Research Data Management
What’s it all about?

Jane Fry, MacOdrum Library
November 21, 2017
Agenda

- Why the Library
- Define key components of RDM
- What is an RDMP?
Acronyms

- **Research data management**
  - aka RDM
  - aka data management
    - *aka DM*

- **Research data management plan**
  - aka RDMP
  - aka data management plan
    - *aka DMP*
Data

Home » Find » Data

Need Canadian data? Start here if you're on campus

Find data in Anywhere

Off-campus or need more search options? Visit the <odesi> website for full access.

Need International data? Start here

Search ICPSR or access ICPSR resources.

Contact
Data Services
Email

Related Links
<odesi>

Related Help
Citing Data and Statistics
Statistical Literacy Tutorial
Surveys: Which survey is right for you?

Related Services
Statistical Consulting

Related Policies

https://library.carleton.ca/find/data
Why the Library?

- Research partner
- Support the research endeavor
- RDM expert
- Partner with CU Research Office
- The scholarly life-cycle
- Discipline-agnostic
Why the Library? (cont’d)

- **Our role**
  - Information
  - Consultation

- **Challenge**
  - Determine how we can help researchers advance their research

References: Rambo Neil; Shorish, Yasmeen
Exercise:

What do you think – what are research data?

*Take 2 minutes to discuss in your group.*

* Someone record answers, please.
“Research data may be defined as the factual records (e.g. microarray, numerical and textual records, images and sounds, etc.) used as primary sources for research, and that are commonly accepted in the research community as necessary to validate research findings. For the most part these data are born digital, and stored and managed electronically, making them easy to share, replicate, and combine with other data. …”

Source:
http://www.carl-abrc.ca/advancing-research/research-data-management/
Why are research data important?

Sharing research data

Check out the following examples …
Example: Reproducibility

- Political Persuasion and Attitude Change Study: The Los Angeles Longitudinal Field Experiments, 2013-2014

- Principal Investigator:
  - Michael J. LaCour

Reference: https://www.openicpsr.org/openicpsr/project/100037/version/V8/view
Irregularities in LaCour (2014)

David Broockman, Assistant Professor, Stanford GSB (as of July 1), dbroockman@stanford.edu
Joshua Kalla, Graduate Student, UC Berkeley, kalla@berkeley.edu
Peter Aronow, Assistant Professor, Yale University, peter.aronow@yale.edu
May 19, 2015

Summary

We report a number of irregularities in the replication dataset posted for LaCour and Green (Science, “When contact changes minds: An experiment on transmission of support for gay equality,” 2014) that jointly suggest the dataset (LaCour 2014) was not collected as described. These irregularities include baseline outcome data that is statistically indistinguishable from a national survey and over-time changes that are unusually small and indistinguishable from perfectly normally distributed noise. Other elements of the dataset are inconsistent with patterns typical in randomized experiments and survey responses and/or inconsistent with the claimed design of the study. A straightforward procedure may generate these anomalies nearly exactly: for both studies reported in the paper, a random sample of the 2012 Cooperative Campaign Analysis Project (CCAP) form the baseline data and normally distributed noise are added to simulate follow-up waves.

Timeline of Disclosure

- January - April, 2015. Broockman and Kalla were impressed by LaCour and Green (2014) and wanted to extend the article’s methodological and substantive discoveries. We began to plan an extension. We sought to form our priors about several design parameters based on the patterns in the original data on which the paper was based, LaCour (2014). As we examined the study’s data in planning our own studies, two features surprised us: voters’ survey responses exhibit much higher test-retest reliabilities than we have observed in any other panel survey data, and the response and reinterview rates of the panel survey were significantly higher than we expected. We set aside our doubts about the study and awaited the launch of our pilot extension to see if we could manage the same parameters. LaCour and Green were both responsive to requests for advice about design details when queried.


Reference: http://stanford.io/2bzRWFo
When contact changes minds: An experiment on transmission of support for gay equality

Michael J. LaCour¹, Donald P. Green²

Author Affiliations

Science 12 Dec 2014:
Vol. 346, Issue 6215, pp. 1366-1369
DOI: 10.1126/science.1256151

This article has been retracted. Please see:
Is retracted by - June 05, 2015

Reference: http://bit.ly/1NxWG5M
“New Study Links Vaccines To Autism. There's Just One Tiny Problem With It”

“... one of its own co-authors claimed that figures in the paper were deliberately altered before publication. The data had been tampered with. ...”

Source: http://bit.ly/2zSwAxo
RETRACTED: Subcutaneous injections of aluminum at vaccine adjuvant levels activate innate immune genes in mouse brain that are homologous with biomarkers of autism

Dan Li a, Lucija Tomljenovic a, Yongling Li a, Christopher A. Shaw a, b, c, d, e

Show more

https://doi.org/10.1016/j.jinorgbio.2017.08.035

Under a Creative Commons license
“Researchers from the University of British Columbia are retracting their scientific paper linking aluminum in vaccines to autism in mice, because one of the co-authors claims figures published in the study were deliberately altered before publication — an issue he says he realized after allegations of data manipulation surfaced online.”

“…original data cited in the study is inaccessible, which would be a contravention of the university's policy around scientific research.”

“…the original data is in China, with an analyst who worked on the paper.”

(October 16, 2017)

Exercise:

Research data management (RDM) - What do you think it is?

(*2 minutes in your group)
(*Have a recorder!)
What is RDM?

- **What is it?**
  - “…describes the activities researchers perform as they create and save their research data.”
    - Source: [http://researchdata.library.ubc.ca/learn/](http://researchdata.library.ubc.ca/learn/)

- **Includes:**
  - Sound practices
  - Data curation
  - Data stewardship
Benefits of RDM

- Confirmation of original findings
- Further research
- Planning follow-up studies
- Bonus …
Why RDM Now?

- **Requirement by funders**
  - Tri-Council (SSHRC, CIHR and NSERC)
  - CFI
  - Genome Canada

- **Tri-Agency Statement of Principles on Digital Data Management**

- We should be ahead of the curve in this

- You are at the beginning of a research career
Managing research data entails the many activities dealing with the *operational support* of data across the stages of the research lifecycle. This involves the “*what*” and “*how*” of research data.

Data Stewardship involves *assigning responsibility* for ensuring data management activities are performed to best practice levels and standards across the complete lifecycle. This addresses “*who*” is responsible for specific data activities.

Source: Moon, J. & Fry, J., September 2017
What do you think it is?
What is it
- Information about the data
- Usually in a standardized and structured format

Explains ...

Why is it important

Who enters it
Why keep metadata
- Researchers re-use data
- Good research practice

When to record it

What to keep
- What do you think?
What to keep - everything!

- Research design
- Data collection
- Data preparation
- Questionnaires
- Interviewer instructions
- Details of decisions made
  - Why certain decisions were made
    » e.g. if data collection not to be done on a certain date (Easter)
- Keep all processes
  - What worked
  - What didn’t work
  - Changes made after pilots conducted
    - Why they were made
    - Was another pilot conducted after changes made

- Any and all changes that were made or not made

- End goal
What documentation will be needed

- For the data to be read
- For the data to be interpreted correctly
- For the data to be replicated in the future, if necessary

What do you think?
- **Study background**
  - Purpose
  - Time frame
  - Geographic location
  - Creator
  - Sampling design
    - *Description*
    - *Size*
  - Any changes that were made
  - ...
• **Study description**
  
  Describes all aspects of the data collection and processing
  
  - *Data collection methodology*
  - *Data preparation procedure*
  - *Instruments used*
  - *Geographic coverage*
  - *Temporal coverage*
  - *Date of file creation*
  - *Description of codes used*
Data description & collection

- **Data description**
  - A brief summary of the overall project

- **Data collection**
  - Collection method
  - File formats
  - File Names
Exercise:

- Choose one of the data types identified in the previous exercise and draw a lifecycle model representing the steps through which the data would flow in a research project.

- Focus on high-level, generalized steps in the research process – aim for six to eight steps.

*Take 10 minutes, and then we’ll report back…*
UKDA RDM Lifecycle

Source: http://www.data-archive.ac.uk/create-manage/life-cycle
Creating data

- design research
- plan data management (formats, storage etc)
- plan consent for sharing
- locate existing data
- collect data (experiment, observe, measure, simulate)
- capture and create metadata

http://www.data-archive.ac.uk/create-manage/life-cycle
Processing data

- enter data, digitise, transcribe, translate
- check, validate, clean data
- anonymise data where necessary
- describe data
- manage and store data

http://www.data-archive.ac.uk/create-manage/life-cycle
Variables

- Names
- Labels
  - Comprehensible
  - Unique
- Description
- Value labels
  - Comprehensible
  - Complete
- Associated question

Missing values

- Codes used
  - Should be consistent
- Reasons for missing values
Data storage

- Where will it be stored?
- Other tips
Analysing data

- interpret data
- derive data
- produce research outputs
- author publications
- prepare data for preservation

http://www.data-archive.ac.uk/create-manage/life-cycle
Preserving data

- migrate data to best format
- migrate data to suitable medium
- back-up and store data
- create metadata and documentation
- archive data

http://www.data-archive.ac.uk/create-manage/life-cycle
Data preservation

- Preservation
  - Where

- Where will all information be backed-up
  - If at your institution

- How often will you back-up

- Long term-preservation and access
Giving access to data

- distribute data
- share data
- control access
- establish copyright
- promote data

http://www.data-archive.ac.uk/create-manage/life-cycle
Data sharing

- Why share?
- What data will you be sharing?
- How will you be sharing your data?
- Describe any restrictions placed on your data for access when they are made available
Confidentiality

■ What procedures will be taken to ensure confidentiality
  ■ Not possible to identify any individual

■ Record all decisions made
  ■ Why this decision made
Re-using data

- follow-up research
- new research
- undertake research reviews
- scrutinise findings
- teach and learn
You have learned

- What research data is
- What RDM is
- What a research lifecycle is
Questions are guaranteed in life; Answers aren't.

http://taitegallery.net/wp-content/uploads/2012/02/unanswered-questions.jpg
What’s next?

- Need an RDMP

- Why an RDMP?
  - Safety
  - Efficiency
  - Quality

- If no RDMP?
  - Potential problems
DMPs can help researchers

a. identify institutional services that can support their data during and after a project ends,

and

b. determine the process for transferring their data.
Platforms & Templates

- Web-based data management platforms
  - Tools
  - Software

- Data stewardship templates
  - Frameworks
  - Used for planning
  - Within a platform
  - Could also be called a form
Exercise:

- Choose one step in the research lifecycle.
- List **three** data management activities that would be conducted for this step.

* 5 minutes! *
UKDA Research Data Lifecycle

- Creating Data
- Processing Data
- Analysing Data
- Preserving Data
- Giving Access to Data
- Re-Using Data
§ “As we are at the beginning of our project, the DMP process really helped us to plan what we want to do with our data and how we want to proceed. We found it very useful. However lots of points will need to be clarified.”

§ “The process of writing the DMP really pointed out the great diversity of the data we are dealing with within our project, and showcased the importance of having a distinct data management approach for each type of data, and the challenges that come with it.”

§ “I came to see the DMP process as about planning for the preservation and archivization of data, not about its ideal presentation or optimal accessibility.”
Foster a **community of practice** for research data management (RDM)

Facilitate and provide leadership in the development of RDM infrastructure

Engage and advocate for research data management with **stakeholder communities**
**THE PORTAGE NETWORK** is dedicated to the shared stewardship of research data in Canada through:

- Developing a national research data culture
- Fostering a community of practice for research data
- Building national research data services and infrastructure

Launched in 2015 by the [Canadian Association of Research Libraries](https://www.curl.ca/), Portage works within the library community to coordinate expertise, services, and technology in research data management.

Reference: [https://portagenetwork.ca/](https://portagenetwork.ca/)
What’s next?

- the Data Management Plan and
- the Data Management Plan Assistant
DMP Assistant is a bilingual tool for preparing data management plans (DMPs). The tool follows best practices in data stewardship and walks researchers step-by-step through key questions about data management.

**Step 1** Sign up with DMP Assistant

**Step 2** Sign in and select a template under Organizations. The Portage template is the default.

**Step 3** Answer the questions that are relevant to your work. Guidance and examples are provided.

**Step 4** Revisit the tool throughout your research to review or revise your answers.

Source: https://portagenetwork.ca/
DMP Assistant is a bilingual tool for preparing data management plans (DMPs). The tool follows best practices in data stewardship and walks researchers step-by-step through key questions about data management.

- **Step 1**: Sign up with DMP Assistant
- **Step 2**: Sign in and select a template under Organizations. The Portage template is the default.
- **Step 3**: Answer the questions that are relevant to your work. Guidance and examples are provided.
- **Step 4**: Revisit the tool throughout your research to review or revise your answers.

Sign in
If you have an existing account with DMP Assistant or previous version of DMP Builder.

Sign up
New to DMP Assistant? Sign up today.

Please note that we are currently working on single sign-in authentication. For now, please create a new DMP Assistant account. You will have the option to link your DMP Assistant account to your campus ID when that feature becomes available.
Sign up

New to DMP Assistant? Sign up today.

Email *

Organization

My organisation isn't listed

Password *

Password confirmation *

I accept the terms and conditions *

Sign up

https://assistant.portagenetwork.ca/
Create a new plan

Please select from the following drop-downs so we can determine what questions and guidance should be displayed in your plan.

If you aren't responding to specific requirements from a funder or an institution, you can choose the Portage Data Stewardship Template. The Portage Data Stewardship Template is based on internationally accepted standards and best practices. It has been prepared and is maintained by a group of research data management experts from research libraries across Canada.

To see institutional questions and/or guidance, select your organization.

You may leave blank or select a different organization to your own. If you leave blank, default Portage own template will be used.

Create plan
Editing the plan details

Please fill in the basic project details below and click 'Save' to save

<table>
<thead>
<tr>
<th>Plan details</th>
<th>Portage Data Management Questions</th>
<th>Share</th>
<th>Export</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plan name</td>
<td>My first DMP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>20170731</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grant number</td>
<td>CU#7524</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Principal Investigator/Researcher</td>
<td>Jane Fry</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Principal Investigator/Researcher ID</td>
<td>9235</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plan data contact</td>
<td><a href="mailto:jane.fry@carleton.ca">jane.fry@carleton.ca</a></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Description

Questions to consider:
- What is the nature of your research project?
- What research questions are you addressing?
- For what purpose are the data being collected or created?

Guidance:
Briefly summarise the type of study (or studies) to help others understand the purposes for which the data are being collected or created.
My first DMP

This page gives you an overview of your plan. It tells what your plan is based on and gives an overview of the questions that you will be asked.

<table>
<thead>
<tr>
<th>Plan details</th>
<th>Portage Data Management Questions</th>
<th>Share</th>
<th>Export</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Plan name</th>
<th>My first DMP</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>20170731</td>
</tr>
<tr>
<td>Grant number</td>
<td>CU#7524</td>
</tr>
<tr>
<td>Principal Investigator/Researcher</td>
<td>Jane Fry</td>
</tr>
<tr>
<td>Principal Investigator/Researcher ID</td>
<td>9235</td>
</tr>
<tr>
<td>Plan data contact</td>
<td><a href="mailto:jane.fry@carleton.ca">jane.fry@carleton.ca</a></td>
</tr>
<tr>
<td>Description</td>
<td>-</td>
</tr>
</tbody>
</table>

This plan is based on:

<table>
<thead>
<tr>
<th>Institution</th>
<th>Portage</th>
</tr>
</thead>
</table>
Portage DMP Assistant
- Data Collection
- Documentation and Metadata
- Storage and Backup
- Preservation
- Sharing and Re-use
- Responsibilities and Resources
- Ethics and Legal Compliance
Data collection

- Types of data
- File formats
- Conventions and procedures
Documentation and metadata

- Documentation
- Consistency
- Metadata standard and tools
Storage and backup

- Storage requirements

- Storage and backup

- Access to data
Preservation

- What data
- Where will you deposit your data
- Preservation ready
Sharing and reuse

- What data
- How
- End-user license
- Promotion
Responsibilities and resources

- Who
- How to handle change
- Resources
Ethics and legal compliance

- Sensitive data
  - Primary use
  - Secondary use

- Legal, ethical and IP issues
### My first DMP

**Tips**

Not all questions will apply to all research projects. Researchers are encouraged to answer the questions relevant to their work.

Researchers should revisit the tool throughout their research to review or complete their responses.

<table>
<thead>
<tr>
<th>Plan details</th>
<th>Portage Data Management Questions</th>
<th>Share</th>
<th>Export</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Collection</td>
<td>(3 questions, 0 answered)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Documentation and Metadata</td>
<td>(3 questions, 0 answered)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Storage and Backup</td>
<td>(3 questions, 0 answered)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preservation</td>
<td>(2 questions, 0 answered)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sharing and Reuse</td>
<td>(3 questions, 0 answered)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Responsibilities and Resources</td>
<td>(3 questions, 0 answered)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethics and Legal Compliance</td>
<td>(3 questions, 0 answered)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Data Collection

What types of data will you collect, create, link to, acquire and/or record?

Guidance
Portage Guidance
Examples: numeric, images, audio, video, text, tabular data, modeling data, spatial data, instrumentation data.

What file formats will your data be collected in? Will these formats allow for data re-use, sharing and long-term access to the data?

Guidance
Portage Guidance
Proprietary file formats requiring specialized software or hardware to use are not recommended, but may be necessary for certain data collection or analysis methods. Using open file formats or industry-standard formats (e.g. those widely used by a given community) is preferred whenever possible.

Read more about file formats: UBC Library or UJK Data Archive.
What conventions and procedures will you use to structure, name and version-control your files to help you and others better understand how your data are organized?
You can give other people access to your plan here. There are three permission levels.

Users with "read only" access can only read the plan.
Editors can contribute to the plan.
Co-owners can also contribute to the plan, but additionally can edit the plan details and control access to the plan.

Add each collaborator in turn by entering their email address below, choosing a permission level and clicking "Add collaborator".

Those you invite will receive an email notification that they have access to this plan, inviting them to register with DMP Assistant if they don't already have an account. A notification is also issued when a user's permission level is changed.

Collaborators

<table>
<thead>
<tr>
<th>Email address</th>
<th>Permissions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jane Fry</td>
<td>Owner</td>
</tr>
</tbody>
</table>

Add collaborator

Email

Permissions:
- Co-owner
- Co-owner
- Edit
- Read only
- Add collaborator
The export tab
- allows displaying a plan in full or selectively for specific themes and their questions

Export formats include:
- pdf
- csv
- html
- json
- text
- xml
- docx
DMP title

Project Name My first DMP
Project Identifier 20170731
Grant Title CU#7524
Principal Investigator / Researcher Jane Fry
Project Data Contact jane.fry@carleton.ca
Institution Portage

Data Collection
What types of data will you collect, create, link to, acquire and/or record?
Question not answered.

What file formats will your data be collected in? Will these formats allow for data re-use, sharing and long-term access to the data?
Question not answered.

What conventions and procedures will you use to structure, name and version-control your files to help you and others better understand how your data are organized?
Question not answered.

Documentation and Metadata
What documentation will be needed for the data to be read and interpreted correctly in the future?
Question not answered.
## Exporting (cont’d)

### Settings (Using template PDF formatting values)

<table>
<thead>
<tr>
<th>Plan title</th>
<th>DMP title</th>
</tr>
</thead>
</table>

### Included Elements

#### Admin Details
- Project Name
- Project Identifier
- Grant Title
- Principal investigator / Researcher
- Project Data Contact
- Description
- Funder
- Institution

#### Sections

**Data Collection**
- What types of data will you collect, create, link to, acquire and/or record?
- What file formats will your data be collected in? Will these formats allow for data re-use?
- What conventions and procedures will you use to structure, name and version-control?

**Documentation and Metadata**
- What documentation will be needed for the data to be read and interpreted correctly?
- How will you make sure that documentation is created or captured consistently throughout?
- If you are using a metadata standard and/or tools to document and describe your data?
From here you can download your plan in various formats. This may be useful if you need to submit your plan as part of a grant application. Select what format you wish to use and click to "Export".

**Format**

- pdf
  - **Export**

**Settings (Using default PDF formatting values)**

<table>
<thead>
<tr>
<th>Font</th>
<th>Margin (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Face</strong></td>
<td><strong>Size (pt)</strong></td>
</tr>
<tr>
<td>Arial, Helvetica, Sans-Serif</td>
<td>12</td>
</tr>
</tbody>
</table>
DMP title

Project Name My first DMP
Project Identifier 20170731
Grant Title CU#7524
Principal Investigator / Researcher Jane Fry
Project Data Contact jane.fry@carleton.ca
Institution Portage

This document was generated by DMP Assistant (https://assistant.portagenetwork.ca)
Anyone

Step-by-step

The length

Different agencies

Remember …
Some tips

- Mark it down!
- It is not written in stone!
- Easy!
Exercise:

- Choose one of the sections from the DMP template.
- Read over the questions for your section.
- Answer one of the questions listed.
- Bonus
  - Are there any other questions that you think were omitted?

* Hint: use your research experience to help answer these questions 😊
* You have 5 minutes!
Still don’t believe me?

- What could happen if you don’t practice good RDM?

  https://www.youtube.com/watch?v=N2zK3sAtr-4&t=17
RDM help

- **Help with RDM**
  - [https://library.carleton.ca/services/research-data-management](https://library.carleton.ca/services/research-data-management)
  - Consultations

- **Help with RDMPs**
  - Portage: [https://assistant.portagenetwork.ca/](https://assistant.portagenetwork.ca/)
  - Word template: [https://library.carleton.ca/services/research-data-management#how](https://library.carleton.ca/services/research-data-management#how)
You are now able to:

- Define the key components of RDM
- Define an RDMP
- Create an RDMP
- RDM at Carleton
  https://library.carleton.ca/services/research-data-management

- Portage DMP Assistant
  https://portagenetwork.ca/

- Research Data Lifecycle (UK Data Archive)
  http://www.data-archive.ac.uk/create-manage/life-cycle

- *Tri-Agency Statement of Principles on Digital Data Management*
https://portagenetwork.ca/about/documents-and-presentations/


http://acrl.ala.org/techconnect/post/the-library-as-research-partner


UK Data Archive. “Create & manage data: Research Data lifecycle”. Retrieved 13 October 2013 from
http://data-archive.ac.uk/create-manage/life-cycle
Sometimes the questions are complicated and the answers are simple.

Thank you!

Jane Fry
Data Services Librarian
Rm 122
MacOdrum Library
613.520.2600 x1121
jane.fry@Carleton.ca

http://www.library.carleton.ca/find/data