



A U S S D A


AUSTRIAN  
SOCIAL SCIENCE  
DATA ARCHIVE

# SELF-DEPOSIT MANUAL

Self-depositing data at AUSSDA

**Version 1.0 (17.12.2020)**

Heider, Veronika; Butzlaff, Iris; Bodlos, Anita; Hönegger, Lisa

<b>Date</b>	17.12.2020
<b>Version</b>	1.0
<b>Status</b>	draft – commenting phase – testing phase – <b>implemented</b>
<b>License</b>	 <p>This work is licensed under a <a href="https://creativecommons.org/licenses/by/4.0/">Creative Commons Attribution 4.0 International Licence</a> (CC BY 4.0). This document contains material from the "Access policy" attributed to Kaczmirek L. and Hönegger L. and the document "Data deposit guideline" attributed to Butzlaff I. available at <a href="https://aussda.at/ueber-aussda/downloads/">https://aussda.at/ueber-aussda/downloads/</a>.</p>
<b>Access</b>	Public
<b>Suggested citation</b>	Heider, Veronika; Butzlaff, Iris; Bodlos, Anita; Hönegger, Lisa (2020). Self-Deposit Manual. Vienna: The Austrian Social Science Data Archive.
<b>Contact</b>	University of Vienna Vienna University Library and Archive Services AUSSDA -The Austrian Social Science Data Archive Universitätsring 1 1010 Vienna Austria  T +43 1 4277 15323 <a href="mailto:info@aussda.at">info@aussda.at</a>

## Table of contents

---

1.	Scope and purpose of this document .....	3
2.	The individual steps of the deposit process: an overview.....	3
3.	Legal questions .....	4
4.	Preparing your data and documentation material.....	5
4.1.	Check compliance with the GDPR and pseudonymise the data.....	5
4.2.	Data cleaning .....	7
4.3.	Recommended documentation material: What should be included?.....	8
5.	Self-depositing your data in Dataverse.....	9
5.1.	Login .....	10
5.2.	Create a dataset entry.....	11
5.2.1.	Metadata .....	11
5.2.2.	Files in Dataverse: Add your documentation and data .....	25
5.2.3.	Terms.....	32
5.3.	Publishing.....	34
5.4.	Updates .....	34
6.	Checklist Box for whole deposit process .....	36

## Version control

Version	Date	Description/Reason	Written/Revised by
1.0	17.12.2020	First published version	Authors

## 1. Scope and purpose of this document

---

Depositing and publishing data at a professional archive offer several advantages: Most funding agencies require data management plans from applicants for grants which include a description of the whole research data lifecycle, from data collection to data archiving and sharing.

Moreover, researchers who deposit their data, metadata and documentation material in the AUSSDA Dataverse increase the chances for receiving visibility, academic credit and increased citation counts.<sup>1</sup> In addition, the AUSSDA Dataverse provides data with persistent identifiers (a DOI) and allows for the preservation and storage of research data in a way that accessibility is ensured in the future. Thereby, the research community has the possibility to reuse data for research or teaching and replicate results.

Self-depositing data at AUSSDA is the fastest way for researchers to benefit from all these advantages of publishing data at a professional and certified repository. This document offers guidance on the entire deposit process and provides detailed step-by-step instructions on how data and documentation material need to be prepared in order to be ready for the self-deposit in the AUSSDA Dataverse. Please note that the self-deposit service is limited to quantitative data from the Social Sciences with a connection to Austria. Self-depositors can also find supporting materials at <https://aussda.at/>. If you are interested in depositing other types of data or from other disciplines, please contact us, we can refer you to a suitable repository.

If you are familiar with the individual steps of the data deposit process, you can jump directly to the checklist of the process to make sure that you do not miss any step (Section 6).

## 2. The individual steps of the deposit process: an overview

---

The deposit process starts with the decision to deposit data at AUSSDA and ends with the publication of the data. The following list provides an overview of the steps and decisions necessary to take for depositing data at AUSSDA.

- Clarify the legal framework for the published data sets and documentation material including licenses for each item (see Section 3).
- Prepare the data and documentation material for the deposit. Section 4 provides useful comments and processes for this step.
- Get to know Dataverse (see Section 5) by creating an entry for publishing your data (see Section 5.2)
  - Section 5.2.1 tells you more about metadata you need to add to your dataset in Dataverse. This step also creates a DOI for you.
  - Section 5.2.2 guides you through the next steps: file formats, adding the DOI and version as variables to the dataset, naming your files using the DOI and uploading everything to Dataverse.
  - The third component to every dataset are terms (see Section 5.2.3).
  - After checking everything, you can publish your dataset (see Section 5.3).
- If you need to update your dataset, see Section 5.4 on updates and versions.

---

<sup>1</sup> <https://dataverse.org>

### 3. Legal questions

---

Legal questions related to the deposit process concern three aspects:

- a. The agreement with AUSSDA that allows depositors to self-deposit data.
- b. License agreements for data, metadata and documentation material.
- c. The AUSSDA Non-Compliance Policy and potential consequences.

#### a) The agreement with AUSSDA that allows depositors to self-deposit data

In order to be authorised to self-deposit data at AUSSDA, each depositor is asked to sign an agreement which specifies the rights and obligations for the use of this service. This standard agreement is concluded once between the depositor and a legal representative of the University of Vienna and is the legal basis for all data deposits by this depositor. The agreement template is available on the AUSSDA website via:

[https://ausdda.at/fileadmin/user\\_upload/p\\_ausdda/Documents/AUSSDA\\_Self-deposit\\_Vertrag\\_SU\\_1\\_0.pdf](https://ausdda.at/fileadmin/user_upload/p_ausdda/Documents/AUSSDA_Self-deposit_Vertrag_SU_1_0.pdf)

#### b) License agreements for data, metadata and documentation material

License agreements determine how users of the AUSSDA Dataverse can use the available data and documentation. Self-depositors using the AUSSDA Dataverse need to be aware of three licenses under which their content will be published:

- All metadata in the AUSSDA Dataverse are published under the public domain dedication (CC0 1.0 Universal), a license that allows for open re-use without attribution.<sup>2</sup>
- All documentation material accompanying datasets (e.g. questionnaires or codebooks) are published open access under a Creative Commons Attribution 4.0 International license (CC BY 4.0). This means that documentation material is openly available and can be downloaded by all users of the website (after agreeing to the terms and conditions). A CC BY license allows for reusing and sharing the material under the condition that attribution is given to the original authors.<sup>3</sup> We recommend that you include a notice on the licence in the documentation material (see Section [4.3](#)).
- The “License for use for scientific purposes only” (SUF) is used for all datasets published by self-depositors.<sup>4</sup> This licence restricts the use of the material to specific scientific purposes. In addition, in order to access the data, users need not only agree to the terms and conditions, but they also have to identify themselves either by creating accounts or by login via their institutional account. **When uploading the datasets in the AUSSDA Dataverse, you need to set the proper restrictions for each data file to guarantee that users have to login before they can download data.** Section 5.2.2. “[Steps in Dataverse: Upload Files](#)” contains detailed instructions on how to set these restrictions.

---

<sup>2</sup> <https://creativecommons.org/publicdomain/zero/1.0/>

<sup>3</sup> <https://creativecommons.org/licenses/by/4.0/>

<sup>4</sup> <https://ausdda.at/en/contracts-and-licenses/scientific-use/>

### c) AUSSDA Non-Compliance Policy and potential consequences

Users of the AUSSDA Dataverse must agree to specific terms and general guidelines before using AUSSDA services. The [AUSSDA Non-Compliance Policy](#)<sup>5</sup> serves to inform depositors and users of social science data provided by the repository about the possible consequences in case of non-compliance with the agreed upon terms. In case the repository staff detects any forms of non-compliance or the repository is informed about the non-compliance of any of its users, AUSSDA may take certain measures and draw consequences according to this policy and at the discretion of the AUSSDA staff.

#### Take aways

Get in contact with AUSSDA and **sign the agreement** that allows you to self-deposit data.

Understand the license agreements for data, metadata and documentation material.

Read and understand the AUSSDA Non-Compliance Policy.

## 4. Preparing your data and documentation material

Data are of course the centrepiece of the archiving process and hence likely require the lion's share of the effort preceding publication. Before archiving, we recommend completing the following steps that are described in more detail below:

- 1) Check your data for compliance with the General Data Protection Regulation (GDPR)<sup>6</sup> and pseudonymise the data (see Section [4.1](#))
- 2) Data cleaning: double-check data for any errors, plausibility, clarity of labels and consistency with the documentation material (see Section [4.2](#))

Also have a look at our recommendations about documentation material(s), as comprehensive documentation accompanying the data is crucial for the re-usability of data. The more documentation material is made available, the better is the users' understanding of the data.

- 3) Recommended documentation material: What should be included? (see Section [4.3](#))

### 4.1. Check compliance with the GDPR and pseudonymise the data

Since most of the studies archived with AUSSDA contain information about individuals and therefore comprise private and/or sensitive information, pseudonymisation processes before the publication are inevitable. Pseudonymisation protect research subjects against identification, fulfil the requirements of data protection, and meet ethical research standards. By signing the self-deposit contract and by accepting the AUSSDA terms of service, self-depositors agree that the data that will be deposited will comply to the highest data protection standards.

<sup>5</sup> <https://aussda.at/aussda-non-compliance-policy>

<sup>6</sup> <http://data.europa.eu/eli/reg/2016/679/oj>

As hardly any raw data fulfils these criteria, carefully checking all variables and eventually recoding or deleting some of them is absolutely necessary.

**More detailed information on data protection guidelines is available in the [Data Deposit Guideline](#)<sup>7</sup>: We strongly recommend depositors to carefully read the Data Deposit Guideline before every publication as any violation of data protection regulations has consequences.**

The following list contains several examples for potential violations of data protection and countermeasures but is by no means an exhaustive list:

- **Remove all direct identifiers**

Such as social security number, ID number from data collection institute, full name, email address, phone number, postal code, data of birth, workplace/employer, vehicle registration number, bank account number, IP address, student ID number, passport/identity card number etc.

- **Thoroughly check or delete answers to open questions**

Answers to open questions may contain detailed information of respondents that allow for their re-identification. Therefore, all open answers have to be either deleted or carefully checked for any information that may violate the GDPR.

*Example*

**Q:** Do you have any experience in working in politics?

**A:** I have been elected as delegate in the parliament for party XYZ for 12 years in a row.

This information is so detailed that (even without the combination of the constituency) the respondent could be identified easily.

- **Check all standard demographic variables and recode the data to broader categories if variables contain less than 20 observations for certain values:**

*Examples*

**Age:** aggregate to cohorts if less than 20 observations with certain age (e.g. 15-20 years, 21-25 years, etc.)

**Nationality:** categorize into groups with min. 20 observations per nationality (group together to regions if less than 20, e.g. according to UN geoscheme<sup>8</sup>)

- Although all the pseudonymisation steps mentioned above have to be completed in any case, we would like to raise awareness that even **more extensive measures** may be appropriate if the data you gathered contains information on **especially sensitive population** groups (e.g. children, minorities or victims of crimes) or if your population is small and hence the re-identification risk is high (e.g. a small town or a group of public interest like MPs).

<sup>7</sup> [https://www.aussda.at/fileadmin/user\\_upload/p\\_aussda/Documents/AUSSDA-Data-Deposit-Guideline.pdf](https://www.aussda.at/fileadmin/user_upload/p_aussda/Documents/AUSSDA-Data-Deposit-Guideline.pdf)

<sup>8</sup> <https://unstats.un.org/unsd/methodology/m49/>

## 4.2. Data cleaning

In the following paragraphs, we outline some of the processes that proved useful for checking and cleansing data.

### a) Making data easier compatible with other formats

Submitted data should comply to requirements defined in AUSSDA's [Data Deposit Guideline](#)<sup>9</sup> to make it easier for users of other analytical software to use the data.

Technical setting	Reason
System missing  should be coded numerical (e.g. using negative values)  (e.g. -9 no answer; -8 do not know)	In Stata, system missing can be implemented as ".a" or ".b" etc.. After conversion of the dataset to SPSS, these system missings are not displayed correctly anymore.
Variable labels should not be longer than 79 characters	Variable labels are truncated after 79 characters in Stata and remain incomplete. They may be not understandable (e.g. rephrase questions for the labels).

**Table 1 Requirements to increase cross-software usage.**

### b) Recommended steps for cleaning the data and labels:

- Checking whether values are labelled clearly, correctly and consistently.
- Checking the accuracy of missing values.
- Spellcheck of variable names, labels, value labels and string variables (e.g. by exporting all labels to Excel and conducting a spell check).
- Checking whether information is attached to the dataset that you would not like to archive (e.g. notes attached to the dataset or preliminary comments).
- Scanning data for unlabelled values.

**Stata users** can find useful commands in the accompanying \*.do-file "recommended\_datachecks.do" available here:

[https://aussda.at/fileadmin/user\\_upload/p\\_aussda/Documents/recommended\\_datachecks.do](https://aussda.at/fileadmin/user_upload/p_aussda/Documents/recommended_datachecks.do)

### c) Comparing the data and documentation material.

We recommend to double-check if all labels are assigned correctly and if the labels assigned in the dataset correspond to the labels in the documentation material and vice versa. Frequent errors are e.g. a lack of completeness and comprehensibility of labels.

#### Examples

- The variable f1\_4 has the options "don't know" and „not applicable" in the dataset, but the codebook does not contain these options.

<sup>9</sup> <https://aussda.at/aussda-data-deposit-guideline>



- The variable capturing the respondents' age is named "age\_respondent" in the dataset, but "Age" in the codebook. This discrepancy requires either renaming the variable or adapting the codebook.

#### d) Double-checking the plausibility of the data.

We recommend checking the data for outliers and plausibility, hence, whether the observations are plausible.

##### *Example*

- If a respondent is 15 or younger, he/she could not have casted a vote in the last general election.
- If a respondent indicates to be single, the observation for the spouses' occupation should be either missing or not applicable.

## 4.3. Recommended documentation material: What should be included?

Comprehensive documentation accompanying the data is crucial for the re-usability of data. The more documentation material is made available, the better is the users' understanding of the data. The following list provides an overview which files we consider as mandatory or recommended to enhance re-use:

### **I. The following documents are mandatory for publication:**


- Data files
- Documentation material; at least one of the following documents have to be included:
  - Instrument of data collection (questionnaire with interviewer instructions, information material for respondents, data collection guidelines)
  - Codebook
  - Methods report
  - Field report

### **II. The following documents enhance the re-usability of the data:**

- Project report
- Tabulation report
- Data Management Plan (DMP) of project proposal
- Interviewer guidelines
- Interview cards
- Documentation about incentives, contacts...
- Recoding protocol
- Informed consent form template
- Any further documents that help users to understand the data

We recommend including a **suggested citation** and **information about the license** of the document in each document, preferably on the first page. As outlined above, documentation material is published under a CC BY license. The information on this license consists of two aspects: the CC BY logo can be downloaded from the Creative Commons' website (<https://creativecommons.org/about/downloads>) in several formats. The second aspect is a text component containing the title of the document, the information of the author(s), the year, and a link to the Creative Commons' website explaining the terms of the license <https://creativecommons.org/licenses/by/4.0/>.

Below, you find an example how this information can look like:



Questionnaire for the EOSC-Pillar "National Initiatives" Survey by Anita Bodlos, Lisa Hönegger, Lars Kaczmirek, Volker Beckmann, Vincent Breton, Geneviève Romier, Jos van Wezel, Achim Streit, Uros Stevanovic, Fulvio Galeazzi, Federica Tanlongo, Inge Van Nieuwerburgh (2020) is licensed under a [Creative Commons Attribution 4.0 International Licence](#).

## Take aways

### I pseudonymised my data, following the examples and AUSSDA's Data Deposit Guideline.

I cleaned the data as suggested, e.g. regarding variable and value labels and missing values, among others.

I double-checked the data with regard to plausibility and outliers and they are fine.

I compared the data with the documentation material and they match.

I gathered all documentation material for my data.

I included a suggested citation and information about the license of the document in each document, preferably on the first page.

## 5. Self-depositing your data in Dataverse

The following pages contain a detailed step-by-step guide for uploading your data in Dataverse and finally publishing your work.

**Be aware, that after you publish your dataset, this step cannot be reversed.** If you discover a grave error, please contact us immediately. Together, we can devise a plan of action. In the worst case, this may include deaccessioning your dataset so that users cannot access the uploaded files anymore. A reason for deaccessioning your dataset would be if a high risk of identification of data subjects existed. However, the metadata of your study, including the DOI, will always be findable<sup>10</sup>, openly available and harvested by other data catalogues.

AUSSDA Dataverse is the digital archive of AUSSDA - The Austrian Social Science Data Archive. It is based on the open source application Dataverse which has been developed by Harvard University.

**Dataverse** is the name of our archival software but it is also a "folder" for researchers, projects or journals to share and make their data available. As a self-depositor, your data will be stored in the "Self-Deposit Dataverse". This Self-Deposit Dataverse contains all **datasets** curated by accredited self-depositors. A dataset consists of metadata, files (research data, documentations, code, etc.) and terms.

<sup>10</sup> If you want to know more about why metadata are still available, see more in CESSDA's PID Policy (<https://doi.org/10.5281/zenodo.3611327>).

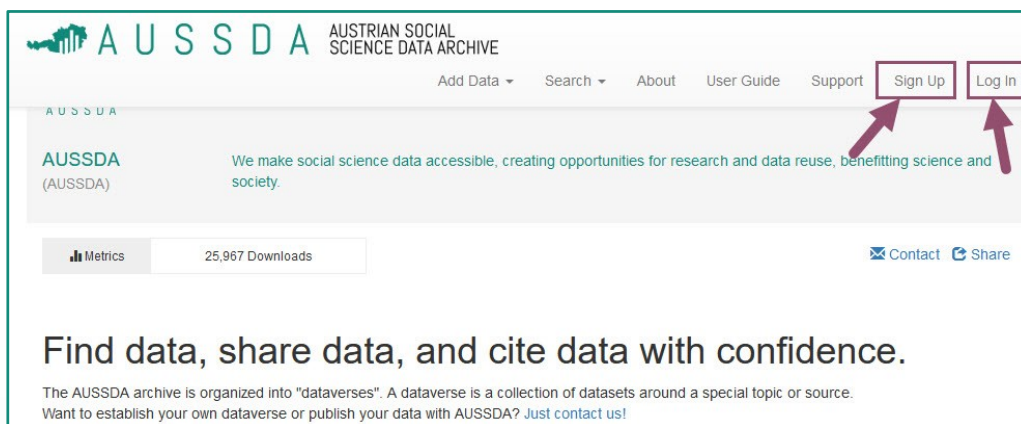
Some settings in Dataverse are set by AUSSDA and cannot be changed: Metadata fields that are marked as “required” need to be filled. Also, the settings in the Self-Deposit Dataverse allow all logged-in users to download account-based scientific use (SUF) files.

Other self-depositors in the Self-Deposit Dataverse see your drafts, and can even edit them, but they cannot publish them. So please show the courtesy not to edit drafts/entries by other self-depositors (any changes are recorded in log files).

## 5.1. Login

➡ Go to “[data.aussda.at](https://data.aussda.at)”.

➡ “Sign up” or “Log in”.



**Figure 1 Dataverse Log in and Sign up.**

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” (Figure 1), you can choose a username and password and create a local account in our Dataverse. The next time you visit our 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 (you have to look for the English name of your institution) or by typing the English name in. You are then forwarded to the login page by your institution, where you log in using your institutional username and password. If you encounter any problems, please contact us.

If you cannot find your institution in the list, please contact us and choose “Sign up” in the meantime.

If at first you used the “Sign up” option and then you want to use institutional login with the same e-mail address, the system will ask you if you want to merge the two accounts. This cannot be undone. If you merge, we cannot reset your password anymore, as it is your institutional password.

**If this is your first self-deposit: send the AUSSDA-Team a quick message including your username so they can add you to the Self-Deposit Dataverse.**

## 5.2. Create a dataset entry

---

As mentioned above, the **Self-Deposit Dataverse** contains all **datasets** curated by accredited self-depositors. A dataset consists of:

- 1) **Metadata** (Section [5.2.1](#))
- 2) **Files** (Section [5.2.2](#))
- 3) **Terms** (Section [5.2.3](#))

Below we will guide you through entering your metadata, data and documentation in the Self-Deposit Dataverse, starting with Metadata.

### 5.2.1. Metadata

---

Metadata describe available data resources of an archive to facilitate searching for and cataloguing of data. Metadata offer a systematic overview of the data resources and contain e.g. information on the author(s) or keywords on the data. Open and unrestricted access to metadata is essential for effective data use and re-use. To this end, metadata in the AUSSDA Dataverse are always published under the public domain dedication (CC0 1.0 Universal) and may thus be freely and openly accessed and used by the public.

You prepare the metadata that make the data **Findable, Accessible, Interoperable and Re-usable** (FAIR). By following the FAIR principles, you make research data attractive to other researchers. The description of the deposited data follows international standards set by the Data Documentation Initiative (DDI)<sup>11</sup> and the Consortium of European Social Science Data Archives (CESSDA ERIC)<sup>12</sup>. For several metadata entries, we use established "controlled vocabularies" of standardised items. This standardisation allows for comparability of the metadata in the entire AUSSDA data holdings and also for comparability within the international CESSDA data catalogue. Therefore, using the controlled vocabularies is mandatory for all publications in Dataverse. Please do not use thousands separators, as they may cause problems when metadata are harvested by other data catalogues. For example, instead of writing "12,300 participants answered the survey", do not use the separator and add the text "12300 participants answered the survey".

Please be aware that the working language in our Dataverse is English.

**We recommend that you read the entire section on metadata and prepare all metadata accompanying your dataset (see Table 2) before you actually create a dataset entry and start entering the metadata in Dataverse.**

Steps in Dataverse: Add Metadata

➡ Go to the "[Self-Deposit Dataverse](#)".<sup>13</sup>

➡ On the right, middle side of the page, click „+ Add Data" > "New Dataset" (see Figure [2](#)).

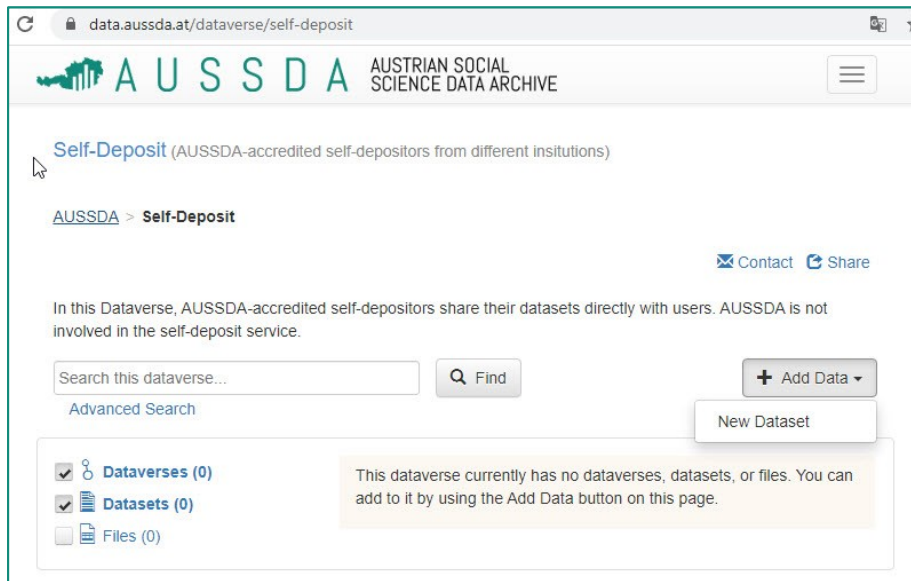
If you cannot see this button, please contact us.

---

<sup>11</sup> <https://ddialliance.org/>

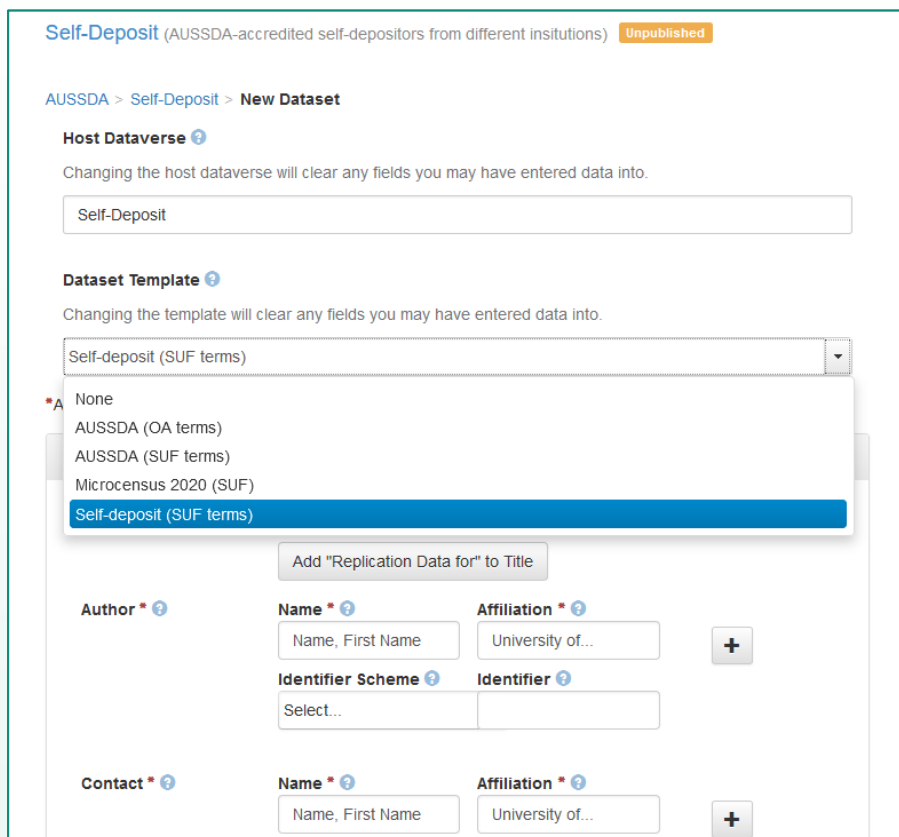
<sup>12</sup> <https://www.cessda.eu/>

<sup>13</sup> <https://data.aussda.at/dataverse/self-deposit>



**Figure 2 Adding a „New Dataset“ to the “Self-Deposit Dataverse”.**

After clicking on “New Dataset”, a new page appears on which, you can enter all required metadata fields using the “Self-deposit: Scientific Use” template (see Figure 3).



**Figure 3 Using the Self-Deposit template, enter all your metadata.**

➡ “Host Dataverse”: no changes needed, Self-Deposit is correct.

➡ Under “Dataset Template”, choose the “Self-deposit: Scientific Use” template.

Here are some tips:

- Some fields are prefilled in the template. In Table 2, you can find more information where you need to replace information and where no changes are needed.
- **Required** fields that you need to fill are marked with a red asterisk \*.
- If we ask you for information that you do not have or is not applicable to you, fill the field with “n/a”.
- If you are not sure which category of a controlled vocabulary applies, choose the higher level one.

Metadata table with examples

Table 2 gives an overview over the different metadata fields. The table also presents some examples and contains a column that serves as a checklist. We provide links to the related vocabularies in the text and the footnotes.

You can find an example of a dataset in the AUSSDA Dataverse here:

<https://data.aussda.at/dataset.xhtml?persistentId=doi:10.11587/QDETRI>

Some of the links in the table below are outgoing links to controlled vocabularies that we use to follow CESSDA ERIC standards. Here are some tips on how to navigate these pages:

### ELSST – The European Language Social Science Thesaurus

The European Social Science Thesaurus (<https://thesauri.cessda.eu/elsst/en/>) is a very extensive resource. In the upper, right corner you can type in a word you would like to use as a keyword or you can navigate on the left side by either scrolling alphabetically or in the hierarchical view.

Important for you:

- Please choose from *preferred terms, broader or narrower or related concepts* (clickable terms). Other terms, like *entry terms*, should not be used to index your dataset. In the **example** in Figure 4 the search word was “political science”. As this is an entry term, the user has been forwarded to the *preferred* term “Politics”.
- When entering the “Term” in Dataverse, do not use all capital letters. See examples in Table 2.
- If you cannot find a keyword in ELSST, add the term to your abstract text to make it findable in Dataverse.

Choose a **maximum of 14 keywords**. If you need more, try to choose a broader concept.

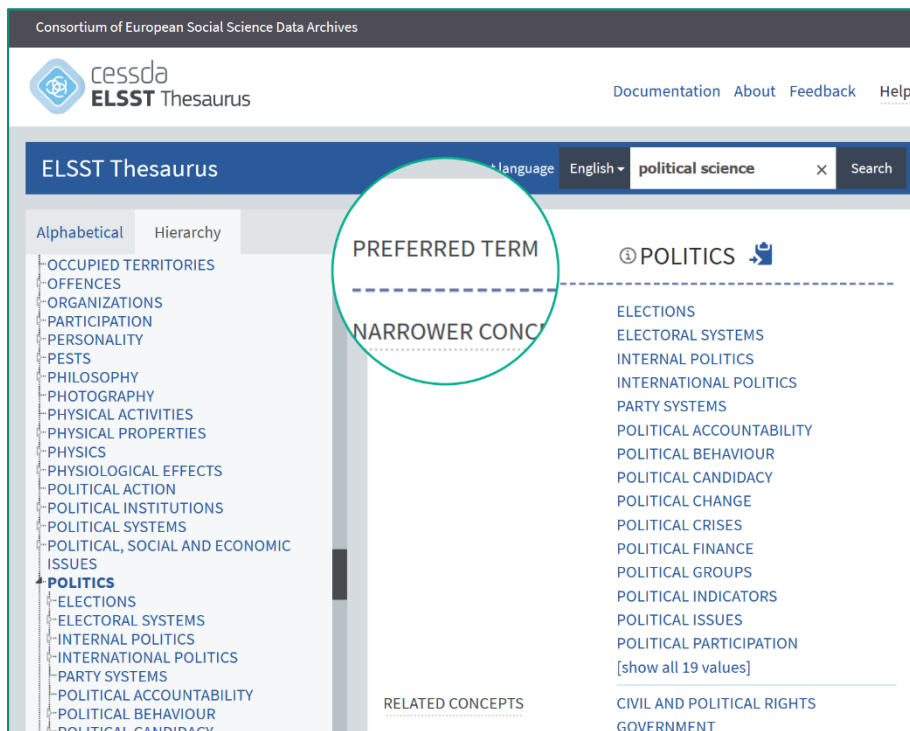


Figure 4 Controlled vocabulary for Keywords: ELSSST.

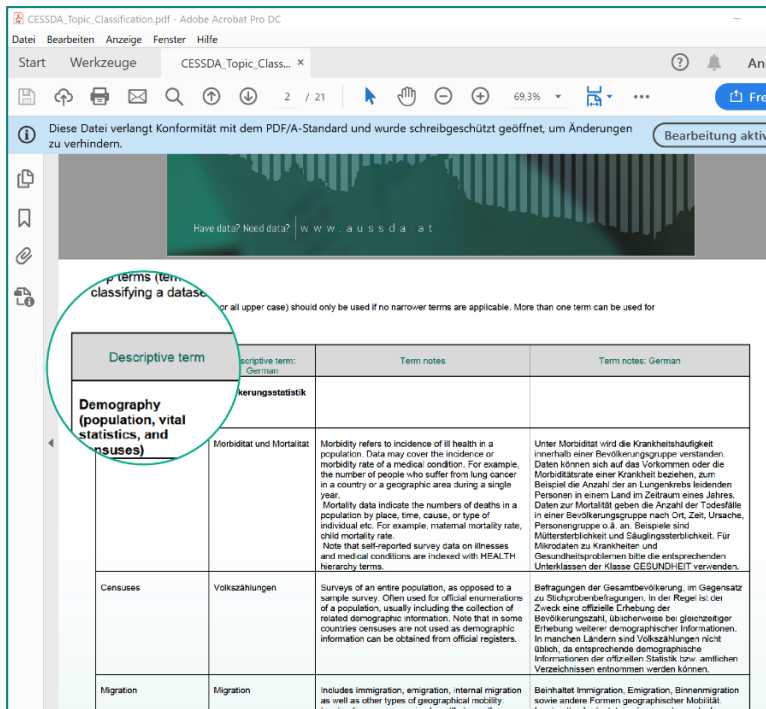
### CESSDA Topic Classification

The CESSDA Topic Classification (<http://aussda.at/cessda-topic-classification>) is not as extensive as ELSSST. In the column "Descriptive term" you see broader topics in bold font and sub-topics in normal font.

Important for you:

- Please choose from the column "Descriptive term".
- In the column "Term notes" you can see how each term is described.
- Searching for a topic: Use the search function of your PDF reader to find matching topics (e.g. by starting the search with Strg./Ctrl.+F) or scroll through the document. You may need to download the document first in order to be able to use the search function.

Choose a **maximum of 7 topics**. If you need more, choose a broader topic.



Descriptive term	Descriptive term: German	Term notes	Term notes: German
Demography (population, vital statistics, and censuses)	Bevölkerungsstatistik	Morbidity refers to incidence of ill health in a population. Data may cover the incidence or morbidity rate of a medical condition. For example, the number of people who suffer from lung cancer in a country or a geographic area during a single year. Mortality data indicate the numbers of deaths in a population by place, time, cause, or type of individual etc. For example, maternal mortality rate, child mortality rate. Note that self-reported survey data on illnesses and medical conditions are indexed with HEALTH hierarchy terms.	Unter Morbidität wird die Krankheitshäufigkeit innerhalb einer Bevölkerungsgruppe verstanden. Daten können sich auf das Vorkommen oder die Morbiditätsrate einer Krankheit beziehen, zum Beispiel die Anzahl der an Lungenkrebs leidenden Personen in einem Land im Zeitraum eines Jahres. Daten zur Mortalität geben die Anzahl der Todesfälle in einer Bevölkerungsgruppe nach Ort, Zeit, Ursache, Personengruppe o.ä. an. Beispiele sind Müttersterblichkeit und Säuglingssterblichkeit. Für Mikrodaten zu Krankheiten und Gesundheitsproblemen bitte die entsprechenden Unterklassen der Klasse GESUNDEHEIT verwenden.
Censuses	Volkzählungen	Surveys of an entire population, as opposed to a sample survey. Often used for official enumerations of a population, usually including the collection of related demographic information. Note that in some countries censuses are not used as demographic information can be obtained from official registers.	Befragungen der Gesamtbevölkerung, im Gegensatz zu Stichprobenbefragungen. In der Regel ist der Zweck eine offizielle Erhebung der Bevölkerungszahl, üblicherweise bei gleichzeitiger Erhebung weiterer demographischer Informationen. In manchen Ländern sind Volkszählungen nicht üblich, da entsprechende demographische Informationen der offiziellen Statistik bzw. amtlichen Verzeichnissen entnommen werden können.
Migration	Migration	includes immigration, emigration, internal migration as well as other types of geographical mobility	Beinhaltet Immigration, Emigration, Binnenmigration sowie andere Formen geographischer Mobilität.

Figure 5 Controlled vocabulary for Topics: CESSDA Topic Classification.

## DDI – Data Documentation Initiative

For some metadata fields, we use terms from the DDI controlled vocabularies. Each of these terms has a website including a table with three columns. The terms that you can copy and paste to your Dataverse entry can be found in the column in the middle with the name “**Descriptive Term of the Code**”. In the column on the right, you can see a description of each term.

Value of Code	Descriptive Term of the Code	Definition of the Code
Numeric	Numeric	Data consisting largely of values expressed as digits from 0 to 9 and, optionally, signs for negative values, decimal points, or letters only when intended to represent numbers (for example, A-F or a-f in hexadecimal).
Text	Text	Data consisting largely of text, including letters, numbers, and special characters or symbols used in writing for punctuation, abbreviation, etc. For example, interview transcriptions, narratives or essays written by study participants, newspaper articles, etc.
Still image	Still image	Static images, such as graphs, drawings, photographs, diagnostic/medical images like X-rays, etc.

Figure 6 Example of a DDI controlled vocabulary: Kind of Data.



Metadata item	Description	Example for survey data	Example for content analysis data	Dataverse Tips & Tricks	Yeah, I got this!
Title * <sup>14</sup>	<p>Full title by which the dataset is known + "(SUF edition)".</p> <p>Replace "Your title" with the name of your dataset. The second part of the title, "(SUF edition)" enhances the visibility of the access mode.</p> <p>Below the tile, there is a "Add: 'Replication data for' to Title" button, this is for AUSSDA's replication service. If you want to deposit the replication data for a journal article, please contact us.</p>	Your title (SUF edition)	Your title (SUF edition)	It is possible to also enter "Alternative title" (e.g. German translation), see Section " <a href="#">Steps in Dataverse: Add additional Metadata</a> "	☒
Author *	<p>Dataverse will use the same order of the authors in the metadata for creating the recommended citation. Author's family name, given name or the name of the organization responsible for this dataset as well as the organization with which the person/author is affiliated are required fields. Identifiers are optional, for example, ORCID is a persistent digital identifier (PID) that distinguishes researchers from each other and ensures that research is recognised.</p>	Random, Jane, University of Vienna, ORCID <sup>15</sup> 123456)	Random, Jane, University of Vienna, ORCID <sup>16</sup> 123456)	<p>You can add more authors with the "+" button.</p> <p>If you list an organisation in the field "author name", this could later lead to an error after publishing the dataset. Just contact us if you do not have natural persons as authors.</p>	☒

<sup>14</sup> Asterisks indicate required fields \*

<sup>15</sup> <https://orcid.org/>

<sup>16</sup> <https://orcid.org/>

Metadata item	Description	Example for survey data	Example for content analysis data	Dataverse Tips & Tricks	Yeah, I got this!
Contact *	<p>The contacts' surname, first name and organization or the name of the organization who should be contacted for questions about the dataset.</p> <p>The button "contact" at the top of the dataset page allows users to reach out to you with questions (see Figure 7).</p>	Random, Jane, University of Vienna, E-Mail	Random, Jane, University of Vienna, E-Mail	Dataverse will automatically use your account-details as contact information. If you would like to change them, you can manually edit the fields name, affiliation and email address. User can only see the "contact" button, but not the underlying contact details.	☒
Description *	An abstract describing the purpose, nature, and scope of the dataset. The text is required here, the date field is optional.			Please note that HTML elements are NOT allowed here. You can add more descriptions with the "+" button.	☒
Subject *	The scientific discipline of the dataset, most likely Social Sciences.	Social Sciences	Social Sciences	Please choose "Social Sciences".	☒
Keywords * <sup>17</sup>	Key terms that describe important aspects of the dataset as available in the controlled vocabulary <a href="https://thesauri.cessda.eu/elsst/en/">ELSST (European Language Social Science Thesaurus)</a> . You can use up to 14 keywords from ELSST to describe your dataset. If you need more keywords, try to choose broader concepts.	Trust in government Elections Politics	Electoral issues	<p>Enter the correct "Term", do not use all capital letters, see examples to the left.</p> <p>You can add more keywords with the "+" button. If you add more keywords, do not forget to copy the Vocabulary "ELSST - The European Language Social Science Thesaurus" and URL <a href="https://elsst.cessda.eu/">https://elsst.cessda.eu/</a>.</p>	☒

<sup>17</sup> <https://thesauri.cessda.eu/elsst/en/>

Metadata item	Description	Example for survey data	Example for content analysis data	Dataverse Tips & Tricks	Yeah, I got this!
Topic Classification * <sup>18</sup>	The classification field indicates the broad important topic(s) and subjects that the data cover. You can use up to 7 classifiers from the CESSDA topic classification (controlled vocabulary). If you need more, try to choose broader terms. Here is the list of topics: <a href="http://aussda.at/cessda-topic-classification">http://aussda.at/cessda-topic-classification</a>	Public health Censuses Education	Media	Enter the correct "Term", do not use all capital letters, see examples to the left.  You can add more topics with the "+" button. If you add more topics, do not forget to copy the Vocabulary "CESSDA Topic Classification" and URL <a href="https://vocabularies.cessda.eu/vocabulary/TopicClassification">https://vocabularies.cessda.eu/vocabulary/TopicClassification</a>	<input type="checkbox"/>
Related Publication	The information of related publications consists of multiple fields (citation, unique identifiers and URLs) and is optional.	The data are related to the paper "Random, Jane. (YYYY). Impact of X on Electoral Behavior". doi: 10.01234/ABC4E. ( <a href="https://doi.org/10.01234/ABC4E">https://doi.org/10.01234/ABC4E</a> )	The data are related to the paper "Random, Jane. (YYYY). Why content analysis is great. doi: 10.01234/ABC4E. ( <a href="https://doi.org/10.01234/ABC4E">https://doi.org/10.01234/ABC4E</a> ).	"Citation": Enter the full bibliographic citation here. "ID type": Type of unique id (e.g. doi, ISBN) "ID number": number of the unique identifier URL: Link to the paper (preferably using the link of the unique identifier (e.g. <a href="https://doi.org/10.01234/ABC4E">https://doi.org/10.01234/ABC4E</a> )  You can add more publications with the "+" button.	<input type="checkbox"/>
Notes	Please do not change the prefilled information.	This dataset has been curated and deposited by one of our accredited self-depositors.	This dataset has been curated and deposited by one of our accredited self-depositors.		<input type="checkbox"/>
Language *	Language(s) of the dataset	German, English	German	You can choose multiple languages.	<input type="checkbox"/>

<sup>18</sup> <http://aussda.at/cessda-topic-classification>

Metadata item	Description	Example for survey data	Example for content analysis data	Dataverse Tips & Tricks	Yeah, I got this!
Grant information: Grant agency * and grant number *	The agency that provided funding and the grant number of the project.	Federal Ministry of Science, Research and Economy: HRSM - XYZ4	H2020: FA123456	You need the Grant Agency and the Grant Number. If this field does not apply to you, please enter "n/a". You can add more grant information with the "+" button.	<input type="checkbox"/>
Distributor *	Please do not change the prefilled information.	The Austrian Social Science Data Archive	The Austrian Social Science Data Archive		<input type="checkbox"/>
Depositor *	The person(s) (family name, given name) or the name of the organisation(s) who own(s) the data.	Surname, first name	University of X: Department of Communication	There is only one field, if you have more than one depositor, you can separate them with a semicolon (";") between the names.	<input type="checkbox"/>
Deposit Date	Prefilled with date of draft curation. Keep or delete as Dataverse generates an additional publication date.				<input type="checkbox"/>
Date of collection: Start *	Date when the data collection started in the format (YYYY-MM-DD).	2019-10-22	2019-10-22	If you have several waves for a dataset, we recommend that you add the very first start date and the last end date of your data collection period. You can add more information about the different dates later as additional metadata information.	<input type="checkbox"/>
Date of collection: End *	Date when the data collection ended in the format (YYYY-MM-DD).	2020-03-30	2020-03-30	See above.	<input type="checkbox"/>

Metadata item	Description	Example for survey data	Example for content analysis data	Dataverse Tips & Tricks	Yeah, I got this!
Kind of Data * <sup>19</sup>	Choose from list: Numeric, Text, Still image, Geospatial, Audio, Video, Software, Interactive resource, 3D, Other.  See descriptions here: <a href="#">Controlled vocabulary Kind of Data (DDI)</a>	Numeric	Text; Other	You can add more fields with the "+" button.  For all DDI fields: <b>Choose a category that is predominant</b> , for <b>example</b> , if you have one string variable and five numerical variables – choose "Numeric" or if the string is very important choose both "Numeric" and "Text".	☒
Data Sources * <sup>20</sup>	A typology of data source that has been used to generate the data. The total list can be found here: <a href="#">Controlled vocabulary Data Sources (DDI)</a> . Please note that "Research data" should only be chosen if your data relies on pre-existing research data (e.g. you merge or combine research data). „Population group" is most likely the right choice if you survey individuals.  <b>Examples:</b> Population group, Communication: Public, Research data, Registers/records/accounts, Other, ...	Population group	Communication: Public	You can add more fields with the "+" button.	☒
Geographic Coverage *	Information on the geographic coverage of the data. You have to enter information in the field "Country/Nation". The fields "State/Province", "City" and "Other" are optional.	Austria	Vienna, Austria	You can add more countries with the "+" button.	☒

<sup>19</sup> [https://ddialliance.org/Specification/DDI-CV/KindOfDataFormat\\_1.0.html](https://ddialliance.org/Specification/DDI-CV/KindOfDataFormat_1.0.html)

<sup>20</sup> [http://www.ddialliance.org/Specification/DDI-CV/DataSourceType\\_1.0.html](http://www.ddialliance.org/Specification/DDI-CV/DataSourceType_1.0.html)

Metadata item	Description	Example for survey data	Example for content analysis data	Dataverse Tips & Tricks	Yeah, I got this!
Unit of Analysis * <sup>21</sup>	<p>Basic unit of analysis or observation that this dataset describes.</p> <p>Choose from list: Individual, Organization, Family, Family: Household family, Household, Housing Unit; Event/Process, Geographic Unit, Time Unit, Text Unit, Group, Object, Other</p> <p>See descriptions here: <a href="#">Controlled vocabulary Unit of Analysis (DDI)</a></p>	Individual	Text Unit	You can add more fields with the "+" button.	☒
Universe *	<p>Description of the population covered by the data in the file; the group of people or other elements that are the object of the study and to which the study results refer. Age, nationality, and residence commonly help to delineate a given universe, but any number of other factors may be used, such as age limits, sex, marital status, race, ethnic group, nationality, income, veteran status, criminal convictions, and more. The universe may consist of elements other than persons, such as housing units, court cases, deaths, countries, and so on. This term is also known as the universe of interest, population of interest, and target population.</p>	Austrian resident population from age 14	Any news articles in the respective media outlets.	You can add more fields with the "+" button.	☒

<sup>21</sup> [http://www.ddialliance.org/Specification/DDI-CV/AnalysisUnit\\_1.0.html](http://www.ddialliance.org/Specification/DDI-CV/AnalysisUnit_1.0.html)

Metadata item	Description	Example for survey data	Example for content analysis data	Dataverse Tips & Tricks	Yeah, I got this!
Time Method * <sup>22</sup>	The time method or time dimension of the data collection, the total list can be found here: <a href="#">Controlled vocabulary Time Method (DDI)</a>  <b>Examples:</b> Longitudinal, Longitudinal: Trend/Repeated cross-section, Longitudinal: Panel, Time series, Cross-section, Other	Longitudinal: Panel	Other	Single entry field but you can add more time methods by separating your entries with a semicolon (",").	☒
Sampling Procedure * <sup>23</sup>	Type of sample and sample design used to select the survey respondents to represent the population, the total list can be found here: <a href="#">Controlled vocabulary Sampling Procedure (DDI)</a>  <b>Examples:</b> Total universe/Complete enumeration, Probability, Probability: Stratified, Non-probability, Non-probability: Quota, Mixed probability and non-probability, Other	Non-probability: Quota	Probability: Stratified	Single entry field but you can add more sampling procedures by separating your entries with a semicolon (",").	☒
Collection Mode * <sup>24</sup>	Method used to collect the data, the total list can be found here: <a href="#">Controlled vocabulary Collection Mode (DDI)</a>  <b>Examples:</b> Telephone interview: CATI, Self-administered questionnaire: Web-based, Focus group, Content coding, Transcription, Compilation/Synthesis, Other	Self-administered questionnaire: Web-based	Content coding	Single entry field but you can add more methods of data collection by separating your entries with a semicolon (",").	☒

<sup>22</sup> [http://www.ddialliance.org/Specification/DDI-CV/TimeMethod\\_1.2.html](http://www.ddialliance.org/Specification/DDI-CV/TimeMethod_1.2.html)

<sup>23</sup> [http://www.ddialliance.org/Specification/DDI-CV/SamplingProcedure\\_1.1.html](http://www.ddialliance.org/Specification/DDI-CV/SamplingProcedure_1.1.html)

<sup>24</sup> [http://www.ddialliance.org/Specification/DDI-CV/ModeOfCollection\\_3.0.html](http://www.ddialliance.org/Specification/DDI-CV/ModeOfCollection_3.0.html)

Metadata item	Description	Example for survey data	Example for content analysis data	Dataverse Tips & Tricks	Yeah, I got this!
Type of Research Instrument * <sup>25</sup>	Type of data collection instrument used, the total list can be found here: <a href="#">Controlled vocabulary Type of Research Instrument (DDI)</a>  <b>Examples:</b> Questionnaire, Structured questionnaire, Interview scheme and/or themes, Data collection guidelines, Programming script, Other	Structured questionnaire	Data collection guidelines	Single entry field but you can add more used time methods by separating your entries with a semicolon (",;").	☒

**Table 2 Metadata fields, examples and links.**

**Do not yet upload your data and documentation files as they still need some changes outlined below in Section 2)5.2.2.**

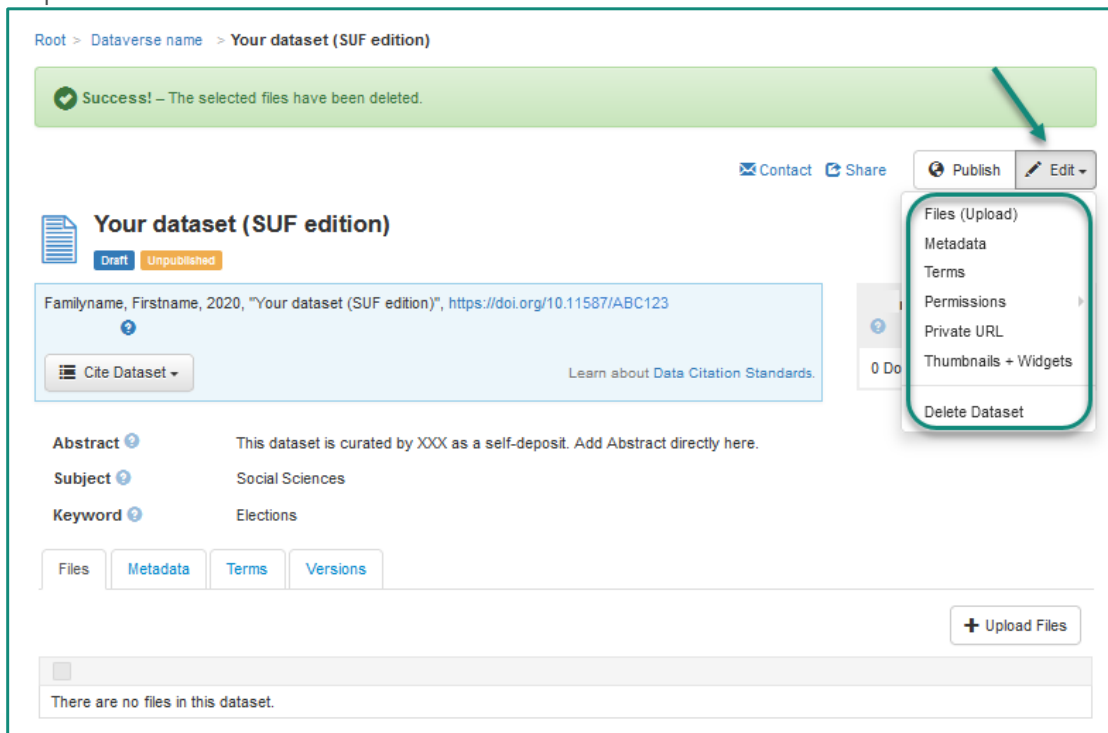
 Click „Save” as a next step.

You can revise and edit your metadata also at a later point in time.

<sup>25</sup> [https://www.ddialliance.org/Specification/DDI-CV/TypeOfInstrument\\_1.1.html](https://www.ddialliance.org/Specification/DDI-CV/TypeOfInstrument_1.1.html)



## Steps in Dataverse: First Draft saved



**Figure 7 Edit your dataset in Dataverse.**

Here (Figure 7), you can see, how your dataset draft looks like after saving the required metadata. The "Edit" button in the right corner is very important. There, you can

- 1) Upload and restrict files: "Files (Upload)" (see Section 5.2.2).
- 2) Add additional information: "Metadata" (see next topic below).
- 3) See the Terms of Use and Terms of Access: "Terms" (see Section 5.2.3).
- 4) Delete your dataset draft (learn more about this and Publishing in Section 5.3).

## Steps in Dataverse: Add additional Metadata

➡ You can add additional metadata by clicking on "Edit" > "Metadata".

There you can see additional fields, that are not mandatory for a deposit.

We recommend also checking for and eventually correcting trailing blanks (a space after the last digit) and spelling errors.

Some fields you could add:

- Alternative title: If you want to increase the search hits for your dataset, use this field to enter the German name of your dataset here.
- Contributor(s): You can select different types of contributor(s) to the data set. It is possible to mention data curator(s), data collector(s), researcher(s) here. You can add more contributors with the "+" button.
- And many more optional fields...

➡ Either "Save changes" or "Cancel".

## Take aways

I gathered all relevant metadata and filled my dataset draft in Dataverse with the relevant, correct information.

### 5.2.2. Files in Dataverse: Add your documentation and data

In this section, we outline in detail the necessary steps for preparing the data and documentation material to upload them in Dataverse.<sup>26</sup> This includes:

- File formats
- Saving DOI and version as variables in the dataset
- File naming
- Uploading files in Dataverse

#### File formats

AUSSDA encourages the self-depositors to use the recommended file formats to enhance re-use of data. Using the recommended formats also ensures that your files are secured long-time as they are fully compatible with our long-time archival and preservation strategy.

**During the upload, Dataverse automatically generates a tab-delimited version of many data files, also 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. When clicking the “Download” button, the original file format is still the first download option for users, open formats are provided for usage in other software. By conversion to \*.tab-format, long-term re-use is secured. For more details on this topic and how Dataverse treats other formats, please see the Dataverse [user guide](#)<sup>27</sup>.

#### Documentation

We recommend saving the documentation material as PDF/A, a secure long-term format. You can find step-by-step instructions [here](#)<sup>28</sup> (German), [here](#)<sup>29</sup> (English) or [here](#)<sup>30</sup> (English). These are different solutions since the conversion steps depend on the software you use.

#### Data

The most frequently used formats in the AUSSDA Dataverse are SPSS (\*.sav) and Stata (\*.dta). We recommend using the unicode format for all files and to avoid diacritics if possible.

#### *For Stata users*

We recommend saving Stata files in version 14 as unicode is the default file format since Stata 14 and users may not always have the most recent version of Stata at their disposal.

<sup>26</sup> See also the [Data Deposit Guideline](#).

<sup>27</sup> <http://guides.dataverse.org/en/latest/user/tabulardataingest/index.html>

<sup>28</sup> <https://zid.univie.ac.at/computer-rooms/anleitungen/pdfa-erstellen/>

<sup>29</sup> [https://library.princeton.edu/special-collections/sites/default/files/Creating\\_PDFa.pdf](https://library.princeton.edu/special-collections/sites/default/files/Creating_PDFa.pdf)

<sup>30</sup> <https://blogs.adobe.com/acrolaw/2011/05/using-save-as-to-to-conform-to-pdf/>

### For SPSS users

We recommend running SPSS in unicode mode (default mode since SPSS version 21), data files will then be stored in unicode encryption.<sup>31</sup>

### For Excel users

Do not use Excel as a data format, we recommend saving datasets as \*.csv (UTF-8 encoded):

For EXCEL 2016:

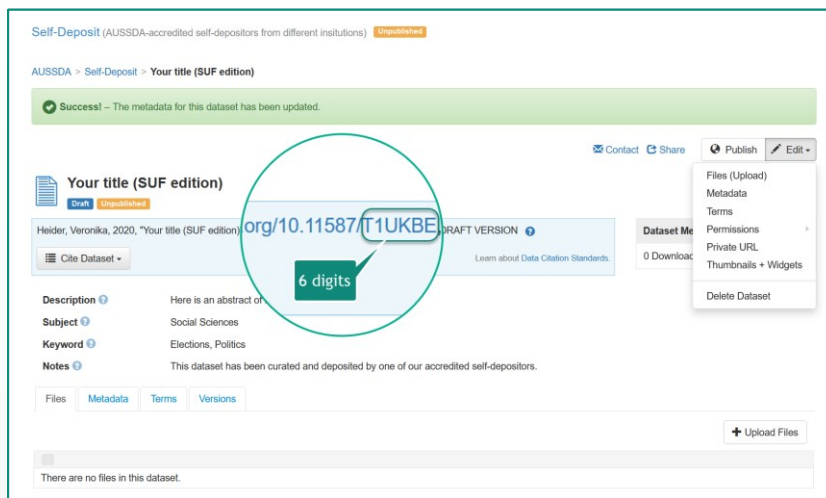
Go to "File" -> "Save As" -> select "CSV UTF-8 (comma delimited) (\*.csv)"

If it is really necessary to publish Excel files, please double zip the file before uploading. Dataverse can ingest Excel files, but e.g. the usage of more than one Excel-sheet will produce ingest errors.

For the best reuse of your data, we recommend storing your data also in other file formats that are used by the community. If you want to learn more, do not hesitate to contact us.

Save the DOI and the version as variables in the dataset

The DOI is part of the file naming (see section below). In addition, the DOI of the dataset and the version should also be stored as variables in the dataset. Thereby, users can see which version they use in their analyses at a glance and can cite the data easily.



**Figure 8** You can find the DOI of your Dataverse entry in the blue citation field and in the tab "Metadata". Please note the last 6 digits as suffix of the DOI.

### Variable version

The recommended format of the variable "version" is:

- variable name: version
- content: "vmajor.minor (YYYY-MM-DD)", *example* "v1.0 (2020-08-26)"
- variable label: "AUSSDA archive version"

<sup>31</sup> [https://www.ibm.com/support/knowledgecenter/en/SSLVMB\\_26.0.0/statistics\\_mainhelp\\_ddita/spss/base/faq\\_unicode.html](https://www.ibm.com/support/knowledgecenter/en/SSLVMB_26.0.0/statistics_mainhelp_ddita/spss/base/faq_unicode.html)

### Variable *doi*

The recommended format of the variable "doi" is:

- variable name: doi
- content: "doi:10.11587/ABC123"
- variable label: "digital object identifier"

As the DOI is generated after creating and saving the Dataverse entry, these steps can only be undertaken shortly before the upload of the dataset. We recommend ordering the version and the DOI as the first variables of the dataset. The accompanying do-file contains exemplary commands for these steps.

### File Naming

AUSSDA encourages self-depositors to use the recommended file naming scheme which enables users to quickly understand the content of the files and also makes the file names consistent over different data deposits. Also, Dataverse does not accept files with identical names which makes the usage of our recommended file naming scheme even more important.

We recommend naming the files using the following pattern:

`DOIsuffix_description_language_version.fileExtension (e.g. ABC123_mr_en_v1_0.pdf)`

### **DOI**suffix

A DOI consists of the prefix (in the case of AUSSDA, the prefix is *10.11587*), followed by a slash ("/") and a 6-digit suffix. Hence, the suffix is the combination of letters and numbers in the DOI after the "/" (e.g. in the DOI 10.11587/ABC123, "ABC123" is the suffix). As these 6 digits of the suffix are unique to your dataset, use them to name all your files. You can find the DOI that has been reserved for you in several places in Dataverse (example: as part of the citation box at the top in Figure 7).

### **description**

AUSSDA usually uses two letters for describing the content of a file. The conventionally used abbreviations of the content of files are shown in Table 3.

If you upload more than one file per category (e.g. two codebooks), we recommend adding numbers after the abbreviation (e.g. *ABC123\_co01\_en\_v1\_0.pdf*).

### **language**

AUSSDA conventionally uses two letters for abbreviating the language in filenames for documentation material or datasets. The most frequently used abbreviations are "de" for German and "en" for English. Follow the [ISO 693-1 code](https://de.wikipedia.org/wiki/Liste_der_ISO-639-1-Codes)<sup>32</sup> if you need abbreviations for other languages.

---

<sup>32</sup> [https://de.wikipedia.org/wiki/Liste\\_der\\_ISO-639-1-Codes](https://de.wikipedia.org/wiki/Liste_der_ISO-639-1-Codes)

## version

The version is usually preceded by the letter "v". We recommend using major and minor versions that are separated by underscores ("\_") rather than a full stop ("."). E.g. "version 1.1" is given in the style "v1\_1".

We recommend avoiding any blanks, diacritics or uppercase letters in filenames.

Abbreviation	Description English	Description German
da	data	Daten
qu	Questionnaire	Fragebogen
im	Interviewer manual	Interviewerhandbuch
fr	Field report	Feldbericht
co	Codebook	Codierliste
rr	Research report	Forschungsbericht
mr	Method report	Methodenbericht
sx	Code/ syntax	Code/ Syntax
ta	Tabulation report	Tabellenband
om	Other material, specify in Dataverse description	anderes Material
vi	Variable identifiers and descriptions	Variablenidentifikation und - beschreibung

**Table 3 Descriptions and abbreviations used for the file names of documentation material**

## Documentation

**Example 1** After creating the Dataverse entry, the DOI "10.11587/ABC123" is created. The recommended file name for version 1.0 of the questionnaire in English is *ABC123\_qu\_en\_v1\_0.pdf*.

**Example 2** After creating and saving the Dataverse entry, the DOI "10.11587/DEF456" is created. You have two German and two English method reports. The recommended file names for the English version 1.0 of your method report are *DEF456\_mr01\_en\_v1\_0.pdf* and *DEF456\_mr02\_en\_v1\_0.pdf*. The recommended file names for the German method reports are *DEF456\_mr01\_de\_v1\_0.pdf* and *DEF456\_mr02\_de\_v1\_0.pdf*.

## Data

Data files are conventionally abbreviated as "da". For an explanation of all other elements, please see the section on file naming of the documentation material.

DOI suffix *da\_language\_version.fileExtension* (e.g. *DEF456\_da\_en\_v1\_0.dta*).

**Example 1** Imagine, you have one German and one English Stata dataset that are identical besides the language. After creating and saving the Dataverse entry, the DOI "10.11587/GHI789" is created. The recommended file name for the English version 1.0 of your data is *GHI789\_da\_en\_v1\_0.dta* and for the German data *GHI789\_da\_de\_v1\_0.dta*.

**Example 2** After creating and saving the Dataverse entry, the DOI "10.11587/JKL123" is created. You have two English SPSS datasets. The recommended file names for the English version 1.0 of your datasets are *JKL123\_da01\_en\_v1\_0.sav* and *JKL1236\_da02\_en\_v1\_0.sav*.

## Take aways

I saved my documentation material as PDF/A to ensure long-time archive format.

I added two variables to my dataset (“version” and “doi”).

I named my data files according to AUSSDA's suggested file naming scheme.

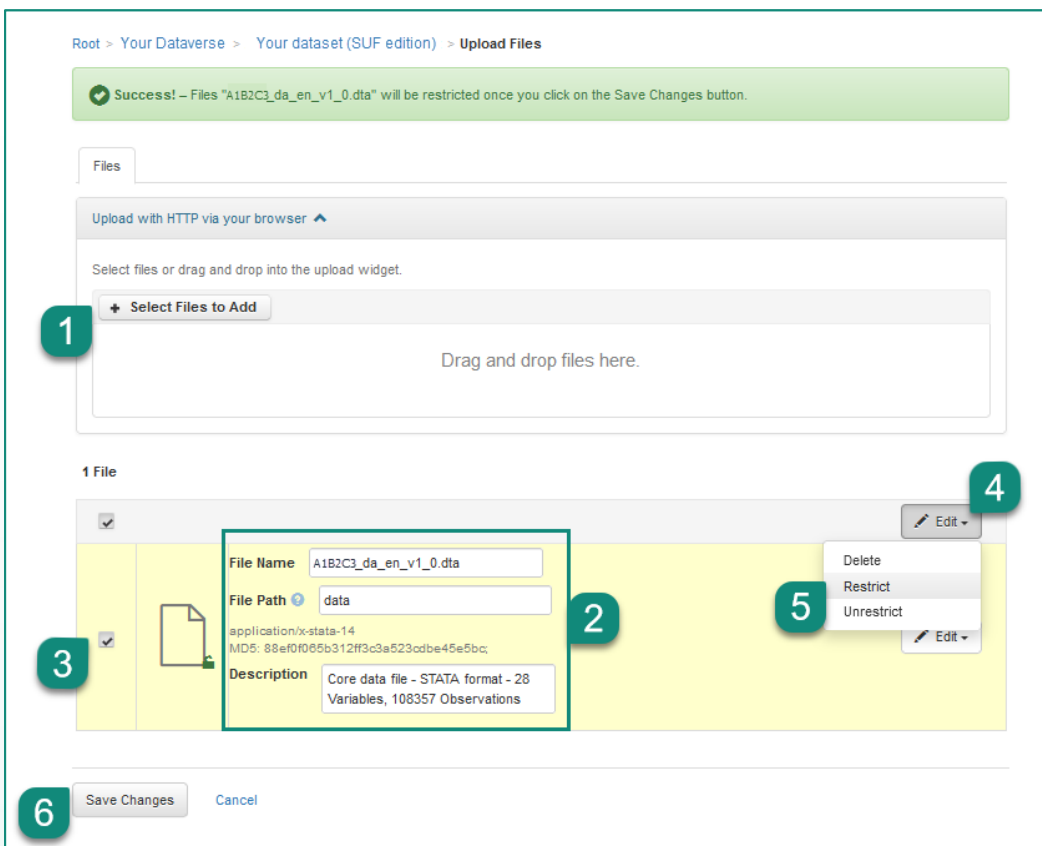
I generated other file formats to increase the reuse potential.

## Steps in Dataverse: Upload Files

Next, you can upload your files to your dataset draft.

➔ As shown in Figure 7, click on “Edit” > “Files (Upload)”.

Figure 9 and the list below contain all the individual steps:



**Figure 9** Add files, file metadata and restrictions in Dataverse. Note that the edit button here (marked with a “4”), is not the same edit button as in Figure 10.

- 1) Select all files you want to make available to users. In the next step, add more information about your files (also called “file metadata”).

- 2) The file metadata consist of the correct "File Name", a "File Path", a short "Description" of the content of the file and "Tags".<sup>33</sup>
  - I. File name\*: You can find information about file naming in the section above.
  - II. File Path: Use either "data" or "documentation", depending on the file.
  - III. Description\*: Use a short descriptive text to describe the content of your file  
**Examples:** Codebook  
Method report  
Core data file – STATA format – XX variables, XXX observations
  - IV. Tags: See Figure 10 for more details on how to further describe your files.
    - a) Click "Edit".
    - b) Choose from the three existing tags ("Data", "Documentation", "Code").
    - c) Add custom tags (choose from tags that relate to Table 3 "STATA", "SPSS", "Questionnaire", "Codebook", "Method report", "Research report", "Results", "Field report").
    - d) "Apply custom tags".
    - e) "Save changes".

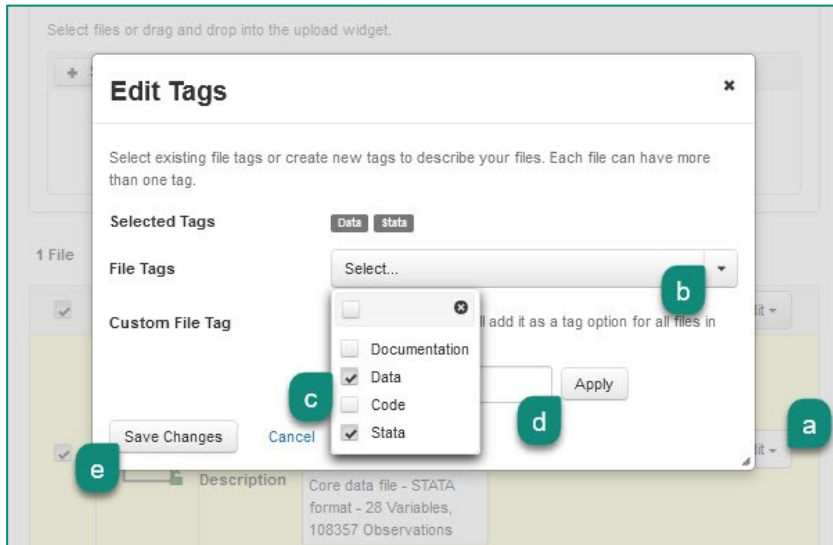
Figure 9 shows the most important steps:

- 3) **Click the checkbox on all files that will be provided under the scientific use licence (i.e. the data files).**
- 4) Go to "Edit".
- 5) **Click on "Restrict" to restrict your data files.**
  - I. Add "Terms of Access": Copy and paste to Terms of Access:

```
<p><b>Ich akzeptiere die für die Datei gültigen Lizenzbedingungen (Lizenzvertrag). Falls ich alle Dateien auf einmal herunterlade, akzeptiere ich alle Lizenzbedingungen.</b></p><p>Die Lizenzbedingungen befinden sich oben im Abschnitt "Terms of Use".</p><hr />For each file you download, you declare:<p><b>I accept the license terms valid for the file (license agreement). If I download all files at once, I accept all license terms.</b></p><p>The license terms can be found above in the section "Terms of Use".</p>
```
  - II. Tick "Enable Access Request".
  - III. Save Changes.

**If you see a green, open lock next to the file thumbnail, the restriction is in place. If you see no lock, try again. If you see no lock after a second attempt, please contact us.** The green lock signifies that the data file is restricted for users that are not logged in (they see a red, closed lock) and available for logged in users (they also see a green, open lock).
- 6) Save the changes.

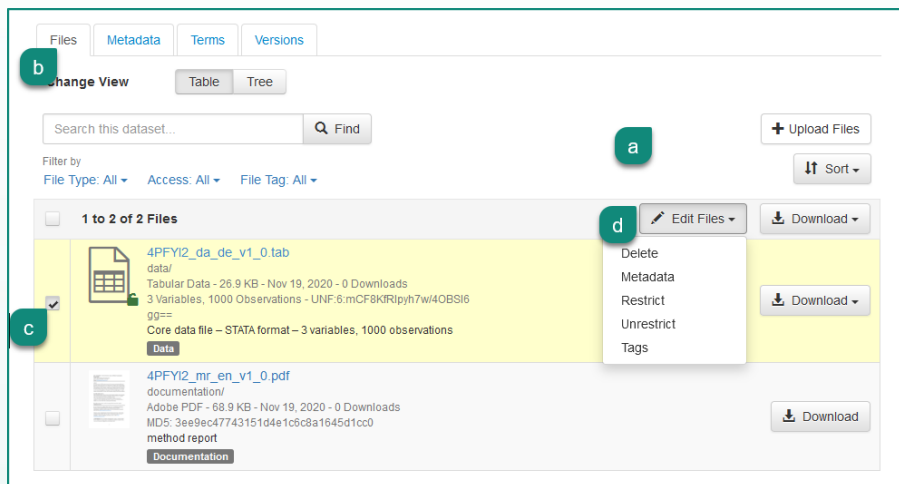
<sup>33</sup> Consider „File Name“ and „Description“ required fields, whereas „File Path“ and „Tags“ are optional.



**Figure 10** Add tags to a file to further describe your file. Note that the edit button here (marked with an “a”) is not the same edit button as in Figure 9.

If something is not working while uploading your files, try the following (see Figure 11):

- a) Save any changes.
- b) Go to the tab “Files”.
- c) Mark the checkbox of the file you want to edit.
- d) “Edit”.
- e) Make your changes.
- f) Save changes.



**Figure 11** In the tab “Files” you can edit metadata, restrict or delete files and add tags.

After you saved your files, Dataverse “ingests” your files. Dataverse will automatically try to convert your data files to \*.tab format. Ingesting can take a while, small files in minutes, larger files in eight hours or more. While Dataverse processes your data files, you cannot edit the dataset entry.



After the ingest is finished, check the numbers of variables and observations, these should correspond to the original numbers. If there is a difference, contact us. You can also add additional information to your metadata or files after the ingest is done.

## Take aways

I used the “Self-deposit: Scientific Use” template to create my dataset and replaced the pre-filled fields with my information if necessary.

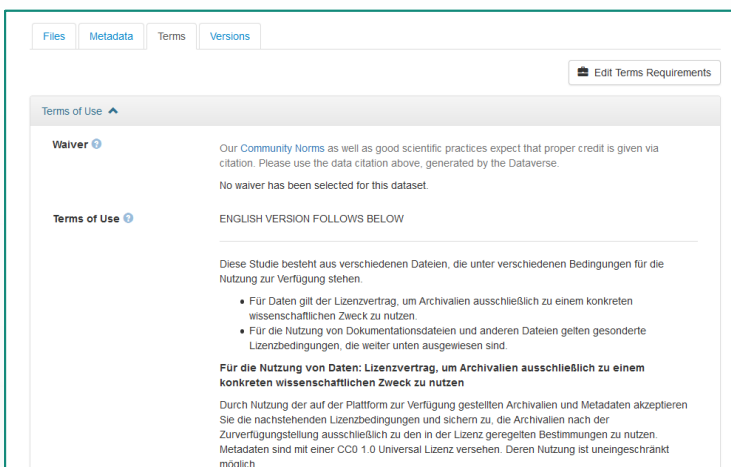
I uploaded my files with the correct file naming and a short description text.

**All my data files are restricted (I see a green, open lock next to my data file(s)).**

## 5.2.3. Terms

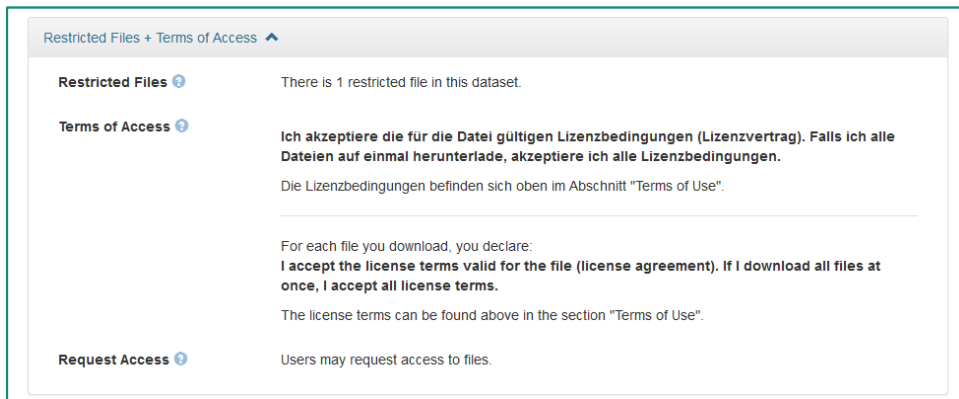
As next step have a look at the Terms and check if everything has been added from the “Self-Deposit (SUF)” template you used.

The “Terms of Use” may not be changed as they are part of the agreement you signed with AUSSDA and look like displayed in Figure 12.



**Figure 12 Terms of Use.**

The “Terms of Access” should look like displayed in Figure 13, “Request access” should be enabled, stating “Users may request access to files.”



**Figure 13 Terms of Access with enabled access request.**

If this field is empty, copy and paste this text to the Terms of Access and enable Access Request, see Figure 14:

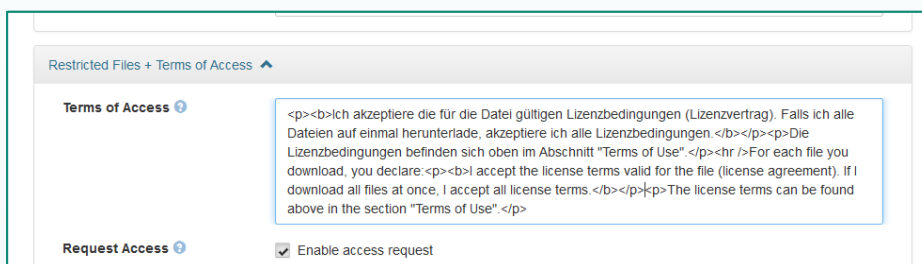
➡ "Edit Terms Requirements".

➡ Copy and paste to Terms of Access:

```
<p><b>Ich akzeptiere die für die Datei gültigen Lizenzbedingungen (Lizenzvertrag). Falls ich alle Dateien auf einmal herunterlade, akzeptiere ich alle Lizenzbedingungen.</b></p><p>Die Lizenzbedingungen befinden sich oben im Abschnitt "Terms of Use".</p><hr /><p>For each file you download, you declare:<p><b>I accept the license terms valid for the file (license agreement). If I download all files at once, I accept all license terms.</b></p><p>The license terms can be found above in the section "Terms of Use".</p>
```

➡ Tick "Enable Access Request".

➡ Save Changes.



**Figure 14 Adding Terms of Access and enabling access request.**

## Take aways

**I checked that the Terms of Access are filled with the correct text, provided above.**

## 5.3. Publishing

In this section you learn more about publishing your draft entry. Also, we explain how you can delete a draft if you do not need it anymore.

### Check your dataset

Thoroughly check your dataset in Dataverse before the publication.

Your draft is now finished, congratulations!

If this is your first upload, contact the [AUSSDA Team for feedback on the metadata and your Dataverse entry in general](#). After the feedback, you can publish the dataset yourself. If this is not your first data deposit, you can publish right away.

You can go ahead and publish your entry!

- ➡ Click "Publish".
- ➡ Confirm by clicking "Continue".

If you do not get a "Success" message at the top of the page, please contact us with further details and a screenshot of the error message.

Publishing a dataset may take some time as Dataverse processes your data files, so please be patient.

### Delete drafts

It is possible to delete your draft, if you really do not need it anymore, but it cannot be restored once deleted (see Figure 15).

- ➡ Click "Edit".
- ➡ Click "Delete Dataset".
- ➡ Confirm by clicking "Continue".

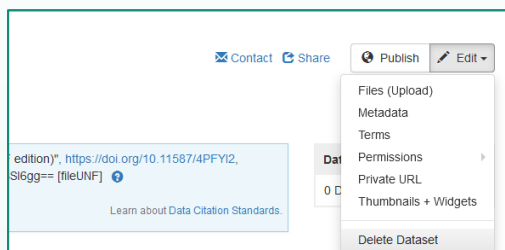
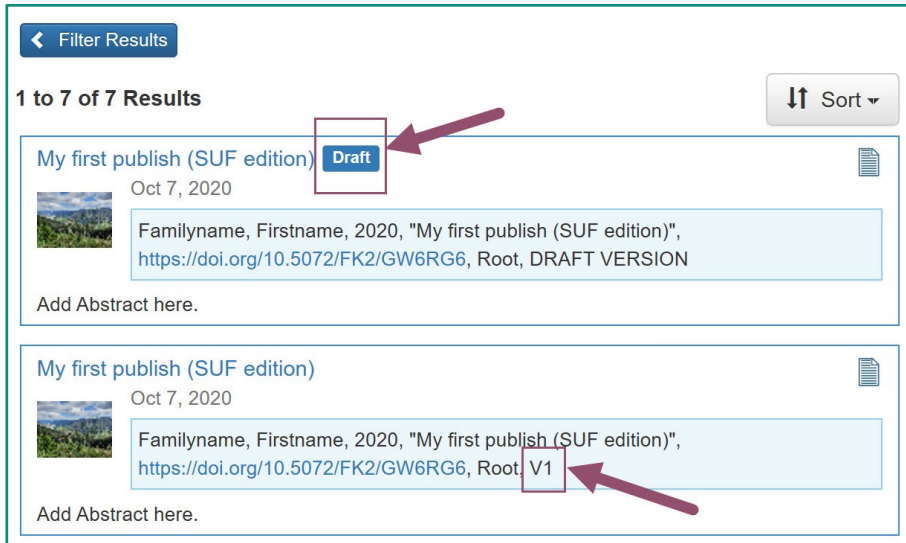


Figure 15 Deleting a dataset draft.

## 5.4. Updates

If you discover that you forgot to add something or want to upload additional files, you can update your existing dataset.

As soon as you edit something in an already published dataset, Dataverse creates a new "draft" version of your dataset (see Figure 16). Previous versions including all documents, data and files always remain unchanged and accessible for users (see tab "Versions" in Figure 17). Deaccessioning published files should only be done by AUSSDA and is foreseen for legal reasons (e.g. violations of data protection or intellectual property rights) and requires an explanation and a justification. Therefore, please check your files carefully before publishing.



Filter Results

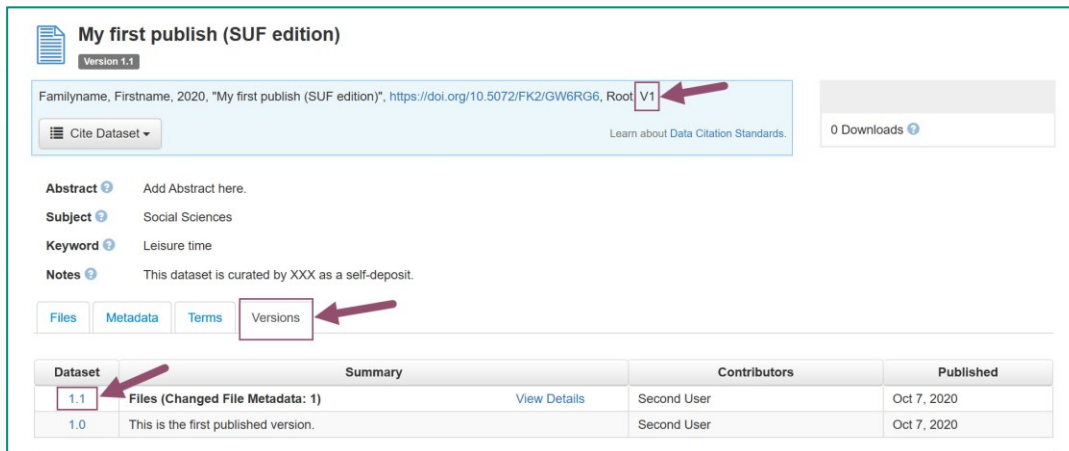
1 to 7 of 7 Results Sort ▾

My first publish (SUF edition) **Draft** 📄  
 Oct 7, 2020  
 Familyname, Firstname, 2020, "My first publish (SUF edition)",  
<https://doi.org/10.5072/FK2/GW6RG6>, Root, DRAFT VERSION  
 Add Abstract here.

My first publish (SUF edition) 📄  
 Oct 7, 2020  
 Familyname, Firstname, 2020, "My first publish (SUF edition)",  
<https://doi.org/10.5072/FK2/GW6RG6>, Root\_V1  
 Add Abstract here.

**Figure 16 Draft vs. Published versions in Dataverse.**

When all changes are completed and you want to publish your new version, you may have to make a decision whether it is a major or minor update, see Section “Minor vs. major changes” below. For a new version, you should also change the versioning in the names of your files, see Section “File Naming” for more on file naming.



My first publish (SUF edition)  
Version 1.1

Familyname, Firstname, 2020, "My first publish (SUF edition)", <https://doi.org/10.5072/FK2/GW6RG6>, Root\_V1 0 Downloads

Cite Dataset Learn about Data Citation Standards.

Abstract Add Abstract here.  
 Subject Social Sciences  
 Keyword Leisure time  
 Notes This dataset is curated by XXX as a self-deposit.

Files Metadata Terms **Versions**

Dataset	Summary	Contributors	Published
<b>1.1</b>	Files (Changed File Metadata: 1) <a href="#">View Details</a>	Second User	Oct 7, 2020
1.0	This is the first published version.	Second User	Oct 7, 2020

**Figure 17 Versioning of dataset in Dataverse.**

## Minor vs. major changes

If there are **major** changes to data files (files deleted, new files uploaded, etc.) Dataverse decides for you: it will be a major update to the next full number.

If you made **minor** changes, for example in the metadata, Dataverse asks you if this is a major update (changing to the next full number) or a minor update (changing the decimal place, e.g. 1.1, 1.4, etc.). The changes made to your data entry can be followed under the tag “Versions” (see Figure 17), click “View Details”. You can also find more information in the AUSSDA [User Guide on major/minor](#)<sup>34</sup>.

<sup>34</sup> <https://aussda.at/userguide/#c291053>

## 6. Checklist Box for whole deposit process

### Legal questions

Yeah, I got this!

Get in contact with AUSSDA and <b>sign the agreement</b> that allows you to self-deposit data.	<input type="checkbox"/>
Understand the license agreements for data, metadata and documentation material.	<input type="checkbox"/>
Read and understand the AUSSDA Non-Compliance Policy.	<input type="checkbox"/>

### Preparation of data & documentation material

<b>I pseudonymised my data, following AUSSDA's Data Deposit Guideline.</b>	<input type="checkbox"/>
I cleaned the data as suggested, e.g. regarding variable and value labels and missing values, among others.	<input type="checkbox"/>
I double-checked the data with regard to plausibility and outliers and they are fine.	<input type="checkbox"/>
I compared the data with the documentation material and they match.	<input type="checkbox"/>
I gathered all documentation material for my data.	<input type="checkbox"/>
I included a suggested citation and information about the license of the document in each document, preferably on the first page.	<input type="checkbox"/>

### Metadata

I gathered all relevant metadata and filled my dataset draft in Dataverse with the relevant, correct information.	<input type="checkbox"/>
---	--------------------------

### Files

I saved my documentation material as PDF/A to ensure long-time archive format.	<input type="checkbox"/>
I added two variables to my dataset ("version" and "doi").	<input type="checkbox"/>
I named my documentation files according to AUSSDA's suggested file naming scheme.	<input type="checkbox"/>
I generated other file formats to increase the reuse potential.	<input type="checkbox"/>
<b>I used the "Self-deposit: Scientific Use" template to create my dataset and replaced all pre-filled fields with my information.</b>	<input type="checkbox"/>
I uploaded my files with the correct file naming and a short description text.	<input type="checkbox"/>
<b>All my data files are restricted (I see a green, open lock next to my data file(s)).</b>	<input type="checkbox"/>

### Terms

<b>I checked that the Terms of Access are filled with the correct text, provided above.</b>	<input type="checkbox"/>
---	--------------------------