DATA MANAGEMENT PLANS (DMPS) IN A NUTSHELL

Table of Contents

Data Management Plans: what librarians should know

This document discusses several key concepts around Research Data Management and Data Management Plans (DMPs) for librarians who do not work in this field. These concepts are useful to know and help librarians to support researchers more extensively with their data management.

By understanding these concepts, librarians can provide essential support to researchers, fostering a robust and efficient research data management environment. This not only enhances the quality and impact of the research but also ensures compliance with funding agencies' requirements and facilitates future data sharing and preservation efforts. This document briefly explains some key concepts around DMPs and refers to practical applications where relevant.

Librarians who have developed their professional skills in different areas, such as cataloguing, acquisitions, training or even circulation, are increasingly willing to get engaged in new tasks. Those devoted to RDM must be enabled to play a role at the core of the research process. Therefore library professionals are facing reskilling and upskilling challenges to be better prepared for services regarding RDM. While librarians change their roles, sometimes applying to new job categories, organisations are transformed and new teams and interactions with researchers arise.

The LIBER Strategy 2023-2027 understands libraries as “engaged and trusted hubs of their user communities” and identifies “advancing Open Science”, providing “state-of-the-art services” and “upskilling the library workforce” as some of its spearheads (LIBER, n.d.). Research Data Management is definitely an area where libraries can develop “a new leadership role” (LIBER, 2020) guiding and supporting researchers during the whole research process and the research data lifecycle.

Introduction to the concepts of RDM and the RDM lifecycle

Research Data Management (RDM) is a process or a set of practices for handling research data: how you collect, produce, store and share the data during and after the research project. RDM practices are thus embedded in the research and research data lifecycles, which are not identical but intertwined. The **research lifecycle** includes, in a way, the research project as a whole from the first ideas to funding and planning the project to publishing the results, whereas **research data lifecycle** focuses on part of the project, namely the data. Examples of data management actions include defining user rights, producing data privacy documents, backing up, documenting procedures, describing the contents, disposing of, publishing, archiving, etc.

Paying attention to data management supports good scientific practices. Good data management ensures high-quality and reliable research results, minimises risks of e.g. data loss and data breach, and enhances the reuse of data.

In RDM practices, the FAIR principles (Findable, Accessible, Interoperable, Reusable) are important (Wilkinson, 2016). These principles provide guidance for research data management and are relevant to all stakeholders (researchers, funders, research support staff) in the research data ecosystem. Research libraries can use the FAIR Principles (GO FAIR, n.d.) as a framework for fostering and extending research data services. The LIBER Working Group for Research Data Management has provided a factsheet for libraries to get started (LIBER Europe, 2024).

Planning the research process: Data Management Plans

Data Management Plans (DMPs) are documents in which researchers write down their plans for managing their data, from the data collection to the archiving and publishing phases. DMPs typically include all the topics outlined below. DMPs are often required by grant providers, such as Horizon Europe and many national funders. Funders usually have templates for DMPs, as do most universities. Many research libraries also offer support for writing DMPs by dedicated staff, such as RDM specialists, data librarians or data stewards. Some universities may also offer this support on the faculty level or in the research office. These RDM teams usually also support researchers with writing and submitting DMPs. Many institutions offer a tool for writing DMPs such as DMPonline, DS Wizard, Argos.

Active data management

The phase of active data management encompasses the ongoing processes of collecting, analysing, and collaborating on research data. This is crucial for ensuring the integrity, accessibility, and usability of data throughout its lifecycle. For librarians supporting researchers in this domain, this involves understanding several core concepts and tools.

- **Data Collaboration Tools (storage, sharing, writing, management)**: Collaboration is a crucial aspect of modern research, frequently involving multidisciplinary and multi-institutional teams. It is therefore essential for researchers to be familiar with platforms that enable secure data sharing and collaborative analysis, such as cloud storage solutions, collaborative coding environments (e.g. Github) and data repositories (e.g. Zenodo). Furthermore,

collaborative writing (e.g. Office 365, Google Docs, Overleaf) and reference tools (e.g. Zotero, Endnote, Citavi) are important for researchers' everyday work. Librarians can play a pivotal role in being aware of and recommending these tools. The mentioned tools are usually provided by an institution's IT-infrastructures.

- **Data Security, Backup and Ethics**: The protection of sensitive data is a fundamental aspect of active data management. Librarians should ensure that researchers are aware of institutional policies and best practices for data security, including encryption, access controls, and compliance with ethical standards and regulations. The actual security and backup measures are usually implemented by an institution's IT-infrastructures.
- **Electronic lab notebooks (ELNs) and laboratory information management systems (LIMS)** are two distinct types of software that facilitate the documentation, organisation, and sharing of experimental data. ELNs (e.g. Jupyter) are digital versions of traditional lab notebooks, while LIMS (e.g. Labware) are more comprehensive systems designed to manage laboratory samples, associated data, and workflows in laboratory environments.
- **Data Versioning**: As with software development, the tracking of changes and the retrieval of previous versions, if necessary, needs to be done for research data as well. This is vital for the reproducibility and transparency of research. Librarians should therefore promote best practices for data versioning, including the use of repositories and tools that support version control (for example Git or SVN). The mentioned tools are usually provided by an institution's IT-infrastructures, or set up individually by research groups.

Documentation and metadata

Librarians often have extensive experience with bibliographic metadata and can help researchers with adding metadata to their project, specifically on publication level. In the research domain, metadata are similar to bibliographic metadata, but can also describe the data on a variable level. Another important way to describe data is through documentation.

Data documentation is essential for understanding and re-using research data. It involves describing the data throughout the research process and is best created by the researchers themselves. Without documentation and metadata, data are meaningless. Documentation helps outsiders and the researchers themselves understand and use the data effectively. Researchers should document their work on three levels: project level, file level, and variable level.

Examples of documentation include:

- Code books, lab books, field diaries, notes
- Descriptions of instrument settings and calibrations
- Methods used
- Readme files describing data origin and contents
- Administrative documents like research plans, data management plans, contracts, research permits, scientific publications, and licences

Documentation practices vary across disciplines and depend on project needs. Good documentation makes the research understandable, minimises risks of misinterpretation, and facilitates smooth archiving. It is also essential for validating results and replicating studies.

As with publications, *metadata*, or "data about data," include information necessary for

understanding, interpreting, and using the research data, such as data origin, collector, time, place, methods, and subject words. Metadata are a crucial part of documentation and must be structured (i.e. use controlled thesauri) and machine-readable to comply with the FAIR principles, enabling data transfer between different services. Documentation and metadata furthermore contribute to the FAIR principles.

The Consortium for European Social Science Data Archives gives a more detailed description of documentation and metadata in the social sciences (CESSDA, 2020). Similarly, the ELIXIR RDM Kit gives some guidelines for the life sciences domain (ELIXIR, 2021). There are more guidelines like these.

Preserving and sharing/publishing data

After the research project ends, data need to be preserved for a number of years. This is usually required in university and funder policies, but according to national law, may also be legally required, for example in the case of medical research. The preservation term can differ according to the situation.

When preserving data, it is important that data can be found and retrieved, thus contributing to the FAIR principles. When datasets receive a persistent identifier (PID), such as a DOI, this makes data more findable. It is also important that data integrity is guaranteed, that is, that data cannot be modified after depositing, so that there is a snapshot of the data that were actually used in the research. However, data do not always need to be openly accessible, and in the case of confidential data, should not be. When data are confidential, they can be preserved as a closed or restricted dataset, but in that case, the metadata should be available, so that the dataset can still be found. This ensures that data can be verified. If data are restricted, the metadata should also specify the conditions in which data can be accessed, and by whom. Additionally, openly accessible data should have a licence that gives information about how and under what conditions data can be reused.

Most universities and research-performing institutions will offer repositories that enable FAIR data preservation, and have people that support researchers with choosing and using them. Domain-specific data repositories also exist. RDM experts in the library can advise researchers what repository suits their needs best.

Another way to publish data, besides an institutional or domain-specific repository, is as a data paper in a data journal. A data journal is dedicated to publishing (descriptions of) datasets and how they may be used or reused. Data journals are not common in all fields, but can offer researchers an opportunity to put their dataset in the spotlight.

Here are some commonly-used terms that come up in data preservation talk:

- Open: data that are openly available to the public, with minimal reuse restrictions.
- Restricted: data that (usually due to confidentiality concerns) are not openly available but may be accessed and/or reused under certain conditions. Metadata can be open.
- Closed: data that (usually due to confidentiality/sensitivity concerts) are not openly available and also cannot be accessed, except for verification purposes. Metadata can often be open.

- Preserving: storing data in a fixed state for the medium (10 years) term or for long-term (50+ years) archiving.[1]
- Publishing data: making data and metadata openly accessible for the public, e.g. as a deposit in a repository or as a data paper in a data journal.
- Deleting: destroying data, for example when the preservation term expires or according to contractual obligations.
- Persistent Identifier (PID): a unique code that can refer to an object (e.g. a dataset), in the case of a DOI or handle, or to a person, in the case of an ORCID. PIDs are unique, so there can be no confusion, and they will reliably resolve to the current page where the information lives.

RDM Support in Libraries

Libraries are key partners in Research Data Management since they act as providers of support services, infrastructures and training. Many research libraries also employ dedicated staff that can support researchers with writing their DMPs. Therefore libraries are increasingly liaising with researchers; they are key agents in the scientific ecosystem as they connect to other services at the institution, such as the research offices, the IT offices or the legal services, as well as to the management boards at various levels. This role is becoming more strategic, and is a responsibility of the library directors and the heads of the library units.

The LIBER-ADBU toolkit for RDM support services (2023) provides an overview of possible stages of development for RDM support in research libraries. It focuses on a number of aspects of RDM support, including the following:

- **Organisational structures**. In initial stages libraries tend to start including RDM services in generic research support services. In the development stage, some libraries choose to create a dedicated Open Science team where RDM supports the whole research community. Others prefer a different approach, integrating RDM in the tasks of all librarians. In more advanced stages, RDM becomes discipline-oriented where there is a central team for coordination purposes and support by embedded data stewards. These data stewards might not be librarians but people employed directly by research-performing units.
- **Job categories**. The following categories are not necessarily specific jobs, since the same data specialist can perform a variety of tasks:
    - Data librarians or stewards, sometimes still considered as information specialists, focus on advisory roles, whether they work at central units at the libraries or get embedded within research teams.
    - Data managers' roles go beyond generic support and take operational roles within research teams freeing researchers from these tasks.
    - Data curators take care of the quality of metadata and data before they are available on the repository; they also check compliance with data protection regulations and make sure licensing is correctly applied. FAIR compliance is under their responsibility.
    - Educators/trainers. RDM training is increasingly important.
- **Competences**. Beyond generic competences, RDM experts must achieve (and get trained for):

[1] Archiving and preserving are not always very precisely defined and their use can differ in institutions, tools and contexts.

- Knowledge on the institution, the research process, the planning and design phases and the understanding of data management and description, infrastructures, data protection and licensing, among others. The focus here is on an advisory and educational role.
- Know-how of technical and operational tasks regarding technology, formatting, storage, metadata, analysis and visualisation, and FAIR compliance, among others.
- Inter-personal skills and soft skills, from communication and teamwork, to pedagogical competences and the ability to set, monitor and evaluate protocols and procedures.

References

CESSDA (2020). *Data Management Expert Guide.* Documentation and Metadata. https://dmeg.cessda.eu/Data-Management-Expert-Guide/2.-Organise-Document/Documentation-and-metadata

ELIXIR (2021). *Research Data Management Kit. A deliverable from the EU-funded ELIXIR CONVERGE project (grant agreement 871075).* https://rdmkit.elixir-europe.org

GO FAIR (n.d.). *FAIR Principles*. https://www.go-fair.org/fair-principles/

LIBER Europe (n.d.). *Strategy 2024-2027*. https://libereurope.eu/wpcontent/uploads/2022/01/LIBER_STRAT_A5_digital-final-1.pdf

LIBER Europe (2020). *Libraries & Research Data: Towards a New Leadership Role*. https://libereurope.eu/wp-content/uploads/2020/11/LIBER-Libraries-and-Research-Data-factsheet1-1.pdf

LIBER Europe, ADBU (2023). *Research Data Management Support Services by Libraries in Collaboration with Other Stakeholders*. https://libereurope.eu/wp-content/uploads/Liber_ABDU1920X1080_FINAL.pdf

LIBER Europe (2024). *A Guide to FAIR Practices In Research Libraries*. https://libereurope.eu/wp-content/uploads/A-Guide-to-FAIR-Practices-in-Research-Libraries_2024.pdf

Wilkinson, M. D. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3(1), 160018-160018. https://doi.org/10.1038/sdata.2016.18

These guidelines have been compiled by LIBER's Research Data Management Working Group. We are grateful to the Information Sciences students of the Data Management course at the Fachhochschule Graubünden and our colleagues at different institutions for providing their feedback.