



User Guide for Linguistics Data

Submit your study and
deposit data

For more information, please contact:

swissubase@ub.uzh.ch

<i>Version:</i>	1.1	<i>Persistent link:</i>	https://resources.swissubase.ch/linguistics_user-guide_en/
<i>Last modified:</i>	08.09.2022	<i>License:</i>	CC BY
<i>Introduced:</i>	04.07.2022	<i>Publisher:</i>	Language Repository Switzerland (LaRS), Florian Steurer, swissubase@ub.uzh.ch

Table of contents

1. Create a 'Linguistics' study	2
2. Create a SWISSUbase dataset linked to a study	2
3. Add a 'FILE COLLECTION'	3
4. Add Metadata.....	4

User guide for linguistic data

Preamble: This guide focuses on linguistics data and metadata and addresses data depositors who wish to use SWISSUbase as their data repository for publication and archiving. Please consider consulting the following documents and websites:

- > [Linguistic Metadata Guide](#)
- > [Policy Language Repository of Switzerland](#)
- > [Deposit Contract Language Repository of Switzerland](#)
- > [General SWISSUbase User Guide](#)

i INFO: Validate and save information / contact

- > Whenever you enter metadata, data or change information, please click the 'Validate' or 'Save' button in the respective input form.
- > If you need help or have suggestions, do not hesitate to contact us:
swissubase@ub.uzh.ch

1. CREATE A 'LINGUISTICS' STUDY

- 1) Go to the 'My studies' view and choose 'Create new study'.
- 2) For 'Scientific domain', select 'Linguistics'. This ensures that the metadata scheme specific to Linguistics is used. The 'University of Zurich' is automatically assigned as the data curating institution.
- 3) Please proceed as suggested in the [General Guide](#).

Create new study

Scientific domain • *Cannot be updated later*

Social sciences ⓘ

Linguistics ⓘ

General ⓘ

2. CREATE A SWISSUBASE DATASET LINKED TO A STUDY

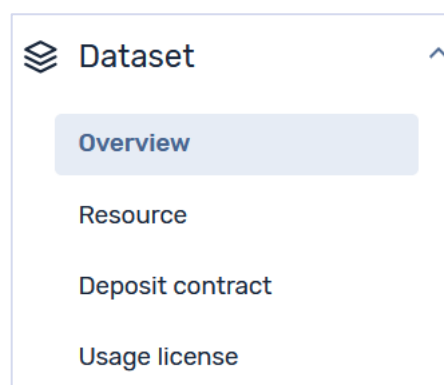
📘 INFO: Creating datasets and files

- **Recommended:** The easiest way to upload a coherent file collection is to provide a ZIP file. To do so, you need to select the 'File Collection' option in the 'Files' block.
- In the study you created, you will need to upload at least one 'SWISSUbase dataset', and then add linguistics-specific metadata.
- You can create as many SWISSUbase datasets as needed in a study. Within a dataset, you can create as many 'Single Data File', 'File Collection' and 'Documentation' entries as needed.

At least one dataset is needed for each study. It can include 'Single Data File', a 'File Collection' and/or 'Documentation' units. Dataset options:

Single Data File	A 'Single Data File' is a single text, audio, image, video, or any other file type that is not a ZIP file (or similar) or any other packed archive.
File Collection	A 'File Collection' is a ZIP file container that holds a linguistic file collection. This is the recommended way to upload and archive linguistic data and metadata.
Documentation	If you plan to upload different documentation files such as Readme files or codebooks, you have to create for each one its own 'Documentation' unit. If you need to provide more than one file for documentation reasons, you can e.g. upload a ZIP archive. The Language Repository Switzerland (LaRS) also accepts documentation material that is included in the 'File Collection': Thus, there is no need to create a separate SWISSUbase 'Documentation' unit if you prefer to keep everything in one place (e.g. uploaded in a ZIP archive).

- 1) In the dataset view, create a dataset by clicking on the **'Add dataset'** button.
- 2) Fill in all metadata that are needed for the dataset in the sections **Overview, Resource, Usage License**. In **'Resource'** choose the most appropriate **'Resource type'** for your data collection.



- 3) In **'Usage license'**, select one of the available licenses to make your dataset directly available in the catalog once it is published. If you select the **'Closed contract'** for Linguistics, your data will not appear for download in the catalog. There are two options available:
 - o **No prior agreement selected in the contract condition:** Users need to login, accept the 'closed contract' first, and can then download the data.
 - o **Prior agreement selected in the contract condition:** Users need to login, accept the 'closed contract' first, send a request to the data depositor, and have the data made available for download later.

i INFO: Linguistics metadata on dataset level

- For more specific details on **'Resource'** metadata, please consult the [Linguistics Metadata Guide](#).
- For more details on general metadata fields for each dataset, and the contract and licenses, please consult the [SWISSUbase General User and Metadata Guide](#).

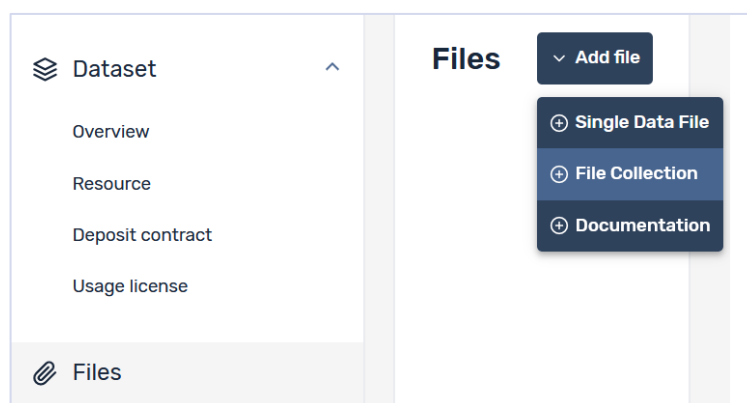
- 4) After completing the metadata, click on **'Files'** to upload your files and documentation.

3. ADD A 'FILE COLLECTION'

i INFO: Uploading your data as a **'File Collection'** is recommended.

- The easiest way to upload a coherent file collection is to provide a ZIP file container. To do so, you need to select the **'File Collection'** option.
- You can put all your files in a ZIP container and upload it in one go.
- Put your data on top level: 'corpus.zip/your_data'. Avoid a nested structure like 'corpus.zip/corpus/your_data'.
- Please remove hidden files such as **'__MACOSX'** or **'.DS_store'**.

- 1) To upload a file collection, click on 'Add file' and choose 'File Collection'.
- 2) Now you can drag and drop your file collection into the upload field. Please allow sufficient time if you upload a large file collection. If you wish to upload it later, you can skip the upload step for now.

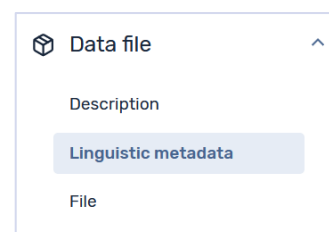


4. ADD METADATA

INFO: The 'Language' metadata block is mandatory.

- The 'Language' metadata block is mandatory – at least one entry must be completed to deposit your data.
- If a field is not applicable or you cannot find an appropriate option, please select 'Not applicable' or 'Other' in the drop down menus.
- All other metadata blocks are optional (e.g. annotation, video, etc.), but once selected, there are some mandatory metadata fields you must complete. These are marked with a red dot.
- You can add as many blocks as needed. Blocks are repeatable. E.g., if two languages ('bilingual') occur in your dataset, you can add a second language metadata block. Please select 'bilingual' for the 'Linguality type' accordingly.

- 1) Enter the metadata needed and go to the 'Linguistic metadata' view. Click on the 'Edit' button.
- 2) Click on the 'Add metadata' pull down menu. Here you can select the different metadata blocks. Please consult the [Linguistics Metadata Guide](#) to find out more about the available fields in each metadata block.



Description of each linguistics metadata block

Language	Metadata concerning the language itself such as language name or linguality type
Annotation	Metadata about the annotation such as annotation type and controlled vocabulary
Audio	Metadata about audio data such as codec information or audio genre

Image	Metadata about image data such as image genre or Media type
Text	Metadata about text data such as text genre or provenience
Video	Metadata about video data such as container format and resolution information
Tools	Metadata about tools used to generate, process, or annotate data.



INFO: Reference paths

- Each metadata block has a **'Reference path'** field. The reference path allows you to specify to which folder the metadata refer.
 - Choose a slash to refer to the whole collection: /
 - Refer to a specific folder: /folder/
 - Refer to a specific subfolder: /folder/subfolder/

INFO: Known issues and limitations

- The file structure of ZIP archives is not shown but will be a future development.
- Reference paths must be entered manually for now. It is planned to add the functionality to allow path selection directly from the ZIP archive.

3) Once you have entered the metadata you can deposit your dataset. For general advice, please consult the [General SWISSUbase User Guide](#).