Chapter 4
Research Data Stewardship for Healthcare Professionals

Paula Jansen, Linda van den Berg, Petra van Overveld, and Jan-Willem Boiten

4.1 Data Stewardship: What, Why, How, and Who?

Data stewardship is the long-term, sustainable care for research data. This has become an indispensable part of clinical research. This chapter provides an overview of the aspects of data stewardship that you should consider when you are involved in clinical research. The majority of these aspects should be addressed *before* you start collecting data. The chapter is a condensed version of the Handbook of Adequate Natural Data Stewardship (HANDS), which is a living document on the website of the Data 4lifesciences programme of the Netherlands Federation of University Medical Centres (NFU). Please consult the full web version of HANDS for more detailed information and a toolbox.

P. Jansen (✉)
Research Office, UMC Utrecht, Utrecht, The Netherlands
e-mail: pjansen9@umcutrecht.nl

L. van den Berg
Washoe Life Science Communications, Rhoon, The Netherlands
e-mail: linda.vandenberg@washoe.nl

P. van Overveld
LUMC, Leiden, The Netherlands
e-mail: p.overveld@lumc.nl

J.-W. Boiten
Lygature, Utrecht, The Netherlands
e-mail: janwillem.boiten@lygature.org

4.1.1 Definitions

Data stewardship involves all activities required to ensure that digital research data are findable, accessible, interoperable, and reusable (FAIR) in the long term, including data management, archiving, and reuse by third parties. The precise definition of data stewardship and its distinction from data management is a topic of ongoing expert discussions. The Dutch National Coordination Point Research Data Management (LCRDM) has developed a glossary of research data management terms.

4.1.2 Why?

Adequate data stewardship is a crucial part of Open Science. Promoting optimal (re)use of research data through open science is one of the goals of the European Union (EOSC Declaration) and corresponding national initiatives. Scientists, patients, and the general public will benefit from new scientific knowledge, treatments, and applications that result from sharing high-quality data. In addition, data stewardship is required to protect the scientific integrity of research and to meet the requirements of research funders, scientific journals, and laws (e.g., the General Data Protection Regulation, GDPR).

As a clinical researcher, you will benefit from adequate data stewardship in several ways. Your data will be robust and free from versioning errors and gaps in documentation and will be safe from loss or corruption. In addition, the data will remain accessible and comprehensible in the future, allowing you to share the final dataset with others, for scientific research, commercial development, validation, or healthcare. Good data stewardship planning also ensures that you will have timely access to resources such as storage space and support staff time.

4.1.3 FAIR Principles

This chapter describes the fundamentals of research data stewardship according to the FAIR Principles [1, 2], which have been adopted worldwide. The FAIR Principles state that research data should be:

- **Findable:** The data should be uniquely and persistently identifiable and other researchers should be able to find the data.
- **Accessible:** The conditions under which the data can be used should be clear to humans and computers.
- **Interoperable:** Interoperability is the ability of data or tools from non-cooperating resources to integrate or work together with minimal effort. Data should be machine-readable and use terminologies, vocabularies, or ontologies that are commonly used in the field.

- **Reusable:** Data should be compliant with the above and sufficiently well-described with metadata and provenance information so that the data sources can be linked or integrated with other data sources and enable proper citation.

4.1.4 Responsibilities

As a clinical researcher, you are the principal data steward. In practice, this means that you are responsible for the complete scientific process: from study design to data collection, analysis, storage, and sharing. Protecting the privacy of study subjects is also your responsibility. The *formal* responsibility for personal data lies with your research institute, which is accountable for having adequate policies, facilities, and expertise around data stewardship. According to the principle of accountability in the GDPR, it is the institute's responsibility to ensure that the fundamental principles relating to processing of personal data are respected, as well as the ability to demonstrate compliance. Your research institute should appoint a Data Protection Officer that monitors GDPR compliance at the institute. Possible consequences of not adhering to these principles include reputation damage, liability, and losing or having to refund a research grant. Some institutions have appointed formal data stewards that promote or can advise on data stewardship. Researchers can delegate tasks to these data stewards. Table 4.1 provides an overview of the responsibilities of the main people involved in data stewardship for clinical research.

Table 4.1 Responsibilities of people involved in data stewardship[a]

Who?	Responsibilities
Researcher	Is accountable for research data; Is in control of the complete research data flow; Reuses existing data when possible; Collaborates with patient organisations throughout the research project; Protects the privacy and safety of study subjects; Applies the FAIR principles; Protects research quality and reproducibility; Uses available expertise and recommended infrastructure; Thinks ahead about intellectual property rights; Shares data responsibly
Research institution	Employs professionals that provide the procedures and technical systems for data stewardship (e.g., data stewards, data managers, IT-specialists, statisticians); Has institute managers, who govern and facilitate the professionals; Has supervisory bodies such as medical-ethical review committees and privacy officers; Engages with patients and citizens from whom data is collected; Offers facilities to protect data according to the GDPR

Table 4.1 (continued)

Who?	Responsibilities
Manager of research institution	Establishes facilities for data stewardship (e.g., data protection, storage, interoperability); Provides financial means for data stewardship and expert employees; Is responsible for organisation, policy, standard procedures, practical measures; Ensures training for employees that work with data
Professional that supports data stewardship	Provides, gives advice on, and supports the use of terminologies, IT-standards, and e-infrastructure which promote data sharing and integration; Gives advice on writing data management sections and plans, metadata standards, repositories, and data handling Supports data curation and archiving

[a]Note that the majority of items in this table constitute guidelines mentioned in HANDS rather than formal rules

4.2 Preparing a Study

Decisions on data stewardship will affect how you can process, analyse, preserve, and share your research data in the future. This section explains what decisions researchers need to make when preparing a study. It is recommended to consult an expert on these topics.

4.2.1 Study Design and Registration

Careful study design is required to ensure that your research question can be answered in the end. For instance, you should select the most appropriate technique and determine the sample size required to get statistically meaningful results. Study design is the domain of specialists, who can be consulted in the design phase of the study. In addition, researchers can follow basic courses on study design, good clinical practice, and research data management. Randomized controlled trials need to be registered before they start, for instance at clinicaltrials.gov. At many institutions, this is also required for observational research.

4.2.2 Re-using Existing Data

Before starting to collect new data, you should ask yourself whether it is possible to use existing data to answer your research question or to enrich your own dataset. Reusing data may be more efficient, reducing inconvenience for study subjects and saving resources. In addition, the chances of getting funded are significantly better

if you show that you have considered reusing data. Potential sources of reusable data include reference data, data on reference cohorts, similar data collected in a previous study, healthcare systems (clinical data), biobanks, the biomedical literature, and digital repositories. The toolbox in HANDS lists several sources of existing data and biobank material. HANDS also addresses what to consider before using existing data or starting a scientific collaboration. You should also consider re-using metadata from other studies as a template for your own study (see Sect. 4.4.3).

4.2.3   Collaborating with Patients

Clinical researchers are strongly encouraged to involve patients and patient organisations in their research, from design until completion. Patient representatives can suggest research questions, help recruit study participants, select relevant outcome measures, help design the informed consent procedure, provide advice on policies (e.g., regarding incidental findings), and help to communicate research results back to study participants [3].

4.2.4   Data Management Plan and Statistical Analysis Plan

A data management plan (DMP) shows that you have thought about how to create, store, archive, and give access to your data and samples during and after the research project. Nowadays, many research funders and academic institutions demand DMPs from researchers. The responsibility for creating a DMP lies primarily with principal investigators. Examples of DMPs and practical tools such as a Data Stewardship Wizard can be found in HANDS' toolbox.

Statistical analysis plans are obligatory for randomised controlled trials. It is preferable to create this plan before collecting data because this facilitates proper study design (e.g., in-and exclusion criteria, number of study subjects needed, decisions with regard to statistical power, choice of data items to be collected). This is discussed further in Sect. 4.5.2.

4.2.5   Describing the Operational Workflow

Clinical researchers should be able to describe the complete operational workflow for their research data, from data capture, to data analysis, archiving, and sharing. They are responsible for answering questions about the origin of their data, data manipulations, the location where the data is analysed and archived, and with whom it is shared under what conditions. A research institution is responsible for providing infrastructure which is compliant with current regulations and guidelines (e.g., on privacy and data integrity) (Fig. 4.1).
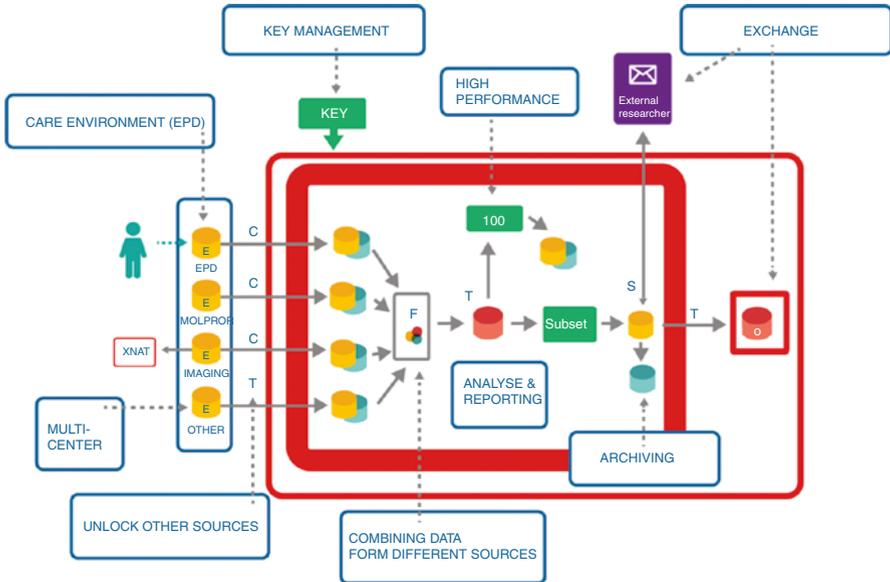
Fig. 4.1 Example of an operational workflow chart. This shows which functionality is involved as well as the typical activities around clinical data, including repositories. (Adapted by the NFU from an original illustration created at Radboudumc, Nijmegen)

In smaller studies, the data capturing system (whether manual data entry from paper, via electronic forms, or sophisticated real-time connections between the primary source data and the study database) should be able to assess and report the logical consistency and the clinical probability of data values. For large datasets, it is important to think ahead about:

- storage capacity;
- when the raw data will become available;
- backups to safeguard against system failure and human error;
- the location where various data processing steps will be carried out (e.g., the capacity of the network should be sufficient if the data must be transported from the measurement location to the analysis location);
- access policies (e.g., whether web-based or multi-user access is required);
- procedures for data documentation and anonymisation or pseudonymisation;
- protection against unauthorised access (see Sect. 4.4.4);
- costs (e.g., for storage and compute capacity).

4.2.6 Choosing File Formats

Ensuring that your data is FAIR requires care in selecting file formats. For instance, it is important to consider how the data can be accessed in 10 years from now: will software still exist that can read the information? Data formats should preferably be

open (i.e., formats that can always be implemented, so not '.doc' and '.xls' or instrument-specific data formats), well-documented (i.e., rigorous like 'xml' with a schema description and not open to multiple interpretations like '.csv' without schema descriptions), flexible (i.e., self-describing formats which can adapt to future needs without breaking old data), and frequently used (i.e., for which conversion tools will be created and maintained if necessary). DANS (Data Archiving and Network Services) has made a useful overview of preferred file formats.

4.2.7 Intellectual Property Rights

Failure to think about Intellectual Property Rights at the start of your study may cause legal dispute and it can lead to limitations to the research, its dissemination, future related research projects, and associated profit or credit. Designing a study may already lead to protectable ideas. Ask yourself questions like 'Is the outcome usable for further research? Is it usable for a product or service? Does it need additional protection (e.g., with a patent or copyright)?' On the other hand, if you wish to allow others to reuse your data, it may be advisable to make this explicit, e.g., through a Creative Commons license, giving the public permission to share and use your work on conditions of your choice. It is advisable to contact a Technology Transfer Office (TTO) at the start of your study and before sharing data. They can help create written agreements on when to share what data with whom under what circumstances. Such agreements should also be included in a consortium agreement.

4.2.8 Data Access

Clinical researchers are responsible for describing the data access and sharing policy of their study. This policy should be tailored to the project and devised prior to collecting data, allowing some room for later adaptations. According to the FAIR principles, all research datasets should at least be *findable* (including non-sensitive data, metadata, and aggregated data about the study) and the conditions under which the data are *accessible* should be clear. Clinical researchers are obliged to share their data with monitoring bodies upon request (e.g., internal audits). A data access policy should take into account a number of considerations (see Sect. 4.7). Many research institutions have their own Data Governance Policy, which may include the instalment of a Data Access Committee that plays a role in the permission of sharing data with third parties.

4.3 Privacy and Autonomy

Clinical research calls for careful attention to the privacy and autonomy of the people involved.

4.3.1   Informed Consent

Informed consent aims at informing potential study subjects of all aspects of participation, including the procedures for data handling, data access, and anonymity. An informed person can freely decide to participate or not. If someone does participate, he or she understands and accepts the risks and burdens involved in that participation. Informed consent also is a crucial aspect of the GDPR. Regarding data management, the informed consent should include the person's wishes about:

- the use and reuse of the data for research in the current and future projects (including the options for data filtering: which data may be used for research);
- notification about incidental research findings (special concern is required for results that cannot be interpreted now, but may be interpretable in the near future);
- which data he/she can access, if applicable;
- the possibility to withdraw certain aspects of informed consent and the consequences;
- data use by commercial parties.

In general, it is very difficult to re-contact patients or study subjects to extend or change the consent. So, it is best to obtain informed consent for storing clinical and personal data for the purpose of *both* healthcare and future scientific research, each with a separate informed consent. In addition, patients should always be able to retract their consent, so your system should allow for data to be removed. Consent should be documented along with the collected data, so subsequent users of the data are aware of the conditions agreed to by study subjects. Most research institutions have access to an ethical committee that can help design your informed consent procedure.

4.3.2   Care and Research Environment

It is important to distinguish between the *care* environment (i.e., data that is used for diagnosis and treatment of patients or self-evaluation of healthcare providers) and the *research* environment (i.e., data that is used to answer scientific questions). Nowadays, these two data environments are increasingly integrated. However, the distinction is important because different laws and guidelines apply to the two environments and these laws may even conflict.

Having said that, healthcare and scientific research can reinforce each other. For instance, data collected in a care environment may be used to answer research questions. Data collected in a research environment may travel back to the care environment as 'unexpected incidental findings' crucial to be communicated to the study subject. Data collected in a research environment may also be used in the clinic to

avoid double data collection (e.g., collection of quality of life data in intervention trials). You should take special measures when you reuse data collected in the care environment for scientific research and vice versa. For instance, research data usually undergoes less stringent quality control than clinical data and extra checks are required before using research data in the clinic, including an extra verification of the identity of the study subject.

4.3.3   Preparing Sensitive Data for Use

Processing your data for scientific research or statistical analysis should be subject to appropriate safeguards for the rights and freedoms of the data subjects, in accordance with the GDPR. Those safeguards should ensure that technical and organisational measures are in place, in particular in order to ensure respect for the principle of data minimisation. Any research data should be anonymised or pseudonymised. Anonymisation means processing data with the aim of *irreversibly* preventing the identification of the person to whom it relates. Pseudonymisation means replacing any identifying characteristics of data with a pseudonym, i.e., a value which does not allow the person to be directly identified. Pseudonymisation only provides limited protection for the identity of data subjects as it still allows identification using indirect means. You may consider involving a trusted third party (TTP) to encrypt and decrypt identifiers. In all cases, the translation table between the research code and the identifying patient information should be stored and managed separately from the research database.

4.4   Collecting Data

Two key principles should guide research data stewardship in the data collection phase: ensuring the scientific integrity of the study and protecting the privacy of study subjects and researchers. This includes ensuring data quality, protecting the data from malicious access, and safeguarding the ability to interpret the data correctly. You can ensure all of this by:

- implementing a suitable data management infrastructure;
- implementing a data validation step after initial data entry;
- including documentation (metadata) to add context to the data;
- taking data protection measures.

In addition, you should use a standardised protocol for data collection in order to allow others to reuse your data in the future, using the terminologies and standards that are accepted your research field. The best time to consider and describe all these issues is at the *start* of your research project.

4.4.1   Data Management Infrastructure

An adequate data management infrastructure can help you work more flexibly, easily, and quickly. It can also simplify version control and collaboration. As soon as (in)direct identification of human study subjects is possible, you should use a professional data management system. The system and its environment should preferably be ISO27001 certified, or at least meet the underlying goals (i.e., protection, accountability, privacy, documentation, risk assessment, quality management). Experts can help you select an appropriate data management infrastructure, which allows for:

- the collection, storage, and analysis of research data; this is often called a 'database';
- sufficient data protection measures (discussed in Sect. 4.4.4);
- accurate management and logging of data access (discussed in Sects. 4.4.4 and 4.7);
- storage of metadata, process flow description, data provenance description, data extraction documentation, and data modification logs (see Sect. 4.4.3);
- support for data interpretation (this crucially depends on knowledge of the data collection process and methodology; see HANDS for information that needs to be documented).

4.4.2   Monitoring and Validation

You can protect the scientific integrity of your study by consistently documenting the data entry process, i.e., who enters or modifies a particular data element at what location and time. This is mandatory for formal clinical trials. You should preferably store this information within the software that you are using. Many software packages do this automatically in the so-called audit trail. In addition, it is advisable to implement a method for validating and cleaning the data after initial entry and to decide when a dataset will be locked for the start of analysis. This may be done by having a second person check entered data, producing data quality reports, extensive internal consistency logic, double data entry, or by comparing the data with the primary source (e.g., an electronic patient file).

4.4.3   Metadata

Metadata is 'data about data', i.e., all information that is required to interpret, understand, and (re)use a dataset [4, 5]. Metadata include:

- the name of the dataset or research project that produced it;
- names and addresses of the organisation or people who created the data;
- identification numbers of the dataset, even if it is just an internal project reference number;

- key dates associated with the data, including project start and end date, data modification dates, release date, and time period covered by the data;
- the origin of all data (i.e., data provenance description; the origin of the data should be verifiable);
- the protocols that were used including experimental aspects and study setup (e.g., persons, standard operating procedures, conditions, instrument settings, calibration data, data filters and data subset selections), since this is all essential for data reuse and data quality verification;
- unambiguous descriptions of all major entities in the study, such as samples, individuals, panels, or genotypes.

Collecting metadata will help you and your collaborators to understand and interpret the data. In addition, other people need metadata to find, use, properly cite, or reproduce the data, ensuring the long-lasting usability of the data. To improve reusability, you should consider collecting more metadata than required for your own research question, such as the geographical area of data collection, instruments used, demographics, and the time between collecting samples and performing measurements. In addition, you should consider interoperability and therefore use standardised terminologies in your metadata. There are many minimal metadata standards for this purpose (e.g., the MIT Libraries' guidelines). Metadata and data should be stored close to each other to make sure that the association between the two is clear. Metadata can be stored as embedded documentation, supporting documentation, or as catalogue metadata.

4.4.4   Security

You should implement state-of-the-art safety measures to preventing unauthorised and unnecessary access to your research data by:

- setting internal and external access policies at the start of your study (i.e., who gets access to which data);
- protecting your data with passwords (use a proper password management system);
- protecting your data from computer viruses (ask your institution's ICT helpdesk);
- using firewalls, encrypted data transport, and backups;
- installing a Data Access Committee to review all data and sample requests.

4.4.4.1   Access Policy

Access policies are part of your DMP, so they should be described before starting data collection. One reason for this is that, in many cases, patients have to give informed consent on data sharing before you start collecting data. In case of a clinical trial, a substantial change in access policies should lead to an amendment of the ethical protocol. Important aspects are:

- never allowing access to personal or clinical data to unauthorised people;
- under no circumstances granting access to (in)directly identifiable data via computer accounts shared by multiple persons;
- verifying the identity of the user logging into a database with (in)directly identifiable data preferably by at least one other method than just password security ('2-factor authentication');
- not providing more information in a data extraction than needed for a particular analysis;
- making sure that access to the database is logged properly.

Any access outside the authorisations in the access policy should be considered unauthorised access. You should be able to detect unauthorised access timely. Note that there is a legal obligation to report personal data leaks in most countries.

4.4.4.2 Protecting Research Data

You should think of these safety measures to protect your data:

- Storage of research data has to be safeguarded primarily under the regulations that apply in your country. The system and its environment should preferably be ISO27001 certified, or at least meet the underlying goals of this legislation.
- A database manager should be able to differentiate data access to parts of the collection per individual via role-based accounts.
- Databases connected to the internet should not contain identifiable data unless the infrastructure has taken sufficient measures to reduce the risk of access to the identity of a subject to an extremely low level.
- Storage that could legally be traced back to a non-EU owner or any non-EU party with access to the data or its physical location requires additional measures such as including it in the informed consent.

4.5 Analysing Data

Properly preparing your research data for analysis and working with a statistical analysis plan will result in a transparent analysis and interpretation process and reproducible results. In addition, it will make your data, intermediate results, and end results suited for archiving and sharing.

4.5.1 Raw Data Preparation

Prepare your research data for analysis by following these steps:

1. Create a data dictionary (i.e., metadata).
2. Create a working copy of the dataset and securely archive the raw data.
3. Clean the data in the working file and document all cleaning steps in a separate file that is archived.
4. Create an analysis file and preserve the cleaned dataset for archiving purposes.
5. Preserve your raw and (if needed) intermediate datasets.

When your data cannot be traced back to individuals (i.e., anonymised data), it is possible to use any decent statistical package as the management tool for your data. However, you should make sure that the entire process is well-documented and that all data manipulations are documented in libraries of syntax files. It is important to name and organise files in a well-structured way because the files can easily become disorganised. A naming convention saves time and prevents errors. If you have a large number of files or very large files, you should keep a master list with critical information. The master list should be properly versioned, so that all changes are registered over time along with their reason.

It is advised to store the raw data and all versions after meaningful processing steps that you cannot easily repeat. At least store the raw data that you use as the basis for your publications, including the descriptions of how you obtained these data and how you processed them (i.e., the metadata). You can consider deleting intermediate files to save storage space and to reduce the risk of inadvertent privacy violations. They can also be excluded from a backup scheme to save time on a possible restore after hardware failure. However, it may be useful to keep intermediate data for trace-back reasons.

4.5.2   Analysis Plan

In more complex studies, you should make a data analysis plan prior to starting the analysis, but it is preferable to already make the plan before you even start collecting data. The plan should at least address the following topics:

- the research question in terms of population, intervention, comparison, and outcomes;
- a description of the (subgroup of the) population that is to be included in the analyses (in-and exclusion criteria);
- which datasets are used and if applicable, how datasets are merged;
- data from which time point (T1, T2, etc.) will be used, if applicable;
- variables to be used in the analyses and how these will be analysed (e.g., continuous or categorical);
- variables to be investigated as confounders or effect modifiers and how these will be analysed;
- missing value treatment;
- which analyses are to be carried out in which order.
- structuring of folders and files, and managing of file version control

You may need to consult a statistician about the choice of statistical methods. You may also consider a workflow system rather than running each analysis step by hand. In addition, you may consider distributed analysis, where data remains at its original location.

4.6   Archiving Data

Scientific data archiving refers to the long-term storage of scientific data and methods. The FAIR principles recommend archiving research data in a trusted and secure environment at your institution or at an external data service or domain repository.

4.6.1   Archiving: What and How?

How much data and methods you must store in a public archive varies widely between scientific disciplines, scientific journals, and research funders. Nowadays, many scientific journals demand open access of the raw research data. The Horizon 2020 programme of the European Commission has recently developed Guidelines to the rules on Open Access to Scientific Publications and Open Access to Research Data (Fig. 4.2). Clinical trial data should always be accessible to monitoring bodies (e.g., internal audits). Research data should be preserved as long as the potential value is higher than the archival and maintenance costs.
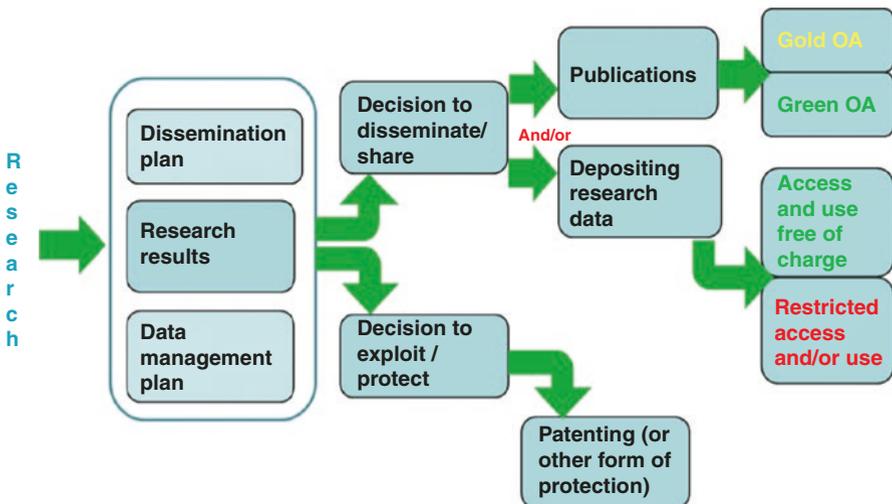
Fig. 4.2 From: Guidelines to the rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020

4.6.2 Archiving: Where?

The existence of research data should be clear to potential re-users. To this end, you should at least archive the data at your home institution. Frequently used data types may be submitted to worldwide archives (repositories). Please consult HANDS for a list of institutions that offer general data repositories as well as domain specific repositories (e.g., for genomics and microarray data, or the BBMRI catalogue for data and sample collections). Data that is archived outside your own institution (e.g., at an international data service or domain repository) should be registered at your home institution and the data should be listed in an open data catalogue.

4.7 Sharing Data

Clinical researchers should always share their data with monitoring bodies upon request. In addition, many research funders request that researchers share some or all of their data with the public and other researchers. Sharing with third parties can range from 'data is findable, but not accessible' to 'data is findable and accessible for everybody for all purposes'. Sharing policies cannot lead to open medical data, unless the data is truly anonymous. The guiding principle is responsible data sharing and protecting the privacy of study subjects.

4.7.1 General Considerations

Your data sharing policy should be tailored to your research project and is affected by the following questions:

- Did the study subjects give permission to share or combine their data? Does the consent mention specific conditions for data sharing?
- How were the data created and how does this affect data sharing (e.g., methodology, protocols, and publications)?
- What type of data will be released? Is there a procedure for data release with, for example, a committee?
- Who would be the recipient of the data?
- What warranties will the recipient give about responsible use of the data?

External access most often means the transfer of datasets under certain conditions (restricted access). If you will obtain the data as part of a research collaboration, the Intellectual Property Rights and openness of the resulting data should be discussed between the partners before you start collecting data. Relevant factors are:

- the consent modality (i.e., is there informed consent and what does it state?);
- the approval of the research by the designated competent body;
- the conditions of the funders of research data;

- the conditions under which data were released by the original creator of the data;
- the conditions of the journal to which the data is submitted (more and more journals demand open access to the underlying data).

4.7.1.1   Anonymity

Anonymity is an important condition of biomedical research, making it impossible to identify the person behind the data. Anonymous data may become identifiable when datasets are combined; you should consider this before sharing data. The solutions to this issue are:

- aggregate the data to such a level that they are never identifiable, irrespective of how you combine the data with other data.
- give access only within the data infrastructure of the original researcher. The new researcher may add data to this infrastructure, but data are only exported when meeting strict, previously determined conditions.
- create a balanced system of Data Transfer Agreements, corresponding to the type of data that are released, legally obligating the receiver to take responsibility to not re-identify the data.

Having said that, complete anonymity seems almost impossible in the age of digital information technology. By combining data from different sets, it is according to some only a matter of time until every individual can be identified in a so-called anonymous set. In addition, personal data sometimes need to be part of a dataset in order to allocate later events to the same person. In that case, you need to take extra measures to secure the privacy of the study subjects to be GDPR-compliant.

4.7.2   Sharing with Commercial Parties

Research data may only be shared with an external commercial party if the patient has provided informed consent for this. You should not hand over exclusive rights to reuse or publish your research data to commercial publishers or agents without retaining the rights to make the data openly available for reuse.

4.8   Conclusion

Adequate research data stewardship has become an indispensable part of clinical research. It is not a goal in itself, but it leads to high quality data and increased data sharing, thus promoting knowledge discovery and innovation. Hence, research funders and scientific journals have formulated guidelines on data stewardship. In addition, adequate data stewardship is necessary to meet legal and ethical

requirements. With the growing role of patients as important stakeholders in clinical research, it is expected that the (re)use of data will become a more transparent and democratic process in the years to come.

References

1. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR guiding principles for scientific data management and stewardship. Sci Data. 2016;3:160018. https://doi.org/10.1038/sdata.2016.18.
2. Wilkinson MD, Sansone S, Schultes E, Doorn P, Bonino da Silva Santos LO, Dumontier M. A design framework and exemplar metrics for FAIRness. Sci Data. 2018;5:180118. https://doi.org/10.1038/sdata.2018.118.
3. Boeckhout M, Reuzel R, Zielhuis G. The donor as partner – How to involve patients and the public in the governance of biobanks and registries. Leiden: BBMRI-NL; 2014.
4. Australian National Data Service. Guide on metadata. 2016.
5. UK Data Service. Document your data.