Research Data Management: 10 things to consider before you begin your research
Hilda Teresa Ayala González

February 26, 2019
Graduate Research and Innovation Center (GRIC)
University of Puerto Rico, Mayaguez Campus

This activity is sponsored by the Transformational Initiative for Graduate Education and Research (TIGER), US Department of Education, Title V, Part B, Promoting Postbaccalaureate Opportunities for Hispanic Americans (PPOHA) Program (#P031M140035).

Goals
• Understand the role of data management in your research
• Recognize good data management practices through the data lifecycle
• Learn about tools, tips, and tricks to enhance your data management skills

Pre-test
Link: https://uprm.libsurveys.com/RDMPretest

Scan QR Code
Before we begin

- You don’t have to do all
- Start with one
- Do it right
- Do it consistently

**What is data management**

*It is the compilation of many small practices that make your data easier to find, easier to understand, less likely to be lost, and more likely to be usable during a project or ten years later.*

*(Briney, 2015, p.7)*
1. Why manage your data?

- **You**
  - Deeper understanding
  - Prevent loss

- **Others**
  - Research continuity
  - Expand collaborations

- **Funders**
  - Use of public funding
  - Data sharing policies

- **Journals**
  - Reproducibility crisis
  - Accountability

2. Data Management Planning

**Data Management Plans** are required by funding agencies

**Data Management Plan**

How you will manage the data during and at the end of the research

*Planning comes first, management is continuous*

**Checklist for a Data Management Plan**

http://www.dcc.ac.uk/resources/data-management-plans/checklist

2. Data Management Planning

Big picture

Policies:
- Privacy - HIPAA
- Retention – Institutional / Funder
- Ownership – Institution / team
- Copyright – Institution / team / subjects
- Sharing – Funders / Journals

Don’t panic… I got you!

3. Organization

Before you begin:
- Think about your project, the data, and documentation you will generate (and need) through the project
  - Types of file & space for storage
  - Versions
  - Retrieval options
  - Location
  - Analog / Digital files
  - Original-Reuse Content

Before you begin:
- Develop a system for organizing the files
  - Folders ➔ Subfolders
  - Don’t go to deep
- Design a file naming convention
  - Simple, up to 25 characters
  - Descriptive
  - No special characters or spaces
3. Organization

Before you begin:

- Select a reference manager for your literature, citations and bibliography
  - Example: Mendeley (Desktop, Web & Plug-in)

- Document your conventions SO YOU DON'T FORGET and to facilitate others navigating your files
  - Read me file

- For retrieval
  - Create an index → excel sheet for sorting
  - Screenshots of directories

- Don’t forget your lab notebooks
  - Cross reference
  - Digitize it!

- Add dates, use ISO Standard 8601
  - YYYY-MM-DD / YYYYMMDD
  - Good for sorting & accuracy

3. Organization: Example

File System
Samples: Project → Site → Date → Time
Lab notes: Project → Researcher → Date

File Naming
Sample: Guayanilla_20181018_AM_V01
Lab notebook: RiosPR_RamirezM_20181018

4. Formats

- Many, messy, unstable → proprietary → obsolete
  - YOU want your data to be available and usable

- Stick to well known, documented, and stable software

- What is common in your field?

- Are you done?
  - Revise constantly, migrate, convert
4. Formats

<table>
<thead>
<tr>
<th>Project and Data Management Planning</th>
<th>Data Acquisition</th>
<th>Data Analysis</th>
<th>Publication and Sharing</th>
<th>Data Preservation</th>
<th>Data Reuse</th>
</tr>
</thead>
</table>

Use this table to learn about formats and their longevity. [https://www.loc.gov/preservation/digital/formats/index.html](https://www.loc.gov/preservation/digital/formats/index.html)


5. Security

Access
- Who has access
- Roles
- Track changes → log

Concerns
- Unauthorized changes
- Data loss / stolen

Compliance
- Sensitive data
- Consent from participants

Keeping your data safe
- Lock pc
- Encryption
- PII → anonymization

Keeping your data safe can also help in compliance with regulations such as GDPR, HIPAA, and others.

6. Documentation

What do you need so your data makes sense?
- For you now
- For other
- For you in ten years

What do you need to validate your research?
- Good notes & spreadsheets, protocols, data dictionaries
- Software code & instruments data

Data needs context in order to be understood and use...data without documentation has no meaning. (Briney, 2015, p. 35)
6. Documentation

• Where do you document?
  • Laboratory notebooks
    • Keep them neat
    • Use a consistent system
    • Always add YYYYMMDD

• Read me txt files
  • With the data (same folder)
  • On a separate sheet (in workbook)

Follow the Metadata Standards


6. Documentation: Examples

Multiple places
• Personal computer
• Lab computer
• External hard drives
• Local server
• External server
• Cloud service

Multiple troubles
• Hard to find
• Hard to backup
• Hard to control
• Unsafe
• Obsolete
• Corrupted

Follow the Metadata Standards
7. Storage/backups

Storage strategy
- 3 copies
- 2 types of storage
- 1 offsite copy

Cloud strategy
Check the cloud before you sign a contract. Use this InterPARES Checklist

Backup strategy
- Manual or automatic
- Full
- Incremental
- Differential
- Cloud provider
- Backup, NOT syncing
- Frequency = daily, weekly, monthly
- Easy to replace?

• Test it
• Always keep two copies

8. Ownership

Before publishing or sharing data, ask yourself:
• Who owns it
  • You
  • Team
  • Funders
  • Institution
• Permission
  • Did you ask → sensitive data → human subjects

Look for:
- Policies – institutional
- Policies – funders

Careful with:
- Copyrights
- Patent
- Trade secrets
- National security

Recommended for data

9. Licensing

A license provides instructions to users

<table>
<thead>
<tr>
<th>License Type</th>
<th>Creative Commons</th>
<th>Open Data Commons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public domain</td>
<td>CC0</td>
<td>Public Domain Dedication and License (PDDL)</td>
</tr>
<tr>
<td>Attribution</td>
<td>CC BY</td>
<td>Attribution License (ODC-By)</td>
</tr>
<tr>
<td>Attribution, share alike</td>
<td>CC BY-SA</td>
<td>Open Database License (ODC-ODbL)</td>
</tr>
<tr>
<td>Attribution-NonCommercial</td>
<td>CC BY-NC</td>
<td></td>
</tr>
<tr>
<td>Attribution-NoDerivatives</td>
<td>CC BY-N</td>
<td></td>
</tr>
<tr>
<td>Attribution-NonCommercial, share alike</td>
<td>CC BY-NC-SA</td>
<td></td>
</tr>
<tr>
<td>Attribution-NonCommercial, noDerivatives</td>
<td>CC BY-NC-N</td>
<td></td>
</tr>
</tbody>
</table>
9. Licensing

Learn about Open Data

Data is open if it can be freely accessed, used, modified and shared by anyone for any purpose - subject only, at most, to requirements to provide attribution and/or share-alike

- **Legally open**: available under an open (data) license that permits anyone freely to access, reuse and redistribute
- **Technically open**: that the data be available for no more than the cost of reproduction and in machine-readable and bulk form

10. Long-term storage

- Put your data to better use
  - Deposit in Repository
- Are they FAIR?
  - Findable
  - Accessible
  - Interoperable
  - Re-usable

Look for repositories per discipline: [https://www.re3data.org/](https://www.re3data.org/)

10. Long-term storage

- Files are organized in a consistent manner
- File formats are open, non-proprietary
- There are no security issues or PII
- Is your data well documented?
- Datasets and documentation are complete
- Ownership and licenses are cleared
- You choose a proper repository
- Your data is FAIR

Options

- Restricted access
- Temporary license
- Intellectual property → embargoes
GRIC’s Data Management Series@GRIC-UPRM. February 26, 2019. By: Hilda T. Ayala González. Sponsored by TIGER (#P031M140035)

Questions?

GRIC’s Data Management Series@GRIC-UPRM. February 26, 2019. By: Hilda T. Ayala González. Sponsored by TIGER (#P031M140035)

Thank you!

• Importance of Data Management
• Data Management Plan
• Organization
• Formats
• Security
• Documentation
• Storage/backups
• Ownership
• Licensing
• Long-term storage

GRIC’s Data Management Series@GRIC-UPRM. February 26, 2019. By: Hilda T. Ayala González. Sponsored by TIGER (#P031M140035)

Which one you’ll do?

GRIC’s Data Management Series@GRIC-UPRM. February 26, 2019. By: Hilda T. Ayala González. Sponsored by TIGER (#P031M140035)

Post-test

Link: https://uprm.libsurveys.com/RDMPPostest

GRIC’s Data Management Series@GRIC-UPRM. February 26, 2019. By: Hilda T. Ayala González. Sponsored by TIGER (#P031M140035)

Scan QR Code

GRIC’s Data Management Series@GRIC-UPRM. February 26, 2019. By: Hilda T. Ayala González. Sponsored by TIGER (#P031M140035)

Where to look for more


GRIC Gestión de Datos de Investigación Libguide

Office of Research Integrity: Responsible Conduct in Data Management Modules

Data One Primer on Data Management: What you always wanted to know but were afraid to ask
Password: 2018

Title: RDM: 10 things to consider before you begin your research
Hilda Teresa Ayala González
February 26, 2019

Find Us
[gric@uprm.edu](mailto:gric@uprm.edu)
[@GRICUPRm](https://twitter.com/GRICUPRm)
[@GricUprm](https://twitter.com/GricUprm)
[GRIC UPRM](https://www.youtube.com/channel/UCm90ZVrYU1G7ZGjXBOB0cSw)

This activity is sponsored by the Transformational Initiative for Graduate Education and Research (TIGER), US Department of Education, Title V, Part B, Promoting Postbaccalaureate Opportunities for Hispanic Americans (PPOHA) Program (#P031M140035).