Data Practices and Management

5

Rea Roje

Abstract

Employing good data management practices is important for enhancing the transparency and validity of research, as well as the reproducibility of research findings. This chapter aims to help early career researchers translate the European Code of Conduct for Research Integrity principles and guidance on data management practices into everyday research. In this chapter we will guide you on data practices and management throughout the lifecycle of research data – data management planning, organizing and storing data, preserving and sharing data, reusing and citing data. You will also learn about the data management procedures relevant to each of the data lifecycle phases – preparation of data management plans, procedures for storing data properly and securely, examples of repositories for preserving and sharing data, licenses for reusing data, etc. The chapter will also outline the FAIR data principles and data protection requirements and safeguards important when handling personal data in your research (GDPR requirements, pseudonymization, anonymization, and deleting data).

Keywords

Data management · Data management plan · Research data · Personal data · Data protection

R. Roje (✉)
Department of Research in Biomedicine and Health and Center for Evidence-based Medicine, University of Split School of Medicine, Split, Croatia
e-mail: rea.roje@mefst.hr

© The Author(s) 2023

A. Marušić (ed.), *A Guide to Responsible Research*, Collaborative Bioethics 1,
https://doi.org/10.1007/978-3-031-22412-6_5

What This Chapter Is About

Employing good data management practices is important for enhancing the transparency and validity of research, as well as the reproducibility of research findings. This chapter aims to help early career researchers translate the European Code of Conduct for Research Integrity principles and guidance on data management practices into everyday research. In this chapter we will guide you on data practices and management throughout the lifecycle of research data – data management planning, organizing and storing data, preserving and sharing data, reusing and citing data. You will also learn about the data management procedures relevant to each of the data lifecycle phases – preparation of data management plans, procedures for storing data properly and securely, examples of repositories for preserving and sharing data, licenses for reusing data, etc. The chapter will also outline the FAIR data principles and data protection requirements and safeguards important when handling personal data in your research (GDPR requirements, pseudonymization, anonymization, and deleting data).

Case Scenario: Data Handling and Record Keeping

This is a hypothetical scenario of a junior researcher who discovers gaps between previously kept records of lab data and what has been published. The original case scenario is developed by the Members of The Embassy of Good Science and is available at the Embassy of Good Science. This hypothetical scenario was adapted from a narrative concerning the links between data management and research integrity. The case below is published under Creative Commons Attribution-ShareAlike license, version 4.0 (CC BY-SA 4.0).

Professor Brown is an epidemiologist who just won a prestigious grant for conducting research about the impact of environmental genetic and clinical factors on the prevalence of obesity in urbanized areas. The research team working on the project is interdisciplinary and includes senior researchers, postdoctoral researchers, and doctoral students. For conducting this research project, the methodology includes collecting data from public Databases (Geographic Information System and Google Street View), data collection from hospital records, studying genetic samples stored in hospital biobank, surveys, and interviews with research participants. While working on the part of the project focused on collecting data from Geographic Information System and Google Street View, one member of the research team insists on sending the datasets to the public repository (making them available to other researchers who want to use the dataset) since the grant agreement requires project members to employ FAIR principles (making data findable, accessible, interoperable, and reusable) in their research. However, not all members of

the research team agree with this. Some members are emphasizing the GDPR requirements that must be respected and intellectual property rights regarding the data collected. Moreover, some members of the research team think that data should be made available to others only after the manuscript publication so no one can endanger their publication plans. For the data collection from hospital records, researchers retrospectively collected laboratory data, histological results, and some personal data of individual patients (age, sex, residential area, occupation). The research team collected these data without obtaining ethical committee approval presuming that it is not necessary to have approval for collecting data from hospital records that were taken a decade ago. Moreover, the informed consent from patients was also not obtained.

Questions for You

1. In light of this case scenario, what data management issues can be identified in this research project?
2. What data management guidelines and practices should be employed in the research described in the case scenario?
3. What data protection practices should be followed in the research described in the case scenario?

Data and Types of Data

Let's first define what data are and what the different types of data are. Data are all unorganized facts that need some processing and organisation to become information. Hence, data are unprocessed information, and similarly, research data can be defined as collected, unprocessed information that will need to be processed, organized, and presented in a certain context to provide information that will support research findings.

Research data can be classified based on different criteria. For example:

- Type: electronic documents, registries, tables, notes and laboratory books, questionnaires, transcripts, codebooks, samples, databases schemes, models, algorithms, protocols, experimental results, metadata, methodologies, etc.;
- Format: textual (word, PDF, XML, etc.), numerical (Excel, SPSS, etc.), audio and multimedia (jpeg, tiff, wav, etc.), software programs, disciplinary-specific (e.g., crystallographic information file, CIF);
- Size and complexity of research data: small, large, simple, complex;
- Research phase: raw, cleared, processed, analysed.

Why Is Good Research Data Management Important?

Data management includes different processes and activities required to manage and preserve data throughout the research lifecycle or research phases (Fig. 5.1):

- planning and preparing the data,
- collecting, organizing, and storing the data,
- analysing, and protecting the data,
- archiving, preserving, and curating the data,
- discovering, accessing, and reusing the data.

Reduced risk of data loss, increased transparency and reproducibility of research results, easier compliance with different requirements (ethical, legal, funders' or publishers' requirements), prevention of errors in research, and research career and reward benefits are just some of the reasons why good data management practices are important and should be considered and applied during every research project.

Good research practice from the European Code of Conduct for Research Integrity:
Researchers, research institutions, and organizations ensure appropriate stewardship and curation of all data and research materials, including unpublished ones, with secure preservation for a reasonable period.

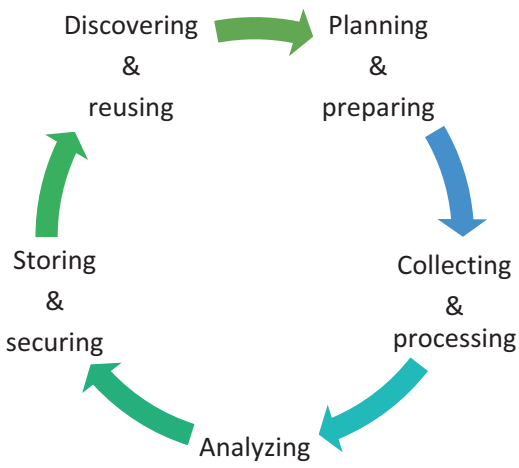

Fig. 5.1 Lifecycle of research data

Planning and Preparing Data

The first step in applying good data management practices in research includes planning before the project has started and even before applying for research funding. The planning process will include thinking about all the types and formats of data that you will collect and process during your project, as well as how the data will be used and by who. Moreover, when you plan and prepare for your research, you should also include getting familiar with the funders' requirements for data management. Research funders usually provide relevant information on data management requirements and helpful guidance on how to fulfil the requirements, so make sure to check it before sending your project application. For example, if you plan to apply to European Commission (EC) research frameworks grant calls, you should check their guidance (e.g., Guidelines on Open Access to Scientific Publications and Research Data for Horizon Europe). Since requirements and conditions may change over time, it is very important to recheck these each time you plan a new research project.

Data Management Plan

Research funders often require writing a data management plan (DMP). DMP is very important for every research project because it helps:

- plan in advance everything concerning your data (e.g., collecting, storing, licensing, sharing, etc.) and how you will deal with your research data during the project;
- anticipate potential issues that may arise during the project;
- enhance data FAIRness (making your data findable, accessible, interoperable and reusable);
- your project collaborators to manage the data in the same way which enhances the integrity of data and ensures proper stewardship.

In your DMP, you should include the information about:

- type of data you will collect, create, or reuse;
- how data will be documented and stored;
- ethical and legal requirements;
- how data will be shared and preserved;
- who will have the access to use or reuse the data;
- data management responsibilities of the team members (project director, research staff, technical staff, supporting staff, etc.);
- data management costs (e.g., costs of transcribing data or costs of long-term preservation).

It is good to bear in mind that in some situations you cannot plan or anticipate all possible scenarios, and hence sometimes you will need to update your DMP if significant changes regarding your data arise (for example, if you decide to collect some new data that was not planned at the beginning). There are many online tools for data management plans that make the work of writing DMP a lot easier. Tools like OpenAIRE guide on DMPs for Horizon Europe and Science Europe Practical Guide to the international alignment of research data management can help you create your DMP.

Collecting, Organizing, Storing, Analysing, and Protecting Data

Once you have planned your data management activities and started with your project, it is time to implement good data management practices in collecting, organizing, storing, analysing, and protecting the data. Usually, you have already defined in your research protocol or plan how your data will be obtained and which procedures and methodology will be used for collecting and analysing the data.

Data Formats

The data format can also be defined in advance since it depends on the type of data you collect (e.g., text files or audio files) and type of preservation (long-term or short-term). There are some general recommendations, for example, the preferred format for images is *tiff*, and for audio files, the preferred format is *wav*. For long-term preservation, it is always better to use open formats, which will ensure accessibility to a wider audience. Moreover, there is a possibility that in the future we will not be able to use some outdated formats or software to open and reuse data, hence the best option is to store data in open formats and formats with widespread use. For example, when dealing with the textual data you can choose between open (e.g., *docx*, *txt*, *pdf*) or closed (e.g., *doc*) formats. Similarly, if you have table data, you can choose between open (e.g., *xlsx*, *csv*, *ods*) or closed (e.g., *xsl*) formats.

Organizing Data

Once you decide on the formats, it is time to organize your data properly. It is important have a clear and consistent naming and organizing both your paper and electronic data. You should use consistent, unique, and descriptive names and develop name conventions that will be followed by everyone involved in your project. This will reduce the risk of losing data, and make research and exchange of data between different project participants a lot easier. Naming conventions should be

written in a separate file and stored properly so that everyone in your research team can access them and check if there are some uncertainties. Moreover, you should track versions of your data by, for example, documenting in a separate file which changes have been made in which version and who made the changes. Do not forget that your raw data should be preserved as they are, so it would be best to store them separately to ensure they will not be lost. There are many options for naming conventions, and you should use those that fit your data and research best. For example, when dealing with structured tabular data, it is very important to write naming and description of used variables and codes. When naming your files, you can follow the convention like "project name/acronym_subject_activity phase" or "version_type of data_researchers initials". Some recommendations for naming conventions include:

- develop naming convention upfront;
- use letters and numbers from A–Z or a–z and 0–9;
- use ISO standards for the date (YYYYMMDD);
- do not use period punctuation mark or special characters;
- use a low dash or CamelCase instead of space (e.g., CamelCase.docx instead of Camel Case.docx or Camel_Case.docx – the name comes from visual "hump" created by a capital letter in the file name).

Another important thing to consider when organizing your data is developing a folder structure or a map. Your folder structure should be properly developed and logical so that you can find all your data easily. For example, you can have a folder named after your project. In that folder, you can have folders called "data", "documentation" and "methodology". In your data folder, you will have files related to your data (raw data and different versions) while in your documentation folder you can keep related data, such as invitation letters and informed consent forms. You can also use tagging which can help you find your data on the computer.

Documentation Describing Data and Metadata

Your data can be accompanied by other relevant information that can help other researchers understand and reuse your data. When sharing your data or depositing the data in the repository, you can also include a README.txt or INFO.txt file in which you will provide basic information regarding your data (Box 5.1). These can include, for example, general information (title of the dataset, author information, date of data collection, geographic location of data collection, information about funding sources), sharing or access information (licenses and restrictions, links to publications, links to other locations of data, recommended citation for the dataset), data and file overview (list of files included in your dataset, explanation of the relationship between the files), methodology, and specific information for certain files. You can include any information that you think is helpful for other researchers to understand and replicate your data.

Box 5.1 Example of the Content in an INFO.txt File

1. Title of the dataset
2. Author(s) information (name, institution, email)
3. Date of data collection (exact date or approximate date; suggested format YYYYMMDD)
4. Geographic location of data collection (city, country)
5. information about funding sources (who funded the research)
6. Licenses/restrictions placed on the data
7. Link to accessible locations of data
8. Recommended citation for the dataset
9. File list (list of all files contained in the dataset with a brief description; relationship between files)
10. Versions of the dataset (if there are multiple versions)
11. Description of methods (for collecting data, methods for processing data; links to publications)
12. Instruments or software used in the analysis
13. Quality assurance procedures applied on the data
14. Specific information for files (number of variables, number of cases/rows, list of variables)
15. Abbreviations used

Metadata are data about your research data, and when, for example, you want to deposit your data in the repository you will be asked to fill in the metadata. Metadata can vary depending on the disciplines and research areas. You can visit Metadata Standards Directory to find more information about metadata in your discipline or research area. Metadata include at least the following: author information and contact, name of the organisation, title, type of data, and keywords. Moreover, when depositing in good and trusty repositories your data will be assigned with the digital object identifier (DOI) or another persistent identifier – a long-lasting reference to a digital resource. Metadata, including the persistent identifier, are intended for machine-reading and they enable retrieval and reuse of the data. Even if you are not sharing your data openly (i.e., in open access), your metadata should be publicly available and hence findable and retrievable, in accordance with the FAIR principles.

Storing and Protecting Data

Many options are available for your data storage and backups. The general recommendation is to have at least three copies of data at three separate places – computer, cloud, and portable device. You can store your data in an infrastructure and storage

space provided and managed by your organisation. This is a good option that can minimize the risk of data loss, as organisations usually do regular backups. Using the institutional infrastructure for storage can ensure adequate security level and easier dissemination with collaborators in your project. For ensuring additional data protection, you can define within your research team who will be in charge of data backups and in which timeframes the backups will be conducted (e.g., on a weekly or monthly basis). Moreover, you can also conduct checksums, i.e., check the similarity between files before and after making backups to ensure that you have identical files and that the backup was done appropriately. To conduct checksums, you can use different IT solutions, such as MD5summer. Data can also be stored in cloud services, but this is not the best option as it usually involves third-party access to data and an appropriate level of security is not ensured. If you decide to use cloud services, you should always check the terms and conditions of cloud providers and their compliance with the GDPR.

Researchers often use different portable devices in their work, but portable devices should be used cautiously as the risk of losing data is quite high. Using encryption for protection is recommended in each case, especially when using clouds and portable devices. Make sure to store and protect your password adequately and have strong passwords that are updated regularly. Besides data in the electronic format, researchers are often dealing with data in the paper format. If you have your data on paper, you should make sure to store these properly – preferably in the safe, under the lock. Moreover, it is recommended that you digitalize your paper data as this can ease the usage and exchange of data and increase data safety. You can create documentation that will help you to get around more easily regarding your data storing and protection actions. For example, you can make a file in which you will write where your data are stored (for example, on a computer of your collaborator), who has the access to the data and who can make changes, who performs backup, and how often, etc. Another thing you should pay attention to is whether you store all your data at the same place or separately. This is especially important when you deal with sensitive personal data. For a qualitive interview study, for example, the interview transcripts, informed consents, and other personal identifiable information should be stored separately and protected adequately (e.g., encrypted).

You must protect research participants and their data, especially when processing special categories of personal data that require even more safeguards (such as health data, data on ethnic origin, religious or political beliefs). You must check applicable laws and guidance (organisational, national, and international) on how to process personal data and employ appropriate technical and organisational measures and safeguards. You should pay attention to the GDPR requirements when processing personal data from European Union (EU) citizens. You can find more information on data protection and privacy legislation worldwide at the end of this chapter.

If there is a Data Protection Officer or a Research Integrity Officer in your organisation, check with them how to ensure proper protection of participants' data in your research. Here are some safeguards:

- Data minimization – collect only data that is needed for your research aims;
- Data anonymization – remove all identifiable information from the data;
- Data pseudonymization – substitute identifiable information with the unique identifier.

Whenever possible, you should use data anonymization, because anonymized data are not considered personal data and hence not under the data protection laws and requirements. However, anonymizing data is not always an easy task. It is important to think about the consequences of under and over anonymization, which affects the further use of the data. When dealing, for example, with qualitative data, it is good to use pseudonyms or descriptive names or tags to change participants' names, and it would be best to make a detailed plan on how qualitative data will be anonymized before the transcription process. In this way, you will assure more accuracy and save time for checking the transcribed data. For more information on how to anonymize different types of your research data check the available resources collected by the FOSTER project (Fostering the practical implementation of Open Science in Horizon 2020 and beyond), EU Data Protection Working Party Article 29 opinion on anonymization techniques and OpenAIRE resources.

Ethics

You should also get yourself familiar with ethical requirements related to your research and the protection of research subjects. Do not forget to obtain an ethics approval (if it is needed) before starting your study and store the document properly. When conducting research with human participants, you are required to develop an information letter in which you will describe the aims and purposes of your research study, what participation in research involves, what the rights of participants are if they decide to participate in research, and what potential risks and benefits are for participating in research. Your information letter should also include items regarding the processing of personal data (which personal data will be processed, how and for what purposes; how the personal data will be stored and protected, and for what period; what will happen to data after the storing period expires; who will have the access to the participants' personal data; what the participants' rights are in regard with the processing of their personal data). Besides providing participants with this information, you will also need to obtain informed consent. In the informed consent, you provide statements that participants have to agree with to participate in research (e.g., statements saying that participants understood the information letter and what is expected of their participation in research, statements that participants are aware of how their data will be collected and processed and that they agree with

it, etc.). The informed consent is usually obtained in the written form and signed by participants.

When collecting data in research studies involving children, in addition to the parental consent, you will also need to ask the assent from children of certain ages, depending on national legislation (see for example Informed Consent for Paediatric Clinical Trials in Europe). Make sure to ask assent in a written or some other form by using language that is appropriate to the child's age, so that child can understand what are you asking.

As informed consent and assent documentation contains personal data, you should take appropriate measures to store these documents. This means in a safe place, with restricted access, and not together with other research data that contains information that can be linked directly to the individual. For example, you should never store your interview transcript together with the research participant's informed consent.

Deleting Data

Once you no longer need data or data storage period was predefined in your study protocol and data management plan, make sure that data are disposed of securely. Just deleting data from your computer may not be enough, as deleted files can be retrieved. Similarly, just tossing your papers in the trash can is not a proper way of disposing data. You should always use appropriate measures such as shredding machines and computer software that will ensure that data is not retrievable.

Good research practice from the European Code of Conduct for Research Integrity:
Researchers, research institutions and organizations provide transparency about how to access or make use of their data and research materials. Researchers, research institutions and organizations acknowledge data as legitimate and citable products of research.

Archiving, Preserving, and Curating Data

Preserving and sharing your data with other researchers has many benefits. We already talked about data storing, which mainly focuses on how you can store your data for your own and your research team's purposes. The preservation of data is more related to long-term availability of data. This means enabling proper storage and preservation of data even if you or your research team do not have the data anymore or you are not reachable by other researchers who would perhaps like to use your data. Good preservation and sharing practices can increase the visibility, impact, and citations of research, ensure validation of the research data, encourage

collaborations and enable reuse for new research findings. However, before sharing the data, you must consider several factors:

– make sure there are no constraints regarding sharing (such as data containing personal information that identifies individuals).
– think about how data might be reused.
– think about the costs of sharing and long-term curation.
– check funder's data-sharing policies and requirements.

To preserve and share your data, you can use project, discipline, national, or international specific repositories. However, before deciding on where to deposit your data, you should check if a repository is reputable and safe and whether there are persistent and unique identifiers that will be added to your data and that will make sharing data easier but also ensure that proper contribution is given to data owners.

Some recommendations for finding a good and reputable repository include checking:

– whether your data formats are acceptable by repository;
– whether the backups of the deposited data are regularly performed;
– whether the repository has a certificate (that ensures it is long-lasting and reputable);
– whether you can track the statistics related to your data in the repository (e.g., how many times your data were downloaded);
– whether the repository is in accordance with the FAIR principles.

To find relevant repositories, you can use the Registry of Research Data Repositories – a database of international repositories for research data. For more guidance on choosing a repository for your data, take a look at the Science Europe Practical guide to the international alignment of research data management, which offers guidance for selecting trustworthy repositories. Some of the well-established repositories are, for example, Figshare and Open Science Framework, where you can share different types of data and preprints. In Open Science Framework you can also register your research protocols (see Chap. 2 on Research Procedures).

Discovering, Accessing, and Reusing Data

One of the important aspects of preserving and sharing research data is also deciding on the terms and conditions on which other researchers will use your data. This may be especially important if you have to share your data before publishing your research (e.g., because of funders' requirements). By licensing your data, you will ensure that data is used under the conditions you set and that appropriate credit is given to you. You can check the Creative Commons licenses or Open Data Commons to learn about different levels of data sharing. When depositing your data in the repository, you will be able to choose between different types of licenses. You should carefully consider which licence is suitable for your data, e.g., how you want

your data to be used, taking into consideration any other intellectual property rights related to your data. You should also bear in mind that licenses with fewer restrictions provide more opportunities for reusing the data. Once you choose the license it cannot be changed, which is another reason to carefully consider the licences for your research.

For example, the Creative Commons (CC) licenses are widespread and commonly used licenses that enable you as an author to copyright your work and set specific conditions under which others can use your work. These licences are based on four main types of reuse:

- Attribution (BY): allows distribution, adaption, and building upon the original work by giving the proper attribution to the creator;
- Non-Commercial (NC): allows distribution, adaption, and building upon the original work only for non-commercial purposes;
- No Derivates (ND): the original work can be used only in its original form and changes are not allowed;
- Share Alike (SA): allows distribution and adaption under the same conditions as they stand for the original work.

Based on these, there are a total of 6 Creative Commons licenses that can be used: CC BY, CC BY-SA, CC BY-ND, CC BY-NC, CC BY-NC-SA, CC BY-NC-ND (Box 5.2). There is also a CC0 license ("No Rights Reserved") which stands for the public domain. This means that you are giving your work in the public domain and it is free for use, there are no restrictions and there is no obligation to provide the attribution to the work.

In 2018, a Plan S was launched as an initiative by Coalition S of national research funders, European Commission, and the European Research Council, dedicated to ensuring full and immediate open access to research publications. The Plan contains ten main principles, and the key principle is that all publicly funded research must be published in open access journals or made immediately available in open access without embargo. In that sense, it is important to mention the Rights Retention Strategy (RRS) which was developed to make sure that researchers retain sufficient intellectual rights on their work, so they could freely share it. The RRS also requires researchers to deposit the Author Accepted Manuscript or the Version of Record into the repository with a CC-BY license and with no embargo.

Citing Data

When using data from other researchers or sources in your research, do not forget to cite it appropriately, as proper credit should be given to data owners. When citing data, you should always:

- be consistent with the referencing style that should, in each case, include authors, title, publication date, publisher, and location (a persistent URL);

Box 5.2 Creative Commons (CC) Licenses

CC0: dedicating the works to the public domain, meaning that the creators wave all their copyright and related rights to their works.

CC-BY: allows distribution, remix, adaptation, and building upon the original work in any format; credit must be given to the creator; should be used if you want maximum dissemination and use of your work.

CC BY-SA: allows distribution, remix, adaption, and building upon the original work in any format but adaptations must be shared under the same terms; credit must be given to the creator.

CC BY-NC: allows distribution, remix, adaptation, and building upon the original work in any format only for non-commercial purposes; attribution must be given to the creator.

CC BY-NC-SA: allows distribution, remix, adaptation, and building upon the original work in any format only for non-commercial purposes; the modified work must be licensed under the same terms and proper credit must be given to the creator.

CC BY-ND: allows copying and distribution only of the original work only, and only in unadapted form; derivates or adaptations of the original work are not allowed; credit must be given to the creator.

CC BY-NC-ND: allows copying and distribution only of the original work, and only in unadapted form for non-commercial purposes only; credit must be given to the creator; the most restrictive license.

– include DOI or another permanent identifier;
– separately cite different datasets.

Good research practice from the European Code of Conduct for Research Integrity:
Researchers, research institutions and organizations ensure access to data is as open as possible, as closed as necessary, and where appropriate in line with the FAIR Principles (Findable, Accessible, Interoperable and Reusable) for data management.

The FAIR principles are the guidance on how to make data findable, accessible, interoperable, and reusable. Making data FAIR enables others to discover, understand, and use data. The Final Report and Action Plan from the European Commission Expert Group on FAIR Data says that data should be made open and FAIR as much as possible and closed as necessary following the ethical and legal constraints and requirements.

Findable To make data findable, they have to be (1) described with adequate and rich metadata; (2) contain a persistent identifier that will permanently link data, metadata, and other relevant material, (3) registered or indexed in the search resources to enable other users to identify and use the data.

Accessible Data are accessible once potential users find the data and know how to access it. Making data accessible does not imply that data are open and free for use, since access may require authentication or authorization. Making data accessible means that users should be able to access the data under certain conditions that need to be transparent and defined clearly. It is also very important to mention that metadata should be retrievable and accessible even when data are no longer available.

Interoperable Data are interoperable if integrated with other data. Both data and metadata must use formal and broadly applicable language for knowledge representation and use vocabularies that follow FAIR principles.

Reusable Making data reusable is considered the ultimate goal of the FAIR principles. To achieve this, several metadata and data requirements should be implemented: (1) metadata and data have to be adequately described and have a clear data usage license; (2) metadata and data have to be associated with detailed provenance and meet domain-relevant community standards.

Good research practice from the European Code of Conduct for Research Integrity:
Researchers, research institutions and organizations ensure that any contracts or agreements relating to research outputs include equitable and fair provision for the management of their use, ownership, and/or their protection under intellectual property rights.

Intellectual property rights in research usually refer to patents, copyrights for data and published research, confidentiality agreements, etc. You should check your organisation's intellectual rights policy to ensure you are properly informed about how you and others can use your research. If you want to use research data or other output from other researchers, check any existing intellectual property rights and use output accordingly. You should also check your funders' policies (especially in industry-sponsored research) and agreements established with your collaborators. In any case, you should ask advice from the appointed university office and staff that deals with the intellectual property issues and industry-sponsored research agreements to ensure that your intellectual property rights as a researcher are protected adequately.

Copyright

Copyright is a legal protection given to some original work, whether literary work, music, artistic work, or research. Having copyright means having the exclusive right to use, copy and disseminate your work and at the same time limiting or enabling others or assigning your copyright to others to use your work under certain conditions or without special requirements. See also the *Discovering, accessing, and reusing the data* section in which licenses were discussed. When publishing your work in a journal, you will be asked to choose or be informed about the type of license for publishing your work. You may also have the option to transfer your copyright to the journal, in which case the journal decides upon licensing and crediting the work. However, you should be careful and consider your funders' requirements, especially if your funder is a part of the Coalition S and you have a right to retention.

If You Want to Learn More

The Embassy of Good Science

A Breach of Confidentiality
A Case Study of Secondary Use of Qualitative Data
Protecting Research Subjects
Anonymity Revisited
Failed Patenting Negotiations in Collaborative Research

Published Articles

El Emam K (2011) Methods for the de-identification of electronic health records for genomic research. Genome Med 3:25. https://doi.org/10.1186/gm239

Wilkinson M, Dumontier M, Aalbersberg I 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

Guidance

European Commission. Collaboration in Research and Methodology for Official Statistics. Anonymisation
European Commission. FAIR guiding principles
European Commission. Final Report and Action Plan from the European Expert Group on FAIR Data. Turning FAIR Into Reality
European Commission, Horizon Europe. Guidelines on Open Access to Scientific Publications and Research Data
i-CONSENT Project. Guidelines for tailoring the informed consent process in clinical studies. 2021
Metadata Standards Directory
OpenAIRE. Amnesia guide
Science Europe Research Data Management
United Nations Conference on Trade and Development (UNCTAD) Data Protection and Privacy Legislation Worldwide)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.