Research Data Managment

Info Sheet I **Data & Research**

May 2026

Good research data management (RDM) practices can help you to make your research more reproducible, plan for infrastructure needs, avoid data security/privacy risks, and facilitate the publication of well-documented and reusable data packages at the end of your project. This Info Sheet serves an introduction to the different aspects of RDM, and to help you get started. **If you want additional help with any of these topics, you can contact the Lib4RI RDM team, and we'll see how we can best support you, or set up a training for you/your team on that topic.**

Contact & Support

In person	Monday-Thursday at the Lib4RI office in Dübendorf.
	Upon request in Birmensdorf+Villigen
Email	data@lib4ri.ch
Web	https://www.lib4ri.ch/research-data-management

Plan before you collect: data management plans

The first step in managing your research data is writing a data management plan (DMP). Many funding agencies require you to submit a DMP for your project (for example: SNSF, Horizon Europe). More importantly, a DMP is a roadmap for the collection, processing, publishing, and archival of your research data. A good DMP should help you avoid preventable risks, anticipate infrastructure and costs, as well as aid you in publishing reusable data packages at the end of your project.

Data management plans provide answers to questions like:

- What data will you collect, how will you collect it, and what **file format and size** will the data have?
- What **metadata** must be collected to make your data understandable to others, and what existing standards will you use?
- Are there any **security or ethical limitations** for your data collection, storage, or publication that must be considered?
- **Who will be responsible for the data**, both during and after the project duration?
- How will you **publish** your data at the end of your project, and how will you **archive** it so it remains available beyond the project end?

License your outputs clearly

Without a licence, others legally can't reuse your data or code. Pick one as open as possible, as closed as necessary.

- **For data** use a **Creative Commons licence:**
  - **CC0** — No Rights Reserved. Maximises reuse.
  - **CC BY** — Free to use, share, and adapt with attribution. The most common choice for open research data.
  - **CC BY-SA** — Same as CC BY, but derivatives must use the same licence ("ShareAlike").
  - **CC BY-NC** — Free to use and adapt with attribution, but not for commercial purposes. Can limit industry collaborations.
  - **CC BY-ND** — Free to share with attribution, but no derivatives.Users cannot modify or build on the data.
- **For code** use a permissive licence like **MIT** — Only restrict reuse where you genuinely must (privacy, prior agreements).
- **Want to explore more options?** Compare options at choosealicense.com

Make your code reproducible

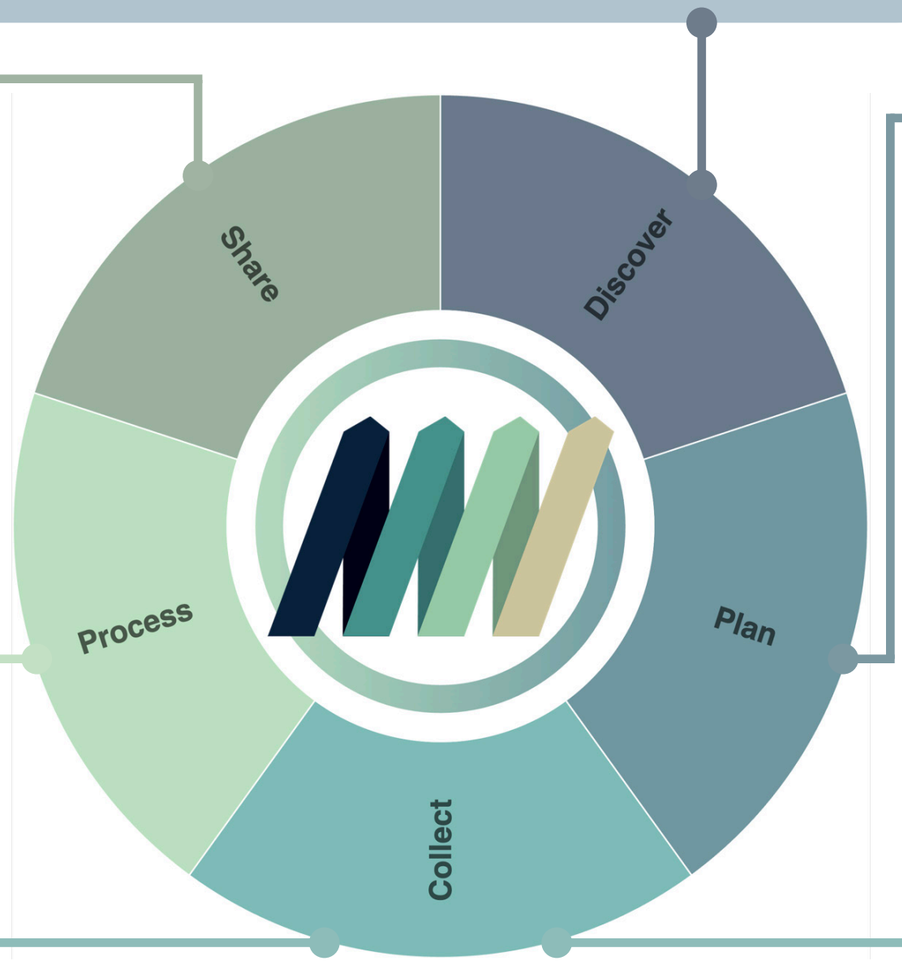
Code that only runs on your laptop isn't really finished. Aim for code another researcher can clone, install, and run.

- **Comment generously** — explain **why**, not just **what** you chose.
- **Use version control** (git) and **push to an institutional or trusted server** (Institutional GitLab, Codeberg, GitHub).
- Pin your dependencies with a **virtual environment**: `renv` for R, `venv` for Python).
- For complex setups, containerise with Docker or Podman.

Protect your data using the 3-2-1 rule for storage

Hardware fails, accounts get locked, laptops get stolen. Redundancy is cheap insurance.

- Keep **3 copies** of your data, on **2 different storage types,** with **1 copy off-site**.
- Make sure **one named person owns the storage setup** — not "the team."
- If you're handling personal or sensitive data, **talk to your institute** before choosing a cloud provider, and follow their guidance on **anonymization, secure storage, and sharing policies**.
- **Restrict access** to sensitive data to authorised users only.



Discover and publish data in a FAIR repository

A repository gives data a permanent home, a DOI, and discoverability.

- **Use your institutional repository** when possible:
  - **ERIC/open** (Eawag, *opendata.eawag.ch*)
  - **SciCat** (PSI, *discovery.psi.ch*)
  - **EnviDat** (WSL, *envidat.ch*)
- **Field-specific or general repositories**
  - Search **re3data.org** for a field-specific repository
  - Use **Zenodo** as a general fallback
- **Crossreference DOI's:** Link the *data DOI* in your paper **and** link the *paper DOI* in your dataset's metadata.
- **Plan for archival separately from publication:** Confirm how long the data must be kept and who is responsible after the project ends.

Generously document every folder

If a colleague opened your project folder tomorrow, could they understand it? Documentation is what makes the answer "yes."

- Add a plain-text (.txt) or Markdown (.md) **README** to every major folder, written for humans.
  - **General information** — title, authors (with ORCIDs and contacts), date and location of collection, funding sources.
  - **Sharing & access** — licence, recommended citation, links to related publications and datasets, any source data it was derived from.
  - **File overview** — list of files and folders with short descriptions, relationships between them, and version history.
  - **Methodology** — how data were collected and processed, instruments and software used (with versions), and quality-assurance steps.
  - **Data-specific details** — for each file: variables, units, row counts, missing-data codes, and any abbreviations.
- Add a structured **metadata** file (JSON, YAML, or XML) listing **variables**, **units**, and **formats** for machines.
- **Update documentation** as the project evolves, not at the end.
- Use an existing template for README or metadata files.

Agree on a file structure early, and enforce it often

A consistent folder layout and naming convention within and between your projects is a gift to collaborators — and your future self.

- **Use top-level folders that match the lifecycle**:

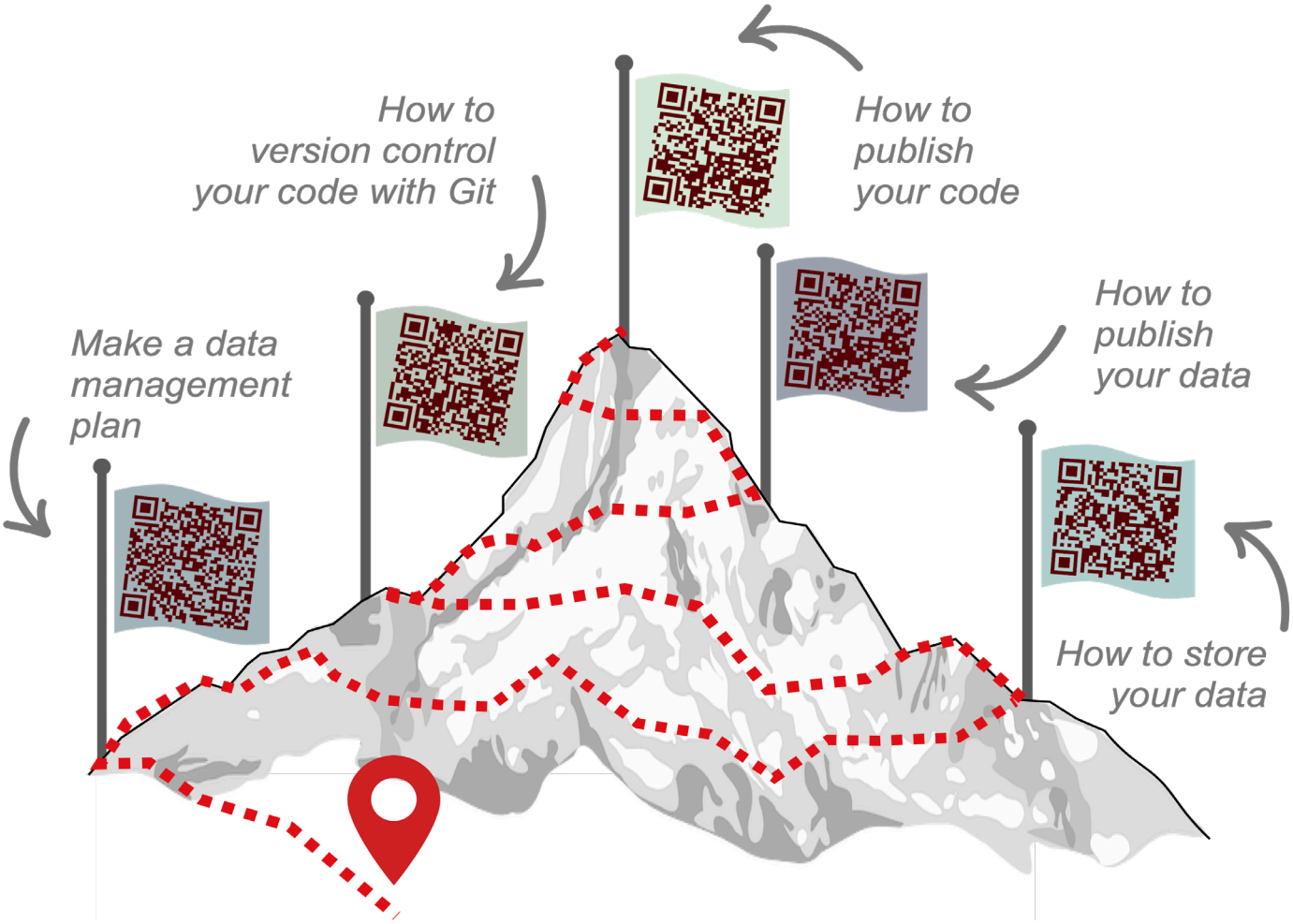
```
project_folder/
├── README.md
├── LICENSE.txt
├── documentation/
├── raw_data/          ← Treat raw data as read-only
├── processed_data/
├── analysis_code/
├── figures/
└── reports/
```

- Name files with a fixed pattern, for example:

  YYYYMMDD_Project_Type_v1.ext
- Prefer open formats:
  - **CSV/TSV** (tables)
  - **TXT/MD** (text)
  - **TIFF/PNG** (images)
  - **PDF/A** (reports)
  - plain **`.R`/`.py`** (code).

Library for the Research Institutes within the ETH Domain: Eawag, Empa, PSI & WSL

Resources



Still have questions?

Research Data Management Courses

Get started with various RDM topics.

On-Demand Trainings

Request a training tailored to your project / group.

One-on-one Consulations

Get expert advice about your specific questions.