



GUIDEBOOK

Making Qualitative Data Reusable

Attribution

This guidebook was produced as part of the [CaRe & DaRe project](#), funded by the NWO Open Science fund under file number 203.001.153 and led by [prof. dr. ir. J. J. Berends](#) (VU)

This guide has been written by

Ricarda Braukmann Ph.D. - Data Station Manager Social Sciences, DANS

ricarda.braukmann@dans.knaw.nl

 [orcid: 0000-0001-9408-3190](https://orcid.org/0000-0001-9408-3190)

Drs. Widia Mahabier - Data Manager, DANS

widia.mahabier@dans.knaw.nl

 [orcid: 0009-0006-0427-9282](https://orcid.org/0009-0006-0427-9282)

Maaïke Verburg M.Sc. - Research Data Management specialist, DANS

maaike.verburg@dans.knaw.nl

 [orcid: 0000-0001-9408-3190](https://orcid.org/0000-0001-9408-3190)

The authors would like to thank the members of the Data Stewards Interest Group for their input for the guide and Hans Berends (VU), Kacana Khadjavi Pour (VU), Eric Haynes (VU) and Jetze Touber (DANS) for their constructive feedback on earlier versions of this guide.

July 2023

Guidebook DOI: [10.5281/zenodo.7777519](https://doi.org/10.5281/zenodo.7777519)

Decision tree DOI: [10.5281/zenodo.7777549](https://doi.org/10.5281/zenodo.7777549)



Table of contents

| | |
|--|-----------|
| About this guidebook | 4 |
| Definition of qualitative data | 4 |
| The importance of making qualitative data reusable | 5 |
| The challenges in making qualitative data reusable | 5 |
| Goals of this guidebook | 6 |
| Planning your project | 7 |
| Preparing informed consent | 8 |
| Organising and documenting qualitative data | 9 |
| Processing your data | 10 |
| De-identification of personal data | 11 |
| Archiving and Publishing qualitative data | 12 |
| Documentation and metadata | 12 |
| File formats | 13 |
| Decision tree for data reuse | 14 |
| Archive open access | 16 |
| Archiving with restricted access | 16 |
| Secure environment | 17 |
| The CaRe & DaRe solution: Decentralised reanalysis | 17 |
| Publish metadata only | 18 |
| Final thoughts | 20 |
| Additional resources | 21 |
| References | 22 |



About this guidebook

This document aims to give an overview of the challenges associated with making qualitative data reusable as well as providing guidance on how reusability can be improved and addressed at all stages of the research data life cycle. The guide includes a [decision tree](#) that researchers and data stewards can use to evaluate the options for making qualitative data reusable that are most suited for their projects.

This guide was created by DANS ([Data Archiving and Networked Services](#)), the Dutch national centre of expertise and repository for research data, in the context of the "[Case-study Research and Data Reuse](#)" (**CaRe & DaRe**) project which focuses on increasing and better facilitating the reuse of qualitative data in science.

Although likely applicable to a broader audience, the main target audience of this document is the **Dutch research community**, in particular within the domain of **social sciences and economics**, working with qualitative data. This guidebook has been informed in part by a small survey that was conducted amongst a group of Dutch data stewards (Verburg et al., 2023).

Definition of qualitative data

Qualitative data refers to information that is not gathered in numerical form, but rather describes qualities or characteristics. Data are expressed in natural language and often in textual or visual form. Collection methods for instance include photography, audio recordings, video, interviews, diary accounts or other unstructured observations (CESSDA, 2022).

Qualitative data is thus quite **heterogeneous** and can take various forms. The types of qualitative data most often encountered by the Dutch data steward survey respondents were interviews, video and audio recordings, transcripts, fieldnotes, focus groups, and observations (Verburg et al., 2023).

Qualitative data often contains **personal** and/or **sensitive** information. Personal information is defined as any information relating to an identified or identifiable natural person. Sensitive information can either be sensitive personal information (e.g. sexual orientation, religious beliefs, health data), but other information such as for instance business strategy data or financial data can also be sensitive.



The importance of making qualitative data reusable

The **Open Science** and **FAIR** (Findable, Accessible, Interoperable, and Reusable, Wilkinson et al., 2016) data movements call for research to be well-documented and available for reuse by others. It is generally proposed that data should be “**as open as possible, as closed as necessary**”, making reuse possible while protecting personal information and intellectual property rights where needed (Landi et al., 2020). In line with these movements, many funders, including the Dutch NWO, require data to be open and ask researchers to outline their data storage and archiving plans as part of a **Data Management Plan** (DMP, see [NWO information page](#)).

Making data available for reuse allows new research questions to be asked and reduces the cost of collecting data. Making data available has also been shown to enhance the **visibility** of existing research (e.g. Colavizza et al., 2020; Piwowar et al., 2007) and allows for research to be **reproducible**, increasing **transparency** of the scientific process (Munafò et al., 2017).

The collection of qualitative data, in particular, is often quite time- and resource intensive, arguably making it even more important that already available data of this kind can be reused. This can reduce the burden of participants in qualitative studies who might be contacted by multiple researchers for similar study goals. In addition, qualitative data is very rich in nature which allows for a multitude of potential research questions that can be answered when this data is shared for reuse.

The challenges in making qualitative data reusable

Qualitative data sharing is often more **complex** or even impossible to facilitate compared to quantitative data. One important reason is that anonymisation of the data can often lead to information loss rendering the data useless for reuse. Sharing non-anonymised data can be difficult or impossible due to **legal** constraints (e.g. GDPR, copyright) and when information is particularly **sensitive** (e.g. sensitive personal data, sensitive business information).

Another aspect that makes the sharing of qualitative data challenging is that it can be hard to capture all the **contextual information** that is essential for the interpretation of qualitative data in documentation or metadata. It can seem difficult to ensure that the data is interpreted in the right way given its variety and richness.

While qualitative data sharing is thus more complex, guidance on **Research Data Management** (RDM) around qualitative data is often lacking (Verburg et al., 2023) and qualitative data is easily overlooked in the developments of Open Science and FAIR data.





It is important to note that making qualitative data available for reuse will not always be possible due to these constraints. Yet, qualitative researchers should carefully evaluate their projects and processes to maximise the reusability of their data. While some qualitative data cannot be made available to other researchers at all, some can be made available openly after de-identification, and some can be made available with restrictions or through alternative analysis processes such as proposed in the CaRe & DaRe project. In addition, documentation and information about the study can likely be made available in all cases, providing fellow researchers information about existing data even if they cannot access the data itself.

The different possibilities and what to consider when evaluating the possibilities for making qualitative data reusable will be further outlined in the next sections of this guidebook.

Goals of this guidebook

The goal of this guide is to give an **overview** of the steps that researchers and data stewards can take in improving the reusability of qualitative data.

Taking into consideration the challenges of making qualitative data available for reuse as mentioned above, we go through each step of the research process and outline how reusability can be considered from the start. In this way, a project can be designed to maximise the potential for reuse.

This guidebook provides **advice** on what to consider with regards to reusability. It also lists existing **open tools** that can be used to process qualitative data. We consider aspects like gathering informed consent, options for de-identification of data, and provide an overview of the different options to archive and publish (parts of the) qualitative data. Lastly, we provide **examples of datasets** available for reuse and alternative reuse practices particularly designed for sensitive data.

We hope that this guide can become a starting point for further discussions and improved guidance for making qualitative data reusable.

Planning your project

The first step in the research data life cycle is the planning of the project. In this phase, researchers are often asked to create a Data Management Plan (DMP) that contains all information about the study that should be collected and described. This is a great moment to already consider what kind of data you will be collecting. The [CESSDA Data Management Expert Guide](#) walks you through this phase of research in more detail and also provides a [checklist](#) with questions to consider. A DMP [checklist tailored to qualitative data](#) is also available from the Qualitative Data Repository.

Specific questions you want to address at this stage which can help you to assess whether and how your data can be made reusable are:

Common view on data sharing amongst collaborators:

- Which other parties are involved in my project?
- How do they feel about making the data reusable?
- Will they be supportive in efforts to improve the reusability of the data?
- Who owns the data? Who would I need to consult before making data available for reuse?
- How can I guarantee a permanent point of contact for the dataset, in particular if I restrict access?

Collecting personal data:

- Does my data contain personal information?
- Is collecting this data in line with the ethical and legal (e.g. GDPR) requirements?
- Can I de-identify the personal data? How would this affect the usefulness of the data?
- Is de-identification needed for my study?
- Do I have permission / consent to share personal data without de-identification?
- Can I minimise the amount of personal data I will collect in any way?
- Can I keep the personal information separate from information that can be shared openly?
- How do the subjects of my research feel about making data reusable?
- Can I include reusability of data in my informed consent forms?

Collecting sensitive data:

- Is the data I am collecting sensitive in any way?
- Can I keep the sensitive information separate from information that could be shared openly?
- Does this level of sensitivity in my data prohibit all types of data sharing, or could I consider sharing under specific circumstances (e.g., restricted access, a secure environment)?
- What additional measures will have to be undertaken to manage the data in line with institutional policies (e.g. encryption, anonymisation, pseudonymisation)

Using open formats and open tools:

- In what formats will I collect my data? Can I use open formats so that it can be read into different analysis programmes?
- What tools will I be using to process my data? Can I use open tools so others can perform the same analyses?
- If you are using specific processing tools for personal and/or sensitive data, do Data Processing Agreements need to be signed?

Preparing informed consent

Informed consent is one of the founding principles of research ethics. Its intent is that participants can enter research freely (voluntarily) with full information about what it means for them to take part, and that they give consent before they enter the research. An important aspect is also to inform participants about their legal rights as specified in the GDPR. Detailed information on the conditions for Informed consent can be found in [Article 7](#) of GDPR. You can also find the [requirements](#) the consent must meet according to **GDPR** rules on the GDPR website.

Consent should be obtained **before** the participant or subject enters the research (prospectively), and there must be no undue **influence** on participants to consent. The minimum requirements for consent to be considered informed are that the participant **must** understand what the research is and **what** they are consenting to (see for example [this guide](#)).

Consent must not only be given for **participation in the study**, but also for the **archiving and sharing of data**. Therefore, it is important to reflect upfront if and how you want to make qualitative data available for reuse and include this in your informed consent form.

Likely your own institution will have guidance for informed consent practices and examples of informed consent forms as well. It is advised to discuss this with your data stewards who can also advise you on the options to include data sharing in your informed consent form.



Organising and documenting qualitative data

When organising and documenting your collected data, you can consider the future reusability of your data as well.

In this process of the research, you want to provide enough **contextual information** with your data so that you and others can still understand how the data is organised and how it was collected, processed and presented at a later point.

The [CESSDA Data Management Expert Guide](#) gives guidance on good practices regarding **file names** and **folder structures** when you are working with your data, and provides examples of documenting qualitative data such as textual data files, audiovisual data, and periodicals or magazines.

When naming your files, it is generally recommended to avoid having personal data (e.g. the name of the person you interviewed) in a data file or folder name. In general, it is wise to minimise the amount of personal data you use and to only keep personal information if it adds to your research question or if it is necessary to understand the research. The latter can for instance be the case if you interviewed celebrities or artists about their work.

If possible, store the parts of your data that contain personal or sensitive information separately from parts of the data that do not. It should be noted that encryption may be recommended for sensitive or confidential qualitative data and specific institutional platforms may be required. Your local data steward can advise you on this. In any case, separating parts of your data that contain personal or sensitive information from other data, can make it easier for you to share (parts of) your data for reuse later. If you consider this distinction from the start of the project, archiving the data will be a lot easier later on.

Processing your data

While this guidebook is not meant as a guide towards processing qualitative data, we do want to encourage researchers and data stewards to consider the softwares and tools that are being used to process the data and whether **free and open alternatives** are available. Using open tools makes it easier for research to be reproduced as other researchers can use the open tools as well.

Adhering to **community standards** is also recommended to increase interoperability and reusability within your community. However, be mindful that programmes used in a certain community may sometimes not be open and could therefore form a barrier for researchers outside of your community.

Some tools and pointers that are known to the authors at the moment of publication are discussed below, but we very much encourage readers of the guide to **contact us**¹ in case you know of other open tools that are used to process qualitative data which we could include here as well.

Tools for transcription:

- [CLARIAH](#), in particular the Centre for Language and Speech Technology, is working on various tools around automatic speech recognition (e.g. for Dutch and English).
- Speech recognition tools are available in Python (e.g. [this guide](#)).
- [oTranscribe](#) is a free web app to transcribe recorded interviews.
- Some community standard tools which are not open, but do include features for transcription are Nvivo, QDA Miner, Otter.ai, amberscript and transcriptonline.

Tools for annotation:

- [ELAN](#) is a free annotation tool for audio and video recordings developed by The Language Archive from the Max Planck Institute for Psycholinguistics.
- [Taguette](#) is a free and open-source tool for qualitative research in which you can import your research materials and highlight and tag quotes.
- [Hypothes.is](#) is an open annotation tool. Together with the Qualitative Data Repository the [Annotation for Transparent Inquiry \(ATI\)](#) was developed to increase transparency in annotations of qualitative data.
- Some community standard tools that are not open, but also include features for annotation are Atlas.ti, Nvivo and QDA Miner.

¹ Please send a message to info@dans.know.nl mentioning "Contribution to guidebook qualitative data"

De-identification of personal data

When working with **personal data**, there are some extra steps you need to take when processing your data, in particular when you want to archive and share the data. There are ethical considerations and legal regulations, such as the [General Data Protection Regulation \(GDPR\)](#) to take into account. It is generally recommended to **consult the privacy officer** of your institutions when you are processing personal or sensitive data.

As has been explained in a previous section, making sure your **informed consent** adequately includes data storage, archiving and sharing will ensure you have a legal base for processing personal data. You can ask participants to consent to the sharing of identifiable personal data. This is recommended when you know in advance it will be difficult to de-identify the data. This can for instance be the case when you are working with video data, or when you are working with data from well-known organisations or individuals that can easily be identified from simple descriptions.

To protect the privacy of your participants, de-identification of the data is an important tool. You can de-identify study participants by **anonymising**² or **pseudonymising**³ the data. When working with qualitative data, however, de-identification is often not possible or desirable as it can result in too much information loss. It is important to find the **balance** between privacy and understandability of the data. If your data cannot be de-identified, you can still share it in some cases. You may want to restrict the access to the data to protect the participants or you may be able to make it reusable through decentralised re-analysis as in CaRe & DaRe as outlined later.

Some useful guides on de-identifying qualitative data have been created by [Erasmus University Rotterdam](#), [UK Data Service](#), and the [Qualitative Data Repository](#). Utrecht university is working on a [Data Privacy Handbook](#) which also covers de-identification of data. All guides recommend to develop a de-identification strategy at the start of your research.

Tools for anonymising qualitative data

- [anonymoUUS](#) is a python package created at Utrecht University to pseudonymise folders and files in your documentation.
- UKDS has a [text anonymisation helper tool](#) that highlights terms in your text that may be disclosive, so you can decide to alter them accordingly.
- [FAMTAFOS](#) is the Free Automated Multi-language Text Anonymisation For Open Science. The tool replaces disclosive information in a way that retains semantics of the data.
- Some community standard tools that are not open, but also include features for anonymisation are Atlas.ti and Nvivo.

² We consider data to be anonymised if the process of removing the identifying characteristics is irreversible, i.e. it is impossible to re-identify the participants / revert to a state in which participants can be identified again. If you are able to fully anonymise your data, it is no longer considered personal data.

³ Whenever a key is available that translates the de-identified data back to the original data, we talk about pseudonymised data. Pseudonymised data is still considered personal data under the GDPR. It can still be useful to perform pseudonymisation to protect the privacy of participants.

Archiving and Publishing qualitative data

After completion of your research project, it is important to **archive** your data to ensure the data is preserved for the long term. If possible, you can also **publish** your data to make it available for reuse. Given the value of qualitative data and the challenges of sharing and preserving it properly, it is important that the data is archived in a domain-specific **trustworthy repository**⁴, like the [DANS Data Station Social Sciences and Humanities](#) (SSH), that will **curate** and **preserve** the data for the long term (Mannheimer et al., 2018). Alternatively, your own institution may have a recommended data repository, for instance [DataverseNL](#). Your local data steward will be able to provide you with more information.

Documentation and metadata

For qualitative data, documentation and metadata detailing the context of the research are essential for fostering understanding and potential reuse. As mentioned before, it can be quite difficult to successfully capture all the needed information.

Many repositories provide **guidance** on what documentation is required and recommended for particular scientific disciplines or research types:

- DANS has developed a template for the metadata that should accompany a transcription of (audiovisual) interview data. The template is listed in the DANS guidance on [preparing data for deposit](#) which also provides information about the required metadata and documentation in general.
- The CESSDA Data Management Expert Guide has a section dedicated to [Documentation and Metadata](#), which contains guidance on project-level and data-level documentation for qualitative data.
- Though the Finnish Social Science Data Archive generally doesn't archive all kinds of qualitative data, it does provide [guidance](#) on organising and documenting background information with useful examples.

When depositing data in a repository, providing as much **metadata** as possible is essential to ensure the data is findable. This will help to inform people about the context of the research.

The use of a **metadata standard** helps to document data in a structured and interoperable way. For the social sciences, the Data Documentation Initiative codebook, [DDI 2.5](#), is most often used. This standard, though created for quantitative data, is also applicable to qualitative and mixed methods research (Mannheimer et al., 2018).

⁴ We refer to a repository as trustworthy if it is certified by an internationally accepted standard that evaluated the processes of that repository. A well established standard is the [CoreTrustSeal certification](#).



Repositories may use several metadata standards tailored to specific communities and scientific disciplines. Repositories, like the DANS Data Station SSH, that are part of the Consortium of European Social Science Data Archives (CESSDA) follow the [CESSDA Metadata Model \(CMM\)](#) to describe their data holdings. Some other examples:

- The [UK Data Service QualiBank](#) uses DDI 2.5, the Qualitative Data Exchange schema (QuDEX), and Text Encoding Initiative schema (TEI) for their data
- The [Qualitative Data Repository](#) uses DDI 2.5 and the DataCite Metadata Kernel 3.1

You do not need to know the ins and outs of such metadata schemas. As long as you deposit in a trustworthy repository that specialises in your domain, the curators will assist you with providing the right metadata for your dataset.

File formats

To facilitate reuse, data should be archived in **file formats** that are easy for other people to interact with. What this comes down to is that there is always a preference for file formats that are **open**, as they are sustainable in the long run and they do not require someone to install or buy a specific software to access, edit, or otherwise process the data.

Sometimes, **community standards** for file formats exist that are not open formats (e.g., because a certain software is so commonly used in that community, or because there is no open alternative available), in which case you can deposit in that format instead. However, do keep in mind that potential reusers of your data may come from very different communities and that they will experience a barrier to access your data. If you do use special software it will be helpful for future users of your data if the software is described including its version.

Some repositories provide lists of **preferred file formats** for their users, which take into account community standards and open formats. Depositing in these formats will ensure your data can be preserved for the long term, as the repository takes responsibility to convert the files as technology advances.

- DANS curates a [list of preferred file formats](#) for the Data Stations with explanations on the reasoning behind these choices. This includes formats for texts, images, audio, video, qualitative data analysis (CAQDAS) and many more.
- DANS and other repositories may transform data for you into preferred formats so that long term accessibility can be ensured.





Decision tree for data reuse

The knowledge that making qualitative data reusable is valuable also comes with certain **challenges** and **choices**, it can be hard to realise what options are available to you and to decide which one suits your situation best. To aid the decision making process, we created a **decision tree** showcasing all the different ways in which qualitative data can be made available for reuse in the current archival landscape. The decision tree is also separately [available on Zenodo](#).

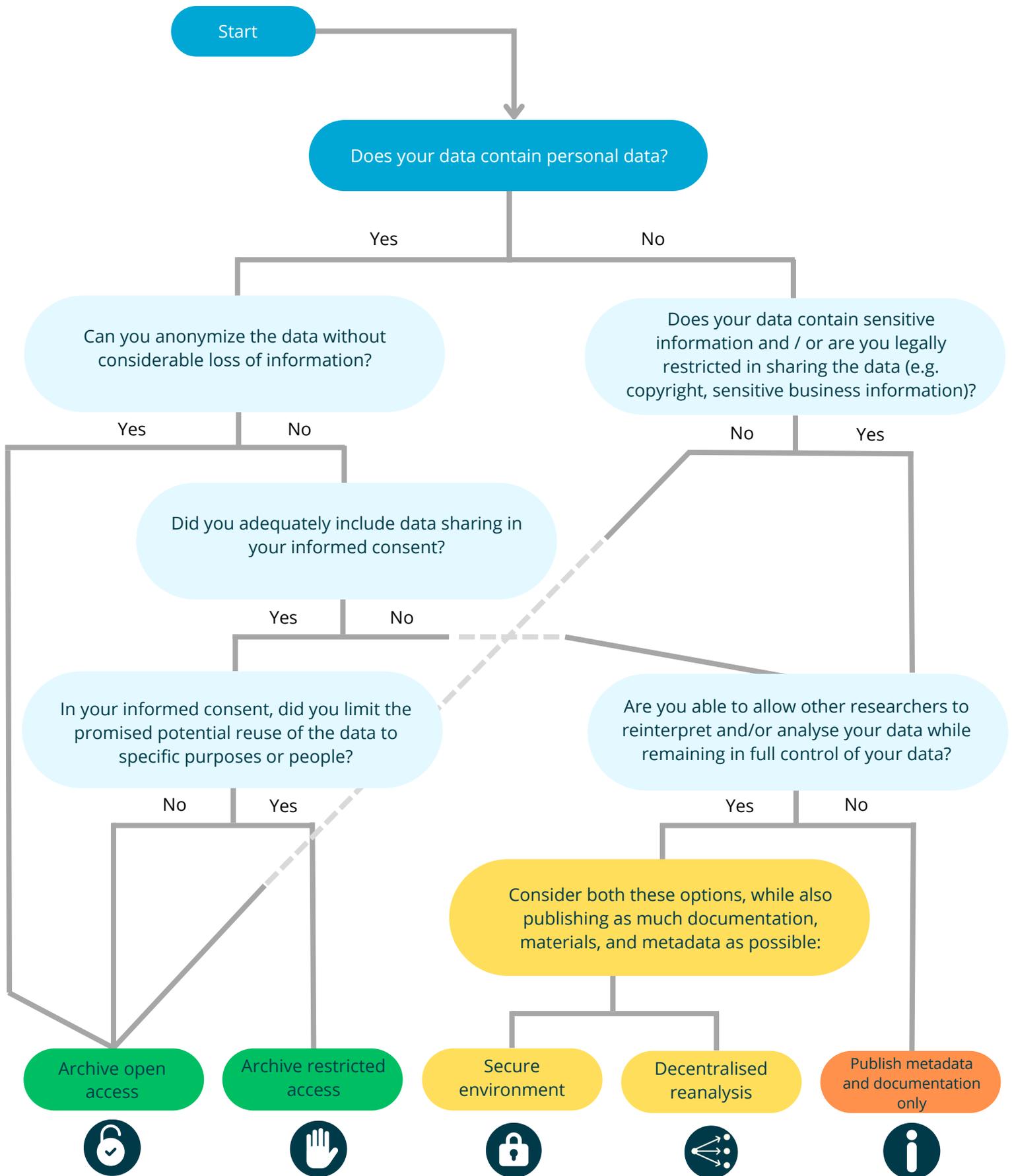
Below, we provide you with some further context for the different reuse options, ordered from most to least open.





Qualitative data reuse decision tree

as open as possible, as closed as necessary





In order to allow others to be aware of your research, we recommend that you in all cases publish your **documentation** and as much **metadata** as possible about your research even when you are not able to share the data itself.

There are some repositories that accept metadata-only records (see below), yet we would advise you to publish any relevant files and documentation, as well as any (anonymised) data that can be shared alongside the metadata as a dataset in a trustworthy repository, like the [DANS Data Station SSH](#).

Archive open access



If you have no personal or sensitive data in your dataset, you can consider archiving the data with open access. It is recommended to use a [CC0 licence](#)⁵ or another Creative Commons licence and choose the least restrictive licence possible. If you manage to fully anonymize your dataset, it is no longer considered personal data under the GDPR. This means that you can theoretically share your data openly, but considering your participants remains important. Read the section on de-identification in this guidebook for more information.

If you consider publishing your data open access, we recommend to archive your data in a repository that is specialised in your domain and certified in trustworthiness, e.g. the [DANS Data Station SSH](#). Alternatively, your own institution may have a recommended data repository, for instance [DataverseNL](#).

Archive restricted access



When depositing data in a repository, your data does not automatically become publicly available. In many repositories, you have a choice of the level of access you want to assign to your dataset. In many trustworthy repositories specialised in the social sciences restricted access to data is a possibility. This can allow you to share data with certain **conditions**, in case you specified those to your participants in the informed consent, or in case such conditions were stipulated by a funder, publisher, or other stakeholder. This could mean that only certain groups of people can access your data, or that reuse can only happen for specific purposes.

In most cases, restricted access means people interested in accessing your data will have to make a **request** that you personally review. It is important that you have a clear Data Access Protocol for accepting or denying requests to the data based on well-founded reasoning, so that you will be able to objectively review requests for the long term.

⁵ If you would like more information about [licences](#) and which one to choose, you can find information in the guidance from [DANS](#)



Another access level that is often grouped under restricted access is the application of an **embargo**, which means that data will become publicly available only after a specified amount of time. An embargo can for instance be used so that data becomes available together with the scientific publication. If you consider publishing your data restricted access, we recommend to archive your data in a repository that is specialised in your domain and certified in trustworthiness, e.g. the [DANS Data Station SSH](#). Alternatively, your own institution may have a recommended data repository, for instance [DataverseNL](#).

Secure environment



In some cases (e.g., when you have to deal with specific legal constraints) you cannot archive your data in a repository but you can enable other researchers to analyse your data in a **secure environment** where you remain in control of all the data that comes in and the results that come out. In some cases you can even require a **“blind” analysis**, where the researcher does not see the original data but their algorithms can perform analyses.

The use of secure environments is a common practice for researchers working with microdata at Statistics Netherlands (CBS) which is highly sensitive. The [Secure ANalysis Environment \(SANE\) project](#) is working on a secure environment infrastructure that builds on this expertise but can be used by various researchers and data providers to analyse sensitive data.

Do consider that these environments require **financial and personal efforts** and **training** for the researchers that want to use them as well as for the data owners who need to prepare the data and potentially evaluate the results that researchers want to export.

Decentralised reanalysis



The CaRe & DaRe solution: Decentralised reanalysis

This innovative option for data sharing has been piloted in the CaRe & DaRe project and is a solution in situations where other forms of data sharing cannot apply. This approach to data reuse leaves the **data in the hands of the original creators**. Potential reusers can approach researchers based on information (metadata and supporting documents) they share about their data and request them to reanalyse their data based on an **analysis protocol** they created. The original researchers will then only share the results of that analysis, not the data itself. In return, participating researchers could become co-author or get another form of recognition for their contribution.





This approach to data sharing can facilitate **cross-case studies** and other reuses of data, while minimising the risk of misinterpretation and decontextualisation because the original researcher performs the analysis.

Do keep in mind that this form of data reuse places a burden of future work on the original researchers that they have to weigh against the benefits for them, and that there is still need for adequate documentation and information about the data to make sure that reusers can correctly interpret their own data (considering you may get requests many years in the future).

The CaRe & DaRe project presented the **pilot** approach for decentralised reanalysis. Initial findings - which were presented during a Data Stewards Interest Group meeting - show a general enthusiasm for the approach with researchers and data stewards from multiple disciplines who otherwise would not be able to facilitate reuse of their data. There is still a lot more to develop in terms of rules of engagement, protocols, and further explaining of the process, but the initial findings are promising. The OPEN-QUAL project, *Innovating Methods for Open Science in Qualitative Management Research*, will focus on further elaborating the process and theory as well as building a platform where the exchange of information and analysis protocols can occur. If you want to learn more about this approach of qualitative data sharing, you can contact the project lead: prof. dr. ir. J.J. Berends⁶.

Publish
metadata only



When none of the options to make your data reusable are applicable to your situation, you can still consider **publishing** documentation, materials, and as much metadata as possible about your research to make it **discoverable** by others. This way, other people will at least know the research has been carried out, which could be useful in reviews or literature studies. Moreover, your materials and documentation might be useful for reuse in other studies regardless of the data.

There are some repositories that accept **metadata-only records**, yet we would advise you to publish any relevant files and documentation as well as any (anonymised) data that can be shared alongside the metadata as a dataset in a **trustworthy repository** like the DANS Data Station SSH.

⁶ For contact on decentralised re-analysis approach, contact jj.berends@vu.nl



Example datasets

For some **inspiration** on how to put the guidance of this booklet into practice, have a look at some of these **qualitative datasets** deposited in the DANS Data Station, which are available with access categories **open** or **restricted**. Consider how these researchers have documented and shared their work and how these can be effectively reused.

Hover over the text to get the link to the dataset

Examples of qualitative datasets available at DANS with access category **Open access**:

- Hanzon, MA C.A. (JSO) (2019): [Leren van herhaald beroep in de JeugdzorgPlus](#)
- Evers, drs. J.C. (Erasmus University Rotterdam/Evers Research & training) (2014): [Thematische collectie: Kwalitatieve analyse: kunst én kunde](#)
- Hoogsteder, dr M.H.H. (Amsterdam UMC, locatie VUmc) (2020): [Samen Gezond Groot - optimalisering van de leefstijladvies in de JGZ 0-4](#)
- Damen, Dr. AM (Universiteit voor Humanistiek) (2022): [Palliatief Landelijk Onderzoek Eerstelijns Geestelijke verzorging \(PLOEG\) deelproject 3: 'Integratie GV eerste lijn vanuit 3 multidisciplinaire praktijken'](#)
- Ducimetière, BA M.D. (University of Amsterdam); Soulioti, BA A.S. (University of Amsterdam) (2020): [Interview Eva Gonggrijp on Grenzeloze Inhoud](#)

Examples of qualitative datasets available at DANS with access category **Restricted access** where users have to request permission from the data owner to use the data:

- Vries, Dr. D.H. de (Universiteit van Amsterdam) (2021): [Sociale impact van fysieke afstand op kwetsbare populaties tijdens COVID-19 \(2020\): herhaalde interviews met kwetsbare ouderen en mantelzorgers](#)
- Hofhuizen, Drs. C.J.M. (Nivel) (2022): [Zorgen en zwaaien: longitudinale studie naar de gevolgen van restrictieve maatregelen vanwege COVID-19 voor naasten van mensen met een verstandelijke beperking en hun behoefte aan ondersteuning](#)
- Hulst, Dr. A.A.L. van der (2015): [Wegbereiders, Roma en Sinti in Nederland en Tsjechië over het profijt van onderwijs, 1950-2020 Interview 05](#)
- Ducimetière, BA M.D. (University of Amsterdam); Voogd, BA M.V. (University of Amsterdam); Stigter, Dr. S. (University of Amsterdam) (2020): [Interview Monique Laros on Roel QoQo and Grenzeloze Inhoud](#)
- Mol, Drs. T. (Rijksmuseum); Stigter, Dr. S. (University of Amsterdam) (2017): [Interview Alessandro Mendini](#)



Final thoughts



This concludes the guidebook on 'Making qualitative data reusable'. We hope this guide helps researchers and data stewards working with qualitative data to consider the sharing and reusability opportunities of their data during the different phases of a project. While qualitative data often seems to not fit well in the current understanding and policies surrounding data management and sharing, data sharing options are available even for the more complex datasets. Since qualitative data is so special in the way it is collected and the value it can have for a wide array of scientific disciplines, it can make a great difference if it becomes more readily available for reuse. The reach and implications of the data may even stretch far beyond the communities the original researchers themselves would consider. The phrase "as open as possible, as closed as necessary" is surely applicable to qualitative data as well.

This guide is a first approach to an overview of resources and information for making qualitative data reusable. If you want to share more information, open tools, or other resources that are relevant to this topic, we would be very appreciative if you contact us with that information⁷.

⁷ Please send a message to info@dans.know.nl mentioning "Contribution to guidebook qualitative data"

Additional resources

Apart from the guides, tools, and tips this guidebook has gathered, there are also some longer-form resources to consider if you want to become better versed in the research data management of qualitative data.

Resources for qualitative data:

- [UK Data Service](#) has some webinar recordings and presentation materials listed on the topic of qualitative data
- [The Qualitative Data Repository](#) has a list of conferences, workshops, and training related to qualitative data on their website.
- The online course [Managing Qualitative Social Science Data](#) was created by the Social Science Research Council and the Qualitative Data Repository.
- The Circular Research Data Coursebook on [Qualitative FAIR data](#) contains the course materials of a workshop created at Maastricht University.

Resources for personal or sensitive data:

- [Research Data Netherlands](#) has created a course on GDPR for data supporters. You can sign up to follow the course formally, or browse the online materials freely (currently only available in Dutch)
- [The Data Privacy Handbook](#) has been created by Utrecht University and contains knowledge, tools, and use cases for working with personal data.
- [EOSC-Future](#) organised an 'ask me anything' session on sensitive data, discussing challenges and frequently asked questions around the topic. You can watch back the recording of this webinar.

References

Verburg, M., Braukmann, R., Mahabier, W. & Khadjavi Pour, K. (2023). Reuse of qualitative research data - community developments and needs (Version 1). Zenodo. <https://doi.org/10.5281/zenodo.7704480>

Colavizza, G., Hrynaszkiewicz, I., Staden, I., Whitaker, K., McGillivray, B. (2020) The citation advantage of linking publications to research data. PLoS ONE 15(4): e0230416. <https://doi.org/10.1371/journal.pone.0230416>

Piwovar, H.A., Day, R.S., Fridsma, D.B. (2007) Sharing Detailed Research Data Is Associated with Increased Citation Rate. PLoS ONE 2(3): e308. <https://doi.org/10.1371/journal.pone.0000308>

Mannheimer, S., Pienta, A., Kirilova, D., Elman, C., & Wutich, A. (2018). Qualitative Data Sharing: Data Repositories and Academic Libraries as Key Partners in Addressing Challenges. American Behavioral Scientist, 63(5). <https://doi.org/10.1177/0002764218784991>

Munafò, M., Nosek, B., Bishop, D. et al. A manifesto for reproducible science. Nat Hum Behav 1, 0021 (2017). <https://doi.org/10.1038/s41562-016-0021>

Wilkinson, M., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018. <https://doi.org/10.1038/sdata.2016.18>

Landi, A., et al. (2020); The "A" of FAIR – As Open as Possible, as Closed as Necessary. Data Intelligence 2 (1-2): 47–55. https://doi.org/10.1162/dint_a_00027

Images credit:

- Animation frontpage: ©Visual Generation via Canva.com
- Open Access icon: ©Barudak Lier via Canva.com
- Halt Icon: ©Icons8 via Canva.com
- Lock Icon: ©Microstd Creation via Canva.com
- Data Flow Icon: SBTS from Noun Project (CCBY3.0)
- Information Icon: ©Pixabay via Canva.com