

THE PREPARATION OF SOCIAL SCIENCE DATA FOR SWISSUbase: IN BRIEF

To deposit your data in [SWISSUbase](#) and make them accessible to other researchers, you must describe your research project and add one (or more) datasets. In addition, the data must meet certain requirements.

Specifically, a dataset must be composed at least of a **cleaned** and **anonymized** data file in one of the **formats listed below**, as well as the necessary **documentation** for the proper interpretation and reuse of the data.

Data cleaning

Quantitative data:

- Variable and value labels must be included in the data file. Labels can take the form of shortened survey questions. The links between variables in the data file and the corresponding questions in the questionnaire must be clear.
- Explicit codes should be assigned to missing values (e.g., don't know, refusal, filter, non-response), where possible.
- Frequencies should be checked, and inconsistencies or detected errors should be corrected or removed (e.g., invalid or highly unlikely values).

Qualitative data:

- Information about the data collection - such as date, location, unique identifier of the interviewer and interviewee - should be available either in the file (e.g., transcripts) or in a separate document.

SWISSUbase

[SWISSUbase](#) is an online platform that allows for the archiving and dissemination of research data and their metadata, with services to assist deposit and access, including data management support

As a collaboration of several partner institutions, SWISSUbase is based on international archiving standards to ensure the long-term preservation and accessibility of data, and to meet the needs of universities and other research institutions.



Anonymisation

It is the responsibility of researchers to anonymize their data before deposit, so that no individual can be identified without “disproportionate effort”. This requires rigorous treatment of direct and indirect identifiers:

- **Direct identifiers**, i.e., any variables or other elements that would directly identify individuals (e.g., names, addresses, telephone numbers) should be removed or replaced with pseudonyms or categories.
- **Indirect identifiers**, which in combination could identify an individual (e.g., age, occupation, postal code), should also be checked and recoded or deleted if necessary.

If it is not possible to anonymize the data, explicit **consent** from the participants is required for the storage and sharing of the data. A copy of the consent form should be provided.

SWISSUbase also allows you to restrict access to data (e.g., exclude access for teaching purposes...) and/or to require a prior agreement from the author.

Formats

Depending on the type of data, the following formats are accepted in SWISSUbase:

- **Tabular data:** SPSS data file & portable data file (.sav; .por), MS Excel (.xls; .xlsx, Comma-separated value (.csv).
Files such as Stata data file (.dta), Statistical Analysis System (.sas), R Data Format (.RData; .rds), OpenDocument spreadsheet (.ods) are accepted as well, but should be converted in SPSS, Excel or csv by the data producer so that we can conduct quality control checks.
- **Textual data:** Portable Document Format (.pdf), Rich Text Format (.rtf), ASCII (.txt), MS Word (.doc; docx), OpenDocument text (.odt)
- **Image data:** JPEG (.jpeg; .jpg; .jp2), TIFF (.tif; .tiff), RAW image format (.raw), Bitmap Image File (.bmp), Portable Network Graphics (.png), Portable Document Format (.pdf)
- **Audio data:** MP Format (.mp3), FLAC (.flac), Waveform Audio Format (.wav), Advanced Audio Coding (.aac)
- **Video data:** MP Format (.mp4), QuickTime video (.mov), Fichier AVI (.avi), Motion JPEG 2000 (.mj2)

The folders and files in the accepted formats can be deposited in zip (.zip, .rar, .7z).

Documentation

The deposit should include all documentation necessary to properly interpret and ensure the quality of the data, namely:

Report describing the following aspects:

- Project: context, problem, objectives, hypotheses
- Method: population, sampling (size, response rate), data collection method
- Data: cleaning, anonymization, coding, validation, description of recoded or constructed variables, weighting

Documents related to data collection:

- Materials sent to respondents in advance (e.g., preliminary letters)
- Instruments used for data collection (questionnaires, grids, protocols)
- Materials presented to respondents during interviews
- Instructions or materials for interviewers

If available, please also add:

- Publications and final reports
- Codebook
- Coding instructions
- Syntaxes
- Data Management Plan

Documentation files must be submitted in **PDF** format (except for syntaxes) and in the different available languages.

More information and resources

- [SWISSUbase](#)
- [SWISSUbase user guide](#)
- [Data Management](#) (FORS-Website)
- [FORS Data protection policy](#) (FORS-Website)
- [The informed consent as legal and ethical basis of research data production](#) (FORS Guide N°05)
- [Data anonymisation: legal, ethical, and strategic considerations](#) (FORS Guide N°11)
- [Data Management Expert Guide: Anonymisation](#) (CESSDA recommendations for data anonymisation)

Need help?

Contact us!

dataservice@fors.unil.ch