

Data-sharing practices in publications funded by the Canadian Institutes of Health Research: a descriptive analysis

Kevin B. Read MLIS, Heather Ganshorn MLIS, Sarah Rutley MLIS, David R. Scott MLIS

Abstract

Background: As Canada increases requirements for research data management and sharing, there is value in identifying how research data are shared and what has been done to make them findable and reusable. This study aimed to understand Canada's data-sharing landscape by reviewing how data funded by the Canadian Institutes of Health Research (CIHR) are shared and comparing researchers' data-sharing practices to best practices for research data management and sharing.

Methods: We performed a descriptive analysis of CIHR-funded publications from PubMed and PubMed Central published between 1946 and Dec. 31, 2019, that indicated that the research data underlying the results of the publication were shared. We analyzed each publication to identify how and where data were shared, who shared data and what documentation was included to support data reuse.

Results: Of 4144 CIHR-funded publications identified, 1876 (45.2%) included accessible data, 935 (22.6%) stated that data were available via request or application, and 300 (7.2%) stated that data sharing was not applicable or possible; we found no evidence of data sharing in 1558 publications (37.6%). Frequent data-sharing methods included via a repository (1549 [37.4%]), within supplementary files (1048 [25.3%]) and via request or application (935 [22.6%]). Overall, 554 publications (13.4%) included documentation that would facilitate data reuse.

Interpretation: Publications funded by the CIHR largely lack the metadata, access instructions and documentation to facilitate data discovery and reuse. Without measures to address these concerns and enhanced support for researchers seeking to implement best practices for research data management and sharing, much CIHR-funded research data will remain hidden, inaccessible and unusable.

To improve health outcomes and research reproducibility, health sciences research has become increasingly focused on the production, management and sharing of research data. Increased interest in making this research more reproducible and reusable has set in motion initiatives in the United States,^{1,2} Europe,^{3,4} Canada⁵ and internationally⁶ to improve data discoverability, accessibility and transparency.

The importance of data sharing in this area is well documented. Sharing research data catalyzes new research discoveries;⁷⁻¹² encourages transparency and holds the research community accountable;¹³⁻¹⁵ and improves the interoperability of data across research communities and systems.¹⁶⁻¹⁸

Canada is at a crucial stage of development with respect to data-management and data-sharing initiatives. The Canadian Tri-Agency has released a policy on research data management that requires researchers to manage and deposit data,¹⁹ Canadian publishers are releasing data-sharing policies,²⁰ the Federated Research Data Repository²¹ has made it possible to discover data that are produced and stored in Canadian repositories, and the New Digital Research Infrastructure Organization has been established to respond to emerging data needs within the Canadian digital research landscape.²²

Although these efforts aim to make data sets more discoverable, valuable data shared alongside publications, in external discipline-specific repositories, via websites or on request are difficult to locate, access and reuse.

With the release of the Tri-Agency Research Data Management Policy¹⁹ and the establishment of new initiatives to locate Canadian research products online, we see value in identifying how and where Canadian research data are being shared, and what steps have been taken to make them reusable. In the present study, we aimed to understand the Canadian data-sharing landscape by reviewing how and where data funded by the Canadian Institutes of Health Research (CIHR) are shared, and comparing the data-sharing practices of CIHR-funded researchers to the Tri-Agency principles for research data management and sharing.²³

Competing interests: None declared.

This article has been peer reviewed.

Correspondence to: Kevin Read, kevin.read@usask.ca

CMAJ Open 2021 November 9. DOI:10.9778/cmajo.20200303

Methods

Design and setting

We used descriptive analysis to identify how and where CIHR-funded researchers share their data, using metadata extracted from CIHR-funded publications. The CIHR is Canada's federal funding agency for health research. For the purpose of this study, metadata are defined as the data used to describe specific elements (e.g., authors, title, provenance) of a publication, data set or research product to make it searchable and interpretable.²⁴ This study began in October 2019 and was completed in October 2020.

Data sources

We searched PubMed and PubMed Central databases to identify CIHR-funded publications indicating that they shared research data underlying the published results. PubMed is a freely available bibliographic database that contains more than 28 million citations of biomedical literature from the MEDLINE database, life sciences journals and online books. PubMed Central is a freely available digital repository that archives open-access full-text scholarly publications in biomedical and life sciences journals. Reciprocal links exist between the full text in PubMed Central and the corresponding citations in PubMed.²⁵ We did not search for data associated with the grey literature because our study focused on CIHR-funded scholarly publications.

We chose PubMed and PubMed Central as our data sources because they provide unique data set search filters²⁶ that identify publications indicating that data underlying the results have been shared. We define research data as

data that are used as primary sources to support technical or scientific enquiry, research, scholarship, or artistic activity, and that are used as evidence in the research process and/or are commonly accepted in the research community as necessary to validate research findings and results.²⁷

Search strategy

Our search strategy, which was built by K.B.R. and peer reviewed by H.G. and D.R.S., identified publications between 1946 and Dec. 31, 2019. Using PubMed Central's data filters,²⁶ 1 of the authors (K.B.R.) searched for all CIHR-funded publications that included a statement on data availability. Data-availability statements contain the authors' description of where and how to gain access to the research data underlying the published manuscript. K.B.R. searched for additional publications in PubMed using its data filter,²⁶ which captures when data have been shared in a repository. One author (K.B.R.) then combined these filters with CIHR-related keywords in English and French, using the grants information field from both databases (Appendix 1, available at www.cmajopen.ca/content/9/4/E980/suppl/DC1).

Publications were included if, in both PubMed and PubMed Central, they had an author who was funded by the CIHR, were published between 1946 and Dec. 31, 2019, and were published in English or French (the 2 official languages of CIHR-funded research). Publications from PubMed were

required to link to a data repository, and publications from PubMed Central were required to include a data-availability statement.

Publications were excluded if, in both PubMed and PubMed Central, they did not include a CIHR-funded author and if, in PubMed, they linked exclusively to a clinical trial registry rather than a data repository.

Metadata extraction

One author (K.B.R.) extracted selected metadata fields from publications that met our inclusion criteria using the PubMed Central Open Access Subset,²⁸ which allows full-text metadata from a publication to be extracted under a Creative Commons licence. When full-text metadata were not available via the Open Access Subset, they were extracted with the use of the minimal level of metadata available in PubMed Central. Publications that were not available in PubMed Central had a limited set of metadata extracted from PubMed (Appendix 2, available at www.cmajopen.ca/content/9/4/E980/suppl/DC1). The Python scripts used to extract the metadata from both databases are available via the Open Science Framework.²⁹

Data abstraction

Examination of data-sharing practices

Using the extracted metadata, we examined each publication to explore data accessibility; how, where and by whom research data were shared; and the inclusion of documentation to support data reuse, including but not limited to codebooks, data-analysis plans, software code and readme files. Each author was assigned a unique set of publications to examine independently. Any uncertainties that arose during the data-collection process were resolved at biweekly team meetings. Our data-collection instrument and descriptive analyses were generated and captured in a REDCap database. The instrument and data dictionary used for our final analysis are available via the Open Science Framework.²⁹

Before starting the complete examination of all publications, each author examined a random preliminary set of the same 150 publications to ensure consistency in how information was interpreted and captured in our data-collection instrument.

Data-sharing status

To frame our analysis, we grouped data-sharing practices into categories representing the most commonly identified data-sharing status types (Table 1); these categories are not mutually exclusive. We examined the frequency of each category across our entire sample and over time.

Data-sharing methods

Using the metadata available (Appendix 2) and building on our data-sharing status categories, we recorded the methods of data sharing evident within each publication. These included but were not limited to sharing data via a repository, within the supplementary files, via request or application, within the publication, via a website or when an author stated that data sharing was not applicable or possible.

Table 1: Data-sharing status categories and their definitions

Category	Definition
1. Data accessible	Research data files could be identified, accessed and downloaded
2. Data available	Authors stated within the manuscript, the data-availability statement or the acknowledgements that research data were available on request or via an application process
3. Data sharing not applicable/possible	Authors stated within the manuscript, the data-availability statement or the acknowledgements that research data sharing was not applicable or possible
4. No evidence of data sharing	Authors made no mention of data sharing within the manuscript, the data-availability statement or the acknowledgements; authors indicated data sharing would be available at a future date; or the publication contained no research data files

If an author’s data-sharing statement indicated that an application was required to access the data, we captured all reasons why authors insisted on this requirement. Similarly, if an author stated that data could not be shared at all, we captured all reasons provided why this was the case.

We examined whether data-sharing statements made by authors aligned with how data were shared in practice. When authors stated that all the research data needed to understand the results were within the publication, we reviewed the publication for evidence that no additional research data files were needed to understand the findings. When authors stated that research data were available in the supplementary files, we attempted to locate and access the data within the supplementary files section. We documented instances of misalignment between author statements and whether and how data were shared, as well as when we were unclear about whether author statements reflected data sharing accurately.

Finally, we categorized institutions and journals where data were shared and ranked them according to their data-sharing status.

Research data documentation

We captured the types of documentation that were included alongside accessible and available research data (Table 1, categories 1 and 2). We identified types of documentation based on the Tri-Agency Statement of Principles on Digital Data Management,²³ which makes recommendations on adherence to standards, data collection and storage, and metadata documentation. We then examined each publication to determine whether documentation such as study protocols, data-analysis plans, software code, data dictionaries, readme files, data-collection instruments, videos and data-management plans was provided. Documentation of this kind has been identified as necessary for improving the transparency, reproducibility and reusability of research results.^{30–33} We also analyzed the frequency of documentation inclusion over time.

Statistical analysis

We performed a descriptive analysis of our results. All data collected during this study were exported from the REDCap database and analyzed with Stata/SE 16.0 software (Stata-Corp). The raw data extracted from PubMed and PubMed Central, the synthesized data exported from REDCap and the analyzed data from Stata along with a summary analysis report are available via the Open Science Framework.²⁹

Ethics approval

Because our study focused on a descriptive analysis of metadata that are publicly available, ethics approval was not required.

Results

Our search identified 4988 publications. After we applied our inclusion criteria, we retrieved metadata for 4144 publications for analysis (Figure 1). Of these publications, 1876 (45.3%) made their data accessible, 935 (22.6%) made their data available (via request or application), 300 (7.2%) indicated that data sharing was not applicable or possible, and 1558 (37.6%) provided no evidence of data sharing even though we had limited our sample to publications that had indicated data sharing of some kind. Many publications shared multiple data sets in different ways. Figure 2 shows the extent of overlap among the categories of data-sharing status, and Figure 3 shows the frequency of these categories over time.

Data-sharing methods

The most frequent methods of data sharing were via a repository (1549 publications [37.4%]) and within the supplementary files (1048 [25.3%]) (Table 2). Notably, 935 publications (22.6%) stated that data were available via request (701 [16.9%]) or application (234 [5.6%]) but provided little detail about how to acquire the data. A total of 538 publications (13.0%) had no evidence of or information about data sharing whatsoever. Some publications shared data in multiple formats and therefore may be represented in more than 1 category.

Among the 1549 publications that reported data sharing via a repository, there were 97 repositories represented (the complete listing is available in the Open Science Framework analysis report²⁹). The most prevalent repositories were the Protein Data Bank (599 [38.7%]), Gene Expression Omnibus (377 [24.3%]) and GenBank (194 [12.5%]). A breakdown of repositories is shown in Appendix 3, Supplementary Figure S1 (available at www.cmajopen.ca/content/9/4/E980/suppl/DC1).

A total of 234 publications indicated that an application was required to access the data underlying the results. The most frequent justification for this requirement was the need to complete a data-access, data-transfer or data-use agreement (78 [33.3%]), followed by general ethics concerns (66 [28.2%]), confidentiality (60 [25.6%]), licence restrictions (28 [12.0%]) and Indigenous considerations (6 [2.6%]). Twenty-three publications (9.8%) did not explain why an application was required. None of the publications that required an application included metadata sufficiently outlining the requirements for access and approval.

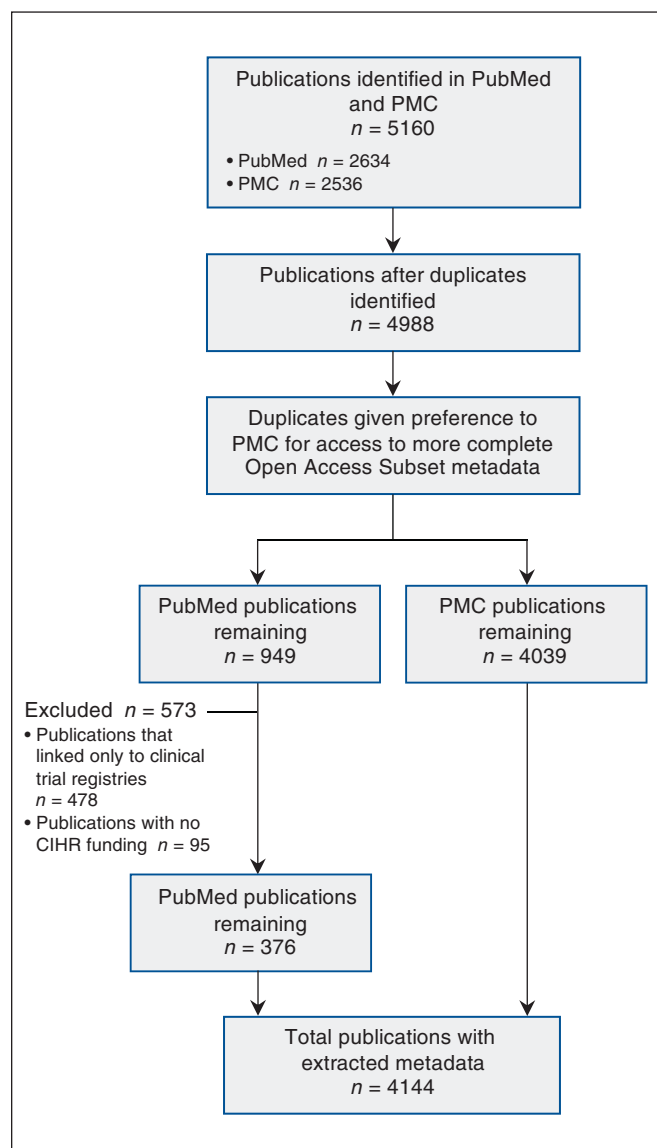


Figure 1: Flow diagram showing metadata extraction. Note: CIHR = Canadian Institutes of Health Research, PMC = PubMed Central.

Among the 300 publications that indicated that data sharing was not applicable or possible, the most common reason cited was confidentiality (109 [36.3%]); 88 publications (29.3%) provided no justification at all (Table 3).

Of the 1048 publications stating that data were available in the supplementary files, 752 (71.8%) did not share data in this way. Similarly, 345 (39.7%) of the 870 publications stating that all data were available within the article shared no research data underlying the results within the publication or supplementary files, although there was clear evidence of data collection.

A breakdown of institutions associated with the publications that shared data is presented in Appendix 3, Supplementary Figure S2. Among institutions associated with more than 10 publications, those with the greatest proportion of publications in which data were accessible or available were

the Structural Genomics Consortium (20/21 [95%]) and the University of Waterloo (12/25 [48%]), respectively.

The journals used most frequently were *PLoS One* (736 [17.8%]) and the *Journal of Biological Chemistry* (208 [5.0%]) (Appendix 3, Supplementary Figure S3A). Among journals with more than 25 CIHR-funded publications, the 3 most commonly used journals that included examples of accessible data were the *Journal of Molecular Biology* (57/59 [96.6%]), *Proceedings of the National Academy of Sciences of the United States of America* (110/116 [94.8%]) and *Nature* (51/54 [94.4%]). Among all journals, the 3 most commonly used that included examples of available data were the *International Journal of Behavioural Nutrition and Physical Activity* (23/26 [88.5%]), *BMC Psychiatry* (13/17 [76.5%]) and *BMC Medical Research Methodology* (17/23 [73.9%]) (Appendix 3, Supplementary Figure S3B and available on the Open Science Framework²⁹).

Research data documentation

The documentation provided alongside publications was varied, with supplementary figures or tables or both, study protocols, research data files and transparent reporting forms most frequently represented (Table 4).

Referring to the recommended documentation types outlined in the Tri-Agency Principles on Digital Data Management,²³ we examined how frequently these were included alongside publications that made data accessible or indicated that data were available over time (Appendix 3, Supplementary Figure S4). The types of documentation required to understand and reuse research data were provided in a minority of publications that shared data (554/4144 [13.4%]). Across all publications regardless of whether they shared data, study protocols were most frequently included (576/4144 [13.9%]), and data-management plans were the least frequently included (4/4144 [0.1%]). Although documentation supporting reuse was scarce, as of 2017, there was increasing availability of data-analysis plans, code and data-collection instruments.

Interpretation

This study highlights substantial room for growth in improving the discoverability, accessibility and usability of CIHR-funded research data. Although, encouragingly, repositories were the most common venues authors chose to share data, the remaining data were made available within the publication or its supplementary files, by request or by a long tail of other means.²⁹ When authors indicated that data were available via request or application (22.6% of publications), they did not provide adequate instructions on how to acquire them. Despite our focus on publications indicating that data had been shared, more than one-third (37.6%) showed no evidence of sharing. The most frequent types of documentation shared alongside data did not generally support their interpretation and reuse. These characteristics conflict with expectations outlined in Canada's Tri-Agency data management²³ and international FAIR (Findable, Accessible, Interoperable, Reusable) guiding principles.¹⁶

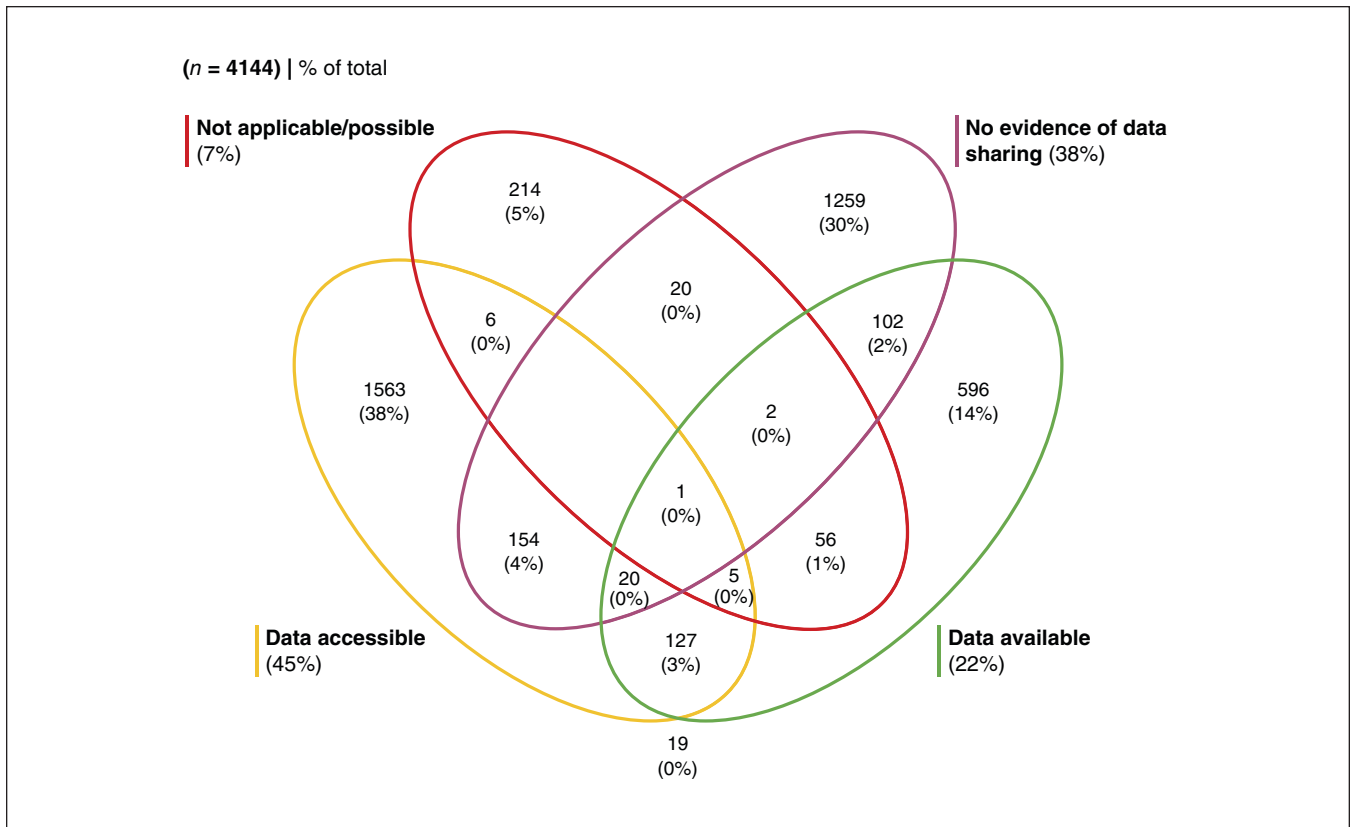


Figure 2: Frequency of publications by data-sharing status.

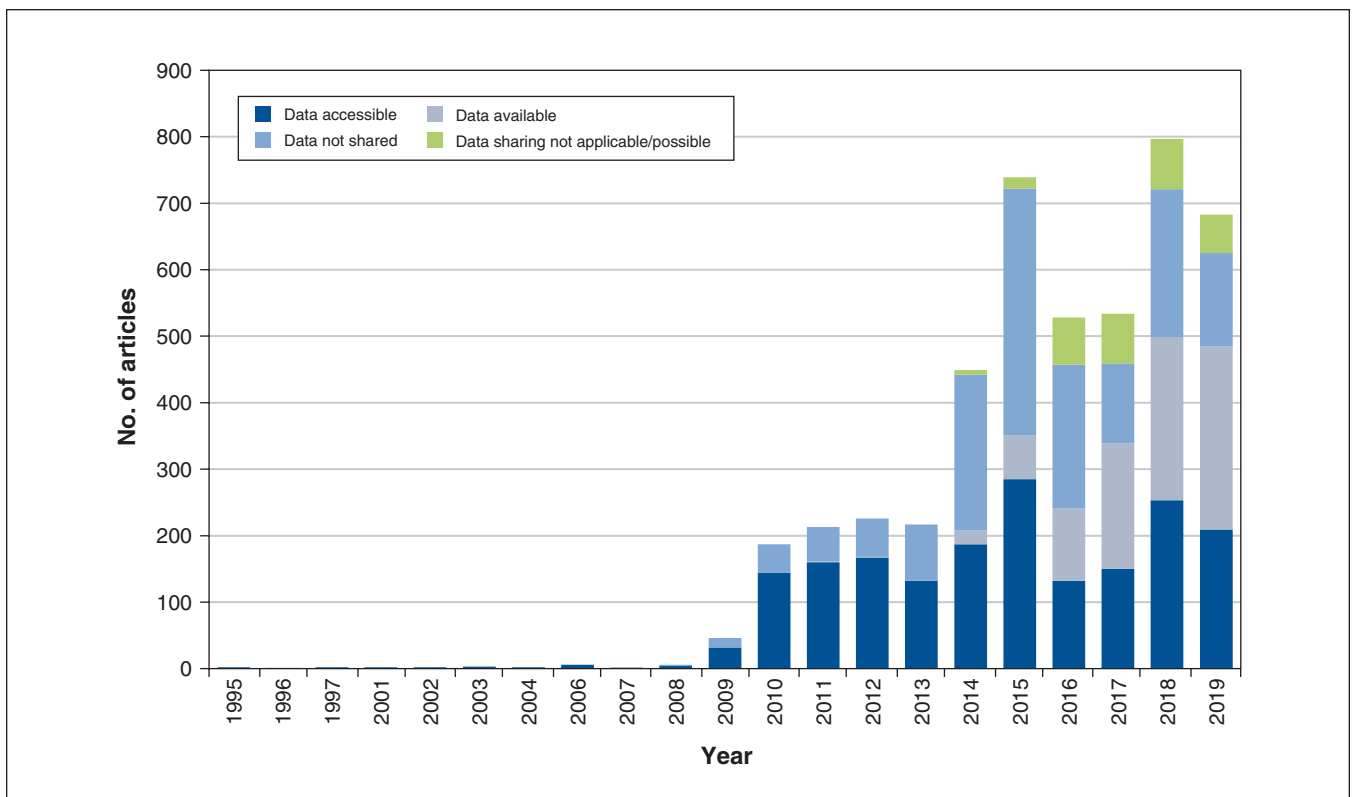


Figure 3: Frequency of data-sharing status categories over time. Categories are not mutually exclusive.

Table 2: Frequency of data-sharing methods

Method*	No. (%) of articles n = 4144
Available via repository	1549 (37.4)
Available in supplementary files	1048 (25.3)
All data available within contents of article	870 (21.0)
Available on reasonable request	701 (16.9)
Data sharing not applicable/possible	300 (7.2)
Available on request via application	234 (5.6)
Available via website	142 (3.4)
Data will be made available at future date	19 (0.5)
Other†	58 (1.4)

*Not mutually exclusive.
†Available via the Open Science Framework.²⁹

Table 3: Reasons for not sharing data

Reason*	No. (%) of articles n = 300
Confidentiality	109 (36.3)
No additional information provided	88 (29.3)
No consent obtained	45 (15.0)
Ethical reasons other than consent	32 (10.7)
No data collected within article	26 (8.7)
Proprietary restrictions	10 (3.3)
Indigenous considerations	4 (1.3)
Other†	20 (6.7)

*Not mutually exclusive.
†Available via the Open Science Framework.²⁹

Table 4: Documentation identified, by data-sharing status category

Type of information	Category				Total
	Data accessible	Data available	No evidence of data sharing	Data sharing not applicable/possible	
Supplementary figure/table	1258	385	881	65	2589
Study protocol	332	94	200	18	644
Data file	504	82	3	10	599
Transparent reporting form	138	57	107	4	306
Data-collection instrument	31	62	52	39	184
Video	56	20	41	1	118
Data analysis plan/documentation	58	25	12	4	99
Image file	23	5	43	0	71
Software code	46	13	8	2	69
Data dictionary/codebook	26	24	10	6	66
Preservation format for structured data	50	4	2	0	56
Readme file	15	5	1	1	22
Data-management plan	2	2	0	0	4
Other*	29	30	28	6	93

*Available via the Open Science Framework.²⁹

Deficiencies in data discoverability, access and usability have been examined in other contexts. A 2015 study of data sharing in publications funded by the US National Institutes of Health showed that 88% of research data were not discoverable.³⁴ Our finding that there is often a gap between data-sharing statements and practices is in keeping with recent studies.³⁵⁻³⁸ We identified several cases in which authors indicated that data were shared but we found no evidence thereof. We speculate that those authors may have

incorrectly considered summary tables and figures to be research data. This finding highlights that many authors may fundamentally misunderstand what it means to make research data discoverable and accessible. Our analysis of reusability practices showed that the most frequent types of documentation shared alongside data rarely support their interpretation and reuse. Sharing descriptive documentation such as codebooks and data dictionaries, and actionable supporting files such as code and software, is increasingly

recognized as best practice,^{30,33,39} and our results indicate that CIHR-funded data sharing can vastly improve in this area.

Inadequate metadata are a recognized problem in the data-sharing landscape,⁴⁰⁻⁴² and the examples of CIHR-funded sharing we encountered are no different. Without adequate metadata to support discovery, data will remain hidden.⁴³ The absence of metadata elaborating on application requirements specifically calls into question the true availability of these data and impedes future research based on them. Other investigators have highlighted the challenges of requesting access to data in relation to the lack of transparency of request processes and a lack of standardization in use agreements for health data.^{41,44,45} Given that data made available by request are often collected from human participants, improving the discoverability of and access to sensitive data can prevent study replication, create opportunities for pooling related data and increase research efficiency.^{41,42,44,45} Our findings indicate that current metadata practices do not provide sufficient information to make successful data requests and secure these outcomes.

Future initiatives should focus on the development of metadata standards that facilitate the discovery of sensitive data and support transparent data-request processes. In particular, we suggest that the Tri-Agency's requirements for data-management plans¹⁹ be extended to include reporting guidelines for making sensitive research data discoverable. These guidelines should require robust descriptions of sensitive data and detailed data-access procedures where applicable, which could be submitted alongside a manuscript for publication. We also suggest that Canadian data repositories explore how to better accommodate sensitive data so they can be made discoverable while honouring access and privacy restrictions. We see value in reestablishing infrastructure that tracks CIHR-funded publications and data sharing, a function that was partially fulfilled by PubMed Central Canada before it was taken offline, in 2018.⁴⁶

Limitations

Our analysis focused solely on CIHR-funded publications indicating that data were shared. We did not deem it logistically possible to identify instances of data sharing from the full body of CIHR-funded literature. We also examined CIHR-funded publications exclusively from PubMed and PubMed Central. Although these are the most comprehensive biomedical databases available, we acknowledge there are other databases where CIHR-funded publications exist. We limited our study to these sources because metadata specific to data sharing were not readily accessible in other biomedical databases, whereas PubMed and PubMed Central provide access to open-source metadata. To manage study feasibility, we limited our review of documentation to that which was shared or stated within the publication and did not extend this analysis to repositories or websites where some research data were shared.

Conclusion

Publications funded by the CIHR largely lack the metadata, access instructions and documentation to facilitate data discovery and reuse. Without measures to address these concerns

and enhanced support for researchers seeking to implement best practices for research data management and sharing, much CIHR-funded research data will remain hidden, inaccessible and unusable.

References

- Collins F. Statement on final NIH policy for data management and sharing. Bethesda (MD): National Institutes of Health; 2020.
- Collins FS, Tabak LA. Policy: NIH plans to enhance reproducibility. *Nature* 2014;505:612-3.
- Durinx C, McEntyre J, Appel R, et al. Identifying ELIXIR core data resources. *F1000Res* 2016;5:ELIXIR-2422.
- Artini M, Atzori C, Bardi A, et al. The OpenAIRE Literature Broker Service for Institutional Repositories. *Dlib Mag* 2015;21:95-104.
- OpenAIRE teams up with Canadian funders to identify research outputs! OpenAIRE; 2019. Available: <https://www.openaire.eu/openaire-joins-forces-with-canada-s-federal-granting-agencies-and-carl> (accessed 2020 Feb. 19).
- Taichman DB, Sahni P, Pinborg A, et al. Data sharing statements for clinical trials: a requirement of the International Committee of Medical Journal Editors. *PLoS Med* 2017;14:e1002315.
- Humphreys GS, Tinto H, Barnes KI. Strength in numbers: the WWARN case study of purpose-driven data sharing. *Am J Trop Med Hyg* 2019;100:13-5.
- McAlister VC, Harvey EJ. The benefits and risks of requiring researchers to share data. *Can J Surg* 2016;59:364-5.
- Owens B. Data sharing: access all areas. *Nature* 2016;533:S71-2.
- Mendelson DS, Bak PRG, Menschik E, et al. Informatics in radiology: image exchange: IHE and the evolution of image sharing. *Radiographics* 2008;28:1817-33.
- Steele Gray C, Barnsley J, Gagnon D, et al. Using information communication technology in models of integrated community-based primary health care: learning from the iCOACH case studies. *Implement Sci* 2018;13:87.
- Vestrup JA, Phang PT, Vertesi L, et al. The utility of a multicenter regional trauma registry. *J Trauma* 1994;37:375-8.
- Dumontier M, Wesley K. Advancing discovery science with FAIR data stewardship: findable, accessible, interoperable, reusable. *Ser Libr* 2018;74:39-48.
- Reiser L, Harper L, Freeling M, et al. FAIR: a call to make published data more findable, accessible, interoperable, and reusable. *Mol Plant* 2018;11:1105-8.
- Open Science Collaboration. Estimating the reproducibility of psychological science. *Science* 2015;349:aac4716.
- Wilkinson MD, Dumontier M, Aalbersberg IJJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 2016;3:160018.
- Persaud N. A national electronic health record for primary care. *CMAJ* 2019;191:E28-9.
- Rathi VK, Strait KM, Gross CP, et al. Predictors of clinical trial data sharing: exploratory analysis of a cross-sectional survey. *Trials* 2014;15:384.
- Tri-Agency Research Data Management Policy. Ottawa: Innovation, Science and Economic Development Canada; modified 2021 Mar. 15. Available: http://science.gc.ca/eic/site/063.nsf/eng/h_97610.html (accessed 2021 Apr. 15).
- Kelsall D. New CMAJ policy on sharing study data. *CMAJ* 2017;189:E1082.
- Wilson L. Exploring the Canadian Federated Research Data Repository service. *Biodiversity Inf Sci Stand* 2017;1:e20185.
- Digital research infrastructure. Ottawa: Innovation, Science, and Economic Development Canada; modified 2019 Aug. 7. Available: <https://www.ic.gc.ca/eic/site/136.nsf/eng/home> (accessed 2020 Mar. 9).
- Tri-Agency Statement of Principles on Digital Data Management. Ottawa: Innovation, Science and Economic Development Canada; modified 2021 Jan. 21. Available: https://science.gc.ca/eic/site/063.nsf/eng/h_83F7624E.html (accessed 2020 Mar. 9).
- Metadata. CASRAI. Available: <https://casrai.org/term/metadata/> (accessed 2020 Nov. 22).
- MEDLINE, PubMed, and PMC (PubMed Central): How are they different? Bethesda (MD): National Library of Medicine; reviewed 2020 Sept. 11. Available: <https://www.nlm.nih.gov/bsd/difference.html> (accessed 2020 June 21).
- Data filters in PMC and PubMed. NLM Technical Bulletin. Bethesda (MD): National Library of Medicine; 2018. Available: https://www.nlm.nih.gov/pubs/techbull/ma18/brief/ma18_pmc_data_filters.html (accessed 2019 Sept. 5).
- Research data. CASRAI. Available: <https://casrai.org/term/research-data/> (accessed 2020 Nov. 22).
- Open Access Subset. Bethesda (MD): National Center for Biotechnology Information, U.S. National Library of Medicine; updated 2019 Mar. 19. Available: <https://www.ncbi.nlm.nih.gov/pmc/tools/openfstl/> (accessed 2019 Oct. 30).
- Read K, Ganshorn H, Rutley S, et al. Surveying the landscape of CIHR-funded research data sharing practices: an analysis of the published literature. Charlottesville (VA): Center for Open Science; updated 2020 Nov. 27. Available: <https://osf.io/n9jv5> (accessed 2020 Feb. 19).
- Bakken S. The journey to transparency, reproducibility, and replicability. *J Am Med Inform Assoc* 2019;26:185-7.
- Holub P, Kohlmayer F, Prasser F, et al. Enhancing reuse of data and biological material in medical research: from FAIR to FAIR-Health. *Biopreserv Biobank* 2018;16:97-105.

32. Miyakawa T. No raw data, no science: another possible source of the reproducibility crisis. *Mol Brain* 2020;13:24.
33. Walters WP. Code sharing in the open science era. *J Chem Inf Model* 2020;60:4417-20.
34. Read KB, Sheehan JR, Huerta MF, et al. Sizing the problem of improving discovery and access to NIH-funded data: a preliminary study. *PLoS One* 2015;10:e0132735.
35. Christian TM, Gooch A, Vision T, et al. Journal data policies: exploring how the understanding of editors and authors corresponds to the policies themselves. *PLoS One* 2020;15:e0230281.
36. Danchev V, Min Y, Borghi J, et al. Evaluation of data sharing after implementation of the International Committee of Medical Journal Editors data sharing statement requirement. *JAMA Netw Open* 2021;4:e2033972.
37. Gorman DM. Availability of research data in high-impact addiction journals with data sharing policies. *Sci Eng Ethics* 2020;26:1625-32.
38. Siebert M, Gaba JF, Caquelin L, et al. Data-sharing recommendations in biomedical journals and randomised controlled trials: an audit of journals following the ICMJE recommendations. *BMJ Open* 2020;10:e038887.
39. Davenport JH, Grant J, Jones CM. Data without software are just numbers. *Data Sci J* 2020;19:3.
40. Kush RD, Warzel D, Kush MA, et al. FAIR data sharing: the roles of common data elements and harmonization. *J Biomed Inform* 2020;107:103421.
41. Learned K, Durbin A, Currie R, et al. Barriers to accessing public cancer genomic data. *Sci Data* 2019;6:98.
42. Vassar M, Jellison S, Wendelbo H, et al. Data sharing practices in randomized trials of addiction interventions. *Addict Behav* 2020;102:106193.
43. Schriml LM, Chuvochina M, Davies N, et al. COVID-19 pandemic reveals the peril of ignoring metadata standards. *Sci Data* 2020;7:188.
44. Miller J, Ross JS, Wilenzick M, et al. Sharing of clinical trial data and results reporting practices among large pharmaceutical companies: cross sectional descriptive study and pilot of a tool to improve company practices. *BMJ* 2019;366:14217.
45. Shabani M, Obasa M. Transparency and objectivity in governance of clinical trials data sharing: current practices and approaches. *Clin Trials* 2019;16:547-51.
46. PubMed Central Canada taken offline in February 2018. Ottawa: Canadian Institutes of Health Research; modified 2019 Aug. 20. Available: <https://cihr-irsc.gc.ca/e/50728.html> (accessed 2021 Apr. 20).

Affiliations: Leslie & Irene Dubé Health Sciences Library (Read), University of Saskatchewan, Saskatoon, Sask.; Taylor Family Digital Library (Ganshorn), University of Calgary, Calgary, Alta.; University Library (Rutley), University of Saskatchewan, Saskatoon, Sask.; University Library (Scott), University of Lethbridge, Lethbridge, Alta.

Contributors: Kevin Read supervised the project and conceived the study. All of the authors designed the study, obtained, analyzed and interpreted the data, drafted the manuscript and revised it critically for important intellectual content, approved the final version to be published and agreed to be accountable for all aspects of the work.

Funding: This project was supported in part by the University of Saskatchewan Faculty Recruitment and Retention Program.

Content licence: This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY-NC-ND 4.0) licence, which permits use, distribution and reproduction in any medium, provided that the original publication is properly cited, the use is noncommercial (i.e., research or educational use), and no modifications or adaptations are made. See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

Data sharing: All raw, processed and analyzed data, as well as accompanying documentation, reports and scripts are available on the Open Science Framework, at <https://osf.io/n9jv5/>.

Acknowledgement: The authors thank the Canadian Hub for Applied and Social Research for its support in procuring and analyzing the research data in this study.

Supplemental information: For reviewer comments and the original submission of this manuscript, please see www.cmajopen.ca/content/9/4/E980/suppl/DC1.