

THE PREPARATION OF SOCIAL SCIENCE DATA FOR SWISSUbase: IN BRIEF

In order to deposit your research data in [SWISSUbase](#) and make them accessible to other researchers, you must describe your research project and add one (or more) datasets. Furthermore, certain requirements related to the data have to be met.

Specifically, a dataset must be composed at least of a **cleaned and anonymized data file** in an accepted format listed below (cf. p. 2), as well as the necessary **documentation** for the proper interpretation and reuse of the data.

Data cleaning

Quantitative data:

- Variables 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).
- 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 dedicated to archiving and dissemination of research data and their metadata. Different services for the management of research projects are also provided.

As a collaboration of several partner institutions, SWISSUbase complies with 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 depositing them, so that no individual can be identified without “disproportionate effort”. This process of anonymisation requires rigorous treatment of direct and indirect identifiers:

- **Direct identifiers** are any variables or other elements that would directly lead to the identification of an individual (e.g., names, addresses, telephone numbers). They should be removed or replaced with pseudonyms or categories.
- **Indirect identifiers** are variable or other element that, when combined, may lead to the identification of an individual (e.g., age, occupation, postal code). They should also be checked and recoded or deleted if necessary.

If the anonymisation of data is not feasible, 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 offers different options to keep control over the access to the data: prior agreement, embargo, use restrictions.

Formats

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

- **Tabular data:** SPSS data file & portable format (.sav; .por), MS Excel (.xls; .xlsx), Comma-separated value file (.csv).
Files such as Stata (.dta), Statistical Analysis System (.sas), R Data Format (.RData; .rds), OpenDocument spreadsheet (.ods) data files are also accepted, but must be converted to SPSS, Excel, or CSV by the data producer so that quality control checks can be conducted by the FORS data service.
- **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 above-mentioned formats can be deposited in ZIP format (.zip, .rar, .7z).

Documentation

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

A 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 beforehand 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 resources](#)
- [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°5)
- [How to draft a DMP from the perspective of the social sciences, using the SNSF template](#) (FORS Guide N°07)
- [Data anonymisation: legal, ethical, and strategic considerations](#) (FORS Guide n°11)
- [Data Management Expert Guide: Anonymisation](#) (CESSDA recommendations for data anonymisation)

Need help?

dataservice@fors.unil.ch