Research Data Management: What’s it all about?

Jane Fry, MacOdrum Library
April 26, 2018
Agenda

- Why the Library

- Define RDM
  - concepts
  - challenges
  - tools and Resources
Acronyms

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

- Research data management plan
  - aka RDMP
  - aka data management plan
    - *aka DMP*
Why the Library?

- Research partner
- Support the research endeavour
- 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
What are Research data?

“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:
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

You are currently viewing the abstract.

This article has been retracted.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

Reference: http://bit.ly/1NxWG5M
Another Example

- “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. …”

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, *,

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)

What is RDM?

“...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
UKDA RDM Lifecycle

Source: http://www.data-archive.ac.uk/create-manage/life-cycle
• Need an RDMP

• Why an RDMP
  | Safety
  | Efficiency
  | Quality

• If no RDMP?
  | Potential problems
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, Portage works within the library community to coordinate expertise, services, and technology in research data management.

Reference: https://portagenetwork.ca/
Goals of the Portage Network

- 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**
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.

Reference: [https://portagenetwork.ca/](https://portagenetwork.ca/)
- Portage DMP Assistant
  - Data Collection
  - Documentation and Metadata
  - Storage and Backup
  - Preservation
  - Sharing and Re-use
  - Responsibilities and Resources
  - Ethics and Legal Compliance
Anyone

Step-by-step

The length

Different agencies

Remember …
Some tips

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

- Researcher buy-in
- Where to get help
Tools and Resources

- **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 concepts of RDM
- Define an RDMP
- Know where to find resources
Resources

- 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


Sometimes the questions are complicated and the answers are simple.
Jane Fry
Data Services Librarian
Rm 122
MacOdrum Library
613.520.2600 x1121
jane.fry@Carleton.ca

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