Research Data Management
What’s it all about?

Jane Fry
MacOdrum Library
November 27, 2019
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

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

[Search bar]

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

Need international data? Start here

[Search bar]

Search ICPSR or access ICPSR resources.

Research Data Management: Manage your data

Dataverse: Store your data

Contact
Data Services
email

Related Links
<odesi>

Related Help
Citing Data and Statistics
Postal Code Conversion File (PCCF)
Statistical Literacy Tutorial
Surveys: Which survey is right for you?

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 time!

- Get together in groups
- Come up with a definition for Research data
- 2 minutes!
- GO!
“Research data are the original sources or material that you have created or collated to conduct your research project. They can be digital or non-digital. The response to your research question is based on the analysis of these research data.”

Source: [https://blogs.ucl.ac.uk/rdm/2015/09/what-is-research-data/](https://blogs.ucl.ac.uk/rdm/2015/09/what-is-research-data/)
“Data are facts, observations or experiences on which an argument or theory is constructed or tested. Data may be numerical, descriptive, aural or visual. Data may be raw, abstracted or analysed, experimental or observational. Data include but are not limited to: laboratory notebooks; field notebooks; primary research data (including research data in hardcopy or in computer readable form); questionnaires; audiotapes; videotapes; models; photographs; films; test responses. Research collections may include slides; artefacts; specimens; samples.”

Source: https://blogs.ucl.ac.uk/rdm/2015/09/what-is-research-data/
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(s):
  - 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. LaCour, Donald P. Green

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
“A top Cornell food researcher has had 13 studies retracted. That’s a lot.”

September 21, 2018
Brian Wansink
“committed academic misconduct,”
“he would retire from the university on June 30, 2019”
“has been removed from all teaching and research,”
“will spend his remaining time … in an “ongoing review of his prior research.””

Wansink refuted these findings. “There was no fraud, no intentional misreporting, no plagiarism, or no misappropriation,” he wrote. “I believe all of my findings will be either supported, extended, or modified by other research groups.”

“In a press release, JAMA said Cornell couldn’t “provide assurances regarding the scientific validity of the 6 studies” because they didn’t have access to Wansink’s original data. So, Wansink’s ideas aren’t necessarily wrong, but he didn’t provide credible evidence for them.”

JAMA Network Retracts 6 Articles That Included Dr. Brian Wansink as Author

FOR IMMEDIATE RELEASE: SEPTEMBER 19, 2018

Media advisory: To contact JAMA Network Media Relations email mediarelations@jamanetwork.org.

CHICAGO – JAMA, JAMA Internal Medicine and JAMA Pediatrics have retracted six articles that included Brian Wansink, Ph.D., of Cornell University, Ithaca, New York, as author. Below is the notice of retraction published online today by JAMA, which references the retracted articles (see references 4-9). Similar notices were published online today in JAMA Internal Medicine and JAMA Pediatrics.

EDITORIAL


Howard Bauchner, MD

Source: https://bit.ly/2OyCH1N
CBC, Sept 3, 2019

“MMIWG final report quietly altered after CBC inquired about errors”

“Report contained errors that were fixed online, but allowed to remain on the official record”

“…the commission left out a word that had appeared in the original Statistics Canada figure.”

"Indigenous women and girls now make up almost 25 per cent of homicide victims" should have referred to their percentage share of female homicide victims …”

“In some cases, the inquiry report's footnotes cite government reports that do not support the footnoted statements.”

What is RDM?

“...describes the activities researchers perform as they create and save their research data.”

• Source: 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
Exercise time again!

- Groups
- What have I given you?
- What can you tell me about it?
- 2 minutes
- GO!
Metadata

- What is it
- 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

- End goal
ICPSR 101: What are Metadata (and why are they so important)?
Is this data set ready for deposit?

- **Dataset**: Attitudes of Pets towards their Owners (Oct 1998)
- **Documentation available**: The following text file: “This survey was conducted by the Pet Researchers of Canada and was analysed by the Acme Research Company. There is no documentation available for this survey. Use basic survey methodology if necessary. There are some interesting results in this survey.”
- **Data available**: A microdata file
  - Example 1: Name of variable: V35
    - Frequency: Yes = 35%, No = 47%
  - Example 2: Name of Variable: Region of Country
    - Frequency: 1 = 12%; 2 = 32%; 3 = 35%; 4 = 15%; 5 = 4%
Questions are guaranteed in life; Answers aren't.

http://taitegallery.net/wp-content/uploads/2012/02/unanswered-questions.jpg
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
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
Giving access to data

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

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

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

http://www.data-archive.ac.uk/create-manage/life-cycle
What’s next?

- 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, seeking to collaborate with other research data management stakeholders.

Research data culture represents widely shared values and principles for digital data management.

Reference: 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.

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
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
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
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*

Sometimes the questions are complicated and the answers are simple.
Thank you!

Contact Information

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

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