

# Tri-Agency Research Data Management Policy Development

Research Data Management Summit for  
Canadian Colleges, Institutes and Polytechnics  
Centennial College  
December 7, 2018

Presented by Kevin Fitzgibbons  
Executive Director, Corporate Planning and Policy, NSERC



**NSERC**  
**CRSNG**

SSHRC  CRSH

# PRESENTATION OUTLINE

- Rationale for Research Data Management (RDM)
- Tri-Agency RDM Policy Development
- Stakeholder Engagement and Consultation

# WHAT ARE RESEARCH DATA?

Research data are contents that are used as primary sources to support research, scholarship, artistic activity or research-creation, and that are used as evidence in the research process and commonly accepted in the research community as necessary to validate research findings and results.

# WHY DATA MANAGEMENT?

Research Excellence

Research Dissemination

Research Impact

Research Best Practice

# RESEARCH EXCELLENCE

Project Management

Support Reproducibility

Avoid Duplication

Journal Guidelines



# RESEARCH DISSEMINATION

Data Sharing

Citation

Interdisciplinarity

OPEN ACCESS Freely available online

PLoS one

## Sharing Detailed Research Data Is Associated with Increased Citation Rate

Heather A. Piwowar\*, Roger S. Day, Douglas B. Fridsma

Department of Biomedical Informatics, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, United States of America

**Background.** Sharing research data provides benefit to the general scientific community, but the benefit is less obvious for the investigator who makes his or her data available. **Principal Findings.** We examined the citation history of 85 cancer microarray clinical trial publications with respect to the availability of their data. The 48% of trials with publicly available microarray data received 85% of the aggregate citations. Publicly available data was significantly ( $p=0.006$ ) associated with a 69% increase in citations, independently of journal impact factor, date of publication, and author country of origin using linear regression. **Significance.** This correlation between publicly available data and increased literature impact may further motivate investigators to share their detailed research data.

Citation: Piwowar HA, Day RS, Fridsma DB (2007) Sharing Detailed Research Data Is Associated with Increased Citation Rate. PLoS ONE 2(3): e308. doi:10.1371/journal.pone.0000308

### INTRODUCTION

Sharing information facilitates science. Publicly sharing detailed research data—sample attributes, clinical factors, patient outcomes, DNA sequences, raw mRNA microarray measurements—with other researchers allows these valuable resources to contribute far beyond their original analysis[1]. In addition to being used to confirm original results, raw data can be used to explore related or new hypotheses, particularly when combined with other publicly available data sets. Real data is indispensable when investigating and developing study methods, analysis techniques, and software implementations. The larger scientific community also benefits: sharing data encourages multiple perspectives, helps to identify errors, discourages fraud, is useful for training new researchers, and increases efficient use of funding and patient population resources by avoiding duplicate data collection.

Believing that that these benefits outweigh the costs of sharing research data, many initiatives actively encourage investigators to make their data available. Some journals, including the *PLoS* family, require the submission of detailed biomedical data to publicly available databases as a condition of publication[2–4]. Since 2003, the NIH has required a data sharing plan for all large funding grants. The growing open-access publishing movement will perhaps increase peer pressure to share data.

However, while the general research community benefits from shared data, much of the burden for sharing the data falls to the study investigator. Are there benefits for the investigators themselves?

A currency of value to many investigators is the number of times their publications are cited. Although limited as a proxy for the

### RESULTS

We studied the citations of 85 cancer microarray clinical trials published between January 1999 and April 2003, as identified in a systematic review by Ntzani and Ioannidis[7] and listed in Supplementary Text S1. We found 41 of the 85 clinical trials (48%) made their microarray data publicly available on the internet. Most data sets were located on lab websites (28), with a few found on publisher websites (4), or within public databases (6 in the Stanford Microarray Database (SMD)[8], 6 in Gene Expression Omnibus (GEO)[9], 2 in ArrayExpress[10], 2 in the NCI GeneExpression Data Portal (GEDP)[gedp.nci.nih.gov]; some datasets in more than one location). The internet locations of the datasets are listed in Supplementary Text S2. The majority of datasets were made available concurrently with the trial publication, as illustrated within the WayBackMachine internet archives ([www.archive.org/web/web.php](http://www.archive.org/web/web.php)) for 25 of the datasets and mention of supplementary data within the trial publication itself for 10 of the remaining 16 datasets. As seen in Table 1, trials published in high impact journals, prior to 2001, or with US authors were more likely to share their data.

The cohort of 83 trials was cited an aggregate of 6239 times in 2004–2005 by 3133 distinct articles (median of 1.0 cohort citation per article, range 1–23). The 48% of trials which shared their data received a total of 5334 citations (85% of aggregate), distributed as shown in Figure 1.

Academic Editor: John Ioannidis, University of Ioannina School of Medicine, Greece

# RESEARCH IMPACT

Within Science

Social Impact

Policy Impact

Data on bushfire fatalities helping to prevent more deaths

**Who did the research?**

Risk Frontiers (based at Macquarie University), RMIT University and Bushfire Cooperative Research Centre.

Melanoma study leads to ban on commercial sunbeds

**What is the research about?**

HILDA survey shapes policy on tax and jobs

**Who did the research?**

The **Melbourne Institute** at the University of Melbourne, funded by the Australian Government.

**What is the project about?**

The **Household, Income and Labour Dynamics in Australia (HILDA) survey** is an annual study of around 8000 households or around 20,000 individuals. It collects information about economic wellbeing, employment, health and family.

HILDA is a longitudinal study, meaning respondents are followed year-on-year to observe changes over time. The data is confidentialised to remove identifying

**What is the real-life data impact?**

The HILDA Survey is **known** as one of the best panel surveys in the world, collecting a broad range of high quality economic and social research data about Australian lives.

The data collected and analysed through HILDA have had a substantial impact on Australian policy development. Its **findings** have been used to help shape many policies, including:

- proposals in **Australia's Future Tax System Review**
- the Government's 2011 Paid Parental leave policy
- decisions about minimum wage rates

EverGraze project boosts Australian farming

**Who did the research?**

**Charles Sturt University** in partnership with **Future Farm Industries Cooperative Research Centre (FFI CRC)** and state departments in Victoria, New South Wales and Western Australia.

**What is the project about?**

**EverGraze** is a collaborative research project testing the innovative use of perennial flora

on farmland in high rainfall zones of southern Australia. It has helped farmers improve soil health (e.g. acidity and salinity), ultimately increasing livestock profitability.

More than 250 scientists, advisers and farmers have contributed to EverGraze since its inception in 2003, including experts in soil science, agronomy, environmental science, hydrology and economics.

With complex modelling across multiple



CIHR IRSC  
Canadian Institutes of Health Research / Instituts de recherche en santé du Canada



NSERC  
CRSNG

SSHRC = CRSH

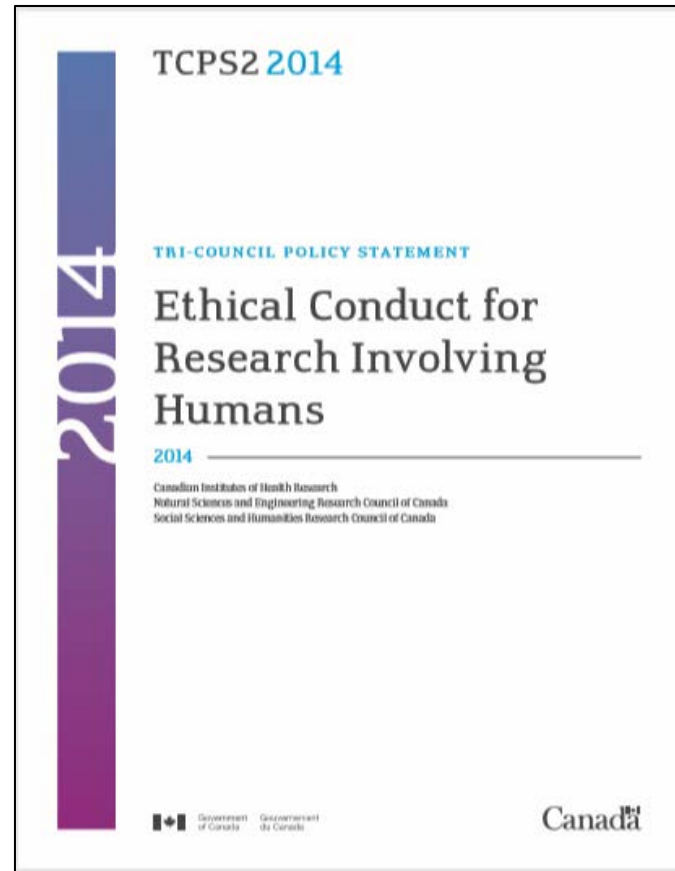
# RESEARCH BEST PRACTICE

Transparency

Trust

Ethics

Responsible Use of  
Public Funds





# INTERNATIONAL DEVELOPMENTS

## Research Funders

- National Science Foundation
- National Endowment for the Humanities - ODH
- National Institutes of Health
- UK Research Councils
- European Commission
- National Natural Science Foundation of China

## Foundations and Charities

- American Heart Association
- Bill and Melinda Gates Foundation
- Alfred P Sloan Foundation
- Gordon and Betty Moore Foundation
- Open Society Foundations
- Royal Society
- Wellcome Trust
- Cancer Research UK

## Research Institutions





OCUL Ontario Council of University Libraries



Libraries

Researchers

compute canada | calcul canada

Grad Studies

IT



Ethics

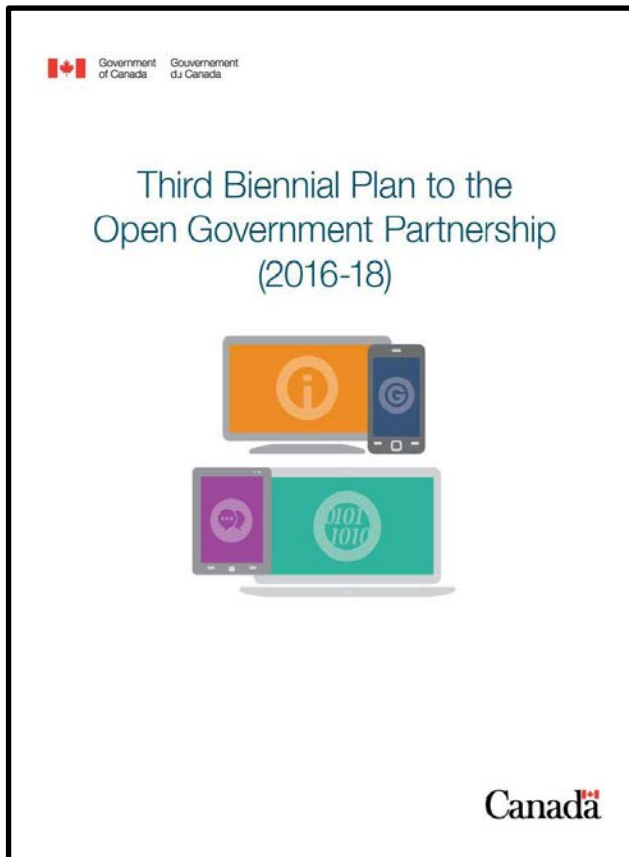
RSO



Courtesy of Chuck Humphrey, former Director, Portage



# GOVERNMENT OF CANADA DIRECTIVE ON OPEN GOVERNMENT



## Table of Contents

- I. INTRODUCTION .....3
- Towards an Open and Transparent Government** .....3
- A Joint Effort** .....4
- The Open Government Partnership** .....4
- II. PROGRESS TO DATE .....5
- III. DEVELOPING CANADA'S THIRD BIENNIAL PLAN .....7
- IV. CANADA'S OPEN GOVERNMENT COMMITMENTS (2016-18) .....8
- A. Open by Default** .....9
  - Commitment 1: Enhance Access to Information .....10
  - Commitment 2: Streamline Requests for Personal Information .....12
  - Commitment 3: Expand and Improve Open Data .....12
  - Commitment 4: Provide and Preserve Open Information .....14
  - Commitment 5: Define an Approach for Measuring Open Government Performance .....15
  - Commitment 6: Develop Open Government Skills across the Federal Public Service .....16
  - Commitment 7: Embed Transparency Requirements in the Federal Service Strategy .....17
  - Commitment 8: Enhance Access to Culture & Heritage Collections .....18
- B. Fiscal Transparency** .....19
  - Commitment 9: Enhance Openness of Information on Government Spending and Procurement .....19
  - Commitment 10: Increase Transparency of Budget and Other Department of Finance Information .....21
  - Commitment 11: Increase Transparency of Grants and Contributions Funding .....22
  - Commitment 12: Improve Public Information on Canadian Corporations .....23
- C. Innovation, Prosperity, and Sustainable Development** .....24
  - Commitment 13: Increase the Availability and Usability of Geospatial Data .....24
  - Commitment 14: Increase Openness of Federal Science Activities (Open Science)** .....26
  - Commitment 15: Increase the Availability of Open Data Exchange (Open Data Exchange) .....27
  - Commitment 16: Align Open Data across Canada (Open Data Canada) .....28
  - Commitment 17: Implement the Extractives Sector Transparency Measures Act .....29
  - Commitment 18: Support Openness and Transparency Initiatives around the World .....30
- D. Engaging Canadians and the World** .....31
  - Commitment 19: Engage Civil Society on Open Government .....32
  - Commitment 20: Enable Open Dialogue and Open Policy Making .....33
  - Commitment 21: Promote Open Government Globally .....34
  - Commitment 22: Engage Canadians to Improve Key Canada Revenue Agency Services .....35
- V. CONCLUSION .....37
- APPENDIX: Summary of Canada's Third Biennial Plan to the Open Government Partnership (2016-18) .....38

# TRI-AGENCY RDM POLICY DEVELOPMENT BACKGROUND

**2013**

Capitalizing on Big Data: Toward a Policy Framework for  
Advancing Digital Scholarship in Canada

**2016**

Tri-Agency Statement of Principles on Digital Data Management

**2017-2018**

Draft Tri-Agency Research Data Management Policy



SSHRC  CRSH

# INTENDED IMPACT OF A TRI-AGENCY RDM POLICY

The agencies aim to contribute to a future research culture that sees:

- strong data management as an accepted signifier of research excellence across disciplines, and a regular feature in the conduct of research;
- more Canadian datasets cited, and valued as a product of research in tenure, promotion and peer review processes;
- Canadian researchers equipped and ready to engage in international research collaboration where data management requirements are becoming the norm;
- Canadian research institutions ready to support the management of the data their researchers produce; and
- increased ability for research data to be archived, found and responsibly reused, to fuel new discovery and innovation.

# DRAFT TRI-AGENCY RDM POLICY

- Consultation feedback will inform final policy
- Proposed policy includes 3 possible requirements:
  1. Institutional Strategy (institutions)
  2. Data Management Plans (researchers)
  3. Data Deposit (researchers)
- Implementation: Phased, incremental

# DRAFT TRI-AGENCY RDM POLICY

## 1. Institutional Strategy

- Each institution administering tri-agency funds could be required to create an institutional research data management strategy. The strategy could outline how the institution will provide its researchers with an environment that enables and supports world class research data management practices.
- The strategy could be posted and made publicly available on the institution's website, with contact information to direct inquiries about the strategy.

# DRAFT TRI-AGENCY RDM POLICY

## Why Require Institutional Strategies?

- Recognizes the role of institutions in providing supports for data management;
- Provides an opportunity for institutions to think through where gaps exist, and how to address them from a campus-wide perspective;
- Could aid institutions in developing an approach that works for them, while encouraging alignment and collaboration with other institutions;
- Could provide information to agencies about data management capacity; and
- Serves as foundation for the potential requirements that follow.

### Example Support

Portage Institutional Strategy Template



# DRAFT TRI-AGENCY RDM POLICY

## 2. Data Management Plans

- Grant recipients could be required to create data management plans (DMPs) for research projects supported wholly or in part by tri-agency funds. Grant recipients could submit these plans to their institution's research office as a condition of the release of grant funds.
- For specific funding opportunities, the agencies could require DMPs to be submitted to the appropriate agency at time of application; in these cases, they may be considered in the adjudication process.

# DRAFT TRI-AGENCY RDM POLICY

## Why Require Data Management Plans?

- DMPs are an emerging international best practice;
- DMPs are an excellent way for researchers to identify opportunities and challenges in managing their data, well before those opportunities and challenges emerge;
- Researchers claim that the process of developing a DMP helps them to improve their research plans and methodologies;
- DMPs could serve the responsible conduct of research and the research ethics approval process; and
- DMPs help identify and mitigate issues related to ownership of data, potential for data sharing, etc.

**Example Support**  
Portage DMP Assistant

# DRAFT

## TRI-AGENCY RDM POLICY

### 3. Data Deposit

- For all research data and code that support journal publications, pre-prints and other research outputs that arise from agency-supported research, grant recipients could be required to deposit these data and code in an appropriate public repository or other platform that will ensure safe storage, preservation, curation, and (if applicable) access to the data.

# DRAFT

## TRI-AGENCY RDM POLICY

### Why Require Data Deposit?

- Methods, expectations and online security will change - storing in a secure location provides better chance for data to be safe and of use to the creator in the future;
- Data deposit helps ensure proper use of public funds;
- Facilitates reproducibility of results; and
- Facilitates data sharing.

#### Example Support

CARL-Portage-Compute Canada's  
Federated Research Data Repository (FRDR)

# COMMUNITY FEEDBACK IS KEY

Research community feedback is essential to inform the final design of the policy and the mode of its implementation.

The agencies consider the draft RDM policy as a proposal through which to advance discussion with stakeholders in the research community, with a tri-agency RDM policy as the desired end product.

# STAKEHOLDER ENGAGEMENT

- Regional stakeholder meetings
  - Vancouver, Calgary, Toronto, Montréal and Halifax;
  - revealed excitement and optimism about the potential for data management to contribute to research excellence; and
  - also demonstrated concern over challenges, such as researcher awareness, capacity and funding.
- Online consultation on draft policy June-September 2018
- Continued discussions with broad array of stakeholders
  - Researchers, scholarly and scientific associations, data management advocacy and support organizations, funding agency colleagues around the globe.

# ONLINE CONSULTATION FEEDBACK

Online consultation June - September 2018

- approx. 150 responses, mostly from colleges & universities
- areas of feedback:
  - strength of policy
  - ethics and privacy
  - implications of the policy
  - capacity
  - supports for education
- feedback will inform further development of the policy over winter 2018-19

# THANK YOU!

Questions or feedback? Contact:

NSERC: [ResearchData-Donneesderecherche@nserc-crsng.gc.ca](mailto:ResearchData-Donneesderecherche@nserc-crsng.gc.ca)

SSHRC: [ResearchData-Donneesderecherche@sshrc-crsh.gc.ca](mailto:ResearchData-Donneesderecherche@sshrc-crsh.gc.ca)

CIHR: [ResearchData-Donneesderecherche@cihr-irsc.gc.ca](mailto:ResearchData-Donneesderecherche@cihr-irsc.gc.ca)



SSHRC  CRSH