Research Data Management:

Jane Fry  
MacOdrum Library  
September 26, 2019
Agenda

- What data resources are out there
- Why the Library
- What are the key components of RDM
- What is an RDMP
Data

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

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

Need international data? Start here

Search ICPSR or access ICPSR resources.

Research Data Management:
Manage your data

Dataverse:
Store your data

"An Introduction to R" Workshop

Want to know more about R? Want to try out the R program? Come to our workshop to get started with R.

https://library.carleton.ca/find/data
What do we offer?

- **Main services**
  - Access to data
    - different data sources and portals
    - different types of data
  - Research Data Management
    - Data management plans
  - Stats Consulting
    - SPSS, Stata, SAS, R

- **One restriction for data use**
  - Can ONLY be used for academic research or teaching
Canadian data sources

- **Statistics Canada**
  - Data Liberation Initiative
  - Includes
    - *All Statistics Canada’s public use data files, databases and geographic files*

- **Canadian Election Surveys**
  - 1965 - 2015
### Public Opinion Polls

<table>
<thead>
<tr>
<th>Poll Title</th>
<th>Year</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assisted Suicide, 2014 [Canada]</td>
<td>2014</td>
<td>Canada</td>
</tr>
<tr>
<td>Baggage Fees, 2014 [Canada]</td>
<td>2014</td>
<td>Canada</td>
</tr>
<tr>
<td>Canadian Football League (CFL), 2014</td>
<td>2014</td>
<td></td>
</tr>
<tr>
<td>Canadian Perceptions of India and Trade 2015</td>
<td>2015</td>
<td></td>
</tr>
<tr>
<td>Concussion and Youth Sports [Canada], 2015</td>
<td>2015</td>
<td>Canada</td>
</tr>
<tr>
<td>Disability and Accessibility, 2015 [Canada]</td>
<td>2015</td>
<td>Canada</td>
</tr>
<tr>
<td>Disability in Canada, 2016 [Disabled Sample]</td>
<td>2016</td>
<td></td>
</tr>
<tr>
<td>Disasters, 2015 [Canada]</td>
<td>2015</td>
<td>Canada</td>
</tr>
<tr>
<td>Election 2015 [Canada]</td>
<td>2015</td>
<td>Canada</td>
</tr>
<tr>
<td>Election 2015: The Politics of Health Care [Canada]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender, Sports, and the Women's World Cup, 2015 [Canada]</td>
<td>2015</td>
<td>Canada</td>
</tr>
<tr>
<td>Global Threats, 2014 [Canada]</td>
<td>2014</td>
<td>Canada</td>
</tr>
<tr>
<td>Hockey Night in Canada Year in Review, 2015</td>
<td>2015</td>
<td></td>
</tr>
<tr>
<td>Hockey Night in Canada, 2014</td>
<td>2014</td>
<td></td>
</tr>
<tr>
<td>Prayer in Canadian Public Life, 2015 [Canada]</td>
<td>2015</td>
<td>Canada</td>
</tr>
<tr>
<td>Sexual Harassment in the Workplace, 2014 [Canada]</td>
<td>2014</td>
<td>Canada</td>
</tr>
<tr>
<td>Truth and Reconciliation Commission, 2015 [Canada]</td>
<td>2015</td>
<td>Canada</td>
</tr>
<tr>
<td>Center for Research and Information on Criminals (CRIC)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
How do you access the data?

- **Different data portals**
  - Search for questions or variables across datasets
  - Perform online analyses
  - Create tables, graphs and charts
  - Download data in SPSS, SAS, Stata, …
Canadian data portals

- **ODESI**
  - Data portal for social science data for Ontario Universities
  - Data from StatCan, public opinion polls, …
  - [http://www.library.carleton.ca/find/data](http://www.library.carleton.ca/find/data)

- **Statistics Canada**
  - Public use microdata files (E/F)
  - Metadata for Research Data Centre Master files (E/F)
<table>
<thead>
<tr>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture</td>
</tr>
<tr>
<td>Business and Financial</td>
</tr>
<tr>
<td>Census of Population</td>
</tr>
<tr>
<td>Communications and Information</td>
</tr>
<tr>
<td>Crime and Justice</td>
</tr>
<tr>
<td>Consumer Surveys</td>
</tr>
<tr>
<td>Demographics and Population</td>
</tr>
<tr>
<td>Education</td>
</tr>
<tr>
<td>Elections and Politics</td>
</tr>
<tr>
<td>Geography</td>
</tr>
<tr>
<td>Government Finances and Economic Indicators</td>
</tr>
<tr>
<td>Health</td>
</tr>
<tr>
<td>Health and Safety</td>
</tr>
<tr>
<td>Natural Resources and Environment</td>
</tr>
<tr>
<td>Public Opinion Polls</td>
</tr>
<tr>
<td>Social Surveys</td>
</tr>
<tr>
<td>Trade</td>
</tr>
<tr>
<td>Travel</td>
</tr>
</tbody>
</table>
How to Use ODESI

Using the <odesi> data repository (Short videos)

- <odesi>: Navigating the <odesi> Repository
- <odesi>: Searching for Variables
- <odesi>: Finding, Subsetting and Downloading a Survey
- <odesi>: Creating a Cross Tabulation
- <odesi>: Downloading a Full Dataset

<odesi> help guides

- Finding, Subsetting and Downloading a Survey in <odesi> (.pdf)
- Searching for variables (.pdf)

Source: https://library.carleton.ca/help/odesi-how-to-use-odesi
<table>
<thead>
<tr>
<th>Agriculture</th>
<th>Indigenous peoples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business and consumer services and culture</td>
<td>Information and communications technology</td>
</tr>
<tr>
<td>Business performance and ownership</td>
<td>International trade</td>
</tr>
<tr>
<td>Children and youth</td>
<td>Labour</td>
</tr>
<tr>
<td>Construction</td>
<td>Languages</td>
</tr>
<tr>
<td>Crime and justice</td>
<td>Manufacturing</td>
</tr>
<tr>
<td>Economic accounts</td>
<td>Population and demography</td>
</tr>
<tr>
<td>Education, training and learning</td>
<td>Prices and price indexes</td>
</tr>
<tr>
<td>Energy</td>
<td>Retail and wholesale</td>
</tr>
<tr>
<td>Environment</td>
<td>Science and technology</td>
</tr>
<tr>
<td>Families and households</td>
<td>Seniors and aging</td>
</tr>
<tr>
<td>Government</td>
<td>Society and community</td>
</tr>
<tr>
<td>Health</td>
<td>Statistical methods</td>
</tr>
<tr>
<td>Housing</td>
<td>Transportation</td>
</tr>
<tr>
<td>Immigration and ethnocultural diversity</td>
<td>Travel and tourism</td>
</tr>
<tr>
<td>Income, pensions, spending and wealth</td>
<td></td>
</tr>
</tbody>
</table>
ICPSR

Inter-university Consortium for Political and Social Research
https://www.icpsr.umich.edu/

World renowned summer program in quantitative methods of social statistics
International Data Portals

- **CESSDA**
  - Consortium of European Social Science Data Archives
  - [http://www.cessda.eu/](http://www.cessda.eu/)

- **GESIS**
  - Leibniz Institute for the Social Sciences
  - [https://www.gesis.org/home/](https://www.gesis.org/home/)

- **ESS ERIC**
  - European Social Survey, European Research Infrastructure
  - [http://www.europeansocialsurvey.org/](http://www.europeansocialsurvey.org/)
COOL Research Data Centre

- Located at University of Ottawa
  - available for all Carleton researchers
- Secure access to detailed microdata from Statistics Canada's surveys, Canadian census data, as well as an increasing number of administrative data sets.
- Data can only be used for research and teaching purposes
  - Check out metadata here:
  - “Statistics Canada metadata for Master Files (RDC)”
- Money for grad students!

Reference: Prof. Jennifer Stewart
Recent RDC Data additions

September 2019
- Canadian Community Health Survey (CCHS) - 2018 Annual Component and Rapid Response on Gambling
- Survey on Sexual Misconduct in the Canadian Armed Forces (SSMCAF) - 2018
- General Social Survey – Family 2017 (GSS Cycle 31)
- National Cannabis Survey (NCS) – 2019 Wave 2

August 2019
- Additional components to the Longitudinal Immigration Database (IMDB)
- Childhood National Immunization Coverage Survey (CNICS) – 2017
- Ontario Ministry of Community, Children and Social Services (MCCSS)
- Historical Ontario Ministry of Health and Long Term Care (MOHLTC)
Consultation
- on the use of data
- Research data management (RDM)
  - RDM Plans

Provide general tips and help on using data
- **Very important**
  - All data should be cited
  - [Why should I cite data?](http://www.statcan.gc.ca/pub/12-591-x/12-591-x2009001-eng.htm) (short video)

- **How to cite StatCan products**
  - [http://www.statcan.gc.ca/pub/12-591-x/12-591-x2009001-eng.htm](http://www.statcan.gc.ca/pub/12-591-x/12-591-x2009001-eng.htm)

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 endeavor
- RDM expert
- Partner with CU Research Office
- The scholarly life-cycle
- Discipline-agnostic
Our role

- Information
- Consultation

Challenge

- Determine how we can help researchers advance their research

References: Rambo Neil; Shorish, Yasmeen
“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/
“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

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

Michael J. LaCour\textsuperscript{1}, Donald P. Green\textsuperscript{2}

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. ...”

“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)

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, Lucija Tomljenovic, Yongling Li, Christopher A. Shaw

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

- You are at the beginning of a research career
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
- **Survey metadata**
  - Questionnaire
  - Data collection
  - Interviewer instructions
  - ???
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
Source: 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.

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

Reference: 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
- 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!
- An example of another DMP
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)
In summary

You are now able to:

- Explore different data sources
- 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*
References


Wednesday January 06, 2016  Three Problems With Spreadsheet

DID YOU SEE ANY ERRORS ON THE SPREADSHEET I PUT TOGETHER?

ONLY THREE.

WHAT ARE THEY?

YOUR DATA, YOUR FORMAT, AND YOUR FORMULAS.

http://dilbert.com/strip/2016-01-06
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