



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101008589

# Guidelines for Sharing Data on Human Participants

Marijana Glavica and Vedran Halamić

2024



10.5281/zenodo.15267610

# Table of Contents

<b>Table of Contents.....</b>	<b>2</b>
<b>Introduction.....</b>	<b>3</b>
About this Guidelines.....	3
Importance of Data Sharing.....	3
Tradition of Data Sharing in Social Sciences.....	4
<b>Part 1: Overcoming Data Sharing Barriers.....</b>	<b>5</b>
Privacy, confidentiality, and ethics.....	5
Participant consent concerns.....	6
Concerns on disclosing personal information.....	8
Sensitive data issues.....	9
Intellectual property rights issues.....	10
Resources constraints.....	10
Priority in publishing results and fear of scooping.....	11
Fear of misinterpretation and misuse.....	12
Perceived lack of interest.....	12
Data quality concerns.....	13
Cultural and institutional resistance.....	13
Limited incentives and recognition.....	14
<b>Part 2. Archiving and sharing data.....</b>	<b>14</b>
Data management and planning.....	14
Steps in data sharing.....	15
Define data sharing objectives.....	15
Select what to share.....	16
Select an appropriate repository.....	17
Prepare data and documentation for archiving and publishing.....	17
Determine access rights and licensing.....	18
Sign the deposit agreement.....	19
Submit data and documentation to a selected repository.....	19
Promote your data.....	19
<b>References.....</b>	<b>20</b>

# Introduction

## About this Guidelines

Many guidelines on data sharing already exist. These often assume that you, as a researcher, are already fully aware of solutions to common concerns about data sharing. They may focus on guiding you through data management and sharing processes in a technical way, presuming you are already convinced of the value of these processes and know why you are doing it. Here, we place greater emphasis on the reasons to share data, by exploring the benefits of data sharing and addressing some doubts you might have about sharing your own data. We attempted to produce accessible reading material designed to dispel common myths about the risks of data sharing. Additionally, we highlight the critical role of data management planning and practices, including data sharing, in ensuring ethical, equitable, and credible research practices.

These guidelines were created within the COORDINATE project and thus were intended primarily for researchers who are producing data about children and youth. The guidelines were originally created to answer the issues that arose during the work on the demonstrator of the developing infrastructure for sharing research data about children and youth, and they were presented at the Conference on data sharing and secondary analysis in child and youth wellbeing research held in Zagreb, Croatia on 10th and 11th December 2024. The guidelines are primarily tailored to academic researchers' concerns and needs, but they can easily be read by expert practitioners from academic, government and NGO sectors, who are dealing with data about people.

## Importance of Data Sharing

Data sharing is becoming an increasingly important practice in the research process. This trend reflects not only a shift in research funders and journals policies but also a growing recognition among researchers of the benefits associated with sharing data. Some of these benefits are:

*Maximising the utility of data.* Researchers are given the opportunity to analyse the data from different perspectives, pose new questions, combine it with other sources, and gain new insights. Students might find lectures and exercises that use real-life examples more engaging. Public policies could be more effective and appropriate if they draw upon credible data sources.

*Enhancing research credibility and accountability.* Sharing data could make research more reliable and trustworthy because results could be checked, reproduced or replicated. Data sharing requires doing research in a careful and clear way, which can contribute to the validity of research.

*Promoting ethical and equitable research practices.* Data sharing ensures that participants' contributions are used responsibly, so that their time and attention invested in research is not wasted. It can also help make sure everyone's voice is heard, especially those who might not usually get the chance to speak up.

*Saving resources.* By reducing duplicate data collection efforts, researchers can accelerate some research, leading to possibly more cost-effective solutions for addressing complex problems.

## Tradition of Data Sharing in Social Sciences

Even before open science gained momentum as a research policy trend, social scientists recognised the advantages of sharing research data and implementing effective data management practices. This awareness has been particularly evident in the context of comparative and longitudinal studies, two important research approaches for addressing complex problems. The establishment of social science data archives several decades ago stands as a testament to this tradition. CESSDA (Consortium of European Social Science Data Archives) has been instrumental in nurturing this tradition as an umbrella organisation for data archives across Europe since the 1970s, facilitating collaboration and promoting standards for data sharing within social science communities.

All kinds of data have already been shared through data archives across Europe and beyond, including sensitive qualitative data. To facilitate this, we require infrastructures such as CESSDA ERIC, which can implement mechanisms for ethical, legal, and secure data sharing.

# Part 1: Overcoming Data Sharing Barriers

Despite the growing recognition of the availability of research data as a vital element of credible scientific work, many researchers -- particularly those collecting data about people, especially children -- continue to express doubts and uncertainties regarding data sharing, and for good reason.

In this part, we will examine some common concerns that researchers face when sharing data and suggest possible approaches to overcoming these concerns, as well as how to balance the benefits and risks of data sharing. Although our primary position is to advocate data sharing, we do not claim that there are no risks involved or that all data should always be openly available. On the contrary, we will demonstrate that researchers should be concerned not only about privacy and ownership issues -- most commonly expressed concerns by researchers -- but also about the technical and organisational environments in which they operate daily, as these can also pose significant risks and potential harm.

Each section will begin with statements expressing reasons why not to share data. The statements were identified using different methods and several sources. First, we took the statements about reasons why not to share data from the materials published by the UKDA in the *Managing and Sharing Data: Training Resources*<sup>1</sup>, adjusted them and supplemented with insights from various interactions of data archive staff and researchers, and conversations with colleagues within data archiving domains. Furthermore, we scanned the literature about researchers' data sharing attitudes, motivations and behaviour (Houtkoop et al., 2018; Kim and Adler, 2015; Kim and Stanton, 2016; Tenopir et al., 2011; Tenopir et al., 2015) to identify other possible concerns and learn how to organise and present them. These sources often also include some possible answers to those concerns.

## Privacy, confidentiality, and ethics

The most prominent concerns researchers have are related to privacy, which is, in general, an important social phenomenon and construct. Over the past few decades, numerous discussions have taken place regarding the use of personal data, leading to the implementation of legislation, such as GDPR, primarily designed to protect individuals' information from being exploited by companies seeking to profit from it and to enable the use of data when it serves public good or individual benefit. However, long before the introduction of the GDPR and other privacy legal regulations, researchers

---

<sup>1</sup> Corti, L., Van den Eynden, V., Bishop, L., and Morgan-Brett, B. (2011). *Managing and sharing data: training resources*. UK Data Archive. <https://www.data-archive.ac.uk/resources/>

had already established ethical principles to guide their work and ensure they avoided causing harm to research participants. What the GDPR did for researchers was merely to formalise procedures that demonstrate these underlying principles are being applied in practice.

Research ethics is often more restrictive on what you can do than GDPR is. For example, legal consent to process personal data is not needed when conducting anonymous surveys, but you would still ask your participants for an ethical consent and explain to them the purpose of data collection, how the answers will be used, and if there are any risks for them. Legally you can do anything with the data if participants have consented. Ethically, you, as a researcher, will not allow for this in all circumstances, because you are more aware of the risks that can be induced by sharing data publicly. With this knowledge, your obligation is to educate participants about the risks.

If responses can be de-identified on a reasonable level, no legal consent is needed from participants as there is no way to link answers to individuals. You would still want to inform your participants about how their answers will be used and how they can have more impact if analysed by many researchers.

### **Enabling access to data is not only about open access to data!**

Not all data should be shared openly with everyone. Data should be anonymized and shared under specific terms and conditions to ensure its appropriate use. Access to the data can be restricted for specific purposes, such as research, and granted only to researchers affiliated with academic institutions. This type of controlled access relies on trust mechanisms, such as formal agreements or contracts between data providers and data users. In some cases, these agreements may also involve the institutions with which the researchers are affiliated. Additionally, access may require researchers to undergo training or education to ensure they understand how to handle the data responsibly and ethically.

### *Participant consent concerns*

- *I am doing highly sensitive research, therefore I cannot possibly make my data available for others to see.*
- *If I ask my respondents for consent to share their data then they will not agree to participate in the study.*
- *I have made promises to destroy my data once the project finishes.*
- *My data have been gathered under complete assurances of confidentiality.*

Participants may alter their responses if they believe their data will be made publicly available. The fear of judgment or repercussions can lead to less candid answers, particularly on sensitive topics.

However, the assumption that participants will not engage in research if they have to consent to data sharing is not always accurate in all circumstances and it should not be an excuse to avoid seeking such consent. We can argue that individuals who agreed to participate in research have already decided to give trust to the researcher, are willing to assist researchers and abide by their rules to contribute to the advancement of science. Following this logic, trust exists from the outset, and it is unlikely to increase or decrease if participants are aware that their answers will be used for a greater good, such as in new research, to improve public policies related to problems that are being researched, or as a proof that research results were calculated correctly. Some people might be even more motivated to participate if they are aware that their responses will be maximally utilised, not only in one study but also in many others.

The likelihood of participants agreeing to data sharing might significantly increase if they are informed that the data will not be shared openly and without any control, but rather that all individuals and organisations involved in the process will have to adhere to certain rules and procedures before accessing the data and commit to handling the data responsibly.

Recommendations:

- Considerable effort should be invested in designing the informed consent form. Simply adopting a ready-made template is not sufficient. It should always be tailored to the specific research. Simple language has to be used. Carefully designing the informed consent form can also serve an educational purpose. Online, people often agree to terms of service for a digital service without reading them. It's understandable that they don't read them because such documents are usually extensive and written in language that is hard to understand.
- A clear distinction should be made between personal data needed for administering the research and participant's responses to research questions. Participants should be assured that their personal data will be stored securely for a specified period of time, and will be deleted when not needed any more. It should be clearly explained that only their responses will be shared.
- Don't promise that you will destroy the data after the project ends. This should apply only to personal data that you have used to administer research, but there is no ethical or legal reason for you to destroy your research data.

- Try to approach confidentiality differently. Rather than believing that confidentiality can only be achieved by promising that information provided by participants will be seen exclusively by you and your team, consider extending that promise to include that the data will be used by you and those you trust. When you promise confidentiality, this does not exclude the possibility to share data as data can be shared in a confidential way. Explain to participants that data will be shared with other researchers responsibly and that data sharing is desirable practice in contemporary research conduct. Benefits of data sharing for other purposes should be clearly explained to participants, assuring them that you are careful not to cause harm while sharing data.
- Participants should be given the option to participate in the research but not consent to data sharing, rather than not being asked about it at all.

### *Concerns on disclosing personal information*

- *I am doing quantitative research and the combination of my variables discloses my participant's identity.*
- *It is impossible to anonymise my transcripts as too much useful information is lost.*
- *I have collected audiovisual data and I cannot anonymise them, therefore I cannot share these data.*

Quantitative data can be anonymised through various techniques to protect individual privacy while maintaining its research utility. These include:

- Aggregation: Grouping specific data, such as age, into broader ranges.
- Top coding: Limiting extreme values by grouping them into a single category.
- Removing variables: Eliminating sensitive and unnecessary variables that are not essential for analysis or relevant to the research.
- Generalisation: Making specific information less detailed, such as replacing an exact address with a broader location like a city.
- Pseudonymisation: Replacing real participant identifiers, such as names, with codes or false names.

By applying these methods, researchers can minimise the risk of disclosing identifying information while preserving the value of the data for analysis.

If too much information would be lost with anonymisation, controlling access to data can be a good solution. This is particularly the case in most of the qualitative research. Less restrictive access might be possible after some period of time has passed.



Visual data can be anonymized by blurring faces or distorting voices; however, this process can be time-consuming and expensive. Additionally, it may result in a significant loss of the data's value. A more effective approach is to seek consent from participants to share the data in its unanonymised form and/or to implement controlled access to the data.

### *Sensitive data issues*

- *I am doing highly sensitive research, therefore I cannot possibly make my data available for others to see.*

When conducting research that involves the collection of sensitive personal data, it is crucial to prioritize their privacy and confidentiality. Public sharing of such information is generally not advisable under any circumstances, even if participants provide explicit consent. This is especially important when the research involves vulnerable individuals or social groups.

If the collected data cannot be anonymized without losing essential information, researchers should consider implementing strict access controls. Different levels of access can be established to ensure that sensitive data remains protected. For instance, data may only be accessible to individuals who have received explicit approval from the researcher. These individuals must sign a contract agreeing to avoid re-identification of the data, to refrain from sharing it without permission, to keep data in a safe environment, and to delete data when not needed any more.

The highest level of control over sensitive data can be achieved through the use of traditional "safe rooms" or modern remote access systems. Safe rooms provide a secure physical environment for data analysis, while remote access systems are contemporary implementations of the same principles, allowing secure access to data in a controlled virtual environment.

A nice example on how sensitive data can be de-identified is available in Campbell et al. (2023).

## Intellectual property rights issues

- *My data collection contains data which I have purchased and it cannot be made public.*
- *There is IPR in the data.*
- *In my research, I used psychological scales that are protected by copyright and measuring instruments that are available only to qualified professionals. I cannot share my research database because it contains information about items from these proprietary scales.*

Intellectual property rights (IPR) encompass various legal protections for creative works. While the data itself cannot be copyrighted as it is considered factual information, unique compilations or databases may qualify for protection under copyright laws.

When using data that has been purchased or sourced from other entities, it is important to ascertain who holds the copyright and what permissions are required for its use. This process should ideally occur at the beginning of a research project to avoid potential legal issues later on. If permissions are not secured upfront, researchers may face challenges if they attempt to publish or disseminate not just the research database, but also findings derived from such data.

## Resources constraints

- *I have not got the time or money to prepare data for sharing.*
- *I don't know how to prepare data for sharing. I don't have a place to put the data. Data sharing requires extra work. Lack of training in sharing data.*

Preparing data for sharing undeniably requires additional time and effort. Maintaining clean, well-documented data is, unfortunately, still not a standard practice in much of the research community.

Incorporating data management practices into your research workflow from the outset can save time and resources in the long run. Well-organised and documented data are easier to analyse, interpret, and share, reducing the need for costly and time-consuming data cleaning and formatting tasks later on. Data management plans, which are increasingly required by research funders, should be perceived as a tool to ensure that you devote the necessary attention to data management. When planning the time required to conduct the project, it is essential to factor in the time needed for data management. Some funders may even encourage you to specifically allocate resources

and budget for data management. Still, even if that is not the case, time and resources for data management should be incorporated in resources needed to do the usual research process.

If you want to share data collected in research conducted in the past, before data management became a requirement for receiving research grants, consider seeking funding for projects aimed at preparing the data for sharing.

## Priority in publishing results and fear of scooping

- *I want to publish my work before anyone else sees my data.*
- *I put so much effort in collecting data, it is my right to publish first.*
- *Someone else can steal my ideas and take advantage of my data before me.*

Research funders should allow a period during which only you and your team can utilise the data for publishing papers. However, in light of recognizing the importance of data sharing, this period should not be too long. Keep in mind that you are the one who knows your data best, and attempts at theft in science are actually quite rare. Science has mechanisms to combat fraud, so fraud and theft can be uncovered.

Trends in scientific communication are moving towards treating the published dataset as a separate product of scientific work, an entity that can be cited. By publishing the data, along with registering your studies and preregistering your hypotheses<sup>2</sup>, you have informed everyone in advance that this was your idea.

Furthermore, it's important to consider the various purposes of data sharing. Journals increasingly require that data underlying the results and conclusions of a paper be published alongside the article. In this case, it is possible to publish only the portion of the data used specifically for that paper. In that case the data is published for the purpose of result verification, i.e. reproducibility and replicability. The complete dataset resulting from a research project can be published later, after you publish all the papers you planned to publish. At that point, the data is shared for the purpose of reuse.

---

<sup>2</sup> <https://help.osf.io/article/330-welcome-to-registrations>

## Fear of misinterpretation and misuse

- *Other researchers would not understand my data at all - or may use them for the wrong purpose.*
- *Somebody will use the data for the wrong purpose. If journalists use the data, they will draw wrong conclusions and spread miss-information through public media.*

Provide comprehensive documentation and metadata alongside your data to facilitate understanding and interpretation by other researchers. Detailed descriptions of data collection methods, variable definitions, and data processing procedures can help users navigate the dataset and make informed decisions about its suitability for their research purposes.

Make sure your data is accurate, reliable, and trustworthy by checking it thoroughly. This builds confidence in other researchers who might want to use your data for their own studies.

Establish formal data sharing agreements with researchers or institutions interested in accessing your data. These agreements should define how they can use the data and make sure they only use it for genuine research. Control who can access your data by only giving access to approved people or research teams.

Keep an eye on how others use your data by tracking its usage. If you spot any misuse, take action to protect the reputation of your data and ensure it's used properly.

## Perceived lack of interest

- *My data is not of interest or use to anyone else.*

It is very difficult or impossible to predict with certainty whether your data will be of interest to others. Data can be used for different purposes. Your data might not have high re-use potential for comparative studies and policy making, but it is still useful because it can improve the credibility of your research. We can attempt to assess the potential for reuse based on simple criteria such as whether the research is based on a nationally representative sample or if it is an international comparative study. If we agree with the assertion that your data are not of interest to anyone, can we then question whether the published results based on these data are interesting to anyone?

Even if your data may not seem immediately relevant, it could inspire new research questions or methodologies for other researchers. Your data, although specific to your

field, might hold value for researchers working across disciplines. It may fill gaps in existing datasets or contribute to meta-analyses and systematic reviews.

Data can also serve as a valuable educational resource for teaching and training purposes. Moreover, accessible data can inform evidence-based policymaking and decision-making processes, driving positive societal impacts beyond the realm of academia.

## Data quality concerns

- *I'm not sure that my data is of sufficient quality to be shared.*

Another significant obstacle to data sharing is the lack of standardisation in data collection, storage, and sharing protocols. Research data were often collected using disparate methodologies, instruments, and formats, making it difficult to integrate and compare findings across studies. Without common standards and formats, integrating and comparing data from different sources becomes arduous, if not impossible. This lack of standardisation has impeded efforts to conduct meaningful cross-study analyses and synthesise findings effectively.

The solution to this challenge is, again, proper data management, but also trying to follow discipline standards if they exist, or engage in initiatives that are aiming to raise awareness about using standards in research data management.

## Cultural and institutional resistance

- *Data should not be electronically available to other researchers. Data sharing is not a common practice in my research field. I prefer to share data only upon a request.*

Researchers may fear that sharing their data could lead to misuse or misinterpretation, which can undermine their work's integrity. Additionally, the perception that data sharing is not standard practice can create hesitancy among researchers to adopt more open approaches.

Institutional frameworks also play a crucial role in shaping attitudes towards data sharing. In many cases, universities and research organizations lack clear policies or incentives that encourage open data practices. Without institutional support for data sharing—such as recognition in performance evaluations or funding opportunities—researchers may feel discouraged from adopting more open practices.

## Limited incentives and recognition

- *My funder, nor my institution require data sharing.*
- *I will not receive any recognition or reward if I share my data.*

Traditional academic reward structures often prioritise individual achievements such as publications, discouraging researchers from investing time and effort into data sharing initiatives that may not directly contribute to their career advancement.

Researchers would be more willing to share their data if secondary users of their data would properly cite them.

## Part 2. Archiving and sharing data

After the barriers described in the previous chapter have been overcome, the next set of uncertainties that researchers may have are related to the specific steps that need to be taken in the process of data archiving and publishing. In the last decade, we have seen an explosion of repository services to publish research data. Researchers have been encouraged and incentivised to manage and share their data, mainly by research funders and journal requirements. At the same time, in some research areas, data sharing practices have been well known, followed by developing research infrastructures that can support such practices, while in other disciplines there was no such tradition.

### Data management and planning

An essential prerequisite for data sharing is the data management process. Managing data is necessary throughout the entire lifecycle of a research project, from the inception of the idea about the research problem to the project's completion and beyond. A data management plan enables you to ensure in time that all steps have been taken in order to be able not only share the data with others but use the data in your own research.

This is necessary not only for other researchers who may use the data in the future but is also crucial for your own sake. Here are some reasons why more attention should be paid to data management:

- During the preparation of data for analysis and the analysis itself, you may heavily transform original data. Without these transformations being

well-documented in a way that allows you to easily and accurately generate the final dataset from the original data, you cannot be sure that your analysis results and conclusions are based on real data. Situations where this is the case are more common than you might think.

- You might already be sharing your data in a way that is not legal. For instance - sending personal data by unencrypted email or posting data to non-European servers.
- You may be keeping unencrypted sensitive personal data on your laptop which you bring everywhere with you and which you may easily lose or it can be stolen.
- You didn't ask for permission to use an instrument in your research, let alone permission to share.
- Your colleague has downloaded a very interesting dataset from a social media platform that is no more available. Can you use this data in your research?

An important aspect of the documentation in research involving human participants pertains to the mechanisms employed to minimize the risk of harm, ensuring that the research is conducted ethically. When informing participants and obtaining their consent, it is essential to explain why it is important that their responses may be used by trusted individuals. Additionally, you should assure participants that their data will not be shared publicly and will only be accessible under restricted conditions.

Many mistakes can occur during data processing, which can significantly impact the validity of your research findings. To mitigate these risks, it is crucial to maintain a clear and reproducible record of your data cleaning steps. This means documenting each action taken, including any transformations, filtering, or corrections applied to the dataset.

A comprehensive resource for research data management in social sciences and related disciplines is the *CESSDA Data Management Expert Guide (DMEG)*<sup>3</sup>. This guide is created and maintained by CESSDA ERIC and its service providers. We recommend consulting this guide for an in-depth exploration of all aspects of data management, including those pertaining to personal and sensitive data. Some of the content presented here was based on the DMEG.

## Steps in data sharing

### *Define data sharing objectives*

Identifying your reasons for sharing data will provide clarity and direction when selecting what data to share, determining how to prepare it, and choosing the most

---

<sup>3</sup> CESSDA Training Team (2017 - 2022). *CESSDA Data Management Expert Guide*. CESSDA ERIC. <https://dmeg.cessda.eu/>

appropriate licence. Two main objectives often guide data sharing: enabling the reuse of data in new contexts and ensuring that research results are reproducible and.

### ***Enabling reuse in new research and beyond***

This is particularly relevant for datasets collected on representative samples, datasets that have historical importance, are unique or non-repeatable, feature innovative approaches, or are collected at significant scale or cost.

- *Representative sample.* Data that is representative of a population or phenomenon can provide foundational insights for further studies.
- *Historical importance.* Data documenting significant events, trends, or conditions over time can inform longitudinal research or historical analysis.
- *Unique or non-repeatable.* Observations that cannot be reproduced due to the specific conditions of data collection (e.g., natural disasters, cultural events) have inherent value for future studies.
- *Feature innovative approaches.* Data resulting from novel methodologies or technologies can inspire further methodological advancements.
- *Collected at significant scale or cost.* Large-scale datasets, or those that require extensive resources to compile, can support research that would otherwise be infeasible.

By sharing data, you make it accessible to other researchers, policymakers, educators, and even the public when possible, allowing them to build upon your work or apply it in different ways. In addition, reusable data can support teaching, helping to educate future researchers and practitioners by providing real-world examples of analysis and methodology.

### ***Facilitating reproducibility and replicability***

Reproducibility ensures that others can achieve the same results by analysing the shared data using the same methods, while replicability allows researchers to test the findings by applying similar methods to new data. By sharing data for reproducibility, you are demonstrating research integrity, transparency, and trustworthiness. Researchers, peer reviewers, and other stakeholders can examine the data to verify your findings, strengthening the credibility of your work.

### ***Select what to share***

It is not always appropriate or practical to share everything that has been collected. Decisions on what to share must align with your data sharing objectives and adhere to ethical, legal, and practical considerations. Striking a balance between openness and



protection ensures the responsible dissemination of data while maximising its potential for future use.

If your goal is to enable reproducibility, ensure the published dataset includes all what is needed to recreate your analysis. Be sure to share the software scripts used in data transformation and analysis. This allows others to validate your findings and build upon your work.

Create two versions of the data: one for future scientific use, containing as much detailed information as possible, and another with redacted and aggregated values to protect individual identities.

Be careful not to share any data that you do not own!

### *Select an appropriate repository*

It is best to select a repository maintained by an organisation that specialises in the domain of your research and meets the needs of potential data users. These are often referred to as domain-specific or disciplinary repositories. They should offer specific advantages, such as tailored metadata standards, access controls, mechanisms to protect sensitive data, and more.

Importantly, if you are dealing with data about individuals, choose organisations that will review your databases before publishing them. This could serve as a strong protective factor against the risk of publishing and publicly sharing a dataset that contains identifying information about individuals.

Some contemporary initiatives advocate for data to be considered a first-class research output. To be counted, research databases should follow a similar publication process to that of an article. They should be properly documented with metadata, reviewed for quality, searchable and discoverable in catalogues and similar services, and citable in publications (Brase et al., 2009).

Choosing a domain specific repository, if available, is recommended for data that are archived and published because of its reuse potential. Choose those who can support you in the process and who will review your data before publishing.

### *Prepare data and documentation for archiving and publishing*

Before depositing, data should be adequately prepared and organised to facilitate understanding and reuse. Address any issues related to data integrity, consistency, and format compatibility.

This involves organising and formatting the data in accordance with established standards and guidelines. Researchers should ensure that the dataset is well-documented, with clear descriptions of variables (data dictionaries, codebooks), methodologies and procedures applied in data collection. The purpose of data collection should be explained. Anything that can contribute to the usability and interpretability of data should be included in the dataset.

Data should be anonymized to protect the privacy of individuals. The level of anonymization may vary depending on the audience and the desired level of openness.

Ensure everything is accurate and complete. Ask for support from data archiving experts.

### *Determine access rights and licensing*

Access categories:

*Open access.* Data that can be accessed by any user whether they are registered or not. Data in this category should not contain personal information unless consent is given (see 'Informed consent').

*Access for registered users (safeguarded).* Data that is accessible only to users who have registered with the archive. This data contains no direct identifiers but there may be a risk of disclosure through the linking of indirect identifiers.

*Restricted access.* Access is limited and can only be granted upon request. This access category is for the most sensitive data that may contain disclosive information. Restricted access requires the long-term commitment of the researcher or person responsible for the data to handle the upcoming permission requests.

Everything does not have to be open to everyone! It should be "*As open as possible as closed as necessary.*"

You don't have to publish everything right now!

### ***Open licences - Creative Commons***

Open licences enable everyone to freely access, use, modify, and share copyrighted work for any purpose.

CC is a set of licences to specify in advance who has what rights to use your work, instead of having to ask for permission each time it is used. You, as the licensor cannot

revoke the freedoms of use you have granted to users, as long as they adhere to the terms of the licence.

Only datasets that do not contain information related to individuals can be made publicly available. However, CC licences are highly recommended for licensing instruments and documentation.

<https://chooser-beta.creativecommons.org/>

### *Sign the deposit agreement*

Once your data and documentation are ready, and you have decided on which conditions you want to share your data, sign the deposit agreement with your chosen repository.

This formalises the deposit process, and outlines your responsibilities and rights as the data owner, as well as responsibilities and rights of organisation that will archive and publish your data.

### *Submit data and documentation to a selected repository*

The process of depositing data typically involves submitting a dataset to the selected repository through an online platform or submission system. Researchers are required to submit the research database accompanied with information about its content, context, and access conditions. Metadata should be provided as well.

Your submission will be reviewed by archive staff. You may be asked to correct some errors or provide more information. Some changes can be done by archive staff and you will need to approve such changes.

You will be asked to verify your dataset and approve it before publishing. After the dataset is published, you may still be able to send corrections in data files or metadata.

### *Promote your data*

Promote your shared data to increase its visibility and reuse. Share links through publications, social media, and academic networks. Be sure to use DOI or any other persistent identifiers. Highlight how others can benefit from your data.

A dataset that includes sufficient documentation for interpretation and has undergone a review process can be regarded as a scientific output equivalent to a publication. In this context, citing data using persistent identifiers (e.g., DOIs) is a crucial practice, as it

enables unique attribution of the work to an author and facilitates the tracking of the dataset's use. This can improve the visibility of your results

## References

- Campbell, R., McKenzie, J., Engleton, J., Fishwick, K., Gregory, K., & Goodman-Williams, R. (2023). Open-science guidance for qualitative research: an empirically validated approach for de-identifying sensitive narrative data. *Advances in Methods and Practices in Psychological Science*, 6(4). <https://doi.org/10.1177/25152459231205832>
- CESSDA Training Team (2017 - 2022). CESSDA Data Management Expert Guide. CESSDA ERIC. <https://dmeg.cessda.eu/>
- Houtkoop, B. L., Chambers, C., Macleod, M., Bishop, D. V. M., Nichols, T. E., & Wagenmakers, E.-J. (2018). Data Sharing in Psychology: A Survey on Barriers and Preconditions. *Advances in Methods and Practices in Psychological Science*, 1(1), 70–85. <https://doi.org/10.1177/2515245917751886>
- Kim, Y., & Adler, M. (2015). Social scientists' data sharing behaviors: Investigating the roles of individual motivations, institutional pressures, and data repositories. *International journal of information management*, 35(4), 408-418. <https://doi.org/10.1016/j.ijinfomgt.2015.04.007>
- Kim, Y., & Stanton, J. M. (2016). Institutional and individual factors affecting scientists' data-sharing behaviors: A multilevel analysis. *Journal of the Association for Information Science and Technology*, 67(4), 776-799. <https://doi.org/10.1002/asi.23424>
- Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M., & Frame, M. (2011). Data Sharing by Scientists: Practices and Perceptions. *PLoS ONE*, 6(6), e21101. <https://doi.org/10.1371/journal.pone.0021101>
- Tenopir C, Dalton ED, Allard S, Frame M, Pjesivac I, Birch B, et al. (2015) Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. *PLoS ONE* 10(8): e0134826. <https://doi.org/10.1371/journal.pone.0134826>