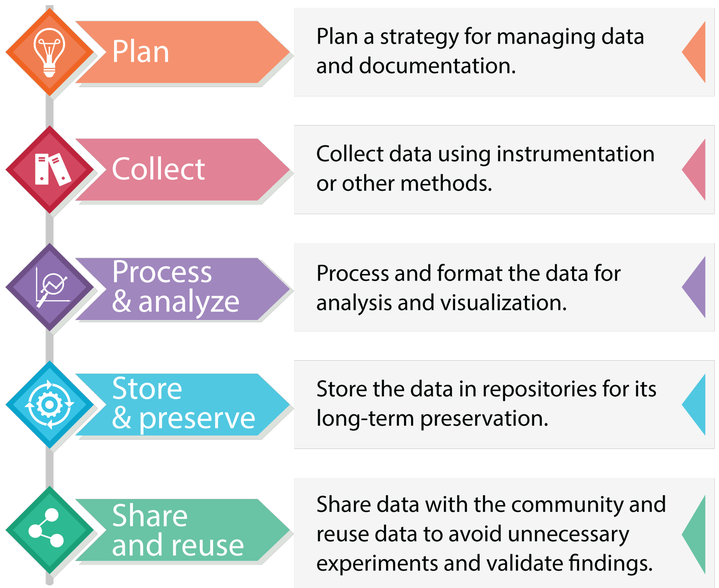
Digital Research Alliance of Canada | Alliance de recherche numérique du Canada

Sharing high-quality research data and reproducible workflows

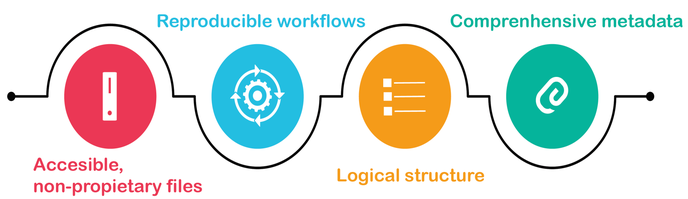
Research data curation team (Federated Research Data Repository)

Keywords — Research data management, Open science, Data sharing

I. Research data life cycle



II. Making research data understandable



Sharing research data comprises systematic procedures to make it discoverable, understandable and reusable for the community. It also ensures reproducible workflows to replicate and validate research findings.

.

i. ***Prioritize open, non-proprietary file formats***

Data Type	Recommended Formats	Notes
Tabular data	CSV, TSV	Widely supported, non-proprietary
Text documents	TXT, PDF/A	PDF/A is a long-term archiving format
Images	TIFF, PNG	TIFF for lossless, PNG for smaller size
Geospatial data	GeoTIFF, Shapefile	Ensure all Shapefile components are included
Audio data	WAV, FLAC	WAV for uncompressed audio
Video data	MP4, MOV	MP4 is widely compatible, but lossy
Code	PY, R, QMD	Provide well-commented code and scripts

ii. ***Implement reproducible workflows***

Use open-source/free software to handle images and data. Employ scripts for batch processing and enable reproducible workflows.

Fiji (Image processing) **QuPath** (Image processing) **R/R-studio** (Data analysis) **Python** (Data/image analysis)

iii. ***Organize your files logically***

The files must be easy to retrieve and identify for any potential reuser.

- Build a logical folder structure hosting raw and processed files, analysis code and final outputs.
- Use descriptive naming conventions to label folders and files.

Example project / Tree generator

TIER Protocol 4.0

```
Project/
├── The Readme file
├── The Report
├── Data/
│   ├── InputData/
│   │   ├── Input Data Files
│   │   ├── Metadata/
│   │   │   ├── Data Sources Guide
│   │   │   └── Codebooks
│   ├── AnalysisData/
│   │   ├── Analysis Data Files
│   │   └── The Data Appendix
│   └── IntermediateData/
├── Scripts/
│   ├── ProcessingScripts/
│   ├── DataAppendixScripts/
│   ├── AnalysisScripts/
│   └── The Master Script
├── Output/
│   ├── DataAppendixOutput
│   └── Results
```

iv. ***Implement comprenhensive metadata***

Describe and contextualize your data the data using **metadata**, **README** files and **codebooks**. This enables the correct re-use of the dataset. Think about the following items when building documentation:

Download a Readme template

- **Authors** and **content of the dataset**: These enable users to understand the dataset independent of any associated research articles. Datasets are stand-alone objects usable on their own.

```
Example from https://doi.org/10.20383/103.0802
The dataset contains characterization and activity data for a series of carbon nitride
electrode samples. Carbon nitride was grown on FTO, plasma-cleaned FTO, and aminosilanized
plasma-cleaned FTO . Pt was also added on top of the CNx for some samples...
```

- **Access/License information:** Describe how the community can reuse the data. Creative Commons offers a wide variety of license options. You can refer to the templates created by Open Brain Consent when working with sensitive Neuroimaging data.

Open Brain Consent

- **Methodological information:** Describe how the data was generated. This contextualizes and allows for data validation.

```
Example from https://doi.org/10.20383/103.0802
Electrochemical measurements were conducted in a quartz electrochemical cell using a
Palmsens4 potentiostat. The electrolyte was fully purged with Ar between scans to create
an inert electrochemical environment...
```

III. Share your data in scientific repositories

These good data management practices make datasets suitable for sharing in repositories like FRDR (https://www.frdr-dfdr.ca/) and Borealis (https://borealisdata. ca/). FRDR is a curated, bilingual, multi-disciplinary repository exclusively for Canadian researchers that specializes in large data deposits. Borealis, the Canadian Dataverse Repository, is a bilingual, multi-disciplinary, secure, research data repository supported by academic libraries and research institutions across Canada.

FRDR Federated Research Data Repository — borealis

Canadian tri-agency RDM policy

The Government of Canada established in 2021 the Tri-Agency RDM policy for responsible conduct of research, including practices to implement data management plans (DMP) and data sharing from publicly funded science.