users, the right collection may be a special collection. For a linguist, e.g., the right collection will usually be TROLLing. If a dataset is created in the **wrong collection**, you curate the dataset in the usual way (see below), but **after** the dataset has been published, you should inform [researchdata@hjelp.uit.no](mailto:researchdata@hjelp.uit.no) about which collection the dataset has to be moved to.

The curation of a dataset is essentially about ensuring that dataset is structured and documented according to best practice, as described in the Deposit Guidelines (see the menu item [Deposit](#)). There are four main areas to be reviewed: files, metadata, terms, and versions. You get to these areas by clicking the corresponding tabs on the landing page, i.e. *Files*, *Metadata*, *Terms*, and *Versions*:



As of today, the notice telling the curator(s) that a new dataset has been submitted for review does not contain any information about whether this is a new dataset or a new version of a previously published dataset. Therefore, it is a good idea to start curating by looking at the **versions tab**. Here you will immediately see whether the dataset is entirely new, or whether it is a new version of an existing dataset. In the latter case, the versions tab will also tell you what changes that have been made since the previous version, and you then only have to review these changes. Read more about versioning in the section *New version of a published dataset* below.

If you are dealing with a new dataset, you should start curating by having a look at the metadata. They will give you an overview of what the dataset is about.

**Note!** If it is obvious that the depositor has not consulted the Deposit Guidelines (e.g. if there is no ReadMe file), it might be just as well not to curate the dataset yet, but rather return it to the depositor (*Return to Author*), and send him/her an email asking him/her to consult the Deposit Guidelines ([in Norwegian](#) | [in English](#)) and re-submit the dataset for review once it is organized and documented in line with the guidelines.

## Metadata



Select the *Metadata* tab, and click the *Add + Edit Metadata* button.

Curation checklist for the most common metadata fields (for a more detailed and extensive overview, see the Word document

“Dataverse\_North\_metadata\_best\_practices\_guide\_v2.0\_ENG.docx” in the DvNO-brukerforum Team):

Are the fields filled in correctly? E.g.:

- **Title:** Quotation marks should not be added around the title as such marks will be added automatically in the reference generated by the system. If the dataset contains background / replication data for a publication, “Replication data for:” may be added to the title of the dataset.
- **Author:**
  - Is/are the name(s) of the author(s) inverted (family name, given name)?
  - Is affiliation provided (e.g. UiT The Arctic University of Norway)?
  - Authors should be encouraged to create an ORCID, and add it to the Identifier Scheme field.
- **Contact:**
  - If the contact is one or several persons, the name(s) have to be inverted (family name, given name). If the contact is an institution, the name is not inverted.
  - Are affiliation and correct email address provided?
- **Description:**
  - A brief presentation of the dataset should be given. To add breaks between to sections, add the HTML tags <p> and </p> around each section. This description should usually also be included in the ReadMe file. One may use (parts of) the abstract of the related publication, in which case the abstract should be entered into a second description field, which can be added by clicking the plus button to the right. NB! If your article is only submitted and not accepted (yet), DO NOT mention the name of the journal it has been submitted to.
  - The *Date* field must be filled in in the format YYYY-MM-DD.
- **Keyword:**
  - Are there reasonable *keyword terms* added? If there are more commonly used keywords already applied in the (sub)archive, you should make the user aware of this.
  - Has each keyword been assigned it own field? If not, ask the user to replicate the field for each keyword.

**Related publication:**

- If the dataset is the basis for a publication, is the reference to the publication provided here?
- If the dataset is used in a article or book manuscript that is submitted for review, but has not been accepted (yet), the name of the journal or publisher should **not** be mentioned in any place in the dataset. Rather, one should use the expression “Submitted for review” or the like. See also the section *Reading access to unpublished dataset* below.

- Also check with the researcher whether the article or book manuscript will be subject to **double-blind review** (both author and reviewers are anonymous). In that case, the researcher cannot share the dataset as it is with the editor, because the researcher's name is displayed in the *Version* tab. How to proceed:
  - The curator curates the dataset as usual, but without publishing it.
  - Once the dataset is ready for sharing, the curator must create a new, identical, but anonymized dataset in a collection which is dedicated to datasets that are part of double-blind review.
  - How to get access to this collection:
    - Send an email to [researchdata@hjelp.uit.no](mailto:researchdata@hjelp.uit.no) and ask for the username and password for the DataverseNO Review Curator.
    - Log into dataverse.no as Review Curator. You have to be logged out as “normal” curator, or open a new window in your web browser in incognito mode. Select the log-in alternative *Other Options – Username/Email*, and log in.
    - Navigate to this collection: <https://dataverse.no/dataverse/review> (Note! The collection is/must not be published.)
    - Create the anonymized dataset (see more information below), don't publish it, but create a private URL (see [Reading access to unpublished dataset](#)).
    - Log out as Review Curator.
  - How to anonymize the dataset:
    - Make sure that all information about the author is removed by either removing identifying information or adding “NN” into the metadata fields that need to be anonymized: *Author (Name etc.)*, *Contact (Name etc.)*, *Producer (Name etc.)* and possible other fields containing author-identifying information. In the default metadata template in the Review collection, “NN” is already added to these fields.
    - Also make sure that the author name and other identifying information is replaced with “NN” in the ReadMe file.
    - In addition, identifying file metadata must be removed from all files. See [this guidance](#) on how to remove this kind of information from Microsoft Office documents. For advice on how to remove personal information from other types of files, see the help pages of the software or information on the web. For any questions, contact DataverseNO.
- It is this **new**, anonymized dataset that the researcher **can share** with the editor, usually via a private URL (see [Reading access to unpublished dataset](#)).

- **But note!:** Once the article is accepted, it will be the **original** dataset created by the author which will be **published**, and the curator will then delete the anonymized version of the dataset. Therefore, it is crucial that the author right from the beginning uses an anonymized **reference** (incl. DOI) to the original dataset in the publication (manuscript).

- 
- **Language:** The field *Language* is not about the investigated language, but about the language of analyses. In a TROLLing dataset about French nouns described/analysed in English, this field should not contain “French”, but “English” or be left empty. However, “French” should be added as a keyword.
- **Producer:** Has there been entered a correct *Producer* of the dataset? In institutional collection, this field is pre-populated. But in special collections of the TROLLing type, you should check whether the user has filled in the right institution (e.g. UiT The Arctic University of Norway or the name of another institution of funder). If the author is not associated with any research organization anymore, the name of the author may be added in inverted order to the *Name*-field, and the other fields may be left empty. For datasets which are deposited by bachelor or master students, there are two alternatives: a) If the student was employed or paid in any other way by a (research) organization to produce the deposited data, the name of the organization is added to the *Name*-field. b) If the student was not employed/paid to produce the dataset, the name of the student may be added in inverted order to the *Name*-field. If – still in case b – the dataset was produced while the student was studying at a HE institution, the name of the institution may be added to the *Affiliation*-field.
- **Contributor:** Here credit should be given to those who have contributed to the dataset, including those who are listed as authors. The role of the contributors can be defined in the *Type* field (e.g. Data Collector).
- **Grant Information:** If there are things indicating that the deposited data are from a project with external funding, the curator should ask the depositor about this, and ask him/her to fill in this field if applicable. Use the full name of the funder, e.g. “The Research Council of Norway”.
- **Time Period Covered:** What time period is the data from or about?
- **Date of Collection:** When was the data collected / generated?
- **Kind of Data:** What kind of data is it? For suggestions, the depositor may hover the mouse over the question mark to the right of the field name, e.g. survey data, experimental data, observation data.
- **Data Sources:** If the deposited data were not generated or collected by the depositor/research project, information should be entered here about the sources. This can for example be an archive, a corpus, or a website from which the data were downloaded.
- **Distributor:** As a rule, this field is pre-filled with the name of the collection (e.g. *UiT Open Research Data* eller *TROLLing*). Make sure the content has not been edited.



- **Distribution Date (embargo):** This field is used for specifying a file embargo. If the author has put any access restrictions on (some of) the files in the dataset, the field *Distribution Date* must contain the date (YYYY-MM-DD) when the file(s) will be made available. In all other cases, this field must be empty. Read more about embargo [here](#). One month before the end of the embargo period, a reminder is sent to [researchdata@hjelp.uit.no](mailto:researchdata@hjelp.uit.no), and the curator will then remind the author about this.
- **Geographic Coverage:** Many datasets are related to one or more geographical places or areas. In case the user has not entered any metadata about this in the metadata section *Geospatial Metadata*, you should recommend him/her to do so.
- In addition, you may ask the depositor to have a look at the **domain-specific** metadata schemas following *Geospatial Metadata*, and add relevant information if applicable.
- Generally:
  - Are the provided metadata sufficient to make the dataset findable in search engines?
  - Are the metadata provided in English or another commonly used communication language in the scientific field in question?
  - **Note!** Metadata fields must not contain certain HTML tags or other special characters (e.g. [ and ]). This applies in particular to the *Description* field. To add breaks between to sections, add the HTML tags <p> and </p> around each section.

## Files



### Curation checklist for files:

- Is the dataset documented in a **ReadMe file**? Note! This is an absolute **requirement**. We recommend researchers using our [ReadMe file template](#). The ReadMe file name must contain the string “readme” (in small and/or capital letters), and not split between the file name and the extension, thus e.g. “ReadMe.txt”, and *not* “Read.me”).
- Is forced numbering applied to the ReadMe file (e.g. “00\_ReadMe.txt”), so that it appears on the top of the file overview?
- Can the files be opened?
- Are the file names consistent and understandable?

- Are the data provided in (a) preferred file format(s)? Consult the [list of preferred file formats](#) in the deposit guidelines. If the file type is not on this list, you take a look at other pages such as [DANS file format list](#), the [Library of Congress information pages](#) on file formats, and the [UK National Archive's PRONOM service](#). If you are unsure about the file format, contact DataverseNO at [researchdata@hjelp.uit.no](mailto:researchdata@hjelp.uit.no). If the data cannot be stored in a preferred format, they can still be published, but with the restrictions this implies for long-term preservation (cf. DataverseNO Preservation Policy: Datasets in non-preferred format(s) will not be migrated to new formats to avoid format obsolescence). New preferred file formats should be discussed in the curator group. Contact DataverseNO at [researchdata@hjelp.uit.no](mailto:researchdata@hjelp.uit.no) in case you want a new file format to be included on the DataverseNO list of preferred formats. NB! Container files are not preferred. In a new version of Dataverse, folders are retained at file upload, and there is therefore no need for container files anymore.
- If appropriate, the data files may also be archived in the original file format(s) **in addition to** preferred format(s).
- Do all files have a file extension, e.g. .txt, .pdf?
- If data is uploaded in both the original and a preferred format, the file name of the original file must be identical with the file name in the preferred format. (Otherwise, creating file overviews for long-term preservation will be very difficult.)
- File names must not contain spaces, commas and other special characters.
- If an **embargo** is applied the curator must ensure that the embargo information on file level is provided as described [here](#).
- File size:
- DataverseNO has **no upper size limit** for a dataset. **However**, below are some advices and procedures for handling uploads of large files.
- A file upload can consist of several files. If a user must split the files on several uploads this can be done by saving the dataset after each upload.
- The following advice, size limits and procedures apply to single files, file uploads and datasets:
  - **The size of individual files should not exceed 5 GB.** Bigger files can create problems for others when it comes to downloading and reusing data.
  - **A file upload should not exceed 10 GB** in total size to minimize the likelihood that errors will occur when transmitting data over the Internet Protocol (http).
  - Upload of single files **bigger than 20 GB but less than 50 GB** has to be agreed on and scheduled with UiT. The curator agrees/clarifies with the researcher and with UiT at [researchdata@hjelp.uit.no](mailto:researchdata@hjelp.uit.no).
  - If the files in a **dataset are more than 50 GB in total**, the handling of the dataset must be agreed on and scheduled with UiT. The curator agrees/clarifies with the researcher and with UiT at [researchdata@hjelp.uit.no](mailto:researchdata@hjelp.uit.no).

You can find detailed information on file naming conventions, preferred file formats, and documentation of research data in the section *Prepare your data for depositing* in the menu item Deposit. There, you will also find detailed guidance on how to save/convert different document types into preferred file formats. If you have questions about this, please contact the support services of your home institution.

Best practice implies saving tabular data as tabulator-separated plain text files, encoded in Unicode UTF-8 without so-called BOM ('Byte Order Mark'). If this is not possible within the spreadsheet software, you may do this in Notepad++ as described here:

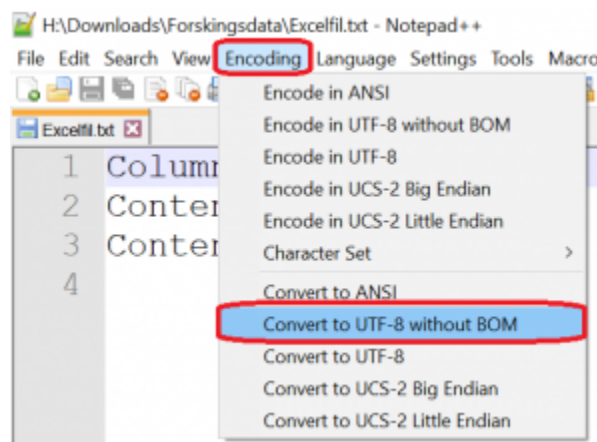
1. Open the (converted) text file (.txt) in Notepad ++ (Notepad++ is based on open source code, and may be downloaded from <https://notepad-plus-plus.org/>. Ask IT support at your institution for help to install the software on your computer).
2. Click *Encoding* in the top menu, and select *Convert to UTF-8 without BOM*:
3. Save the file.

## Statistics data (e.g. R and SPSS)

A useful overview of file formats that are used in various statistics programs is available here.

As for R, their conclusion is as follows:

In conclusion, if you are working with R you should provide a .csv\* file which includes your data and separate .R- or .Rmd-files which include your syntax to ensure long-term availability. Additionally, you may add Rdata-files for easier access to the same information.



(\* We recommend tabulator-separated Unicode UTF-8 .txt.)

The script is in the .R files. .R-files are plain text files (the .R extension may be replaced by .txt). .Rmd:

Rmd RMarkdown files are a great way to combine data documentation, data visualization and data analysis in one single file.

In other words, we want this:

- Basic data as tabulator-separated Unicode UTF-8 text files (.txt) = preferred file format
- The R code as Unicode UTF-8 (.R) = preferred file format
- .rda = non-preferred file format, but works in R, which is an open source based and openly documented software
- Possibly .rmd

## Terms



The default license in DataverseNO is CCo. But if this license can not be used for the dataset in question, then you must change the information in the Terms tab. Any changes of the default terms should be discussed with the superuser(s) at the DataverseNO partner institution at stake. (When the author selects the option of not accepting the CCo terms, he/she is provided with a [Sample Data Usage Agreement](#) by the system.)

For CC BY, we have agreed on the text below. Once the research data support group has decided on the license issue the text can be pasted into the field *Terms of Use* (the quotations marks must be removed):

“This dataset may be reused according to the Creative Commons Attribution 4.0 International (CC BY 4.0) license as described here: <a href=“https://creativecommons.org/licenses/by/4.0/” title=“TermsOfUse” target=“\_blank”>https://creativecommons.org/licenses/by/4.0/</a>.”

For any other license, contact DataverseNO.

Note! DataverseNO only accepts licenses that provide access to data. See the [DataverseNO Access and Use Policy](#):

In line with the intention of DataverseNO to provide maximum public access to unrestricted research data, DataverseNO promotes licenses that are recommended for the re-use of research data, and only accepts licenses providing access to deposited data in one form or another.

If the dataset under curation is (partly) based on data from other sources, one must examine what terms of use these data have, choose an appropriate license, and – if necessary – specify any restrictions in the *Terms* tab.

If the dataset under curation is based on several sources, the most restrictive terms of use determine which license the dataset under curation can be published under. Example: A dataset based on one data set with CC-BY and one with CC-BY-NC can be published under CC-BY-NC.

If the terms of use in the original dataset are formulated as a standard license (e.g. CC-BY), it is usually unproblematic to choose a license for the derived dataset as the license for the original dataset usually specifies under which license one can publish a derived dataset.

If the terms of use are formulated as a tailor-made text, then you must first clarify which license for re-use the terms of use allow for. If they e.g. state that the data can be reused freely and without restrictions as long as you refer to the source, the terms are usually compatible with a CC-BY license.

If a source specifies unclear terms of use, one should clarify with the licensee which standard license these terms of use are compatible with. The answer from the licensee should be added as a PDF/A to the dataset under curation.

If a source does not specify any terms of use, you are basically not allowed to use the data for anything – unless it is clear that (the intended reuse of) the data are not protected by IPR and/or database protection (see the two datasets below for specific examples). In such cases, you should contact the licensee to clarify the conditions that apply to the reuse of the data. You should preferably get a clarification of which standard license the data can be reused under. The answer from the licensee should be added as a PDF/A to the dataset under curation.

Here are two datasets that are based on a number of sources:

<https://doi.org/10.18710/NMKI2B>, <https://doi.org/10.18710/VMUP44>. The rights clarifications documented in the *Terms* tab and in the attached PDFs in these two datasets may be helpful when the curator is to provide guidance on this to the researcher.

## Return dataset to author



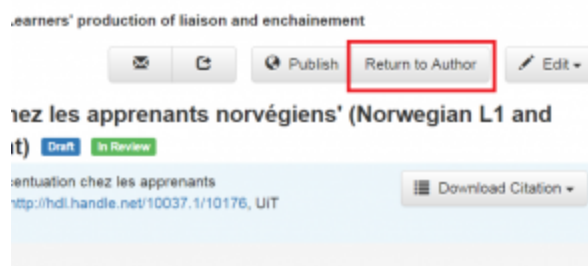
If a submitted dataset has not been appropriately structured and documented, the curator returns the dataset to the author:

**Note!** In addition, the curator sends an email to the author specifying the necessary changes to be made before the dataset can be published.

The author should also be referred to (the relevant sections in) the Deposit Guide (<https://site.uit.no/dataverseno/deposit/>) on the DataverseNO info page

(<https://info.dataverse.no>). It is possible to

provide links to specific sections in the Deposit Guide. To get the right link address, hover your mouse on the link icon at the beginning of the section in question, right-click it, and select “Copy Link Location”:



You should also ask the author to click *Submit for review* once again after having made the necessary changes. The email to the author may be sent in two ways:

1. When you are in the dataset in question, click the button with letter symbol, and write your message in the window that is popping up:

2. When you are in the dataset in question, click *Edit* > *Metadata*, and copy the email address in the field *Contact* > *Email*:

and send the message in your email program (e.g. Outlook).

**Note!** If the curator identifies fundamental nonconformity with the DataverseNO policies and guidelines, and the depositor does not agree to make necessary changes the dataset must not be published. If the curator is in doubt whether the dataset complies with the DataverseNO policies and guidelines the issue should be discussed with the DataverseNO administrator team at UiT The Arctic University of Norway. Contact the team at [researchdata@hjelp.uit.no](mailto:researchdata@hjelp.uit.no). Ultimately, the Board of DataverseNO is to decide on such matters.

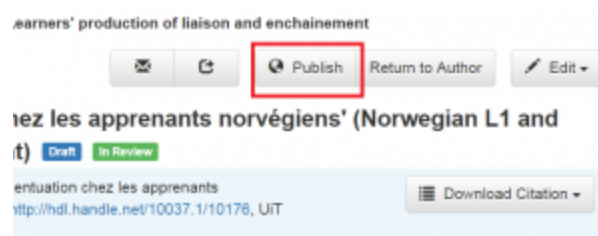
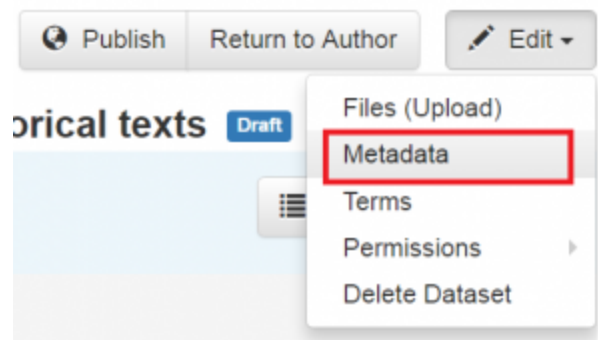
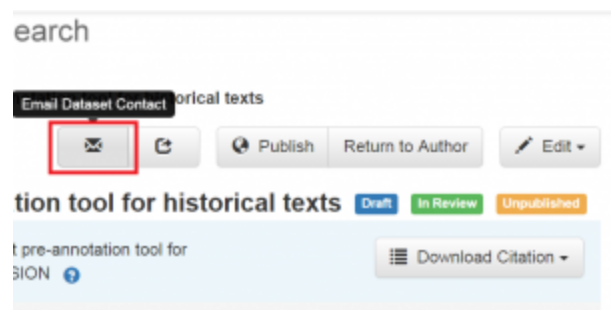
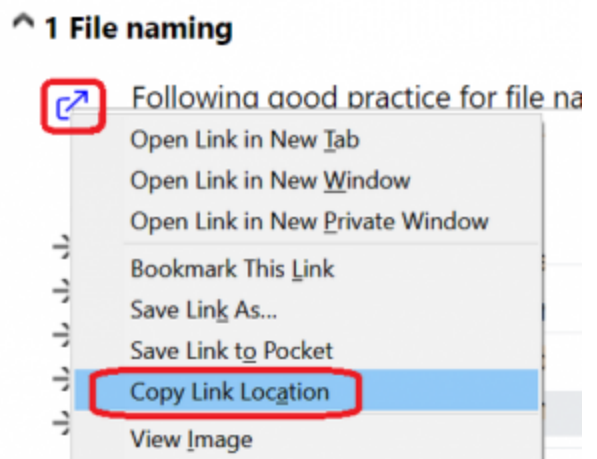
## Publish a dataset

When everything is OK with the dataset, the curator publishes it by clicking the *Publish* button:

The author receives an automatic confirmation by email stating that the dataset has been published.

## Promotion in social media

Some archives are promoted by posting information about new datasets in social media. In TROLLing, the UiT Library post messages on [Twitter](#) and on the [TROLLing Facebook group](#) telling that a new dataset has been published. They also send a email to the author with the following message:



I have now published your dataset. Thanks for sharing your data! You can find an announcement of the upload on our Facebook and Twitter page, and we encourage you to like this in order to get updates about the archive:  
<https://www.facebook.com/TromsoRepositoryofLanguageandLinguistics/>.

## New version of a published dataset (also when removing embargo)



When an author makes changes in a published dataset, a new draft is created. This draft must be submitted for review in order for the new version to be published. The curator(s) will then be noticed that a new dataset is waiting for curation. **Note!** As of today, it is not apparent from this message whether the submitted dataset is an entirely new dataset or a new version of a previously published dataset. Often, a long time may have passed since the previous version was published, and you may not recall that a previous version of this dataset already has been published. It is therefore advisable to start the curation process by checking whether the dataset has more than one version. To see this, click the Versions tab:

| Files | Metadata  | Terms           | Versions           |
|-------|---|-----------------|--------------------|
| DRAFT | Additional Citation Metadata: (1 Changed); <a href="#">View Details</a> | Stefan Hartmann | December 7, 2016   |
| 1.0   | This is the first published version.                                    | Philipp Konzett | September 30, 2016 |

By clicking *View Details*, you get an overview of all changes that have been made between the different versions. As a next step, you should then have a closer look at the changes made in the metadata and/or files. To do this, follow the guidance in previous sections above. When you publish the dataset after having curated it, you are asked to specify the new version number:

As a general rule, the option *Minor Release* should be selected when only the metadata have been changed. In case there have been changes in the data files, the option *Major Release* should be selected. **Note!** When publishing a new version after **removing an embargo** / locks on file(s), the alternative *Minor Release* should be chosen, since we do not want the version number in the dataset reference to be changed.

## Publish Dataset

⚠ Are you sure you want to republish this dataset?

ℹ Select if this is a minor or major version update.

☒ Minor Release (1.1) ☐ Major Release (2.0)

Continue

Cancel

From time to time, the curator(s) should check whether there are unpublished datasets (drafts) that have not been submitted for review. If a dataset has the status *Unpublished* for more than three months, the curator(s) should contact the author and remind him/her that



they have to click *Submit for review* in order for the (new version of the) dataset to be published.

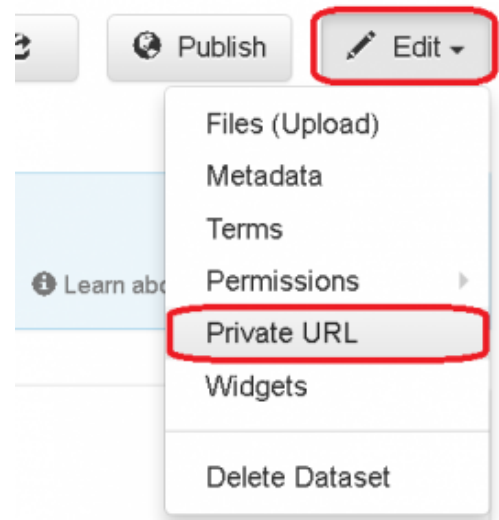
## Reading access to unpublished dataset



Scenario: An author wishes to grant access to a dataset to a collaborator, a peer reviewer, a journal editor or the like before the dataset is published.

Solution:

- Log into DataverseNO, and go to the unpublished dataset.
- Click the *Edit* button to the right, and select *Private URL*:
- Copy the private URL, and send it to the author or, if agreed on, to the person who needs to access the dataset.
- Private URLs can be created of any dataset in DRAFT state, even if there exist previously published versions of the dataset – although the latter case may not be relevant for sharing with editors for peer review of a publication.



## Reading access to locked file(s) in published dataset



Scenarios:

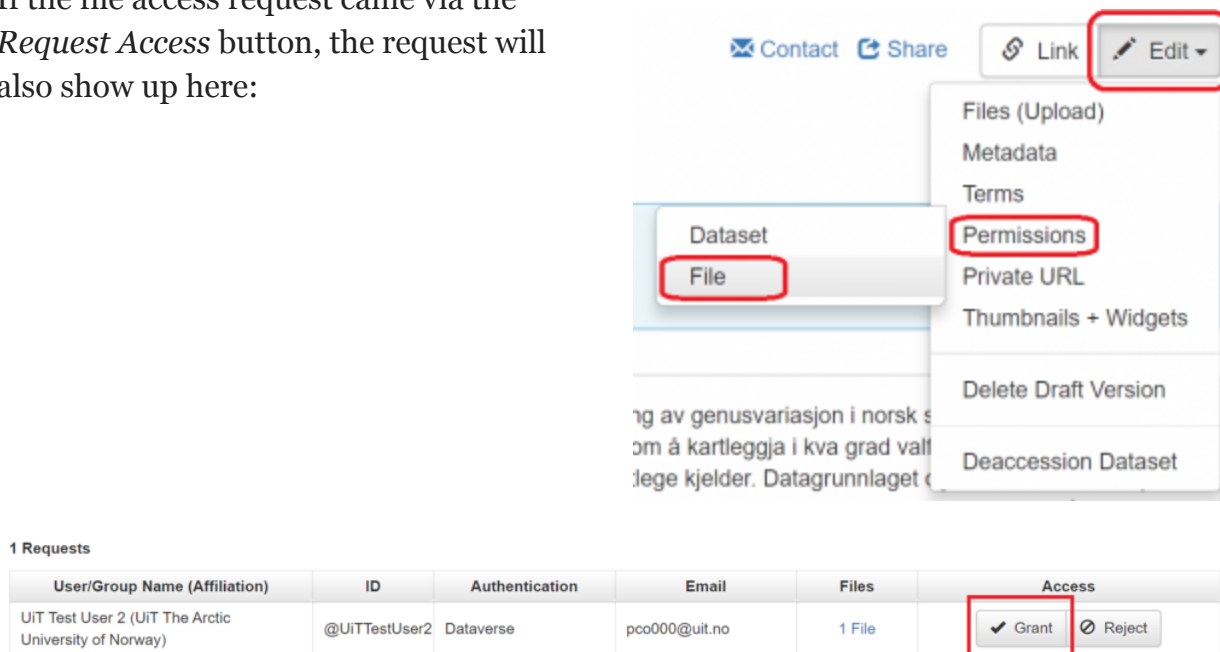
- An author wishes to grant access to (a) **locked file(s)** (= file with embargo) in a **published** dataset to a collaborator, peer reviewer or the like.
- A researcher requires access to (a) locked file(s) by clicking on the *Request Access* button. Note! This request option is only available for logged-in users. This means that researchers who want to get access to a locked file, must have a DataverseNO user account. Instructions on how to create a user account are provided in the [Deposit Guide](#).

Solution:

- If the file access request comes via the *Request Access* button, the curator receives a notification by email. Note! This notification is not sent to the depositor of the dataset (see further explanation below\*). Therefore, the curator must first inform the depositor about the file access request. If the depositor accepts the request, the curator can arrange file access in the following way:



- Log into DataverseNO, and go to the dataset in question (alternatively, click the link you got in the notification by email or in the *Notification* section in Dataverse).
- Click the *Edit* button to the right, and select *Permissions* and then *File*:
- If the file access request came via the *Request Access* button, the request will also show up here:



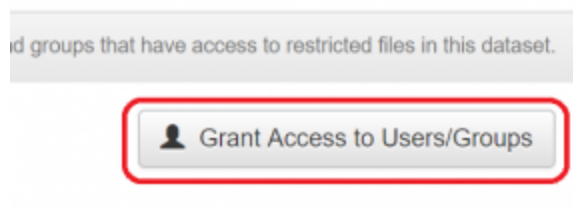
The screenshot shows the Dataverse interface. At the top, there are buttons for 'Contact', 'Share', 'Link', and 'Edit'. The 'Edit' button is highlighted with a red box, and its dropdown menu is open, showing options like 'Files (Upload)', 'Metadata', 'Terms', 'Permissions' (highlighted with a red box), 'Private URL', 'Thumbnails + Widgets', 'Delete Draft Version', and 'Deaccession Dataset'. Below the dropdown, there is a sub-menu with 'Dataset' and 'File' (highlighted with a red box). Below this, there is a table titled '1 Requests'.

| User/Group Name (Affiliation)                         | ID            | Authentication | Email         | Files  | Access  |
|---|---------------|----------------|---------------|--------|---|
| UIT Test User 2 (UIT The Arctic University of Norway) | @UITTestUser2 | Dataverse      | pco000@uit.no | 1 File | <input checked="" type="checkbox"/> Grant <input type="checkbox"/> Reject |

In this case, you simply can grant file access by clicking the *Grant* button. If access only has been accepted for selected files, click the link in the *Files* column, and select the appropriate files.

- If the file access request did not come via the *Request Access* button, the curator must search for the user who wants to get access in the following way:
- Click *Grant Access to Users/Groups*:
- Search for and add the user who should have access to the file(s) in the field *Users/Groups*, select the file(s) the user needs to have access to, and click *Grant*:

#### > Restricted File Permissions



The screenshot shows the 'Restricted File Permissions' section. It contains a text input field for searching users and groups. Below the input field, there is a button labeled 'Grant Access to Users/Groups' with a person icon, which is highlighted with a red box.

(\*The contact for a dataset or dataverse is where email is sent when you click the *Contact* button. When access to restricted files is requested, email does not go to the contact.

Rather, email is sent to the people who have the ability to grant access, which are the people who have a role that contains *ManageDatasetPermissions*. In DataverseNO, these people are usually the curators of the dataverse in question.)

## Edit access to a dataset



When creating a dataset in Dataverse, the depositor is automatically granted edit access to that dataset. However, in some cases it may be appropriate to manually assign edit access. Consider the following scenarios:

- Scenario 1: A curator has created a dataset on behalf of an author (cf., e.g. the pilot project on research data management at UiT in 2016). After the dataset has been created, the author wants to have a look at the dataset and possibly make some changes before it is published.
- Scenario 2: An author has created a dataset and wants other members of the research groups to be able to edit the dataset.
- Scenario 3: An author has created one or several datasets in an institutional archive (e.g. UiT Open Research Data), but has not yet published them. The author quits the institution, and consequently cannot access his/her dataset anymore. Therefore, we the author has to get created a new user account, either via Feide log-in (if s/he works at another institution using Feide), or via local authentication. When the new user account is created, the author must be granted access to his/her “old” datasets.

### Grant File Access

Grant file access to users and groups.

Users/Groups

Selecting 2 of 12 Restricted

| Users                               |         |
|-------------------------------------|---------|
| Testleif Testesen (Test University) | @test01 |

|                                     |                                  |
|-------------------------------------|----------------------------------|
| <input checked="" type="checkbox"/> | Bokmalmateriale_m+n.ods          |
| <input checked="" type="checkbox"/> | Bokmalmateriale_m+n.txt          |
| <input type="checkbox"/>            | Kjeldeoversyn_bokmal.ods         |
| <input type="checkbox"/>            | Kjeldeoversyn_bokmal.txt         |
| <input type="checkbox"/>            | Kjeldeoversyn_nynorsk.ods        |
| <input type="checkbox"/>            | Kjeldeoversyn_nynorsk.txt        |
| <input type="checkbox"/>            | Nynorskmateriale_m+n.ods         |
| <input type="checkbox"/>            | Nynorskmateriale_m+n.txt         |
| <input type="checkbox"/>            | Utgangsmateriale_bokmal_m+n.ods  |
| <input type="checkbox"/>            | Utgangsmateriale_bokmal_m+n.txt  |
| <input type="checkbox"/>            | Utgangsmateriale_nynorsk_m+n.ods |

Please contact the Dataverse administrator at your institution to get changed/assigned access rights on dataset level.

## Moving datasets



As of today, it is not possible to move a dataset between archives via the graphical user interface. If, e.g., a linguist from UiT has created a dataset on linguistics in UiT Open Research Data instead of TROLLing, this dataset should first be curated and published in the archive where it is created, and **after** the publication, the curator gives notice to [researchdata@hjelp.uit.no](mailto:researchdata@hjelp.uit.no) about where the dataset should be moved.

## Deleting published datasets



When a dataset has been published, its DOI has been activated. Through the DOI Agreement and the [DataverseNO Preservation Policy](#), the archive is committed to provide enduring access to the dataset for at least 10 years after its publication. If, after its publication, it turns out that a dataset for ethical, legal or other reasons should not have been published, we may

remove access to the data files in the dataset. However, the metadata entry will still be findable and accessible. Contact [researchdata@hjelp.uit.no](mailto:researchdata@hjelp.uit.no) to get the file access in a dataset removed.

### **Tasks in connection with long-term preservation**



DataverseNO commits to ensure that data published in the archive can be used in the long term. As part of this work, DataverseNO curators have several tasks which are specified in the Preservation Policy and the Preservation Plan, and which they will get assigned by the collection management.