

Where is all the research data?

Opportunities for Libraries and Portage to
Improve the Discovery of Canadian Research Data

Who am I?

Health Sciences Librarian at the University of Saskatchewan

Working in research data discovery for >8 years:

- NYU School of Medicine, Lead of Data Discovery
- National Library of Medicine, Associate Fellow

Current Chair of Portage's Data Discovery Expert Group

kevin.read@usask.ca



The plan for today

Introduce a research project designed to locate Canadian research data

Identify challenges associated with improving the discovery of Canadian research data

Engage in a discussion about:

- Strategies for libraries to improve the discovery of Canadian research data
- Portage's role in making Canadian research data more discoverable

The catalyst

The screenshot shows the Government of Canada website. At the top left is the Canadian flag and the text "Government of Canada" and "Gouvernement du Canada". To the right is a search bar labeled "Search Canada.ca". Below this is a navigation menu with categories: Jobs, Immigration, Travel, Business, Benefits, Health, Taxes, and More services. A breadcrumb trail reads: Home → Collaboration between Federal Research Funding Organizations → Policies and Guidelines → Research Data Management. On the left side, there is a sidebar with a "Research Data Management" header and several links: "Tri-Agency Statement of Principles on Digital Data Management", "DRAFT Tri-Agency Research Data Management Policy – For Consultation" (highlighted), "Public Consultation Summary", and "Frequently Asked Questions". The main content area features the title "DRAFT Tri-Agency Research Data Management Policy For Consultation" and a section titled "1. Preamble". The preamble text states: "The Canadian Institutes of Health Research (CIHR), the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Social Sciences and Humanities Research Council of Canada (SSHRC) (the agencies) are federal granting agencies that promote and support research, research training, knowledge transfer and innovation within Canada. The agencies expect the research they fund to be conducted to the highest professional and domain standards, domestically and internationally. These standards support research excellence by ensuring that research is performed ethically and makes good use of public funds, experiments and studies are replicable, and research results are as accessible as possible."

http://science.gc.ca/eic/site/063.nsf/eng/h_97610.html

The catalyst

The screenshot shows the Government of Canada website. At the top, there is a navigation bar with the Canadian flag, the text "Government of Canada" and "Gouvernement du Canada", and a search box labeled "Search Canada.ca". Below this is a menu with categories: Jobs, Immigration, Travel, Business, Benefits, Health, Taxes, and More services. A breadcrumb trail reads: Home → Collaboration between Federal Research Funding Organizations → Policies and Guidelines → Research Data Management. On the left, a sidebar contains links: Research Data Management, Tri-Agency Statement of Principles on Digital Data Management, DRAFT Tri-Agency Research Data Management Policy – For Consultation (highlighted), Public Consultation Summary, and Frequently Asked Questions. The main content area features the title "DRAFT Tri-Agency Research Data Management Policy For Consultation" and a section titled "1. Preamble". The preamble text states: "The Canadian Institutes of Health Research (CIHR), the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Social Sciences and Humanities Research Council of Canada (SSHRC) (the agencies) are federal granting agencies that promote and support research, research training, knowledge transfer and innovation within Canada. The agencies expect the research they fund to be conducted to the highest professional and domain standards, domestically and internationally. These standards support research excellence by ensuring that research is performed ethically and makes good use of public funds, experiments and studies are replicable, and research results are as accessible as possible."

Data Management Plans

http://science.gc.ca/eic/site/063.nsf/eng/h_97610.html

The catalyst

The screenshot shows the Government of Canada website. At the top, there is a navigation bar with the Canadian flag, the text "Government of Canada" and "Gouvernement du Canada", and a search box labeled "Search Canada.ca". Below this is a menu with categories: Jobs, Immigration, Travel, Business, Benefits, Health, Taxes, and More services. A breadcrumb trail reads: Home → Collaboration between Federal Research Funding Organizations → Policies and Guidelines → Research Data Management. The main content area features a sidebar on the left with links: "Research Data Management", "Tri-Agency Statement of Principles on Digital Data Management", "DRAFT Tri-Agency Research Data Management Policy – For Consultation" (highlighted), "Public Consultation Summary", and "Frequently Asked Questions". The main content displays the title "DRAFT Tri-Agency Research Data Management Policy For Consultation" and the start of the "1. Preamble" section. The preamble text reads: "The Canadian Institutes of Health Research (CIHR), the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Social Sciences and Humanities Research Council of Canada (SSHRC) (the agencies) are federal granting agencies that promote and support research, research training, knowledge transfer and innovation within Canada. The agencies expect the research they fund to be conducted to the highest professional and domain standards, domestically and internationally. These standards support research excellence by ensuring that research is performed ethically and makes good use of public funds, experiments and studies are replicable, and research results are as accessible as possible."

Data Management Plans

Data Deposit

http://science.gc.ca/eic/site/063.nsf/eng/h_97610.html

The catalyst

The screenshot shows the Government of Canada website. The header includes the Canadian flag, the text "Government of Canada" and "Gouvernement du Canada", and a search bar. A navigation menu contains links for Jobs, Immigration, Travel, Business, Benefits, Health, Taxes, and More services. The breadcrumb trail reads: Home → Collaboration between Federal Research Funding Organizations → Policies and Guidelines → Research Data Management. The main content area features a sidebar with links: "Research Data Management", "Tri-Agency Statement of Principles on Digital Data Management", "DRAFT Tri-Agency Research Data Management Policy – For Consultation", "Public Consultation Summary", and "Frequently Asked Questions". The main heading is "DRAFT Tri-Agency Research Data Management Policy For Consultation". Below this is the "1. Preamble" section, which states: "The Canadian Institutes of Health Research (CIHR), the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Social Sciences and Humanities Research Council of Canada (SSHRC) (the agencies) are federal granting agencies that promote and support research, research training, knowledge transfer and innovation within Canada. The agencies expect the research they fund to be conducted to the highest professional and domain standards, domestically and internationally. These standards support research excellence by ensuring that research is performed ethically and makes good use of public funds, experiments and studies are replicable, and research results are as accessible as possible."

http://science.gc.ca/eic/site/063.nsf/eng/h_97610.html

Data Management Plans

Data Deposit

"all digital research data, metadata and code that directly support research conclusions..."

The catalyst

The screenshot shows the Government of Canada website with the following elements:

- Government of Canada / Gouvernement du Canada logo and search bar.
- Navigation menu: Jobs, Immigration, Travel, Business, Benefits, Health, Taxes, More services.
- Breadcrumb: Home → Collaboration between Federal Research Funding Organizations → Policies and Guidelines → Research Data Management
- Left sidebar menu:
 - Research Data Management (selected)
 - Tri-Agency Statement of Principles on Digital Data Management
 - DRAFT Tri-Agency Research Data Management Policy – For Consultation (selected)
 - Public Consultation Summary
 - Frequently Asked Questions
- Main content area:
 - DRAFT Tri-Agency Research Data Management Policy For Consultation**
 - 1. Preamble**
 - Text: "The Canadian Institutes of Health Research (CIHR), the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Social Sciences and Humanities Research Council of Canada (SSHRC) (the agencies) are federal granting agencies that promote and support research, research training, knowledge transfer and innovation within Canada."
 - Text: "The agencies expect the research they fund to be conducted to the highest professional and domain standards, domestically and internationally. These standards support research excellence by ensuring that research is performed ethically and makes good use of public funds, experiments and studies are replicable, and research results are as accessible as possible."

http://science.gc.ca/eic/site/063.nsf/eng/h_97610.html

Data Management Plans

Data Deposit

*"all digital research data, metadata and code that **directly support research conclusions...**"*

Further motivation



Adopt an Open Science approach to federally funded scientific and research outputs.

Develop strategies and tools to implement FAIR data principles to ensure interoperability of scientific and research data.

The Data Strategy Roadmap and the Open Science Action Plan should be aligned.

http://science.gc.ca/eic/site/063.nsf/eng/h_97992.html

My questions

How have Tri-agency funded researchers shared their data in the past (if at all)?

Where can you find Tri-agency funded research data?

How well have researchers utilized RDM best practices when sharing their data?

My questions

How have Tri-agency funded researchers shared their data in the past (if at all)?

Where can you find Tri-agency funded research data?

How well have researchers utilized RDM best practices when sharing their data?

How prepared are researchers for these new policies?

The research project

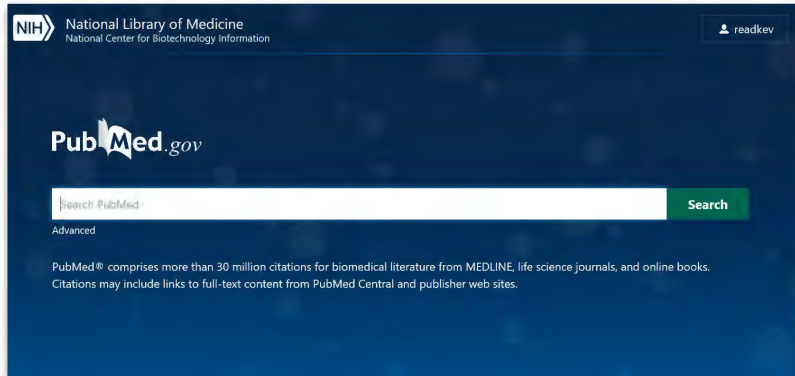
Research goals:

1. Document how and where **CIHR-funded researchers** share their data
2. Compare Canadian researchers' current data practices to the Tri-agency's proposed framework for RDM and sharing

Research team:

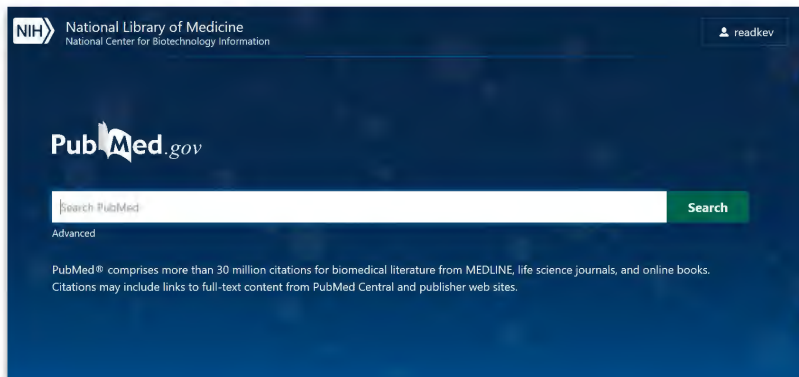
- Kevin Read, MLIS, MAS, University of Saskatchewan
- Heather Ganshorn, MLIS, University of Calgary
- David Scott, MLIS, MA, University of Lethbridge
- Sarah Rutley, MLIS, MA, University of Saskatchewan

What we did

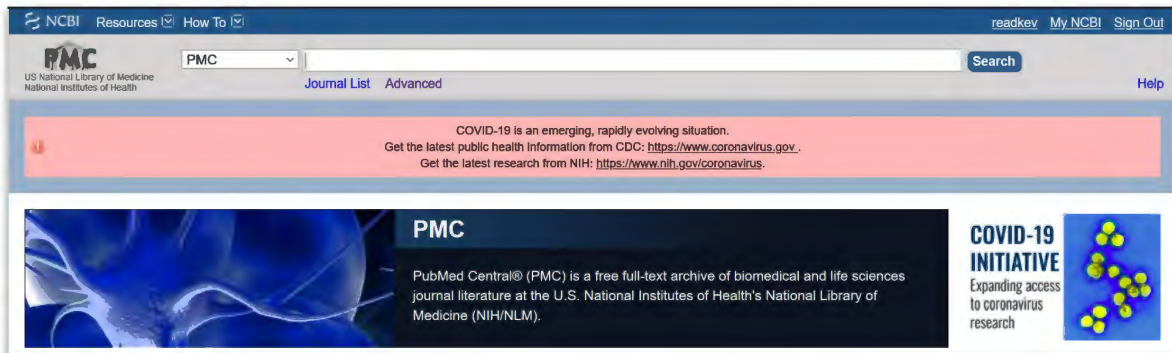


**Extracted metadata from all
CIHR-funded research articles that
indicate data was shared in a
repository**

What we did

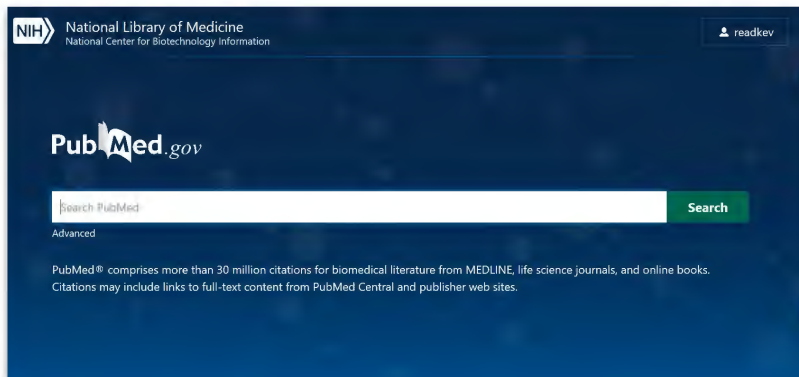


Extracted metadata from all
CIHR-funded research articles that
indicate data was shared in a
repository

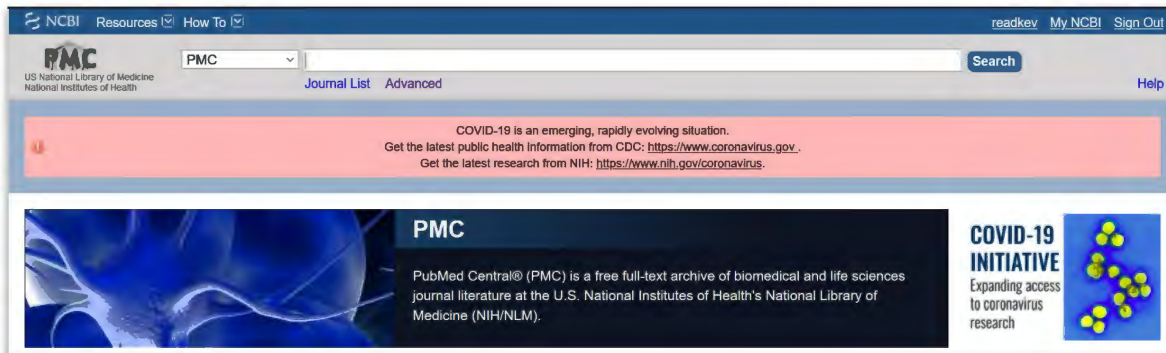


Extracted metadata from all
CIHR-funded research articles that
had a Data Availability Statement

What we did



949 articles



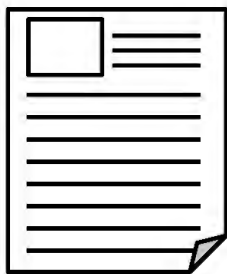
4039 articles

Metadata Element	Definition
Journal title	The journal title abbreviation, full journal title, or ISSN number
Publication date	The date that the article was published.
Publication Type	Used to identify the type of article indexed for MEDLINE
Author affiliation	Institutional affiliation and address (including email address, when available) of the authors of the article as it appears in the journal.
Grant number	Includes research grant numbers, contract numbers, or both that designate financial support by funding sources
Data Availability Statement	Author instructions on if/where data is available
Data Repository	Where data related to the article has been deposited. The complete list of databanks is available at //www.nlm.nih.gov/bsd/medline_databank_source.html .
Acknowledgements	Includes all words in the acknowledgement section of an article (e.g., “figshare[ack]”).
MeSH Major Topic Headings	A MeSH term that is one of the main topics discussed in the article.
Article body - Key Terms	Includes all key terms in the body of an article except for the Abstract and References.

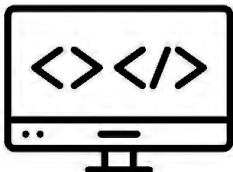
Metadata Extracted

What we did

CIHR-funded article



Article metadata



How/where are the data shared (if anywhere)?

Are researchers' data availability statements informative?

What documentation did researchers share in addition to the research data?

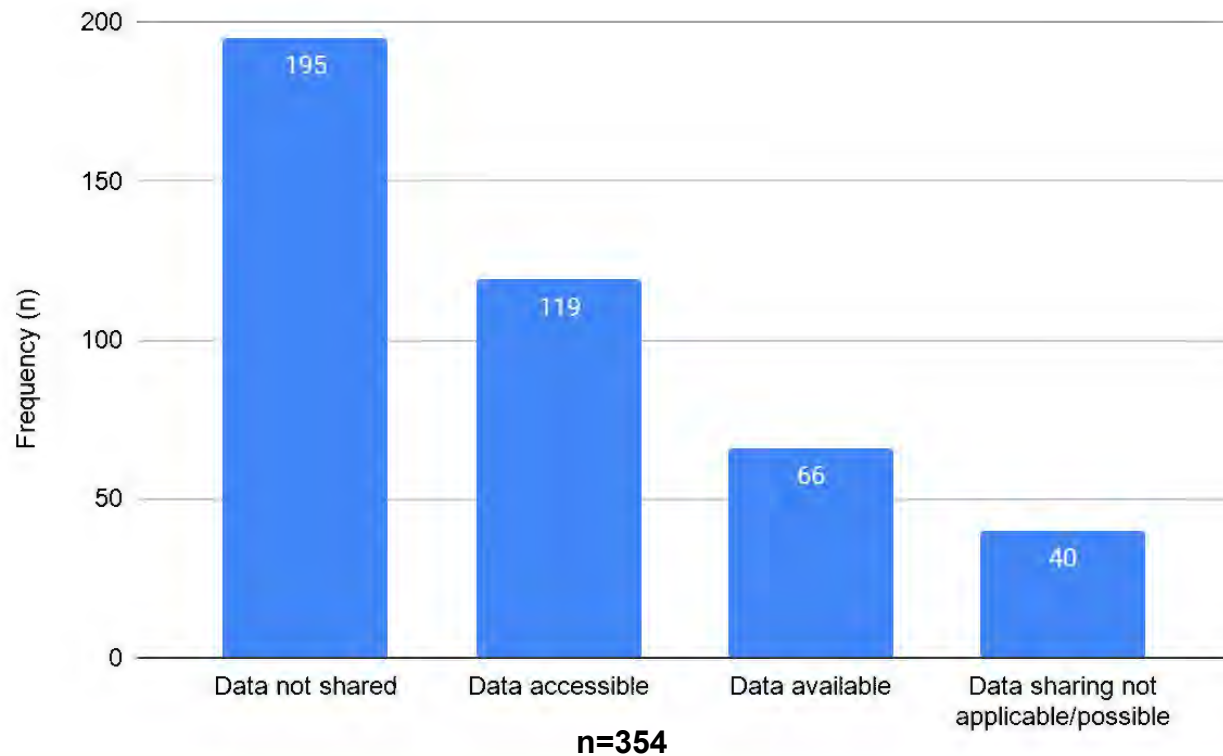
What we've found so far

Outcomes to date

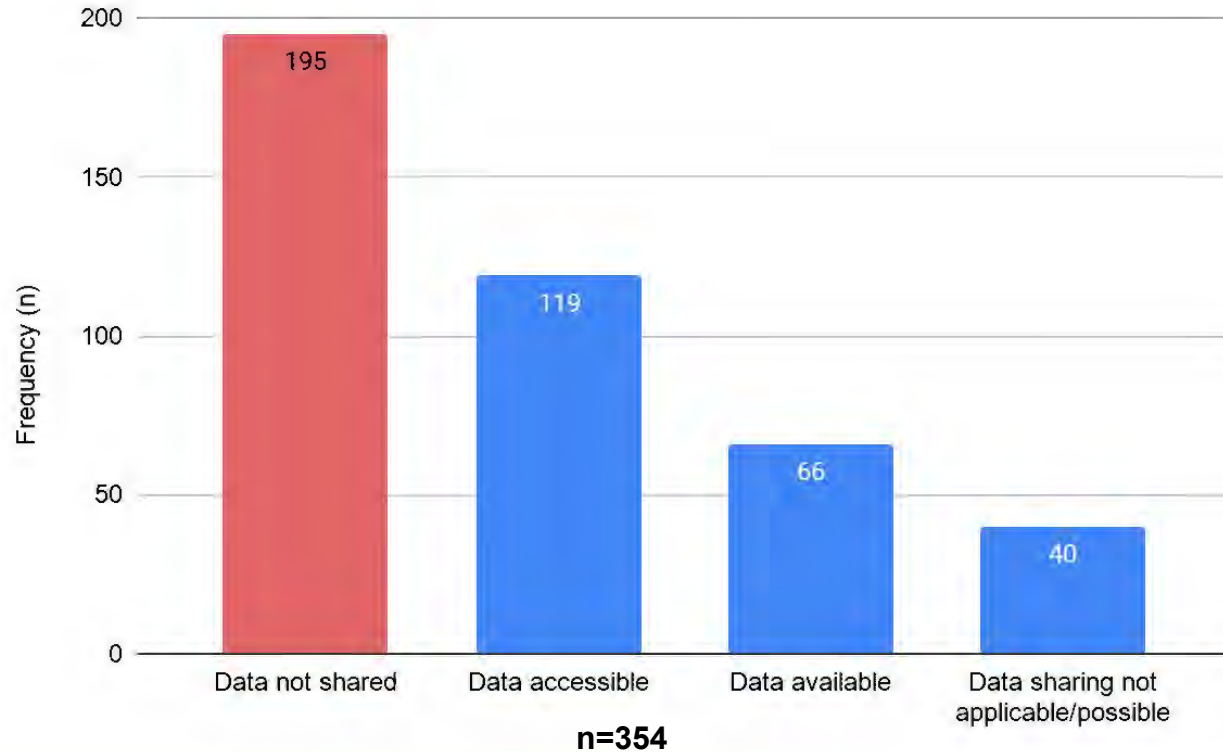
Reviewed a random sample of 354 articles

- Data sharing methods are varied and inconsistent
- Data availability statements are unclear
- Included documentation is limited beyond the tables and figures commonly shared within an article
- 55% did not share data

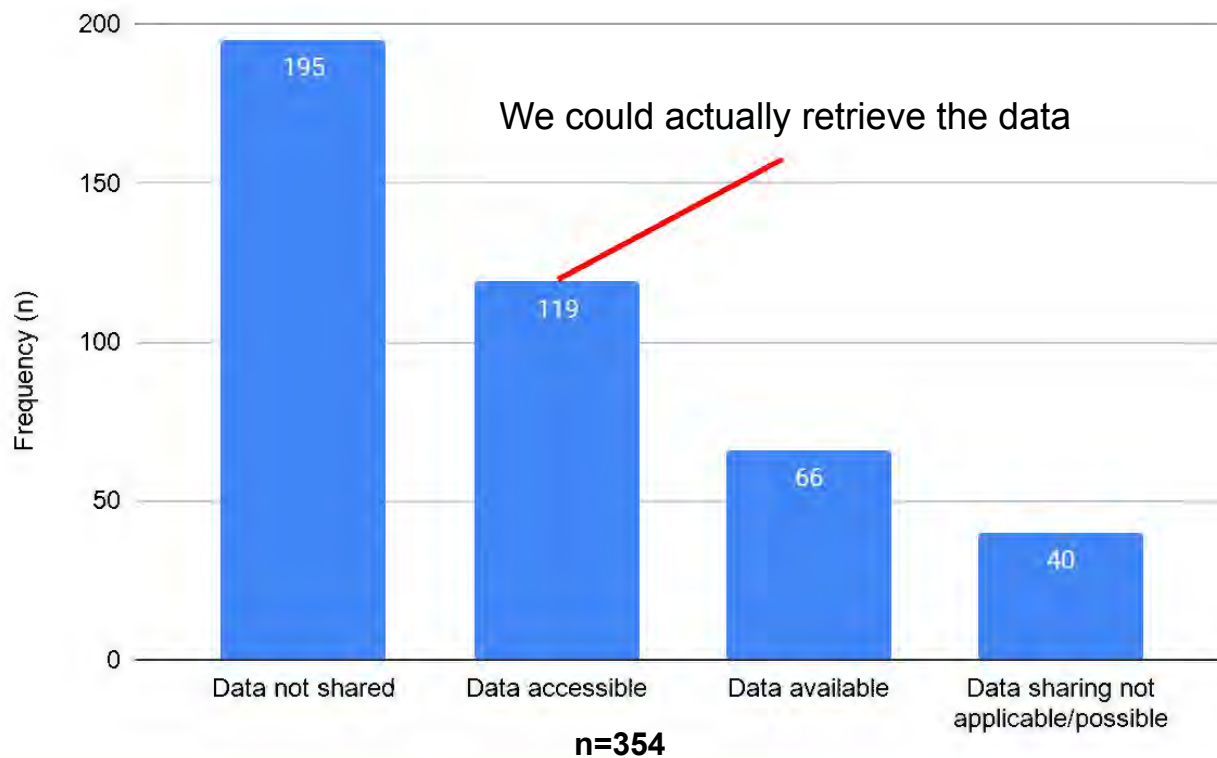
Data sharing status



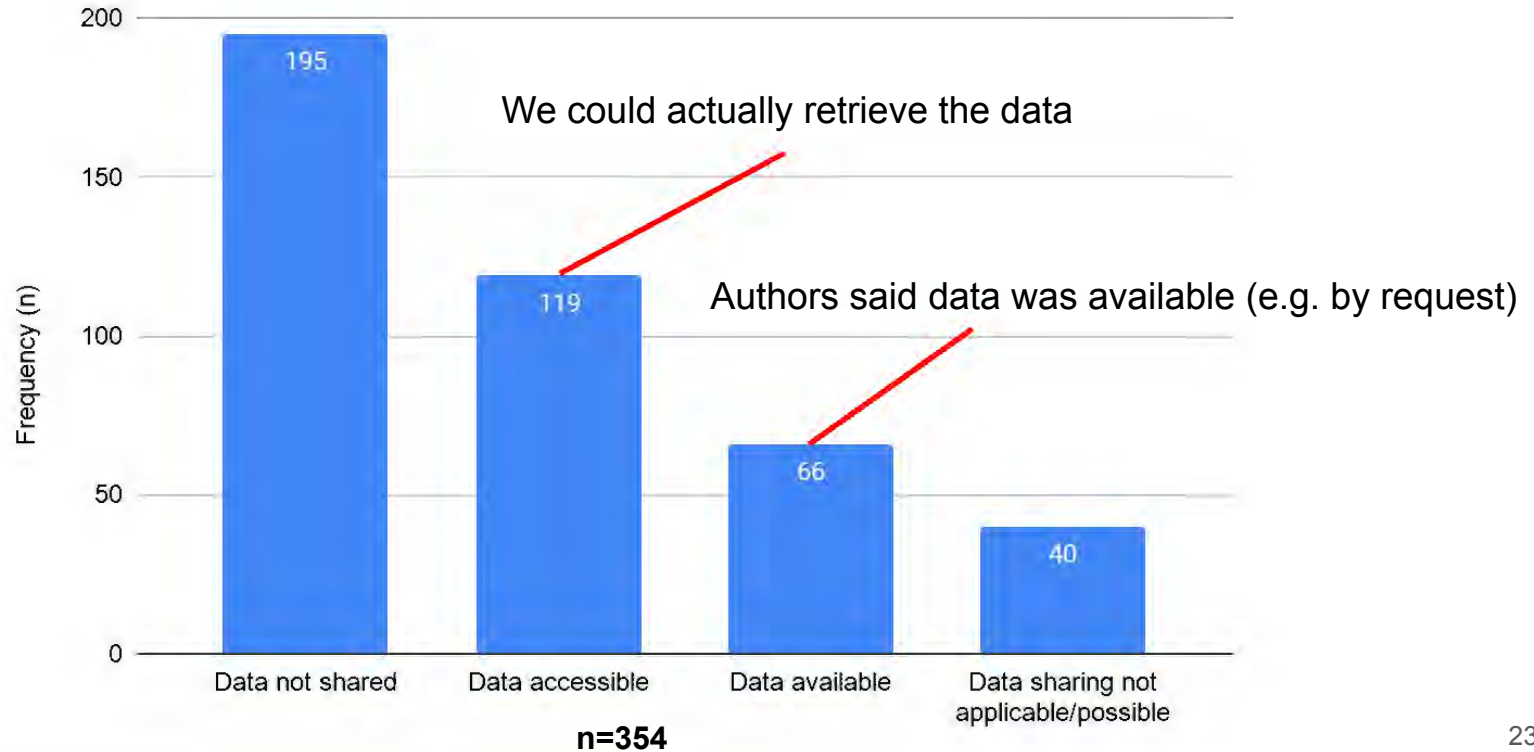
Data sharing status



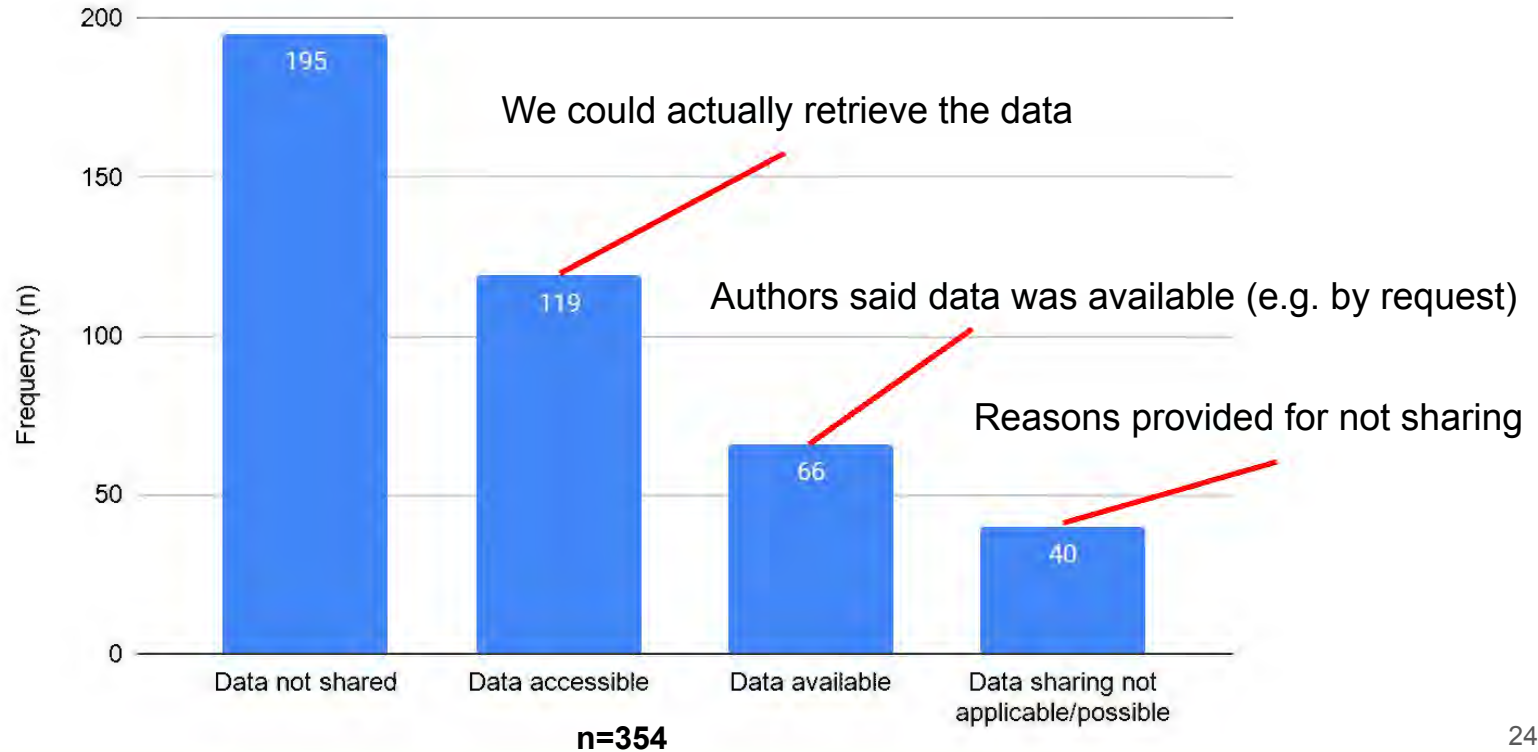
Data sharing status



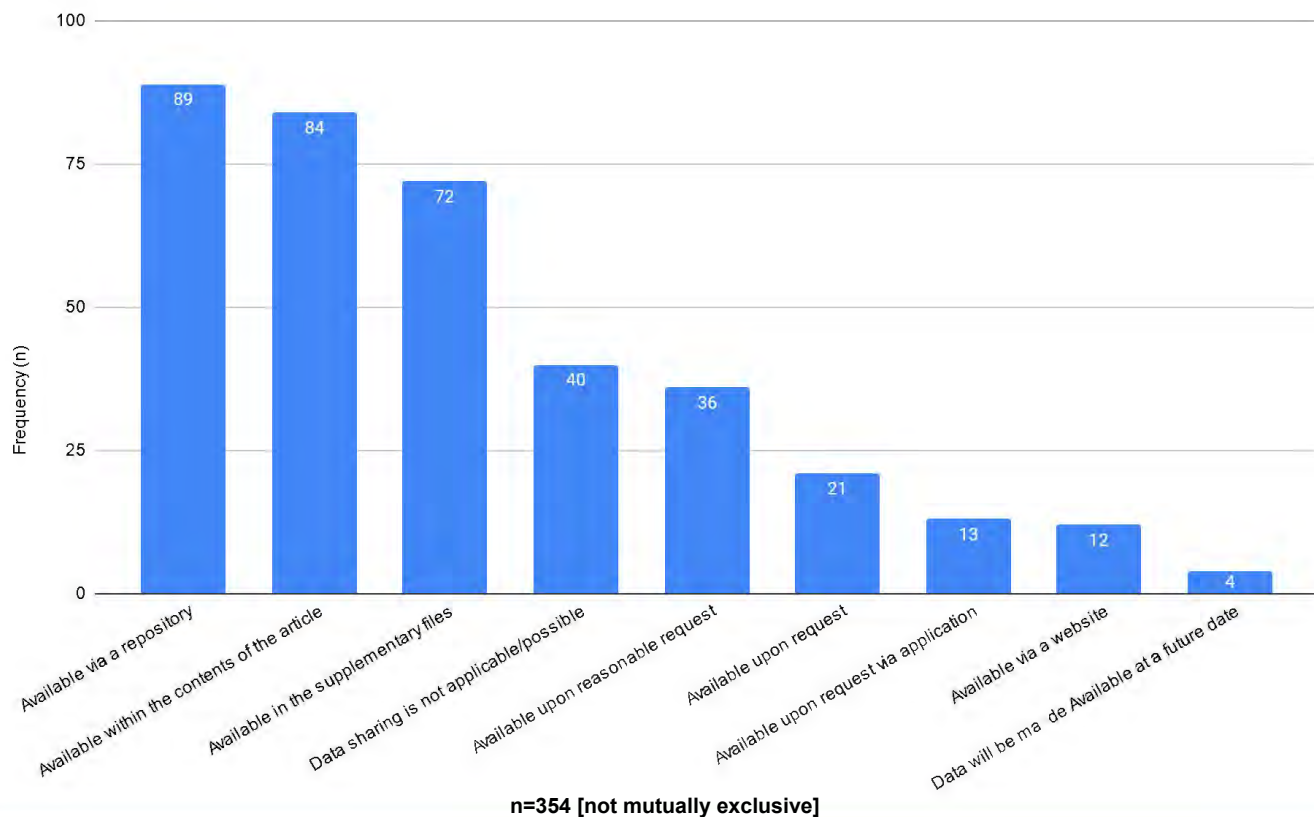
Data sharing status



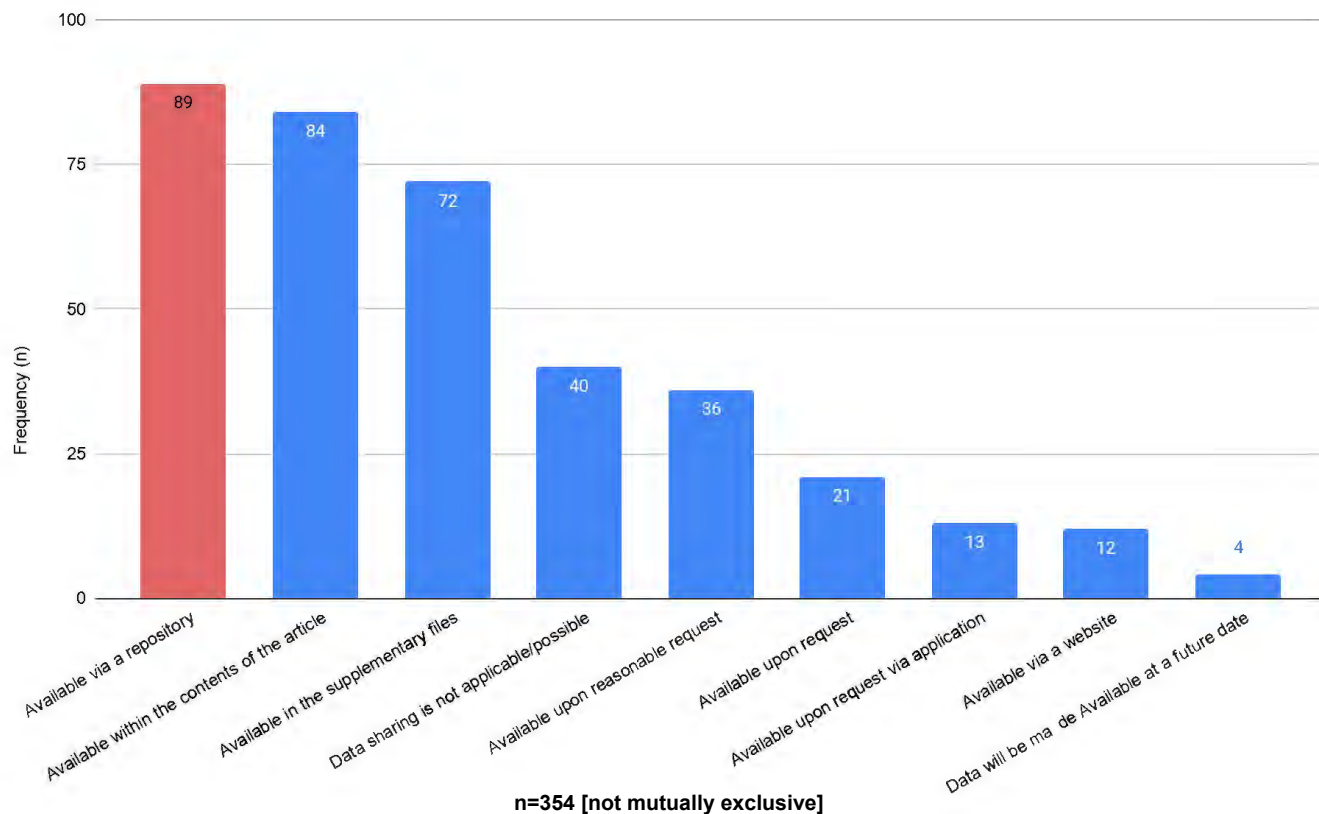
Data sharing status



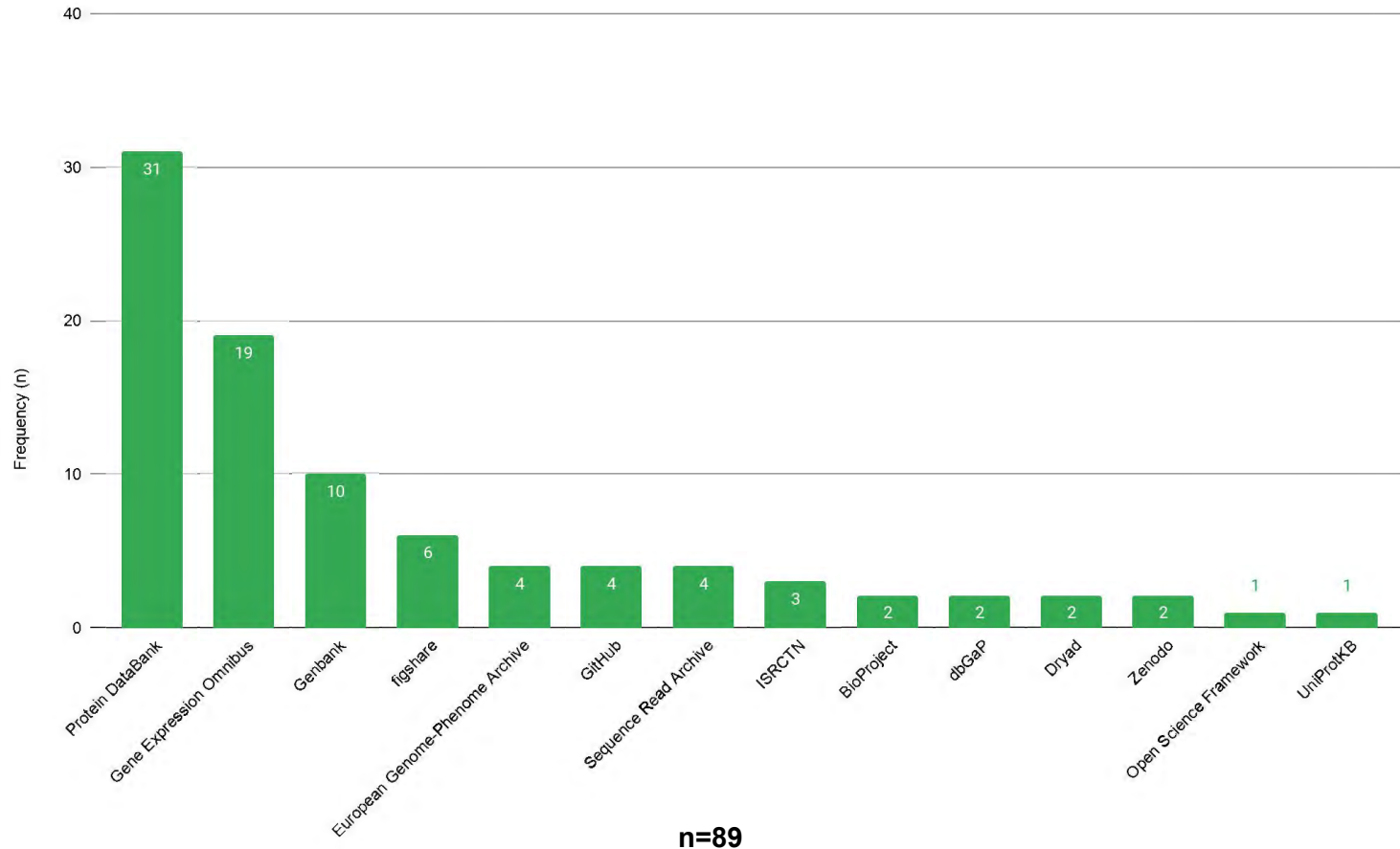
Data sharing methods



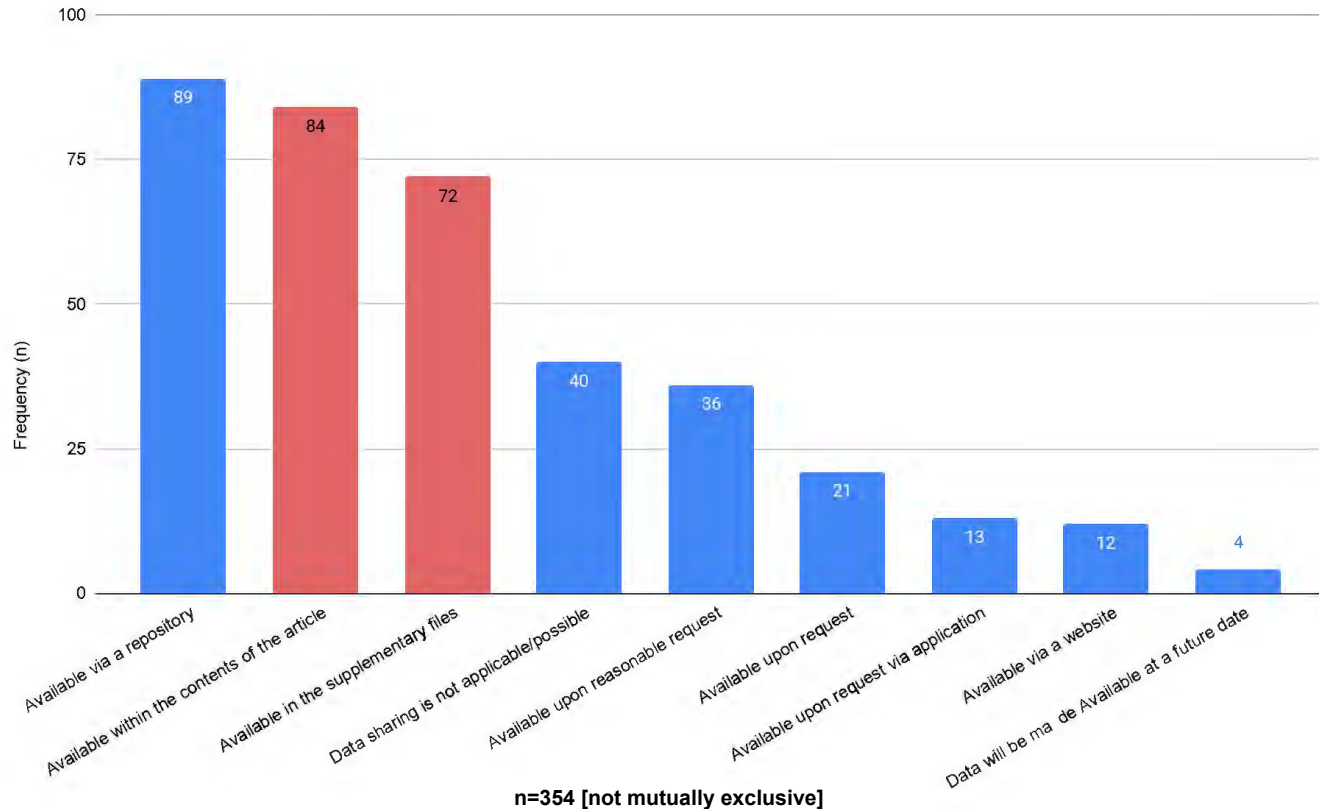
Data sharing method: Repositories



Data repositories used



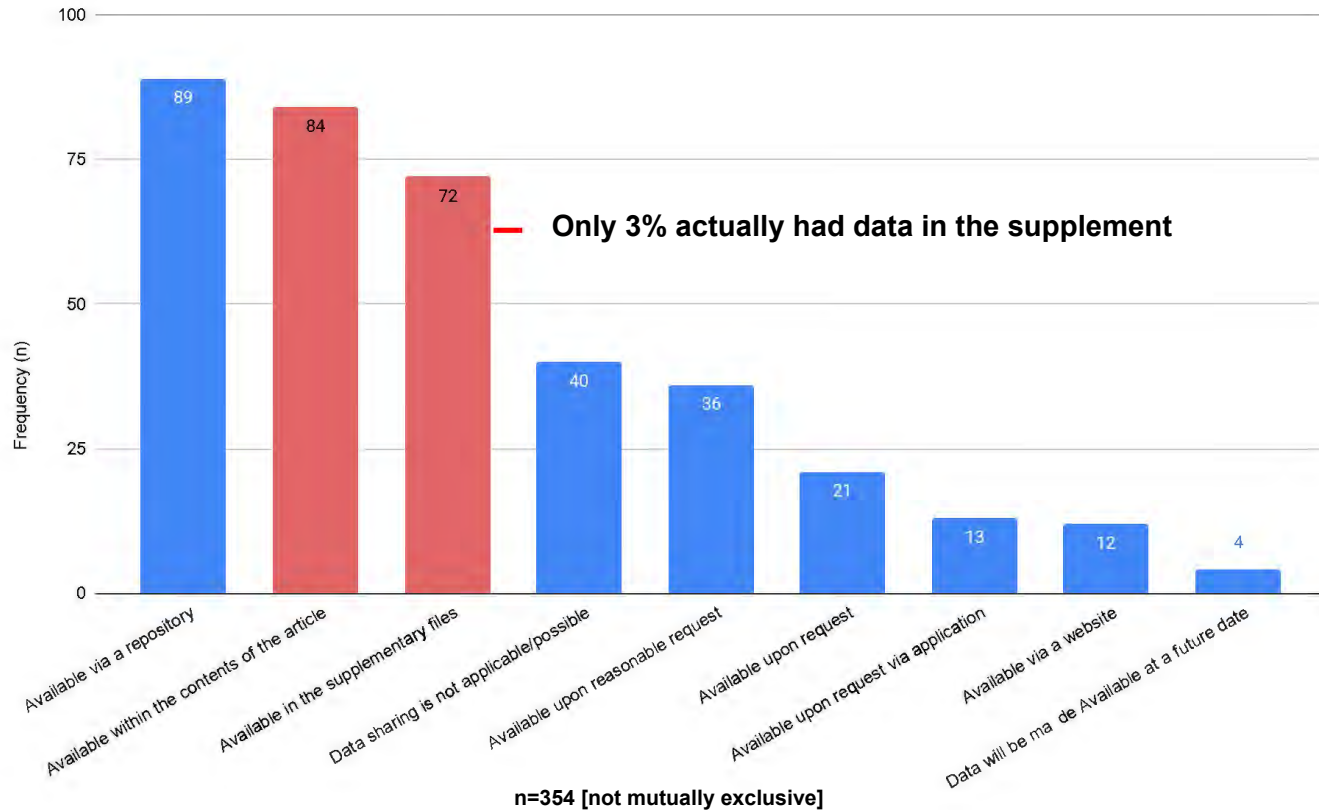
Data sharing method: Available in the article...



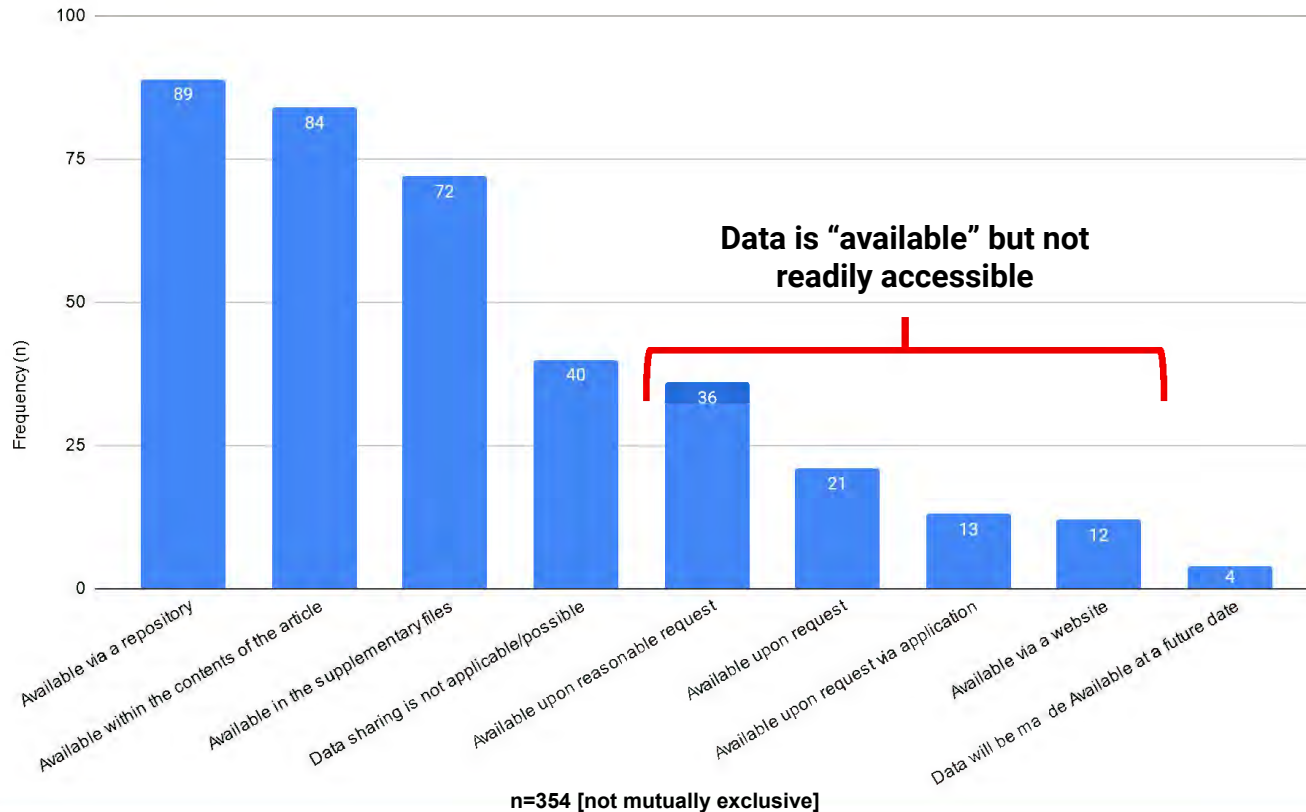
Data sharing method: Available in the article...



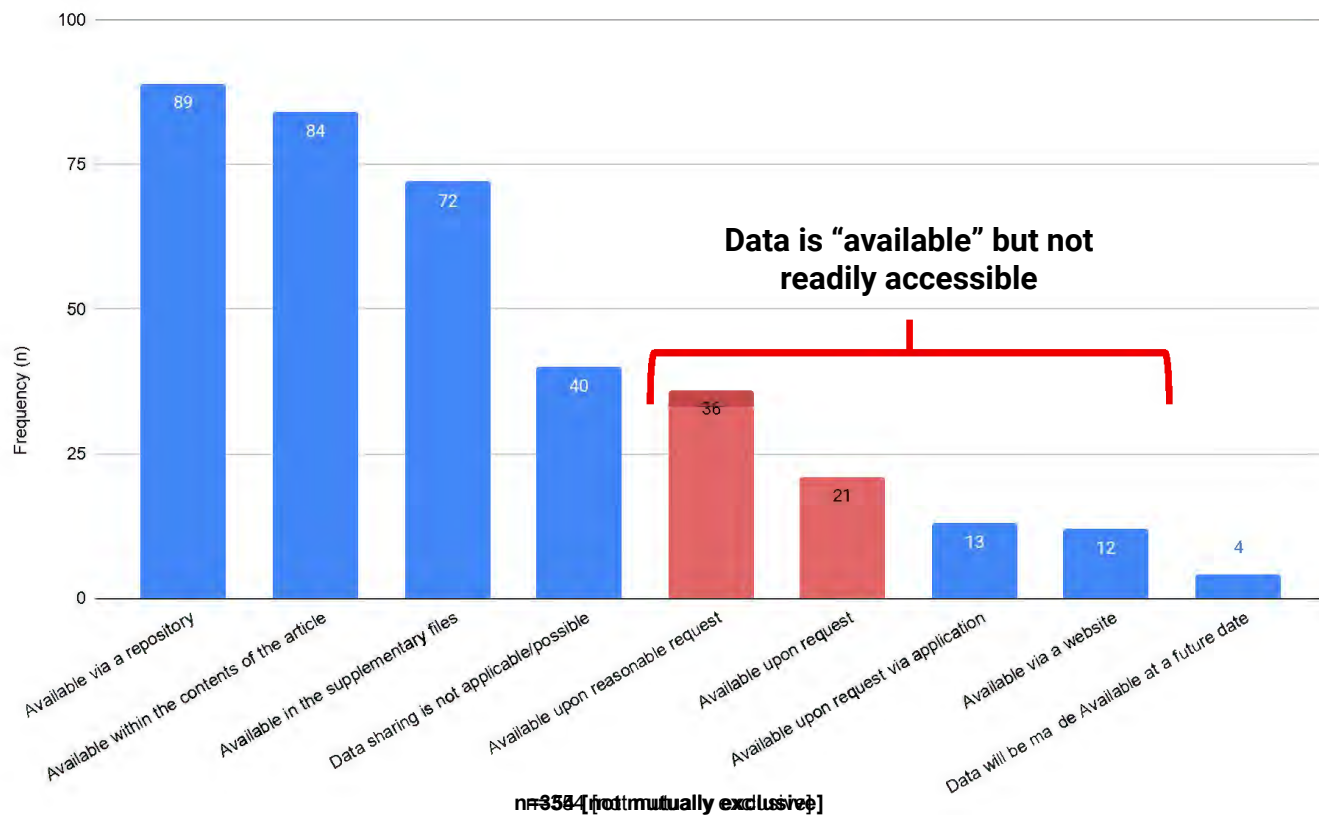
Data sharing method: Available in the article...



Data sharing method: Available data



Data sharing method: Available upon request



Lack of clarity and detail

▼ Data Availability Statement

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Lack of clarity and detail

▼ Data Availability Statement

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Lack of clarity and detail

▾ Data Availability Statement

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

▾ Data Availability Statement

Data availability

The datasets generated and/or analysed during the current study are available from the corresponding author on reasonable request. All data generated or analysed during this study are included in this published article (and its supplementary information files).

Lack of clarity and detail

▾ Data Availability Statement

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

▾ Data Availability Statement

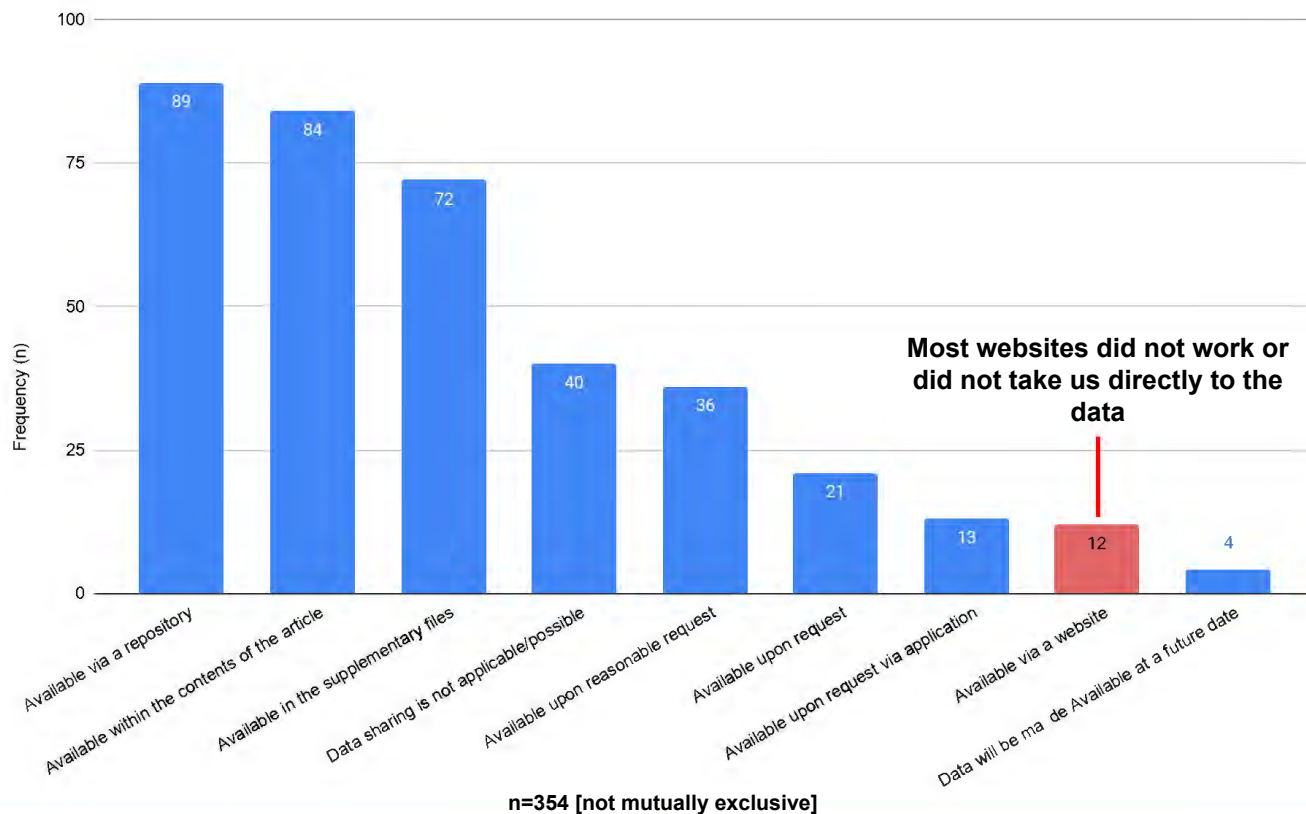
Data availability

The datasets generated and/or analysed during the current study are available from the corresponding author on reasonable request. All data generated or analysed during this study are included in this published article (and its supplementary information files).

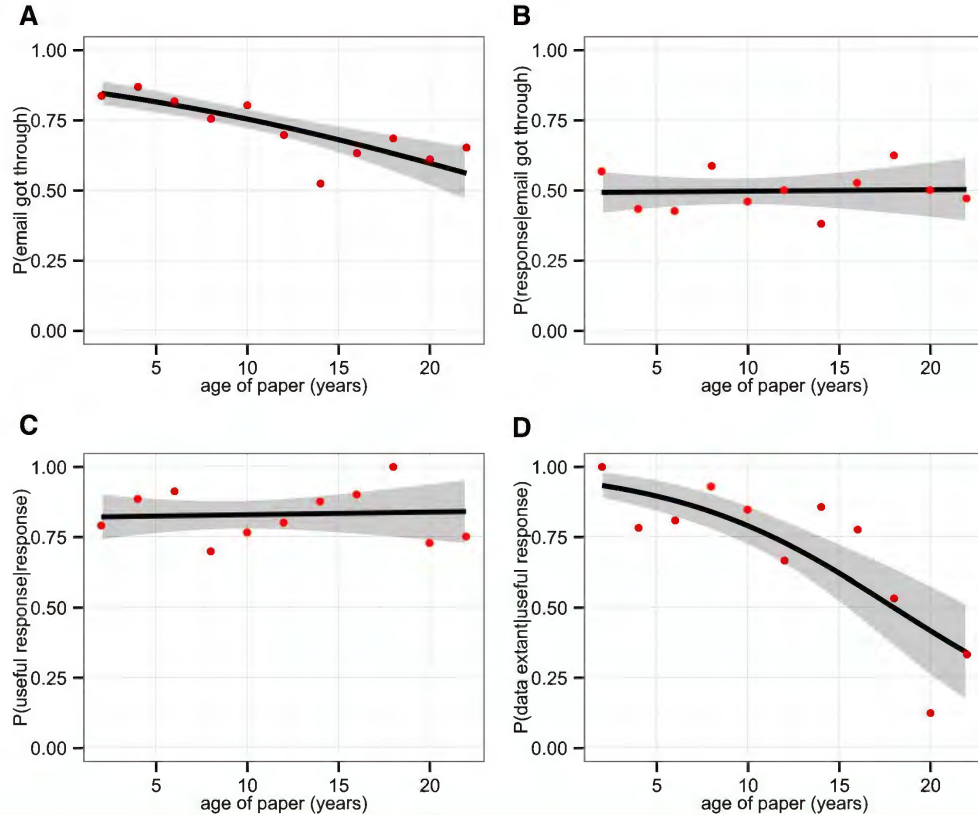
▾ Data Availability Statement

To maintain participant privacy, the minimal anonymized dataset is available by request from the study whose authors may be contacted at csdewa@ucdavis.edu. The results of this manuscript may be replicated in a similar context and participants. There is no data access committee from which the data can be accessed upon request.

Data sharing method: Websites

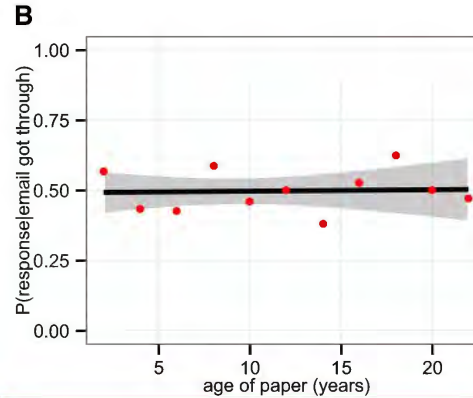
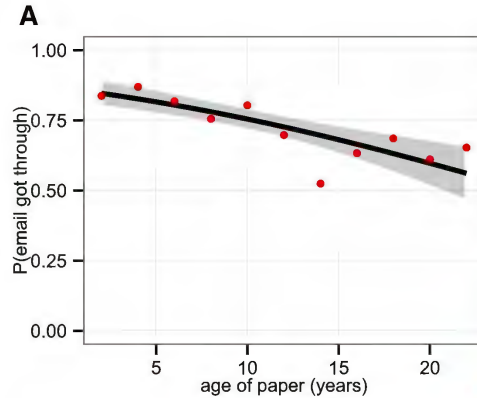


The data availability decline...

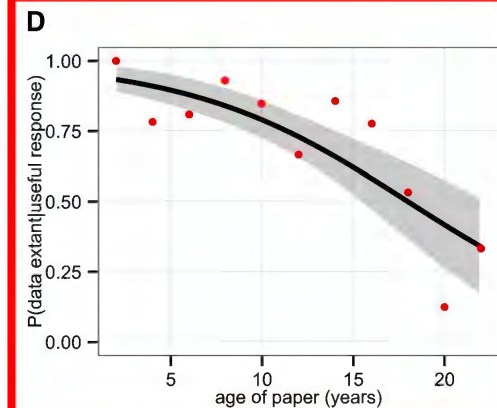
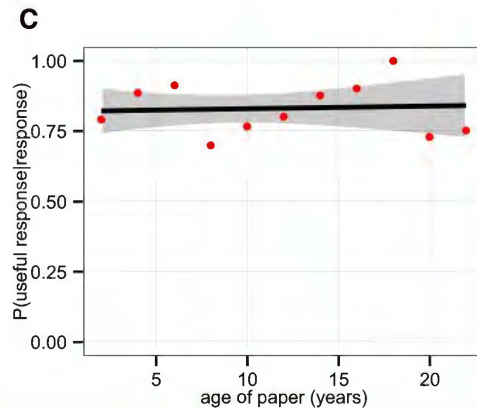


Vines TH, Albert AYK, Andrew RL, Débarre F, Bock DG, Franklin MT, Gilbert KJ, Moore JS, Renaut S, Rennison DJ. The availability of research data declines rapidly with article age. *Curr Biol.* 2014 Jan 6;24(1):94-97. Doi: 10.1016/j.cub.2013.11.014.

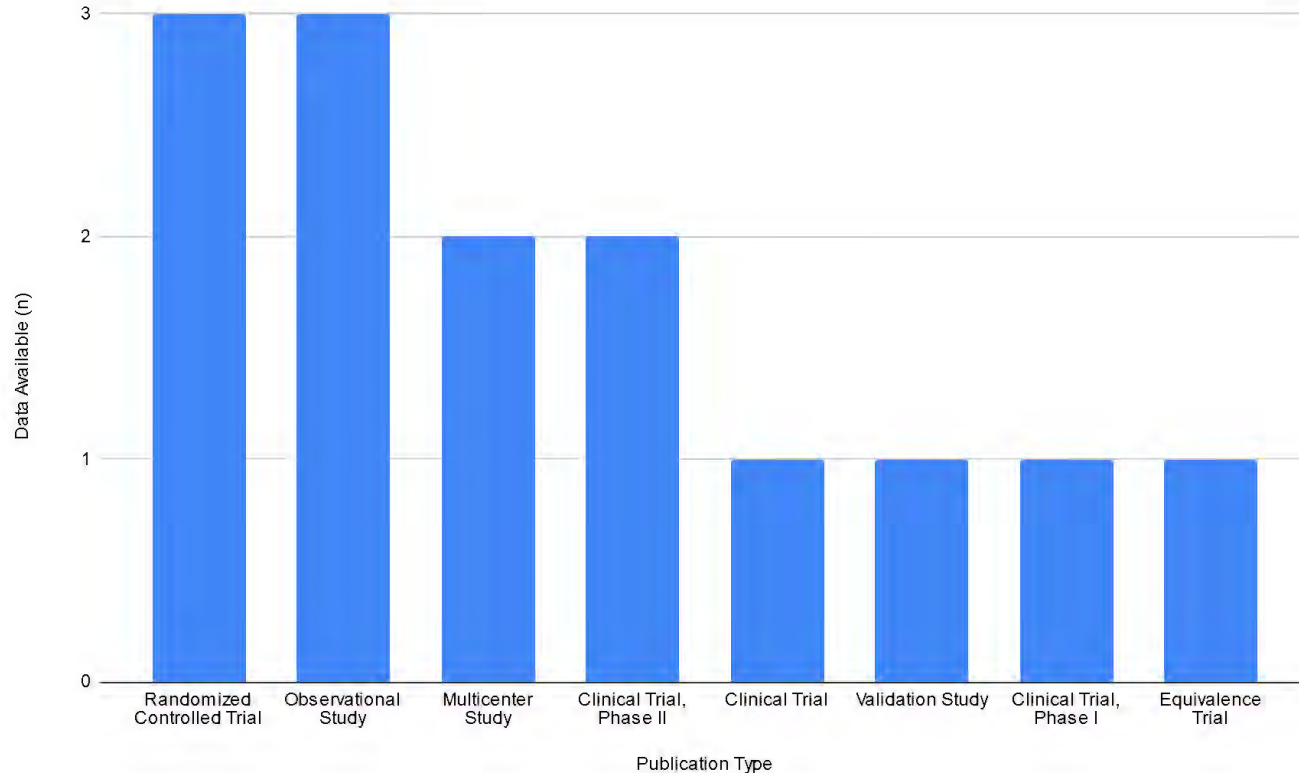
The data availability decline...



Vines TH, Albert AYK, Andrew RL, Débarre F, Bock DG, Franklin MT, Gilbert KJ, Moore JS, Renaut S, Rennison DJ. The availability of research data declines rapidly with article age. *Curr Biol.* 2014 Jan 6;24(1):94-97. Doi: 10.1016/j.cub.2013.11.014.



Restricted or licensed data is still valuable data



The policy says...

The screenshot shows the Government of Canada website with the following elements:

- Government of Canada / Gouvernement du Canada logo and search bar.
- Navigation menu: Jobs, Immigration, Travel, Business, Benefits, Health, Taxes, More services.
- Breadcrumbs: Home → Collaboration between Federal Research Funding Organizations → Policies and Guidelines → Research Data Management.
- Left sidebar menu:
 - Research Data Management (selected)
 - Tri-Agency Statement of Principles on Digital Data Management
 - DRAFT Tri-Agency Research Data Management Policy – For Consultation (selected)
 - Public Consultation Summary
 - Frequently Asked Questions
- Main content area:
 - DRAFT Tri-Agency Research Data Management Policy For Consultation**
 - 1. Preamble**
 - Text: "The Canadian Institutes of Health Research (CIHR), the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Social Sciences and Humanities Research Council of Canada (SSHRC) (the agencies) are federal granting agencies that promote and support research, research training, knowledge transfer and innovation within Canada."
 - Text: "The agencies expect the research they fund to be conducted to the highest professional and domain standards, domestically and internationally. These standards support research excellence by ensuring that research is performed ethically and makes good use of public funds, experiments and studies are replicable, and research results are as accessible as possible."

http://science.gc.ca/eic/site/063.nsf/eng/h_97610.html

Data Management Plans

Data Deposit

*"all digital research data, metadata and code that **directly support research conclusions...**"*

**The value of making harder to access
(and find) research data more
discoverable**

What are the benefits?

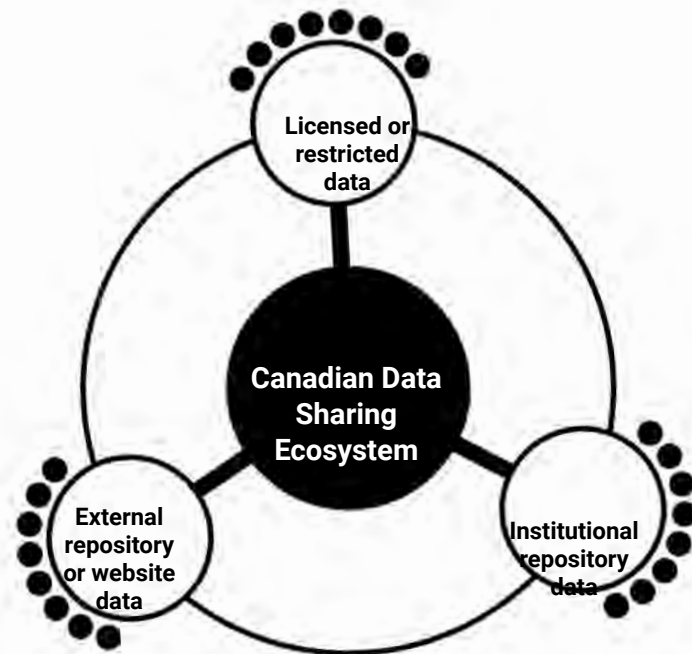
Locate data that may be undiscoverable otherwise

Researchers can remain platform agnostic

Can develop better standards to index:

- Research data available in various locations
- Research data with different access restrictions

Gain a more complete picture of Canadian research data sharing



Challenges

Really hard to do!

How to track data that is everywhere?

Requires:

- Constant monitoring of data
- Close connection to research communities who place restrictions on data access
- Stronger relationships with repositories
- Improved APIs



POLL

Do *you* see value in making Canadian research data that is not easily accessible (e.g., licensed data, restricted data, data available in external repositories/websites) more discoverable?

**Who is responsible for making this
research data discoverable?**

The case for Portage and FRDR: Reporting Tool



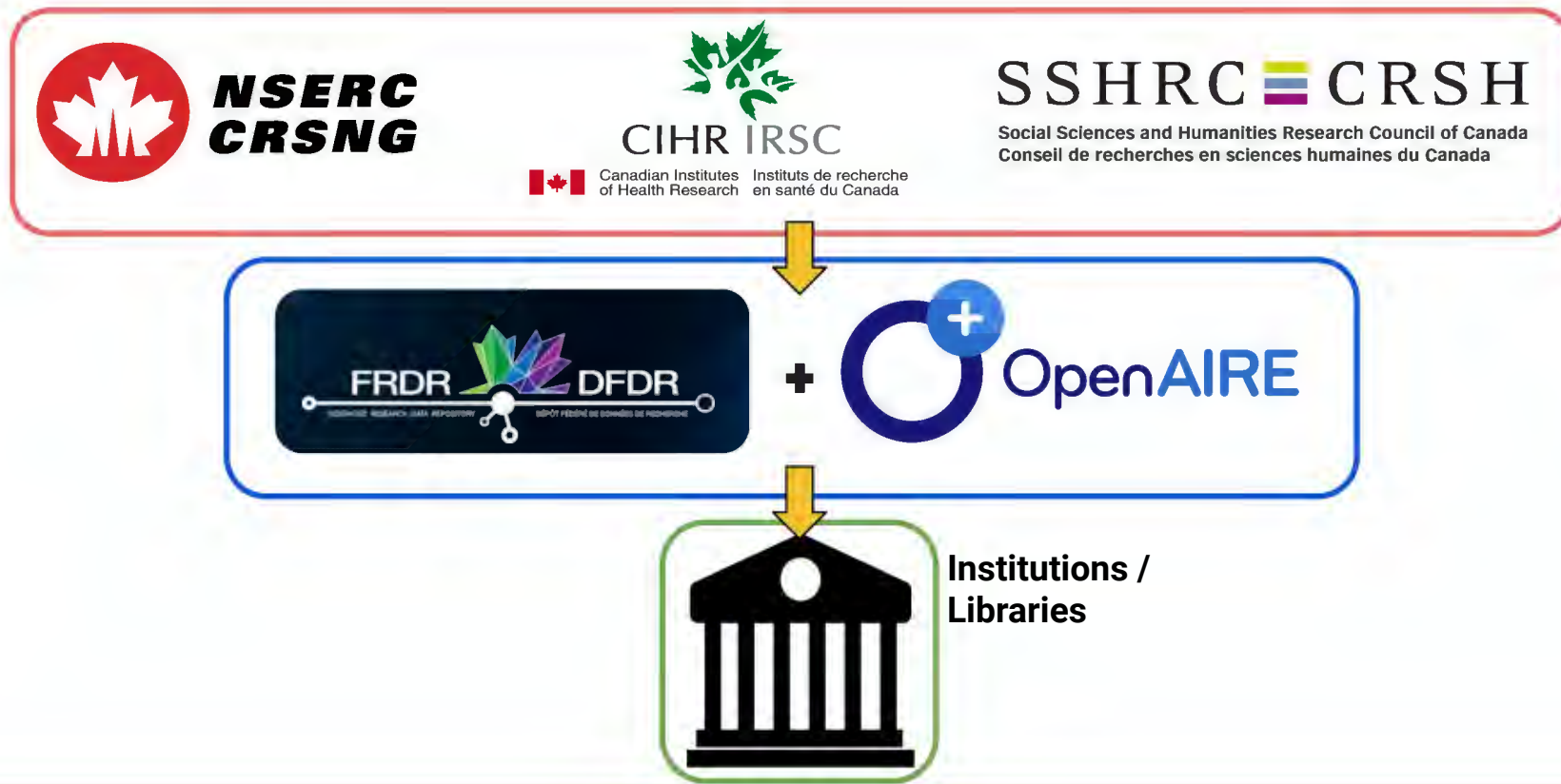
Research Data



The case for Portage and FRDR: More Context



The case for Portage and FRDR: Discoverability

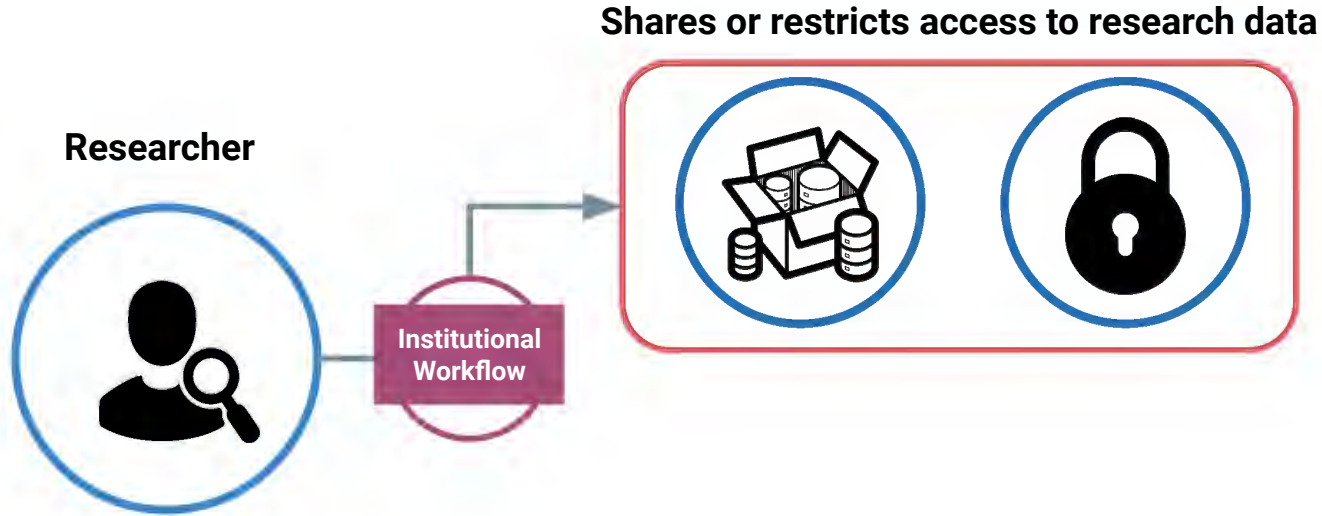


The case for libraries: Front line opportunity

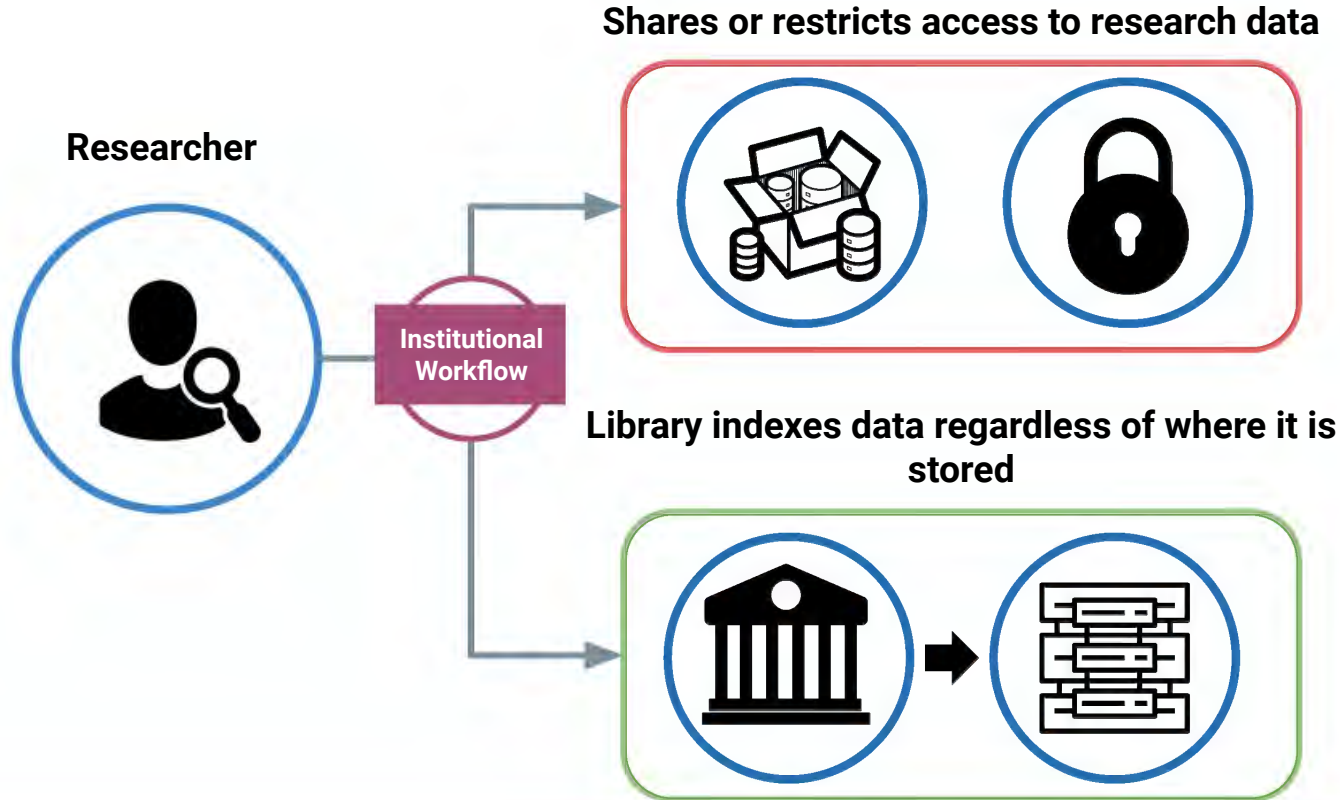
Researcher



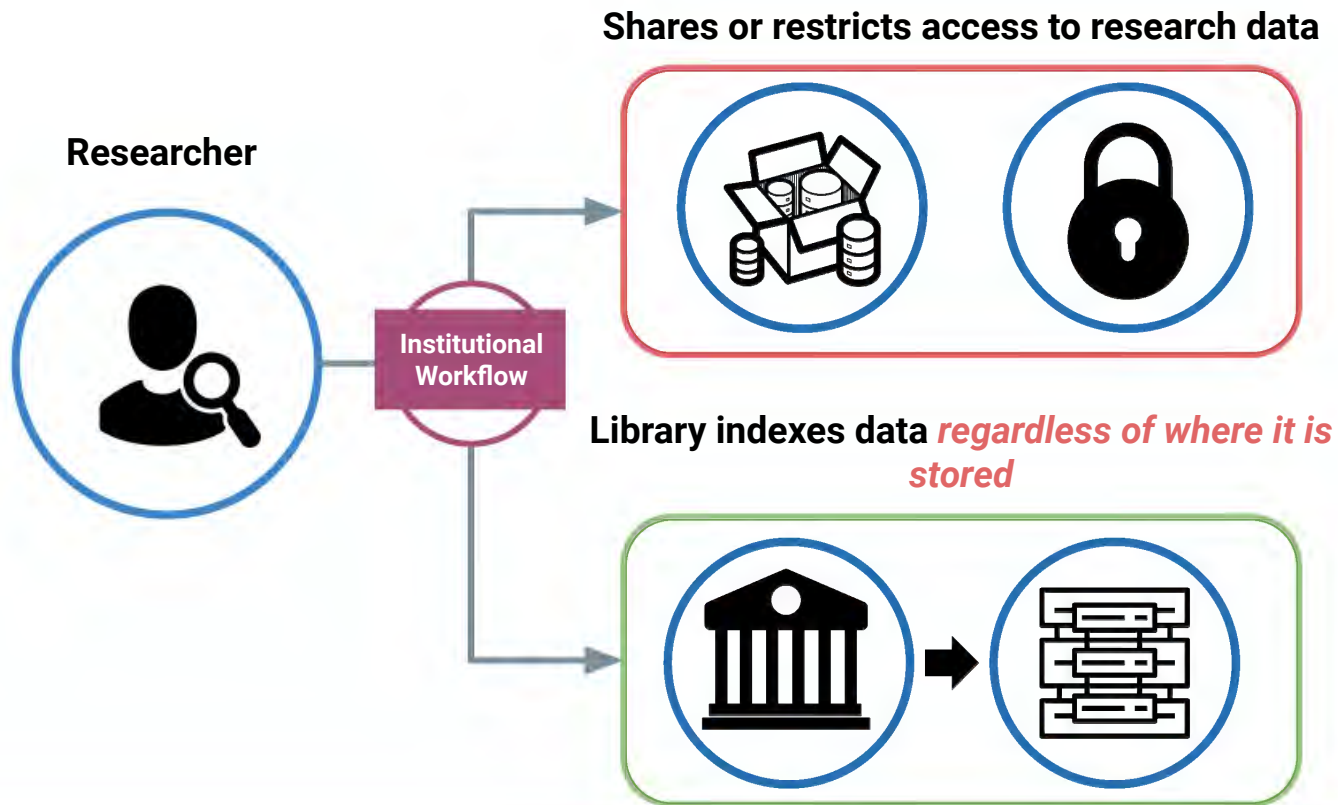
The case for libraries: Front line opportunity



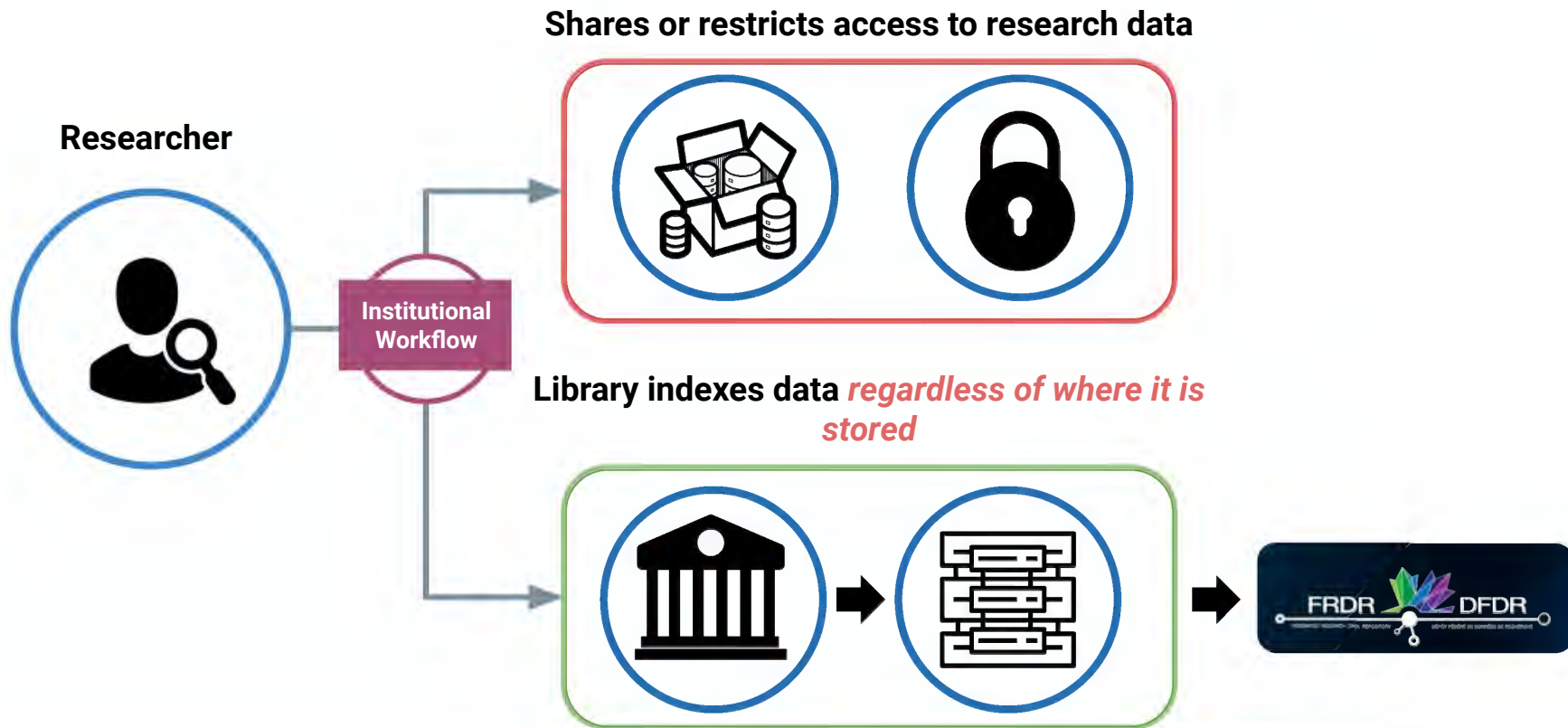
The case for libraries: Front line opportunity



The case for libraries: Front line opportunity



Libraries and Portage working together



POLL

In its current state, would your library have the ability to track all of the research data shared by your research community (regardless of where it is stored)?

Real world example: Data Discovery Collaboration

DATA DISCOVERY COLLABORATION

HOME ABOUT JOIN THE DCCP DCCP BLOG CONTACT US

About

ABOUT US

HOW IT WORKS

PARTNERS

PUBLICATIONS

FUNDING

HOW THE DATA CATALOG WORKS

The Data Catalog provides a simple search interface to discover biomedical research data at the local institution where that data is located. Because we do not require researchers to share their data in a repository, much of the research data you will find in each respective Data Catalog is not available anywhere else.

LOCATE RESEARCH DATA NOT DISCOVERABLE BY OTHER MEANS

The Data Catalog allows researchers to describe their data in order to make it discoverable, but it does not require them to share the data. The Data Catalog allows users to make requests for data directly through the author, an administrator, or a repository.

National Survey of U.S. Smartphone Owner Health App Usage Dataset NYU Dataset

UID: 10052

Author(s): Paul Krebs, Dustin Duncan*

* Corresponding Author



National Survey of U.S. Smartphone Owner Health App Usage Dataset

NYU Dataset

UID: 10052

Author(s): Paul Krebs, Dustin Duncan*

* Corresponding Author

Description

This dataset contains results of a cross-sectional survey of 1,504 smartphone users from the United States. The survey consists of 36 items assessing sociodemographic characteristics, history of health app use, reasons for use, perceived effectiveness of health apps, reasons for stopping use and general health status. Individuals included in the survey data are over 18, speak English and own a smartphone. Respondents in the survey are 50% female, 50% having completed high school or fewer years of education, 60% earning less than \$50,000 per year, 30% Hispanic/Latino, 30% Black, 30% White and 10% Asian or other.

Timeframe

2015 - 2015

Geographic Coverage

National

Subject Domain

Delivery of Health Care
Health Status
Population Characteristics

[Access via Administrator](#)

Access Restrictions

Application Required
Author approval required

Access Instructions

Application is required, using the Study Proposal Form. For access to the form, please email the administrator, Nicole Cortaxis. Once the form is completed and returned, the authors will respond with any additional questions to the data request.

Associated Publications

Krebs P, Duncan DT. Health App Use Among US Mobile Phone Owners: A National Survey. JMIR mhealth health. 2015 Nov 4;3(4):e101.

Data Type

Surveys

<https://www.datacatalogcollaborationproject.org/>

Remember the policy guidance



The screenshot shows the Government of Canada website with the following content:

- Government of Canada / Gouvernement du Canada
- Search Canada.ca
- Navigation menu: Jobs, Immigration, Travel, Business, Benefits, Health, Taxes, More services
- Breadcrumbs: Home → Collaboration between Federal Research Funding Organizations → Policies and Guidelines → Research Data Management
- Left sidebar menu:
 - Research Data Management
 - Tri-Agency Statement of Principles on Digital Data Management
 - DRAFT Tri-Agency Research Data Management Policy – For Consultation**
 - Public Consultation Summary
 - Frequently Asked Questions
- Main content area:
 - DRAFT Tri-Agency Research Data Management Policy For Consultation**
 - 1. Preamble**
 - The Canadian Institutes of Health Research (CIHR), the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Social Sciences and Humanities Research Council of Canada (SSHRC) (the agencies) are federal granting agencies that promote and support research, research training, knowledge transfer and innovation within Canada.
 - The agencies expect the research they fund to be conducted to the highest professional and domain standards, domestically and internationally. These standards support research excellence by ensuring that research is performed ethically and makes good use of public funds, experiments and studies are replicable, and research results are as accessible as possible.

http://science.gc.ca/eic/site/063.nsf/eng/h_97610.html

Data Management Plans

Data Deposit

*“all digital research data, metadata and code that **directly support research conclusions...**”*

POLL

Who should be responsible for tracking the whereabouts of Canadian research data?

How do we improve the reporting of data sharing in the published literature?

Reporting practices are poor

Data Availability Statement

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

BMJ	Nature	Wiley	Taylor and Francis
<p>Please state what the data are (e.g. deidentified participant data), who the data are available from, their publishable contact details (e.g. a generic lab email address or an individual's ORCID identifier – please ensure you have permission) and under what conditions reuse is permitted. Is there additional information available (e.g. protocols, statistical analysis plans)?</p>	<p>The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request</p>	<p>The data that support the findings of this study are available from the corresponding author upon reasonable request.</p>	<p>The data that support the findings of this study are available from the corresponding author, [author initials], upon reasonable request.</p>

Reporting practices are poor

Data Availability Statement

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

BMJ	Nature	Wiley	Taylor and Francis
<p>Please state what the data are (e.g. deidentified participant data), who the data are available from, their publishable contact details (e.g. a generic lab email address or an individual's ORCID identifier – please ensure you have permission) and under what conditions reuse is permitted. Is there additional information available (e.g. protocols, statistical analysis plans)?</p>	<p>The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request</p>	<p>The data that support the findings of this study are available from the corresponding author upon reasonable request.</p>	<p>The data that support the findings of this study are available from the corresponding author, [author initials], upon reasonable request.</p>

Reporting practices are poor

Data Availability Statement

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

BMJ

Please state what the data are (e.g. deidentified participant data), who the data are available from, their publishable contact details (e.g. a generic lab email address or an individual's ORCID identifier – please ensure you have permission) and under what conditions reuse is permitted. Is there additional information available (e.g. protocols, statistical analysis plans)?

Nature

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request

Wiley

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Taylor and Francis

The data that support the findings of this study are available from the corresponding author, [author initials], upon reasonable request.

More clarity = More sharing




Participants who received a data-sharing agreement were more willing to share their dataset

A member of the control group is 24% more likely to share her dataset should she receive the data-sharing agreement

Template data sharing agreements included within publication

[Polanin JR, Terzian M. A data-sharing agreement helps to increase researchers' willingness to share primary data: results from a randomized controlled trial. Journal of clinical epidemiology. 2019 Feb 1;106:60-9.](https://doi.org/10.1016/j.jclinepi.2018.10.006)

Develop better accessibility metadata

Data Availability Go to: 

Data have been provided under agreement with the Canadian Multicentre Osteoporosis Study (CaMos). The CaMos has developed an Ancillary Project and Data Release Policy that governs ancillary project approval and access to the data. The Design Analysis and Publications (DAP) Committee acts to oversee the selection and implementation of ancillary projects. They review each project application considering feasibility, priority and its impact on CaMos and will, after due consultation with the applicants, make a decision regarding the acceptability of the project. Ancillary projects may be undertaken in any one of CaMos' nine regional centres, or as a collaboration among investigators, at least one of which must be a CaMos Centre Director. A formal proposal must be submitted to the DAP Committee for review. Following approval, the authors of the proposal will be notified of the release of data, and will sign an agreement, stating that they will only use the data for the purpose described, will follow the timeline specified for the analysis, and will destroy the data files by a given date. The CaMos principal investigators are Dr. David Goltzman and Dr. Nancy Krieger. To obtain a copy of the Ancillary Project and Data Release Policy and the requirements for proposal submission, or to obtain further information from the CaMos principal investigators about data access, please send an e-mail to info@camos.org.

<>Access type</>

<>Data use agreement required</>

<>Application workflow</>

<>Restrictions to use</>

<>Responsible party</>

Better reporting: Work with authors & publishers

Make data sharing reporting the norm for authors (pre-publication)

Work with journals to improve reporting requirements (e.g., ICMJE)

Develop a list of required documentation that are necessary to accompany shared research data in publications

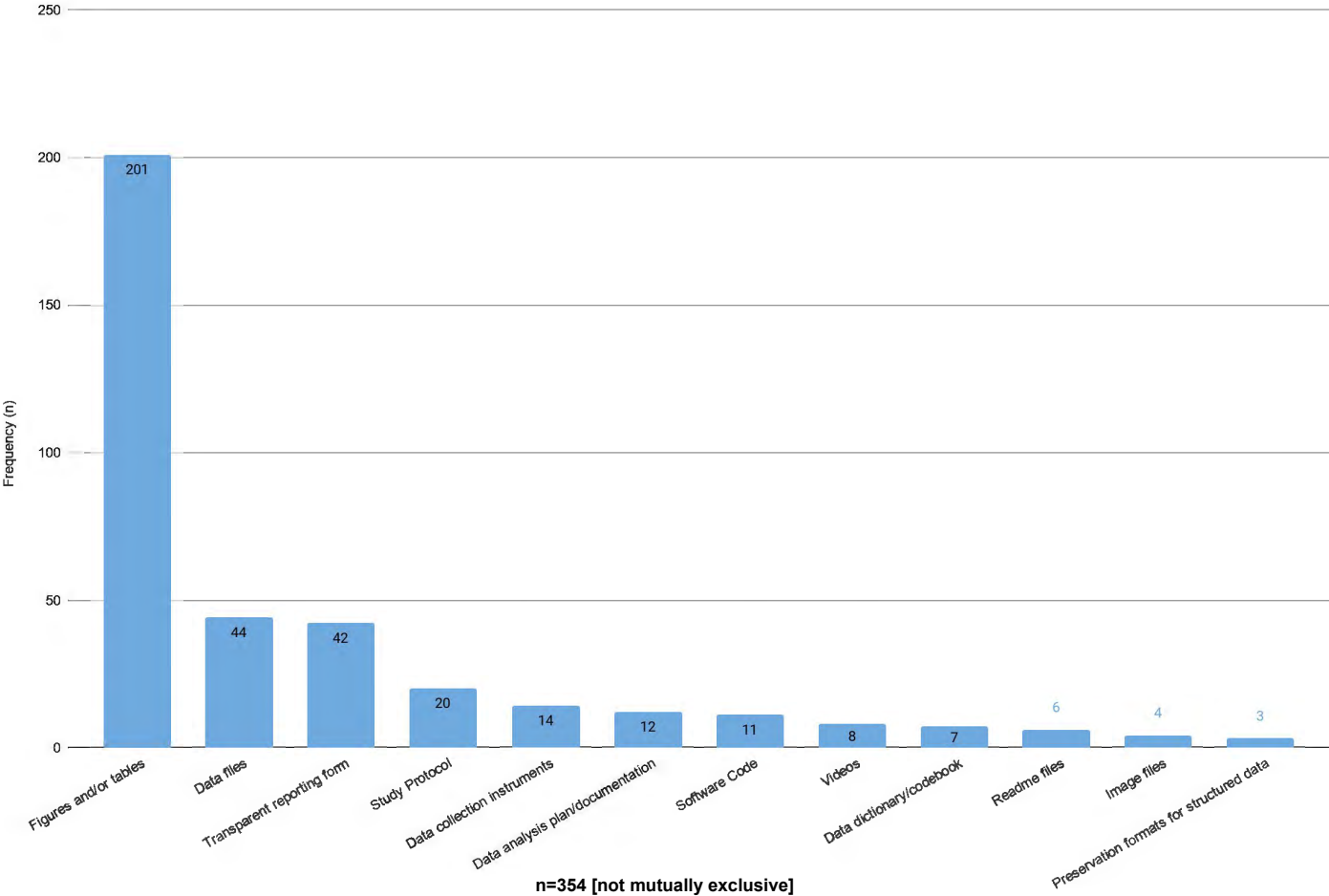
Better reporting: Work with authors & publishers

Make data sharing reporting the norm for authors (pre-publication)

Work with journals to improve reporting requirements (e.g., ICMJE)

Develop a list of required documentation that are necessary to accompany shared research data in publications

Documentation (or lack thereof)



Unusable data...



The screenshot shows the homepage of The New England Journal of Medicine. At the top left is the journal's logo, a red circular seal with the text 'MEDICINE NEW ENGLAND JOURNAL OF' and the years '1827' and '1828'. To the right of the logo is the text 'The NEW ENGLAND JOURNAL of MEDICINE'. Below this is a navigation bar with links for 'HOME', 'ARTICLES & MULTIMEDIA', 'ISSUES', 'SPECIALTIES & TOPICS', 'FOR AUTHORS', and 'CME'. A horizontal banner of various medical images is positioned below the navigation bar. The main content area features the word 'Perspective' in a large font, followed by the article title 'Avoiding Data Dumpsters — Toward Equitable and Useful Data Sharing'. Below the title, the authors 'Laura Merson, Oumar Gaye, M.D., Ph.D., and Philippe J. Guerin, M.D., Ph.D.' are listed, along with the journal information 'N Engl J Med 2016; 374:2414-2415 | June 23, 2016 | DOI: 10.1056/NEJMp1605148'. A comment icon and the text 'Comments open through June 29, 2016' are also present. Below this is a navigation bar for the article with tabs for 'Article', 'Citing Articles (26)', 'Comments (3)', and 'Metrics'. The main text of the article begins with 'The potential health benefits from sharing participant-level clinical research data for the purpose of secondary analysis or meta-analysis have been widely touted. Although some researchers remain wary about sharing data, recent policies and proposals by funders, scientific journals, research institutions, and international health organizations mean that data sharing, in one form or another, is inevitable. Now is therefore the time to focus on developing practices for data sharing that are effective, efficient, equitable, and ethical. In the process, we may need to question the assumption that more is better. Simply making more data openly available may not lead to analyses that are relevant and that are actually applied to improve health.'

<https://www.nejm.org/doi/full/10.1056/NEJMp1605148>

Reproducibility and the need for documentation

nature International weekly journal of science

Home News & Comment Research Careers & Jobs Current Issue Archive Audio & Video For Authors

Archive Volume 533 Issue 7604 Editorial Article

NATURE | EDITORIAL

Reality check on reproducibility

A survey of *Nature* readers revealed a high level of concern about the problem of irreproducible results. Researchers, funders and Journals need to work together to make research more reliable.

25 May 2016

PDF Rights & Permissions

Is there a reproducibility crisis in science? Yes, according to the readers of *Nature*. Two-thirds of researchers who responded to a survey by this journal said that current levels of reproducibility are a major problem.

The ability to reproduce experiments is at the heart of science, yet failure to do so is a routine part of research. Some amount of irreproducibility is inevitable: profound insights can start as fragile signals, and sources of variability are infinite. But, the survey suggests, there is a bigger issue — and something that needs to be fixed. One-third of the survey respondents said that they think about the reproducibility of their own research daily, and more than two-thirds discuss it with colleagues at least monthly. The survey, of course, probably attracted researchers most interested in these issues. But it would be foolish to pretend that there is not serious concern.

Related stories

- The pressure to publish pushes down quality
- Research data: Silver lining to irreproducibility
- Statisticians issue warning over misuse of P values

[More related stories »](#)

<https://www.nature.com/news/reality-check-on-reproducibility-1.19961>

DATA SCIENCE JOURNAL

Reading: Data Without Software Are Just Numbers Share: f t g+ in

Start Submit

Essays

Data Without Software Are Just Numbers

Authors: James Harold Davenport, James Grant, Catherine Mary Jones

Abstract

Great strides have been made to encourage researchers to archive data created by research and provide the necessary systems to support their storage. Additionally it is recognised that data are meaningless unless their provenance is preserved, through appropriate meta-data. Alongside this is a pressing need to ensure the quality and archiving of the software that generates data, through simulation, control of experiment or data-collection and that which analyses, modifies and draws value from raw data. In order to meet the aims of reproducibility we argue that data management alone is insufficient: it must be accompanied by good software practices, the training to facilitate it and the support of stakeholders, including appropriate recognition for software as a research output.

<https://datascience.codata.org/articles/10.5334/dsj-2020-003/>

nature International weekly journal of science

Home News & Comment Research Careers & Jobs Current Issue Archive Audio & Video For Authors

Archive Volume 533 Issue 7604 News Feature Article

NATURE | NEWS FEATURE

1,500 scientists lift the lid on reproducibility

Survey sheds light on the 'crisis' rocking research.

Monya Baker

25 May 2016 | Corrected: 28 July 2016

PDF Rights & Permissions

-2:03

<https://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970>

How to improve reporting practices?

Incorporate data sharing reporting guidelines into DMP Templates



The screenshot shows the Portage DMP Assistant website. At the top left is the 'portage' logo with the tagline 'Shared stewardship of research data'. To the right is a 'Version française' link and a navigation menu with 'Home', 'About', 'Help', and 'Terms of Use' buttons. The main content area on the left describes the tool and lists four steps: 1. Sign up with DMP Assistant, 2. Sign in and select a template under Organizations, 3. Answer the questions that are relevant to your work, and 4. Revisit the tool throughout your research. On the right, there are two prominent buttons: 'Sign in' (with a plus sign) for existing users and 'Sign up' (with a plus sign) for new users. A disclaimer at the bottom right notes that single sign-in authentication is currently being worked on.

portage
Shared stewardship of research data

[Version française](#) | [Home](#) [About](#) [Help](#) [Terms of Use](#)

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.

Summary

- It is essential that we continue to learn about our community's data sharing practices
 - Would similar studies of NSERC and SSHRC data sharing practices be beneficial to inform future discovery efforts?

Summary

- It is essential that we continue to learn about our community's data sharing practices
 - Would similar studies of NSERC and SSHRC data sharing practices be beneficial to inform future discovery efforts?
- Our work suggests that research data is shared via multiple methods and platforms, with varying levels of access permitted

Summary

- It is essential that we continue to learn about our community's data sharing practices
 - Would similar studies of NSERC and SSHRC data sharing practices be beneficial to inform future discovery efforts?
- Our work suggests that research data is shared via multiple methods and platforms, with varying levels of access permitted
- Indexing research data regardless of where it is stored will improve the discovery of Canadian research data

Summary

- It is essential that we continue to learn about our community's data sharing practices
 - Would similar studies of NSERC and SSHRC data sharing practices be beneficial to inform future discovery efforts?
- Our work suggests that research data is shared via multiple methods and platforms, with varying levels of access permitted
- Indexing research data regardless of where it is stored will improve the discovery of Canadian research data
- Important to establish protocols for reporting data sharing methods and including necessary documentation in the sharing workflow

The research project

Study documentation and raw data available:

- Read KB, Ganshorn H, Rutley S, Scott DR. Surveying the landscape of CIHR-funded research data sharing practices: An analysis of the published literature. 2020. <https://osf.io/n9jv5>

Research team:

- Kevin Read, MLIS, MAS, University of Saskatchewan
- Heather Ganshorn, MLIS, University of Calgary
- David Scott, MLIS, MA, University of Lethbridge
- Sarah Rutley, MLIS, MA, University of Saskatchewan

How do we improve the discovery of data that is available but not immediately accessible?

What methods do you think could improve data sharing reporting practices?

Questions?

kevin.read@usask.ca

References

Baker M. 1,500 scientists lift the lid on reproducibility. 2016;533(7604):452-454.

Davenport JH, Grant J, Jones CM. Data Without Software Are Just Numbers. Data Science Journal. 2020 Jan 22;19(1).

Editorial. Reality check on reproducibility. Nature. 2016;533(7604):437.

Noun Project Images:

- <https://thenounproject.com/search/?q=researcher&i=3011359>
- <https://thenounproject.com/search/?q=metadata&i=2563280>
- <https://thenounproject.com/search/?q=library&i=1244000>
- <https://thenounproject.com/search/?q=metadata&i=1677708>
- <https://thenounproject.com/search/?q=data%20repository&i=1630978>
- <https://thenounproject.com/search/?q=lock&i=3362860>
- <https://thenounproject.com/search/?q=ecosystem&i=1637681>

Polanin JR, Terzian M. A data-sharing agreement helps to increase researchers' willingness to share primary data: results from a randomized controlled trial. Journal of clinical epidemiology. 2019 Feb 1;106:60-9.

Vines TH, Albert AY, Andrew RL, Débarre F, Bock DG, Franklin MT, Gilbert KJ, Moore JS, Renaut S, Rennison DJ. The availability of research data declines rapidly with article age. Current biology. 2014 Jan 6;24(1):94-7.