Digital Research Alliance of Canada:
Project Manuals, Tutorials, and User Guides

Tutorial/Manual: Research Data Management (RDM) Guide

Overview: This manual provides a step-by-step guide to effective Research Data Management, ensuring data is discoverable, accessible, and reusable throughout the research lifecycle.

Contents:

This manual will be used to discuss the following points in detail during our live sessions. The practical examples can be found in our practical GitHub repo.

1. **Introduction to RDM**
   - Definition and importance of RDM.
   - Overview of the data lifecycle: Plan, Collect, Process, Analyze, Preserve, Share, Reuse.
2. **Creating a Data Management Plan (DMP)**
   - Using the DMP Assistant to create a DMP.
   - Key components: data types, metadata standards, storage solutions, and sharing policies.
3. **Metadata and Documentation**
   - Importance of metadata in data discoverability.
   - Standards and best practices for metadata creation.
4. **Data Storage and Preservation**
   - Utilizing repositories like FRDR and Borealis.
   - Strategies for long-term data preservation.
5. **Ethical and Legal Considerations**
   - Understanding data licensing and sensitive data handling.
   - Compliance with the Tri-Agency Research Data Management Policy. Science.gc.ca
6. **Resources and Support**
   - Accessing support through the Digital Research Alliance of Canada.
   - Additional readings and tools.

Prepared By: Mohammad Alnabhan University of Ottawa

Digital Research Alliance of Canada: Project Manuals, Tutorials, and User Guides

1. Introduction to RDM

- Research Data Management (RDM) involves systematic planning and handling of data across its lifecycle.

- It is the structured process of organizing, storing, preserving, and sharing research data. It is vital for:

- Reproducibility
- Ethical compliance
- Maximizing research impact

More information in this video: Intro to Research Data Management

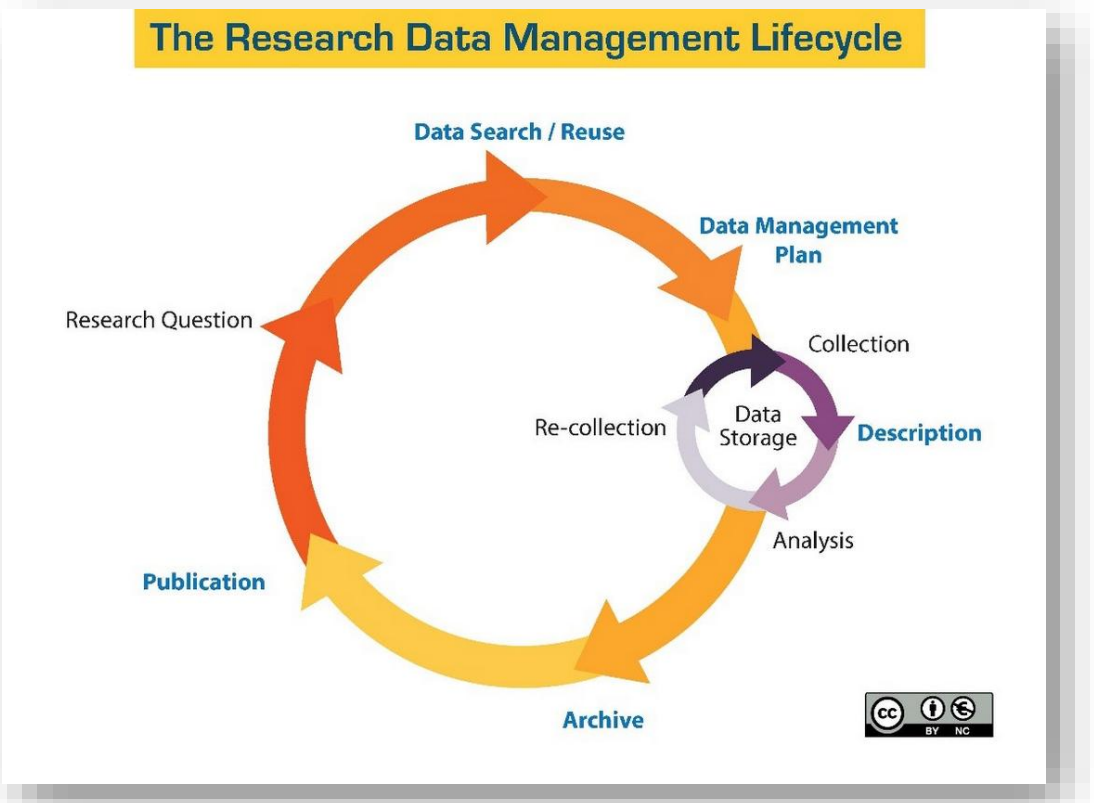
2. The Research Data Lifecycle

RDM occurs in every stage of the research lifecycle, not just at the end where all the data files are simply zipped up in a folder for storage. The image below depicts the RDM lifecycle in conjunction with the different stages of research.

- **Plan**: Identify data needs, tools, and formats
- **Collect**: Acquire or generate data
- **Process**: Clean and prepare data
- **Analyze**: Use tools like R or Python
- **Preserve**: Use repositories like FRDR
- **Share**: Apply licenses, metadata
- **Reuse**: Enable others to validate or extend your work

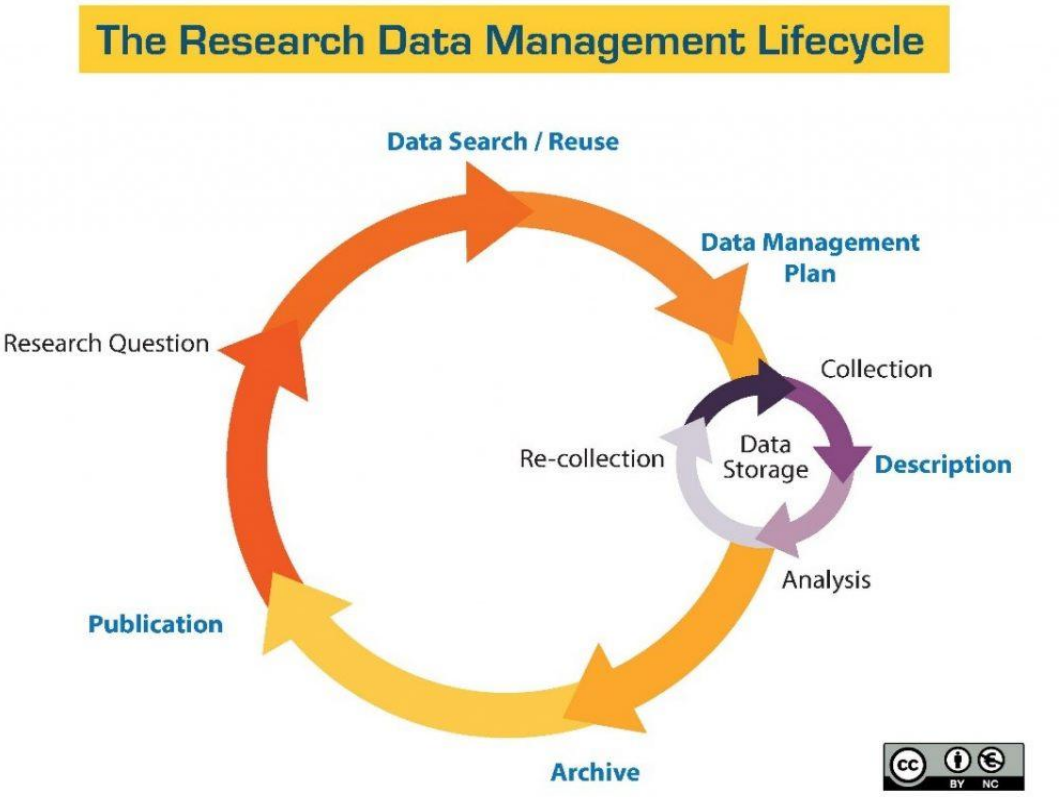
Digital Research Alliance of Canada | Alliance de recherche numérique du Canada

Digital Research Alliance of Canada: Project Manuals, Tutorials, and User Guides



Prepared By: Mohammad Alnabhan University of Ottawa

Digital Research Alliance of Canada: Project Manuals, Tutorials, and User Guides



The University of California, Santa Cruz, Data Management LibGuide, Research Data Management Lifecycle, diagram, viewed 18th April 2017 <http://guides.library.ucsc.edu/datamanagement>

3. Creating a Data Management Plan (DMP)

Use DMP Assistant to generate your plan.

Example Questions:

- What types of data will you collect?
- How will you ensure data security?
- How long will you retain the data?

Digital Research Alliance of Canada: Project Manuals, Tutorials, and User Guides

Sections of a DMP include:

- Types and formats of data
- Metadata standards
- Data storage and backups
- Ethical and legal considerations
- Sharing, licensing, and reuse

Code Snippet: Example README.md Template

```
# Project Title: Climate Change Sentiment Analysis
## Data Description
- Source: Twitter API
- Format: CSV
- Size: 50MB

## File Naming Convention
climate_tweets_YYYY-MM-DD.csv
```

4. Preservation, Storage and Tools

- **DMP Assistant**: Plan creation
- **FRDR**: Archival and publication
- **ORCID iD**: Researcher identification

5. Organizing, Cleaning, and Documenting

- Use script-based cleaning (Python/Pandas)
- Automate file backups with GitHub
- Use standard naming conventions.

Digital Research Alliance of Canada: Project Manuals, Tutorials, and User Guides

```
import pandas as pd

df = pd.read_csv('raw_data.csv')
df_clean = df.drop_duplicates().dropna()
df_clean.to_csv('clean_data.csv')

# Example Naming Convention:
projectname_datatype_YYYYMMDD_v01.csv
```

6. Ethical and Legal Considerations

- Anonymize sensitive data
- Use licenses (e.g., Creative Commons CC-BY)
- Follow Tri-Agency RDM Policy

7. Archiving and Sharing

- Use FRDR or a trusted institutional repository
- Prefer open file formats (CSV, TXT, PNG).
- Assign DOIs to datasets

8. Resources and Templates

- University of Ottawa File Naming Guide
- Cornell README Metadata Guide
- DMP Assistant Video Tutorial Series: Introduction to Data Management Plans
- Writing a Data Management Plan