AUSSDA
AUSTRIAN SOCIAL SCIENCE DATA ARCHIVE

DATA DEPOSIT GUIDELINE FOR QUALITATIVE DATA (Public version) v1.0

Information for Data Depositors

"We make social science data accessible and reusable."

03.04.2023
Theresa Kernecker

Date	03.04.2023
Version	1.0
License	 This work is licensed under a Creative Commons Attribution 4.0 International License.
Access	Public
Suggested citation	Kernecker, Theresa (2023). Data Deposit Guideline for qualitative data v1.0. Vienna: The Austrian Social Science Data Archive.
Contact	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 info@aussda.at

Data Deposit Guideline

Welcome – Archiving Qualitative Data at AUSSDA

AUSSDA – The Austrian Social Science Data Archive is a data infrastructure for the social science community in Austria. We offer advice on the best strategy for sharing your data and provide solutions for sensitive data as well as non-sensitive data. Specifically, we offer solutions for qualitative data that comply with data protection policies. Furthermore, we provide guidance on preparing the data and documentation material in a way that facilitates and promotes reuse.

This guideline provides an overview of qualitative data archiving at AUSSDA. It expands on Butzlaff (2022) and outlines important aspects to consider when depositing qualitative data and specific steps that depositors should take into account (e.g. anonymisation). It also includes templates and examples for anonymisation plans, anonymisation techniques, and codebook and transcript examples.

Depositing Qualitative Data

Qualitative Data

Qualitative data refers to data based on qualitative methodology and encompasses a diversity of methods and tools rather than a single one. Qualitative data can provide a rich source of research material to be analysed, reworked, and compared to other data, and may consist of many different types of research material. These may include transcribed interviews, written texts, still images, or ethnographic diaries. Qualitative data faces different challenges compared to quantitative data, especially regarding data privacy. The nature of qualitative data lends itself to descriptions of the interviewees, their lives and their surroundings, thereby making them more identifiable in many cases.

Despite these challenges, there is a growing trend in the social sciences to archive and share qualitative data. This trend reflects an increasing number of funding institutions, professional associations and journal editors to make data publicly available for reuse[1]. Proponents of qualitative data reuse contend that qualitative data, similar to quantitative data, can contribute to theory building and avoid unnecessary replication[2]. Critics of qualitative data reuse point to the mismatch between data sharing and reuse and epistemological and ethical challenges of qualitative research[3]. The trend to archive and share qualitative data began in the United Kingdom in the mid-1990s. The UK Data Archive and the Finnish Social Science Data Archive are two of the most cited archives given the vast amount of information, material, and guidelines they offer regarding qualitative data processing.

AUSSDA accepts interview transcripts, textual data, and images for archiving, but we currently do not accept audio-visual data due to privacy concerns. Anonymising audiovisual data is time-consuming and costly and additionally creates serious issues regarding confidentiality. Representing audio-visual data into written form is the most typical way of processing interview and discussion data into an analysable format. At AUSSDA, we accept audiovisual data in this format (as a written transcript). We accept images as long as they do not contain any identifying information. See our Transcription template here.

[1] Feldman and Shaw (2019), Chauvette et. al. (2019).
[2] Arzberger et. al. (2004), DuBois et. al. (2018).
[3] Mauthner and Parry (2013), Slavnic (2013), Guishard (2018).

Workflow from data submission to publication

AUSSDA provides advice on how depositors can prepare their data for archiving in compliance with data protection regulations. During the acquisition process, the depositor will coordinate with AUSSDA in order to determine the nature of the data and develop an anonymisation plan. Following data submission, AUSSDA will check the qualitative data to ensure that the data adheres to the anonymisation plan. Following the data checks, we then provide feedback to depositors. If we find any potentially identifying information or discrepancies between the data and the defined anonymisation plan, we will provide feedback and evaluate the next steps. After data curation, we convert files into long-term and community used formats and store them in the archive (preservation). Eventually, the data is ready for publication in our AUSSDA Dataverse[4] and available for reuse by other researchers. AUSSDA ensures access to data by providing the necessary technical and legal infrastructure so that users can access the data via restricted controlled access. Figure 1 illustrates the AUSSDA workflow.

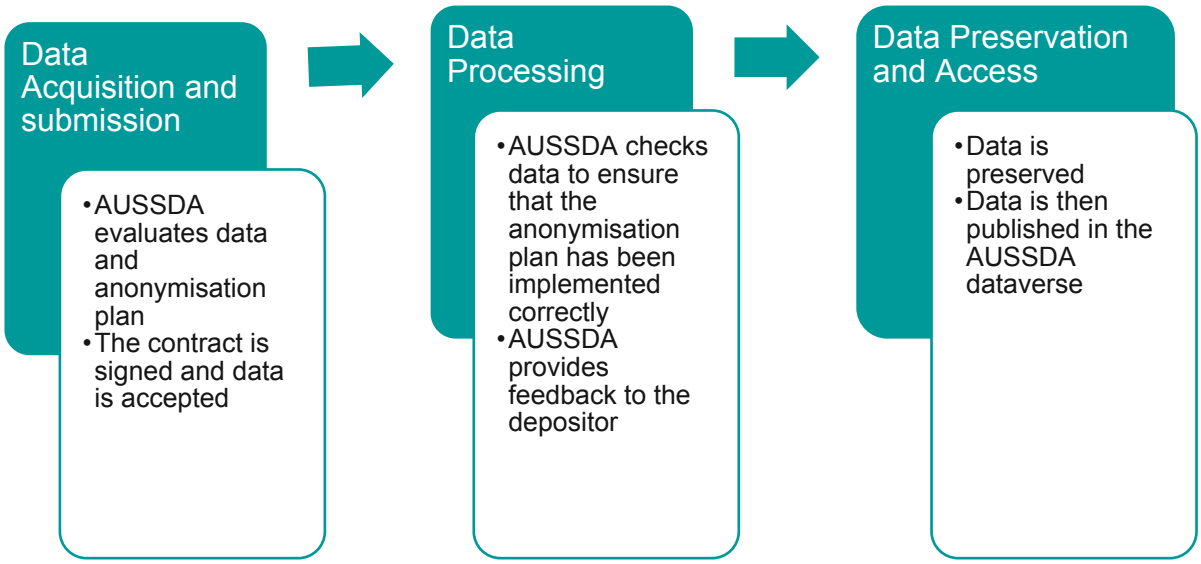


Figure 1 Data depositing process (simplified presentation)

Available licence and access to qualitative data

Given its sensitive nature, the anonymised qualitative data published at AUSSDA is solely available under *restricted controlled access.* The *licence agreement for scientific use (SUF)* is for data that contain personal or other sensitive data, or cover a specifically vulnerable group of research objects. The restricted controlled access additionally requires a login. Users log in to the AUSSDA Dataverse and can then request access to restricted datasets. Data users need to complete and sign a form and prove their legitimate scientific interest in the data. Then, AUSSDA staff verifies the scientific legitimation of the applicants and eventually grants access to the data. After approval and while logged in, users can then download the data.

[4] Find more information on the AUSSDA Dataverse here: https://aussda.at/userguide/.

How to prepare your qualitative data for submission

AUSSDA advises depositors to complete the pre-deposit checklist before beginning the deposit process. This includes specifying how you deal with data anonymisation. A plan documents the issues you face in the context of your data, the changes you will undertake to anonymise them, and the rationale or justification of the latter. Besides providing an anonymisation plan, make sure all submitted materials are saved in preferred formats, include all documentation (e.g. codebooks, questionnaires) and material (e.g. interview transcripts, images) to prepare for a data deposit.

Licence and access for qualitative data

- → The aim of the SUF (scientific use file) licence agreement is the re-use of a dataset by researchers who have a well-defined scientific research purpose.
- → Qualitative data is accessible via restricted controlled access only.

Under all licence agreements and access types, depositors must remove *direct* identifiers from the data before publication. This is a requirement in the EU Data Protection Regulation (GDPR)[5] and the Austrian Data Protection Act.[6] An exception for the scientific reuse of data containing direct identifiers is possible, but only if these direct identifiers have previously been lawfully published and the data stem from publicly available sources (e.g. the first and last name of a politician in a newspaper article). Legal restrictions when depositing data at AUSSDA might further emerge from laws and directives concerning intellectual property rights. Please check our Guideline on Intellectual Property Rights.[7]

Pre-deposit anonymisation plan

AUSSDA places high priority on preserving the confidentiality of participant data. We review all data collections. Protecting the participants' personal data is a multi-layered task summarised below (see Sharing qualitative data). We recommend defining an anonymisation plan prior to collecting the data. First, it is crucial to get permission to share the data from participants via informed consent.[8] Informed consent is the process by which a researcher discloses appropriate information about the research so that a participant makes a voluntary, informed choice to accept or refuse cooperation. Gaining informed consent is crucial to meeting your legal and ethical obligations towards participants while enhancing the value of your data. Another crucial step is risk assessment and understanding which direct or indirect identifiers could potentially lead to identification of your research subjects. Furthermore, in order for you to guarantee the sustainability of your anonymisation plan, it is necessary to define who is responsible for which part of data management and whether extra resources and funding are available to manage data throughout the entire research and depositing process. Researchers should plan ahead as resources for data management are often necessary beyond the scope and length of the project and data collection itself.

---

[5] In Austria: Datenschutzgrundverordnung, DSGVO, according to European Law.
[6] In Austria: Österreichisches Datenschutzgesetz, DSG.
[7] https://aussda.at/aussda-ipr-guideline
[8] See the CESSDA Data Management Expert Guide here for information and examples: https://dmeg.cessda.eu/Data-Management-Expert-Guide/5.-Protect/Informed-consent

How do I define my anonymisation plan? [9]

→ Describe your data
→ Know your data
→ Know the context of your data
→ Understand the disclosure risk
→ Understand the legal and ethical framework and implications
→ Define the steps you will take to anonymise data

What do I need to anonymise?

Two types of data present challenges that could endanger the privacy of participants and render them identifiable: direct and indirect identifiers.

Direct identifiers allow for the re-identification of individuals very easily so their deletion is mandatory: Not only names or telephone numbers render individuals identifiable, but also online identifiers such as IP addresses. A possible exception is when direct identifiers have been lawfully published previously and the data stem from publicly available sources (e.g. the first and last name of a politician in a newspaper article). With regard to indirect identifiers, there can be a likelihood that the combination of "one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person" (GDPR) can identify an individual as well. The extent to which specific indirect identifiers are anonymised depends on the research question and context, and needs to be evaluated on a case-by-case basis. At AUSSDA, we will help you to tackle any doubts regarding your anonymisation plan.

Anonymous versus pseudonymous data

Anonymous data implies that an individual cannot be re-identified with reasonable effort or by combining the data with additional information. Anonymisation is needed to prevent re-identification of individuals, and anonymised data can no longer be associated with an individual in any way. Pseudonymous data implies that individuals cannot be re-identified based on the data without additional information. However, pseudonymous data is not anonymised data. This means that the identifying information could be re-associated with the data in the future. The data are pseudonymous as long as additional identifying information exists, meaning that the individuals in the data are still potentially identifiable. Pseudonymous data become anonymous when the additional identifying information (keys, personal data, and information on techniques to re-identify individuals) is destroyed. As long as this additional data still exists, it is pseudonymous data and therefore subject to the General Data Protection Regulation. According to GDPR, pseudonymous data is still personal data since the process is reversible and individuals are still identifiable with a key. It is worth noting that these terms have specific meanings when applied to the European/UK context. The USA, Canada, and Australia use this terminology different from Europe. Anonymisation (and pseudonymisation) as concepts are understood within a legal framework, so in order to elaborate, it is necessary to look to the GDPR in the European context (Mackey 2019). For an anonymisation template, see Anonymisation plan template. For more

[9] Elliot et. al. (2016)

examples of an anonymisation plan, see our Anonymisation plan example (text) or Anonymisation plan example (table).

The anonymisation process

As part of the pre-deposit process, be sure to remove or modify all identifiers. We recommend facilitating the process by creating an anonymisation plan at the beginning of your project, ideally together with a data management plan. Depositors should document all potential identifiers and determine how these potential identifiers will be handled in the anonymisation process. Specifically, the anonymisation plan should include the type of identifiers found in your data and whether they were removed or categorised, and why this makes sense in the context of your data. Make sure that you allot time and resources for this process (ideally planned in your Data Management Plan). Specific examples of how to deal with different types of identifiers are available (see Anonymisation techniques).

Anonymisation: Best Practice[10]

- → Do not collect identifying data unless it is necessary
- → Communicate to interviewees before the interview that they should avoid expressing confidential information when possible
- → To identify potential confidential information, listen to the entire interview before undergoing the anonymisation process
- → Anonymisation could take place during the interview transcription or be marked for anonymisation at a later point in time
- → Anonymised data should be marked with <> or []
- → Sensitive data marked for subsequent anonymisation could be marked e.g. with $$ or ## in the transcript
- → Create an anonymisation log in which all replacements, aggregations, and deletions are documented, the line and page number of the change, and who performed the process
- → Keep unedited versions of data and anonymisation log for use within your research team only.

[10] Arbor et. al. (2012)

Identifying information and anonymisation techniques

Examples of identifying information

- → **Direct identifiers:** social security number, ID number, full name, e-mail address, phone numbers, vehicle registration number, bank account number, IP address, student ID number, passport or identity card number
- → **Indirect identifiers:** age, gender, education, status in employment, economic activity and occupational status, socio-economic status, household composition, income, marital status, mother tongue, ethnic background, place of work or study and regional variables. Identifiers relating to region of residence include, for example, postal code, neighbourhood, municipality, and major region, postal codes, workplace, employer

How to anonymise data (see also *Anonymisation Techniques for more specific examples*)

- → **Replace personal names** with aliases: replacing proper names with aliases makes it possible for the researcher to retain the internal coherence of the data (instead of replacing them by a letter). Be consistent and track all aliases in a log. Data are not considered anonymous until you delete the original identifiers and original data.
- → **Categorise proper nouns**: Names of people can be replaced with broader categories based on the person's role, e.g. [brother] or [colleague]. This can also be applied to places, e.g. [neighbourhood] or [residential area].
- → **Categorise contextual information**: Detailed background information can be categorised into broader categories, e.g. age 26 can be categorised into a broader group [between age 25-30].
- → **Remove sensitive information**: Sensitive information should be removed or categorised, e.g. [severe illness] instead of cancer.
- → **Remove hidden metadata** from files: Files often contain information about their author or associated geo-location. Make sure these data are removed.
- → **Change values** of data: not recommended – this can change the internal coherence of the data.

Pre-deposit file format checks

AUSSDA prioritises formats that guarantee the long-term availability of the qualitative data and its accompanying documentation. We will make textual data available in our archive in the original format (e.g. .docx, .pdf), and all accompanying documentation in PDF/A format. We accept several image formats, and will evaluate them on a case-by-case basis. Before submitting your data, be sure to remove all (hidden and sensitive) metadata from the files. Such metadata include information such as the photo's geographical coordinates or names of individuals.

List of preferred formats

- → Formats for textual data: PDF, ODT, TXT, XML
- → Formats for accompanying documentation: PDF
- → Formats for images: TIFF, evaluated case-by-case
- → Computer Assisted Qualitative Data Analysis: REFI-QDA Project (QDPX), REFI-QDA Codebook (.qdc)

File naming

We will rename all files according to our file-naming scheme. However, it is crucial upon submission that you name your files in an organised and understandable way so that we can clearly identify the content of the file based on the name.

Pre-deposit metadata checks

Metadata are data that provide information about the data itself. Metadata are required for submission, as these data are necessary for the publishing process. Metadata include information about funding, the principal investigators, the scope of the study, and methodology. Metadata will help make the data findable, accessible, interoperable, and reusable (FAIR). Making data FAIR can contribute to the visibility of researchers' work, increase the likelihood for citations and make the data easier to access and understand. The main elements to be included in the metadata file are detailed below:

Typical metadata elements

- → Principal investigator(s) / authors
- → Year
- → Publisher
- → Title
- → Funding sources and grant number
- → Data collector/producer and contributors
- → Project description
- → Sample and sampling procedures
- → Substantive, temporal and geographic coverage of the data collection
- → Units of observation
- → Related publications/datasets
- → Technical information of files
- → Data collection instruments

Deposit your data

Preparations

The submission of a comprehensive documentation along with the data is crucial for the reusability of data. The more documentation material made available, the easier it will be for researchers to reuse the data. The following list provides an overview of which files we consider mandatory, recommended and optional documentation material. In the List of preferred formats, we provide information about which data formats we prefer for submission to AUSSDA. This guideline ensures that we can preserve your data and guarantee accessibility for re-use in the future. If you have any questions concerning conversion or suitability of your data, do not hesitate to contact us! We update this list on a regular basis to account for software changes or disciplinary-specific format changes.

Gather and submit documentation

Mandatory documents for publication:

- Licence agreement (signed by depositor and AUSSDA)
- Data files
- Instruments of data collection (e.g. questionnaire with interviewer instructions, information material for respondents, data collection guideline)
- Metadata sheet
- Anonymisation plan

Recommended documents for publication:

- Codebook
- Method report
- Informed consent template (in case the data processing is based on informed consent)

Optional documents for publication:

- Project report
- Data Management Plan (DMP) of project
- Guidelines / instructions for interviewers
- Documentation about incentives, contact procedures
- Data recoding protocol or scripts
- Any further document that helps users to understand the data

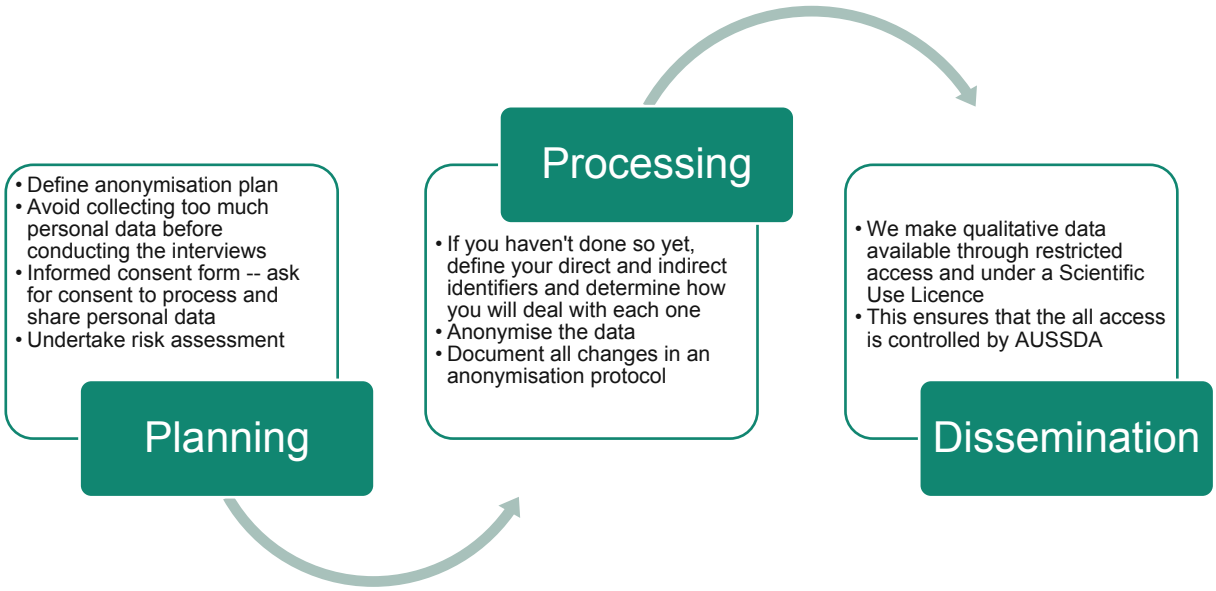
How to submit the data

- Assemble all archive materials (data and accompanying documentation)
- Sign deposit contract with AUSSDA
- Compare documentation materials with data and check for consistency
- Check that all files are available in preferred formats or contact us on other file types
- Check on spelling errors, correct labels
- Document your anonymisation procedure
- Remove all sensitive metadata from files
- Remove restrictions from all documents
- Transfer archive material to AUSSDA

Please make sure that you send us only the anonymous/pseudonymous version of your data. All files should be transferred to AUSSDA in a secure way (e.g. by the Aconet Filesender[11]). As we process all data and related documents (e.g. for data checks and the migration into long-term archive formats), all write protection must be removed before submission. Please do not hesitate to contact us at info@aussda.at, if you have questions or requirements not covered in this document – we will assist you in finding a solution.

[11] https://filesender.aco.net.

Sharing qualitative data



Anonymisation techniques

Technique	Example
Use aliases instead of proper nouns*	Use "Thomas" instead of "Paul"
Categorise proper nouns into a broader category or role	Instead of "Pamela" use [friend], [sister], [colleague].
Remove or replace sensitive information	When referring specific sensitive information, modify the text. For example, a violent interaction could be replaced with [traumatic incident] or a specific illness can be replaced with [severe illness]
Categorise contextual information	Instead of a specific age, replace with an age group and broader categories such as [Age 19-23]. Instead of "Ottakring" use [District with over 100.000 population].
Change the value of identifiers (not recommended)	Changing a specific date, age, or gender of interviewees. We do not recommend this as it may affect interpretation of the data.
Remove hidden metadata from files	Delete location or owner of a device (e.g. files often contain the name of the document owner and photos can also contain the location at which the photo was shot)

* Aliases should be similar to the original name as in the example of "Thomas" and "Paul", i.e. German names should be replaced with German names in order to preserve the original content/context. Similarly, an old name such as "Manfred" should not be replaced with a newer name (e.g. Kevin).

Examples of (in)direct identifiers

Please note that the extent to which the indirect identifiers need to be replaced or categorised depends on the research subject/question and context. This issue has to be addressed and defined in the anonymisation plan on a case-by-case basis.

Identifier	Type of identifier	Example anonymisation technique	Example
Full name	direct	remove	N/A since these are removed
Social security number	direct	remove	N/A since these are removed
E-mail address	direct	remove	N/A since these are removed
Phone number	direct	remove	N/A since these are removed
Vehicle registration number	direct	remove	N/A since these are removed
Passport or identity card number	direct	remove	N/A since these are removed
Student ID number	direct	remove	N/A since these are removed
Bank account number	direct	remove	N/A since these are removed
IP address	direct	remove	N/A since these are removed
Photograph of person	direct	remove	N/A since these are removed
Need for social welfare	indirect	remove	N/A since these are removed
Video file displaying person	direct	transcribe and anonymise	N/A since these are removed
Workplace/employer	indirect	replace/categorise	[Name of Employer] instead of *Universität Wien*
Health-related information	direct/indirect	remove/replace/categorise	indirect example: [Severe illness] instead of *cancer* direct example : fingerprints (remove)
Ethnic group	indirect	replace/categorise	[Name of ethnic group] instead of *Ethnic Slovene*
Position of trust or membership	indirect	replace/categorise	[Position at organization] instead of *Office manager*
Zip code	indirect	replace/categorise	[Zip code in Vienna] instead of *1010*

District	indirect	replace/categorise	[District in Vienna] instead of *Ottakring*
Municipality	indirect	replace/categorise	[Municipality in Styria] instead of *Mariazell*
Region	indirect	replace/categorise	[Western Austria] instead of *Tirol*
Age	indirect	replace/categorise	[Age between 21-25] instead of *22*
Gender	indirect	replace	[Gender] instead of *male*
Marital status	indirect	replace	[Marital status] instead of *married*
Household composition	indirect	categorise	[between 3-6 family members] instead of *5*
Occupation	indirect	categorise	[Administrative manager] or [occupation] instead of sales manager
Industry of employment	indirect	categorise	[Industry of employment] instead of *Energy and utilities*
Employment status	indirect	categorise	[employment status] instead of *unemployed*
Income	indirect	categorise	[500-1000 Euros] instead of *675 Euros*
Education	indirect	categorise	[technical and vocational] instead of *carpentry apprenticeship*
Field of education	indirect	categorise	[field of education] instead of *formal education*
Mother tongue	indirect	categorise	[mother tongue] instead of *German*
Nationality	indirect	categorise	[Central European] instead of Austrian
Crime or punishment	indirect	remove/categorise	[criminal incident] instead of *robbery*
Membership in a trade union	indirect	categorise	[membership of an organization] instead of *works council*
Political or religious allegiance	indirect	replace	[religious affiliation] instead of *Catholic*
Social welfare services and benefits received	indirect	remove/categorise	[Program A] instead of specific program

Anonymisation plan template

Anonymisation Plan for Project [Name of Research Project]

An anonymisation plan provides a description of your project and lays out a plan regarding how sensitive information will be dealt with to avoid the identification of participants.

Research Project Description:

Describe your research project and provide contextual information about the project. What kind of data are we dealing with? State any potential sensitive data that will need to be anonymised (please describe each individual identifier in greater detail below and provide a justification of the technique you chose). Be sure to include all direct and indirect identifiers.

Anonymisation:

Which identifiers exist in your data? Describe how you will ensure the anonymisation of all research subjects. Which anonymisation techniques will be used for which kinds of identifiers? Provide an example of how you will deal with sensitive information in your data and list how you will deal with each one.

Variable A

Here, describe how Variable A will be replaced by [...] or categorised into a more general category (or other anonymisation techniques) as such: [....] and provide a justification.

Justification: Variable A contains sensitive information and could lead to the identification of individuals, especially in combination with Variables B and C.

Variable B

Here, describe how Variable B will be replaced by [...] or categorised into a more general category (or other anonymisation techniques) as such.

Justification: In combination with Variables A and C, this information may lead to re-identification.

Variable C

Describe how Variable C will be removed: [...] and provide a justification.

Justification: Variable C is a direct identifier and will identify individuals.

Anonymisation plan example (text)

Example 1: Anonymisation plan as text

Research Project Description:

Example: This project deals with data reuse amongst researchers. In the interviews, researchers discuss their research and the data they work with. Then, based on the FAIR principles, we asked questions about where they look for data in their field, how they access it, and some of the challenges they face.

We use various anonymisation strategies to ensure anonymisation of interview participants. All anonymisations will be marked in brackets as such: *[anonymised data].* For example, we exchange the name of the interviewers and interviewees with [I] and [P] (short for participant).

Names

The transcripts of the interviews do not contain names. For consistency, [P] (participant) is used for interviewed persons and [I] (interviewer) for interviewing persons. The abbreviation [P] is numbered according to the order of the interviews, so [P] from interview 1 is called [P1], [P] from interview 2 is called [P2], etc.

Justification: N/A (because there are no names in the transcripts)

Location information

We retain information on the country in which an interviewee works, since this information is of interest for the reuse of the data. Any location information beyond this (e.g., the name of a city) is replaced by pre-defined categories: large city: at least 100,000 inhabitants, medium-sized city (20,000 to less than 100,000 inhabitants), and small town (5,000 to under 20,000 inhabitants).

Justification: In combination with discipline or type of employment, this information may lead to re-identification.

Institution

We replace the institution at which the interviewee works by the indication of what type of institution it is: university, non-university research institution, governmental institutions at the federal, state, or local level with a research mandate, company with in-house research. We supplement this information by the number of staff employed at the institution. Instead of providing an exact number, we provide categories: 100 employees, 101-500 employees, 500-1000 employees, > 1000 employees. We retain the country in which the institution is located because this information is of interest for data reuse.

Justification: Specifying the institution is sensitive because academic careers are publicly viewable (CVs are often available online). Thus, participants could be identified. At the same time, it is important in terms of content to know what type and size of institution is given, as this could have an impact on the infrastructure and support services available.

Research projects

When an interviewee talks about a research project or endeavor in which he or she is or has been involved, the name of that endeavor or project is always replaced by the following pseudonym: [name of project] or [name of endeavor]. Since information about projects may be relevant for the follow-up use of the interviews - we categorised the latter as e.g. research project or project at institutional, national, European level.

Justification: The combination of research project/project and indication of country could be used as specific contextual information for re-identification.

Anonymisation plan example (table)

Example 2: Anonymisation plan as a table

Research Project Description:

This project deals with data reuse amongst researchers. In the interviews, researchers discuss their research and the data they work with. Then, based on the FAIR principles, we asked questions about where they look for data in their field, how they access it, and some of the challenges they face.

Anonymisation:

We use various anonymisation techniques to ensure anonymisation of interview participants. All anonymisations will be marked as such: *[anonymised data].* For example, we exchange the name of the interviewers and interviewees with [I] and [P].

Example Identifier	Type of identifier	Change?	Example replacements	Justification
Researcher names	direct	yes	[P1], [P2], etc.	Names are direct identifiers and are therefore removed and replaced with P1, P2, etc.
Nuclear Physics	indirect	no	[Discipline]	We do not categorise or change the discipline given that this is important information for data reuse and will not allow identification of participants together with other available information.
University of Vienna	indirect	yes	[Researcher's University]	The university used in combination with the discipline could lead to identification. We changed the name of the researchers' universities to *[Name of university].*
Austria	indirect	no	[Country in Central Europe]	No change since the country will provide important contextual information and cannot be used to identify interviewees.

Transcription template[12]

Study name: Interviews with Researchers | Interview ID: P8

Depositor: Name of Depositor | Date of interview: 20.01.2021

Information about interviewee: Researcher in the field of Nuclear Physics in Austria

P8 = Respondent/Interviewee

I = Interviewer

[I]: My first question is: Please briefly describe in which scientific discipline you work and what your main research interests are.

[P8]: Right. So, I work as an analyst at the [*university institution with more than 1,000 staff members*]. A large part of my job is research, but also project management and administrative activities. I also teach a course on Nuclear Physics at *[university institution with more than 4,000 staff members*] in *[Name of large city with at least 100,000 inhabitants]*. So my work is mainly, but not exclusively academic, I would say. At *[Name of institution with more than 10,000 staff members]* we are mainly researching new options for nuclear energy. Therefore, we use data that other people generate, but we also produce data that other people reuse.

[I]: Okay. So, what kind of data do you work with?

[P8]: As most other nuclear physicists do, we mainly use data on particle and nuclei interaction probabilities, emitted particle energy spectra, angular distributions, fission product yield probabilities, decay properties and more. We also work with many other kinds of data, for example data on nuclear power reactors in the world. My research team led by *[Name of Principal Investigator]* also uses data on nuclear reaction data that describes cross sections for fundamental collision processes, for example between a neutron and a nucleus.

[I]: Okay. So we are talking about numbers here if I understand correctly, right?

[P8]: No. It's just raw data of numbers stored in some kind of binary format to reduce the size. We also data in different formats, but they are also mostly just numbers representing other quantities that we use. Basically, it is just numerical data.

[I]: Okay. Thank you. Now I would like to ask you several questions about this process of finding, accessing and reusing the data and my first question is on the findability of data. So please think about a current or past research project: How do you search for data or how did you search for data?

[P8]: For this kind of data, there are several well-known databases and repositories. My colleague *[Name of colleague]* works at one of these repositories. They are based in France at *[Name of Institution].* There are other cases where I have looked for data that is not as accessible. In those cases, I mostly use Google. Google makes it easy to find data, as do some of the other repositories or institutions that provide the data. Obviously, the International Atomic Energy Agency serves as a kind of portal with links to the most important data in our field. For the current project we are working on *[Name of project],* we use this portal quite often.

[12] Note: This is a fictional example.

Codebook example

Codebooks contain information on how the interviews were categorized for analysis. Codes can be documented in several different ways (e.g. exported from software, listed in a text-processing document or excel, as a table). This is one example of what a coding scheme can look like.

Code	Description	Example
Findability	We code the interview responses with "findability" when respondents refer to how they go about searching for data or the specific search engines they use, or how easy or difficult it is to locate specific kinds of data/datasets.	"I usually resort to the discipline-specific data repository for nuclear physics when searching for data. If I don't find what I am looking for there, I use search engines such as Google."
Accessibility	We code the interview responses with "accessibility" when respondents refer to how easily accessible data is.	"It is difficult to access the data I need because of legal issues or because of the different legal frameworks across countries".
Interoperability	We code responses with "interoperability" when respondents refer to the possibility of linking data to other sources	"We all use similar codes, so that makes it easier when we merge one dataset with another. However, sometimes it is difficult to use other datasets when the data and specific variables are not properly described".
Reusability	We code responses with "reusability" when respondents refer to data reuse.	"We always submit our data to our discipline-specific repository as soon as we can - in our field, it is important for us to make the data available for others to reuse."

References

Arbor, A., C. Colyer, and D. Donakowski (2012). *Guide to Social Science Data Preparation and Archiving–Best Practice Through the Data Life Cycle.* (6th ed). University of Michigan.

Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., Laaksonen, L. & Wouters, P. (2004). Promoting access to public research data for scientific, economic, and social development. *Data Science Journal, 3*, 135-152.

Butzlaff, Iris (2022). https://aussda.at/fileadmin/user_upload/p_aussda/Documents/Data-Deposit-Guideline_SUF_v2_0.pdf

Chauvette, A., Schick-Makaroff, K., & Molzahn, A.E. (2019). Open data in qualitative research. *International Journal of Qualitative Methods, 18*, 1-6.

DuBois, J.M., Strait, M., & Walsh, H. (2018). Is it time to share qualitative research data? *Qualitative Psychology, 5*(3), 380.

CESSDA Data Management Expert Guide (n.d.) *Consortium of European Social Science Data Archives.* https://dmeg.cessda.eu/Data-Management-Expert-Guide/5.-Protect/Informed-consent

Feldman, S., & Shaw, L. (2019). The epistemological and ethical challenges of archiving and sharing qualitative data. *American Behavioral Scientist, 63*(6), 699-721.

FSSDA Data Deposit Guidelines (n.d.) *Finnish Social Science Data Archive.* https://www.fsd.tuni.fi/en/services/data-management-guidelines/processing-qualitative-data-files/#data-files

Guishard, M.A. (2018). Now's not the time! Qualitative data repositories on tricky ground: Comment on Dubois et al. *Qualitative Psychology 5*(3), 402-408.

Mackey, E. (2020). A best practice approach to anonymization. *Handbook of Research Ethics and Scientific Integrity*, 323-343. https://doi.org/10.1007/978-3-319-76040-7_14-1

Mauthner, N.S., & Parry, O. (2013). Open access digital data sharing: Principles, policies and practices. *Social Epistemology, 27*(1), 47-67.

UK Data Archive (n.d.) https://ukdataservice.ac.uk/learning-hub/qualitative-data/