

THE PREPARATION OF SOCIAL SCIENCE DATA FOR SWISSUbase: IN DETAIL

Sharing research data means preparing it

Research data sharing meaning the deposit of data and metadata generated from a research project to allow access and reuse, contributes to the advancement of science and allows transparency and reproducibility of results.

In the context of Open Research Data, sharing is becoming an obligation. This imperative comes from institutional policies and funding agencies that may require sharing, but also from scientific journals, in which publication of articles is conditional on sharing and publishing the data.

[SWISSUbase](#) enables researchers to meet this sharing imperative and is a reliable and efficient solution for the preservation and dissemination of research data.

The FORS Data Service ensures that the deposited data are of quality, that they are sufficiently documented for reuse, and that the work of collecting and preparing the data is recognized by employing proper data citation practices.

This process of data preparation must be planned from the start (especially when writing the Data Management Plan), and then carried out throughout the project. This requires attention to the various aspects mentioned in this guide, namely: the **organization** of the files, the **documentation** necessary for their proper interpretation, and the **ethical and legal issues**, in particular **consent**, **anonymization** and **copyright**.

SWISSUbase: the free solution for research data

[SWISSUbase](#), the platform used by the FORS Data Service, allows for:

- Deposit of data from different disciplines free of charge, with long-term preservation on servers in Switzerland;
- Control over how research data are accessed and used (e.g., access restrictions and embargoes);
- Visibility and recognition of the investments made in the creation of research data by facilitating their discoverability, re-use and citation;
- Efficient access and re-use of existing research data to avoid duplication of effort and costs.

SWISS  base

Organize and name data files

To facilitate file sharing, certain decisions about **file structure** and **naming conventions** should be made early in a project.

By opting for a structured organization of the various files as well as by applying precise naming conventions, accessing and identifying them becomes easier, which is essential for sharing and archiving.

A good organization of files represents the structure of the information they contain while the application of naming conventions brings logic and makes the structure coherent.

The name of a file is its identifier. It should give clues as to the content and version of the file. Thus, the naming strategy should be systematic and easily interpreted by others.

Quantitative data files

In quantitative research, data are often derived from numerical coding of standardized questionnaires and consist of variables, numbers, matrices, etc.

The names and labels of the variables must be short and respect the requirements and standards of the software used.

Here are some tips for naming variables:

- Start with a letter and not a number and avoid spaces, periods and special characters such as # ! ? &
- Choose a short (no more than 8 characters) and meaningful name using for example:
 - Numerical codes that reflect the position of the variables and the relations between variables (ex: V001, V002)
 - Codes referring to the data collection tool or sources of the data (e.g. question number: Q1, Q2a, Q2b, Q3)
 - Names referring to the content of the variables (ex: AGE, SEX)

Some tips for establishing file naming conventions

- Give short and meaningful names, meaning that they reflect the content of the file;
- Avoid spaces, periods, special characters (& ? ! # etc.) as well as accents and the cedilla;
- Favour the use of capital letters or underscores "_" to separate the elements of a file name (e.g. RuleDenominationFile);
- Keep the 3-letter file extension indicating the file format (e.g. .doc, .xls, .mov, etc.);
- If necessary, indicate the dates in the format YYYYMMDD (year, month, day) which allows the documents to be kept in chronological order;
- Include if necessary the version in the file name (e.g. v01, v02, vf).

Qualitative Data Files

Qualitative data files are derived from various research materials: texts (transcripts of individual interviews or focus groups, newspaper or magazine texts, observation notes, etc.), audio or video files, photographs, etc.

By adopting a consistent naming scheme, the different files related to a data collection event, such as an interview, will be more easily recognizable. Each file should have a common designator to identify the event and indicate the specificity of the file.

The following example illustrates a good practice for naming qualitative files.

- 20200417_Interview2_Audio.wav
- 20200417_Interview2_Trans.pdf
- 20200417_Interview2_Image.jpg

Accepted formats in SWISSUbase

File formats must allow for operability, sharing and long-term preservation of data. Depending on the type of data, the following formats are preferred:

- **Tabulated 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), OpenDocument spreadsheet (.ods) are also accepted but should be converted to SPSS, Excel or CSV by the data producers)
- **Text 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), AVI file (.avi), Motion JPEG 2000 (.mj2)

Folders and files in accepted formats can be uploaded in zip (Zip, 7rm, rar).

Tips for transcriptions

A transcription is a conversion of audio or video recordings into text requiring the adoption of transcription rules adapted to the data and the intended analysis.

It is possible to choose or draw on an established transcription system (such as the "Jefferson Transcription System").

Regardless of the rules adopted, certain information should be systematically entered at the beginning of each transcription:

- The name or initials of the interviewer;
- The interviewee's pseudonym;
- The date and place;
- The event's identification code.

In addition, it is also appropriate to:

- Adopt the same layout;
- Use the same font (e.g. courier);
- Apply the necessary measures to respect the confidentiality of the interviewees.

Here is a [transcription template](#) to be used as is or adapted.

Documentation

Clear, detailed and systematic documentation throughout the project improves the quality of the data and is essential for archiving research data, but also for allowing others to discover, cite and reuse it.

When depositing in [SWISSUbase](#), it is recommended to fill in as many metadata fields as possible in order to increase the discoverability and visibility of the project. Moreover, the addition of documentation in the different languages available and in English increases the chances that research data will be re-used and cited.

Regardless of the format of the data and whether they are quantitative or qualitative, it is important to document the project and the datasets. Both levels of documentation, at the project level and at the dataset level, are provided for in [SWISSUbase](#).

Points to document at the project level

- Describe the project by specifying the **context** (space and time), the **research questions**, the **goals** and **objectives**, the **concepts** and **hypotheses**.
- Describe the **method** and **mode of collection**, including methodology used, collection instruments and software, target population, sampling (size, response rate), and so on.
- Import any publications or a list of publications (in .ris or .txt format).

Points to document at the dataset level

- Describe specific operations and protocols related to **data processing** (data cleaning, coding, data classification, weighting, etc.) and **anonymization**.
- Describe the **instruments** used for data collection (questionnaires, grids, protocols) and upload them to SWISSUbase.
- Describe and download the **instructions** given for data collection and the materials used by interviewers and/or respondents.
- Complete the fields in SWISSUbase related to the **conditions of use and access** to the data (access conditions, embargo, access restrictions, ...).

Documents to be attached

The documentation must allow for interpretation as well as secondary analysis of the data, and should include:

- **Codebook** and/or **coding instructions** and/or **technical report** (quantitative data);
- **Syntaxes** (quantitative data);
- **Collection instruments**, in several languages if available;
- **Instructions** and **materials** for interviewers and/or respondents (information and/or thank you letters, text, image or video related to the questions, interviewer training booklet, etc);
- **Blank informed consent form** (if collected);
- **Publications** or list of publications;
- **Intermediate** or **final reports**;
- Lists of certain **biographical characteristics** of the participants such as age, gender, occupation, and identification details, etc. (qualitative data).

Ethical issues related to data archiving

The information and advice given here does not constitute legal advice.

The collection, use, sharing and archiving of research data must be guided by ethical considerations and legal obligations.

Researchers must ensure that they comply with all national and international laws that may affect their research project and the data they use and/or produce, particularly laws relating to data protection and intellectual property.

Thus, a research project must be designed taking into account the participants, the type of data collected and/or used, and the various laws that govern them. However, data protection laws only apply to personal data and not to anonymous data.

Personal data: Data that directly or indirectly allow identification of person are considered personal data.

Sensitive data: Some personal data are considered sensitive and require special protection. This includes data relating to religious affiliation, political opinions, sexual orientation or racial or ethnic origin.

An increasing number of scientific journals and funding agencies require data to be shared and therefore made available through a data archive or repository. However, sharing and archiving personal and/or sensitive data can be delicate and require some precautions:

- Participation in research must be voluntary and fully informed. **Informed consent procedures** should allow participants to choose whether or not to agree to data sharing.
- **Confidentiality** must be ensured and, if necessary, identifying information can be excluded from data sharing in order to disseminate anonymized data.
- **Access** to the data should be controllable and restricted if necessary.

For more information on this subject, you can consult the FORS Guide n°3 [Ethics in the era of open research data: some points of reference](#).

Overview of Swiss laws

In Switzerland, data protection is regulated at the federal and cantonal levels.

At the federal level, research conducted privately as well as by a federal public body (such as the Federal Institutes of Technology) is subject to the **Federal Act on Data Protection (FADP)**.

In addition, each of the 26 cantons has its own data protection law. Research carried out by universities, colleges, hospitals, etc. is subject to the **laws of the canton** in question.

Research on diseases or on the functioning of the human body involving personal health data or biological material is governed by the **Swiss Federal Human Research Act (HRA)** and must be evaluated by a cantonal ethics commission.

In some universities, committees are responsible for evaluating projects and giving them ethical validation.

Informed consent

Informed consent must be given freely and voluntarily, actively and unambiguously, with full knowledge and through clear affirmative action. Also, before the research begins, potential participants must be given appropriate information about the research so that they can make a voluntary and informed choice about whether to participate.

Different information covered in a consent

- The purpose of the research and the selection criteria;
- Contact information for the researcher(s) or a contact person, as well as the sponsor;
- The measures taken to preserve the anonymity and confidentiality of participants;
- The right to withdraw from research;
- What will happen to the participants' contributions and their data, particularly in terms of **archiving** and **sharing** and therefore **re-use** (specifying the purposes).

The last point, archiving and sharing, is essential for participants to make an informed decision about the use of their data for other research projects.

For more details on consent, see FORS Guide n°5 [The informed consent as legal and ethical basis of research data production](#). Swissethics also offers a guide as well as consent and information sheet templates to download in French, German, and Italian [on their website](#).

What do Swiss laws say about consent?

- The **Federal Act on Data Protection (FADP)** requires **consent** to process **personal data**.
- Furthermore, if a research project falls within the scope of the **Swiss Federal Human Research Act (HRA)**, the **consent** must be in **writing**.

Consent in qualitative research

In qualitative research, it is advisable to obtain written consent. In addition, it may be advisable to offer the participant the possibility of giving differentiated consent depending on the type of data collected, for example:

- I agree that the anonymized transcript of my interview may be archived on SWISSUbase and distributed for reuse by other researchers for research and teaching purposes.
- I agree that the non-anonymized audio recording of my interview may be archived on SWISSUbase and distributed.
- I agree that photographs of myself taken during the interview may be archived on SWISSUbase and distributed for reuse by other researchers for research and teaching purposes.

Consent in quantitative research

For surveys where direct or indirect identifiers are not collected or deleted from the data file, written consent is generally not obtained. However, participants should be provided with the information necessary to make a decision about their participation in the research project through participant information sheets that address more or less the same issues as consent, including the following:

- the **name of the researchers** and the **institution** with which they are affiliated;
- the **name of the project**, the **objectives** of the research and the **source of funding**;
- the potential **risks** and **benefits** of participation and the procedure for withdrawing from the research;
- the **purpose** of the data collection and the intended use (dissemination, archiving, publication);
- the **repository location** of the data as well as details on potential future uses;
- the procedures in place to protect participants in terms of **confidentiality** and **anonymity**;
- ensuring **ethical use** of data throughout the data lifecycle, including **archiving**, **sharing** and **reuse**.

Information sheets, which are also used in qualitative research in addition to consent, must indicate that by agreeing to participate, respondents are consenting to the use of the data for specific purposes and that the data will not be used in a way that would identify them.

Obtain informed consent for data sharing and archiving

In the context of Open Science, obtaining consent for the sharing and reuse of research data is increasingly relevant. Whether the consent is written or oral, addressing the issues of data sharing and archiving allows participants to make an informed decision, and allows researchers to share their research data with confidence, but also to publish in compliance with data protection laws.

The use of information sheets

The information sheets go hand in hand with obtaining informed consent by providing the information needed to decide whether or not to participate in the research. Many of the points to be specified in the information sheets are also included in the written consents. Indeed, the information sheets should stipulate the following aspects:

- Research **objectives**;
- The **name of the project** and various information about the project (source of funding, contact information for the researchers and the institution to which they are affiliated);
- The **benefits** and **risks** of participation;
- The **intended use** of the data;
- The measures taken to maximize the protection of personal information and **confidentiality**, particularly with regard to archiving, sharing and re-use of data.

Anonymization

When the collection of identifiable information cannot be avoided, it is possible to use anonymization techniques to process **direct identifiers** (name, address, telephone number, etc.) that are relevant to the identification of a specific individual and **indirect identifiers** (occupation, salary, place of residence or origin, age) that, when combined, can also reveal an individual. Thus, the anonymization process aims to ensure that no individual can be identified without "disproportionate effort", which requires a rigorous treatment of identifiers that must be aggregated, distorted, generalized or deleted.

Some tips for anonymization in quantitative research

- Aggregate or remove variables;
- Reduce or generalize the textual meaning of variables;
- Reduce the upper and lower ranges of continuous variables to make aberrant values invisible.

Some tips for anonymization in qualitative research

- Use pseudonyms instead of deleting and "lose" the identification information;
- Anticipate anonymization at the time of transcription and apply the use of pseudonyms consistently throughout, and include all members of the research team involved in data processing so that everyone proceeds in the same way;
- Use the "search and replace" feature to systematically make all necessary changes to anonymize data;
- Clearly mark replacements in the text, for example by using square brackets and by establishing or systematically referring to a specific transcription code;
- Create a file containing the anonymization information, i.e. all replacements, aggregations or removals performed and save it separately in a secure manner.

For more details on anonymization, see FORS Guide n°11 [Data anonymisation: legal, ethical, and strategic considerations](#).

Requirements and restrictions for filing in SWISSUbase

- **Anonymization:** The filing process requires confirmation that the dataset does not contain personal data or that all participants have been informed of the publication of the data and/or have given their consent.
- **Audio, image and video files:** Submission of such files is accepted as long as they are relevant for reuse. Access to such files is only granted if participants have consented to the sharing and reuse of their data. In this case, the **blank consent** template must be filed as a documentation file.
- **Restrictions** (for closed contracts only): it is possible to **restrict access** to specific uses (research or academic teaching); to decide on a case-by-case basis whether or not to approve the downloading of data with the option of **prior approval**; or to establish an **embargo** by indicating the date when the data will be available.

Copyright

Copyright is a form of intellectual property right that aims to protect a work and automatically applies, among other things, in research.

Copyright does not tend to protect ideas or concepts related to a research project, but rather the particular ways in which they were expressed. Research results such as spreadsheets, syntaxes, reports, publications, etc. can be protected by copyright. Therefore, these results generally cannot be published, translated, reproduced or adapted without the approval of the owner of the rights, which is why it is important for researchers to pay attention to copyright.

In fact, whether you want to reuse, share or archive data, you must always ask yourself who the copyright holder is, along with the following questions:

- Am I allowed to use the data?
- Am I allowed to archive and/or publish the data in an archive or repository?
- Who owns the rights to the data?

Generally, the person who created the research is considered the rights holder of the research data when depositing the data with an archive or repository. The archive or repository therefore only allows access to and storage of the data.

However, it is important to consult your employment contract. Indeed, some employers stipulate in their contracts that any work created during employment becomes the property of the employer. In this case, the researcher is obliged to request authorization from the employer for the deposit of data and the terms of access and publication.

Special Cases

Multiple researchers:

If research data are created by several persons, they can be considered as joint copyright holders. Therefore, in addition to the research participants, all copyright holders must give their permission for the archiving of the research data.

Data derived from other data:

If research data are derived from data previously created by other individuals or institutions, permission must be sought from the rights holder(s) before depositing the data in an archive or repository.



www.forscenter.ch, dataservice@forscenter.ch, + 41 21 692 46 72