

SAGE is a repository that houses quality data about child and youth development, health, and well-being.
We provide services that promote good data management, increase research productivity, and expedite knowledge translation.



10 - 30%
increase in citation
rate¹



\$240 billion
invested annually in health
research but most of the data
produced is difficult to access^{2,3}



2016
the year the Tri-Council began to
formally support data deposit
as an integral part of research⁴

(1) Piwowar & Vision, 2013 (2) Rottingen et al., 2013 (3) Chan et al., 2014 (4) www.science.gc.ca

Why researchers should share data

Sharing data through SAGE increases research productivity. We actively market datasets and facilitate collaboration with interested researchers.

Data access is vetted by a peer-review process, which may include the original researcher, to ensure scientific integrity and appropriate re-use of data.

Approved users sign a legal agreement which protects the original researchers and participants. A disclosure risk assessment and data access through the SAGE Analytic Environment safeguards participant privacy and confidentiality.

We ensure the contributions of the original researchers are recognized and appropriately acknowledged.

Why researchers should access data

By eliminating the need to collect data, SAGE is a cost- and resource-effective way to enhance your research productivity.

We house rich birth cohort, longitudinal, time-series, and cross-sectional data. Dataset themes include nutrition, developmental disorders, and early childhood development.

The SAGE Analytic Environment allows secure, remote access to data and a variety of analytical software.

We facilitate collaboration and mentorship with original researchers.

Support for data analytics and knowledge mobilization is provided by leveraging PolicyWise' expertise.



\$116,000

average budget of a CIHR
funded project⁵



38%

decrease in CIHR grants
provided to early career
investigators⁶



\$1.2 billion

lost largely due to unforeseen
difficulties in recruitment for
the National Children's Study⁷

(5) www.cihr-irsc.gc.ca (6) http://www.webcitation.org/6h6AM2s21 (7) www.sciencemag.org



Secondary Analysis to Generate Evidence (SAGE)

Research Data Scope Statement

SAGE aims to serve the Alberta research community by providing an environment to share and analyze data for secondary analysis. In particular, SAGE is focused on high quality data with potential for re-use. SAGE seeks to share datasets that:

- Involve primary data collection on human subjects
- Includes content on social, health, human ecology, environment, community, learning, and other related fields. Of special interest are multi-disciplinary studies (i.e. health and environment; community and education)
- Are of sufficient quality for re-use (SAGE can assist in determining quality)

With the above focus, certain datasets are considered out of scope for the SAGE:

- Data collected from non-human subjects
- Data with legal impediments to sharing or deposition (i.e. administrative data, commercially-owned data, etc.)
- Data from literature reviews
- Studies with mainly genomic data
- Clinical drug/intervention trials with limited scope beyond verification of effectiveness

However, these are meant to be guidelines and it is encouraged that all interested parties contact SAGE to discuss the suitability of our services in helping you increase the utility and impact of your data.





SAGE Research Survey: Results Report

July 2016

Amanda Lau; Lucie Richards



Policy Wise
for Children & Families

1. Introduction

The Secondary Analysis to Generate Evidence (SAGE) Researcher Survey 2016 was sent out to the research community in April 2016 to assess, from the perspective of a secondary data user (hereafter referred to as an “accessor”) as well as of a primary data producer (hereafter referred to as a “depositor”), what factors lead to successful sharing and re-use of data. These factors are all, to some extent, features and services that SAGE aims to build, and we hoped that feedback on the relative importance of each feature or service would enable us to prioritize our service development.

Approximately 950 members of the research community were sent an invitation to complete the SAGE Researcher Survey. We received 82 responses to the survey by late May 2016, 8 of which were partial responses. In all cases the partial responses were nearly complete, and so were kept in this descriptive analysis. In three cases, researchers responded more than once, but with different (complete) responses.

Please note these results should be interpreted with caution due to the small sample size and self-selection of respondents.

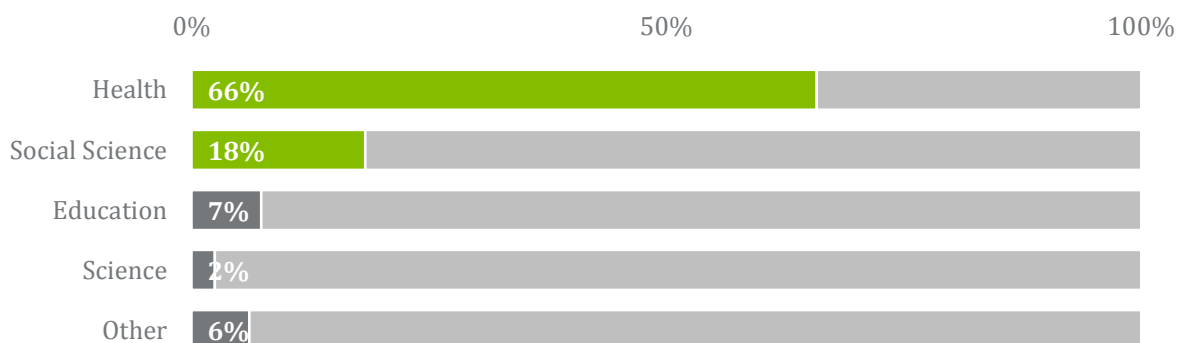
2. Results

2.1. Field of Research

The responding researchers work, for the most part, in health (i.e. mental health, nursing, health promotion, and clinical research) and social science (i.e. psychology, sociology, and kinesiology) fields. To a lesser degree, researchers that work in education or science fields have research that intersected with either health or social sciences.

What is your current field of research?

Percentages of researchers, categorized by field

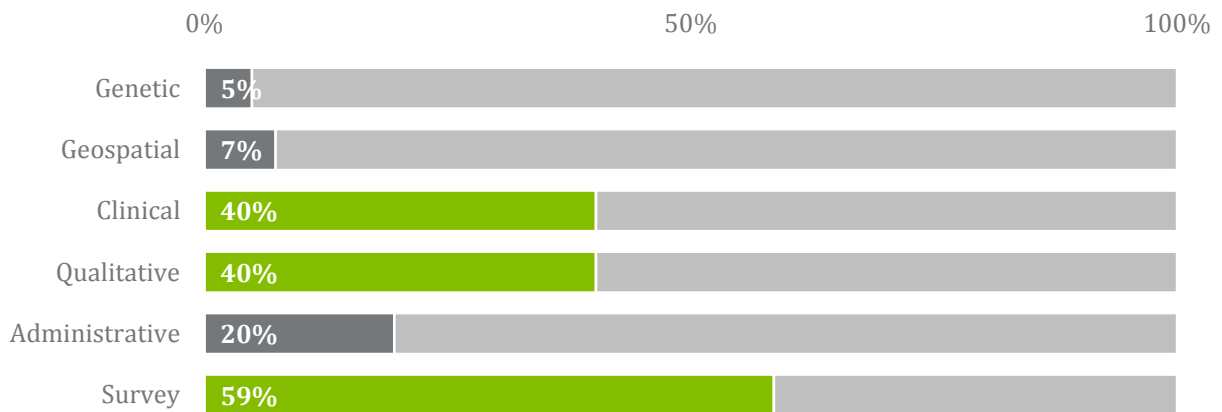


2.2. Data currently being used

Responding researchers use a variety of types of data, although survey, clinical and qualitative (eg. Interviews, focus groups) data make up the majority of listed data types. A moderately high number of respondents also use administrative data.

What types of data do you currently use?

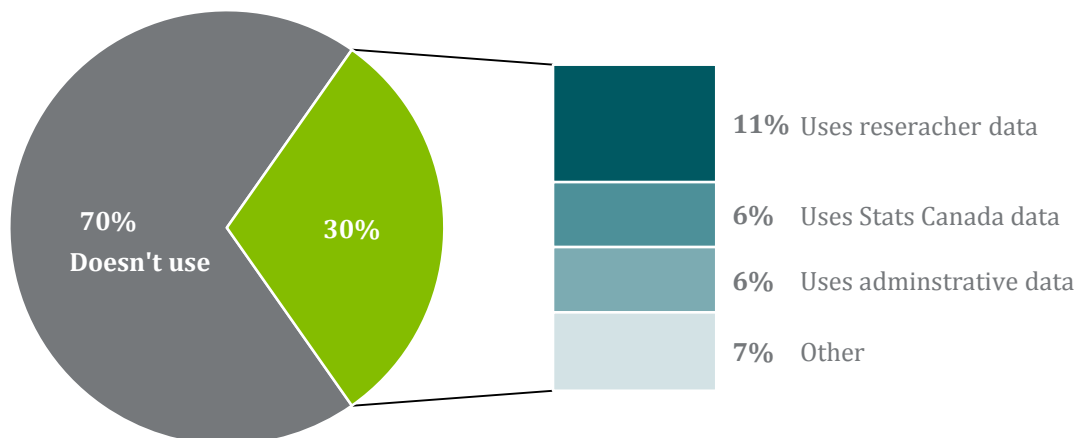
Selective list of research data currently used



2.3. Secondary Data Use

Only 30% of responding researchers indicate currently using secondary data for their research. Of these, approximately one third of researchers work with other researchers to access their data.

Do you currently use secondary data?



For those that did use secondary data, the majority were quantitative researchers followed by mixed methods (i.e. quantitative and qualitative) researchers. None of the qualitative researcher

surveyed used secondary data for their research. However, these results should be interpreted with caution due to the small sample size.

Secondary Use by Research Method

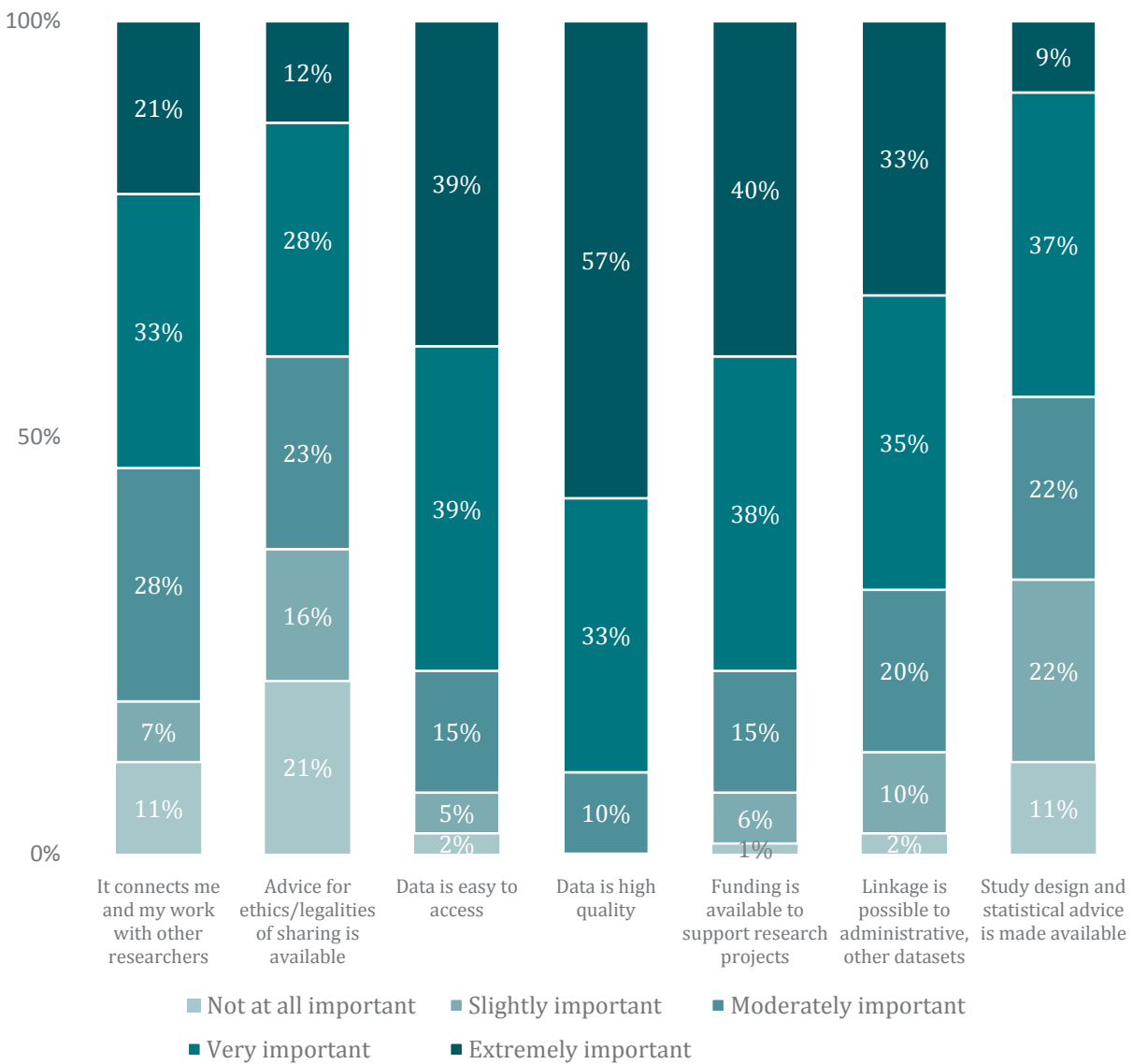
Percentages of researcher, categorized by method



Encouragement for Secondary Data Use

By contrast, respondents were more clearly in agreement on factors that would encourage them to use secondary data at SAGE. By far the most important of these is the presence of high quality secondary data assets. Closely following data quality was ease of access, funding availability and the ability to link to administrative and other datasets. Researchers on several occasions took the opportunity to comment in the “other section” about the importance of quality (or relevance) of data and ease of access as particularly important. In contrast to data management services, individual respondents tended to rate factors much less consistently.

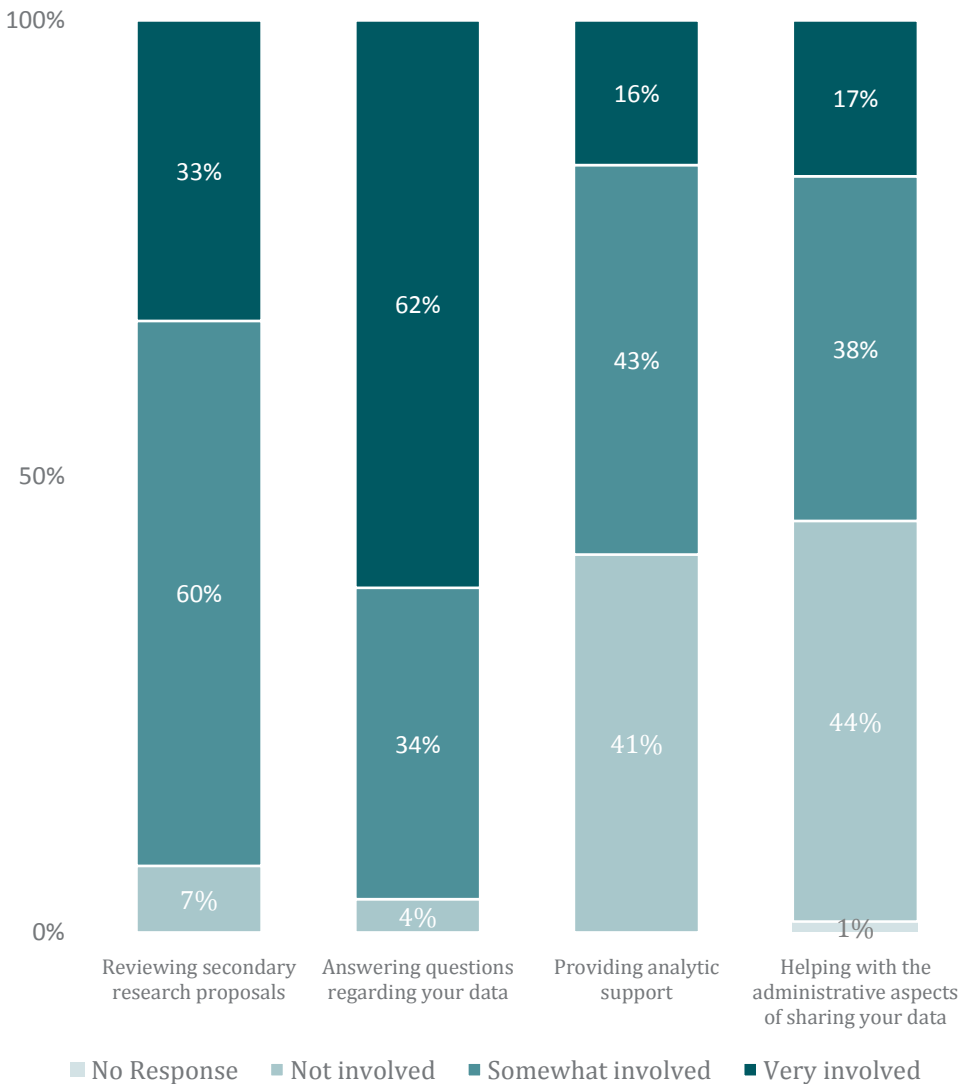
What would encourage you to conduct secondary analysis on data assets at SAGE?



2.5. Depositor's involvement in Secondary Data Use

Respondents asked to imagine themselves as depositors were much more interested in being involved with answering questions about their data and reviewing secondary research proposals than in providing analytic support or dealing with the administrative aspects of sharing data. Nevertheless, there were a fairly high number of researchers who wants at least some level of involvement in all aspects of sharing.

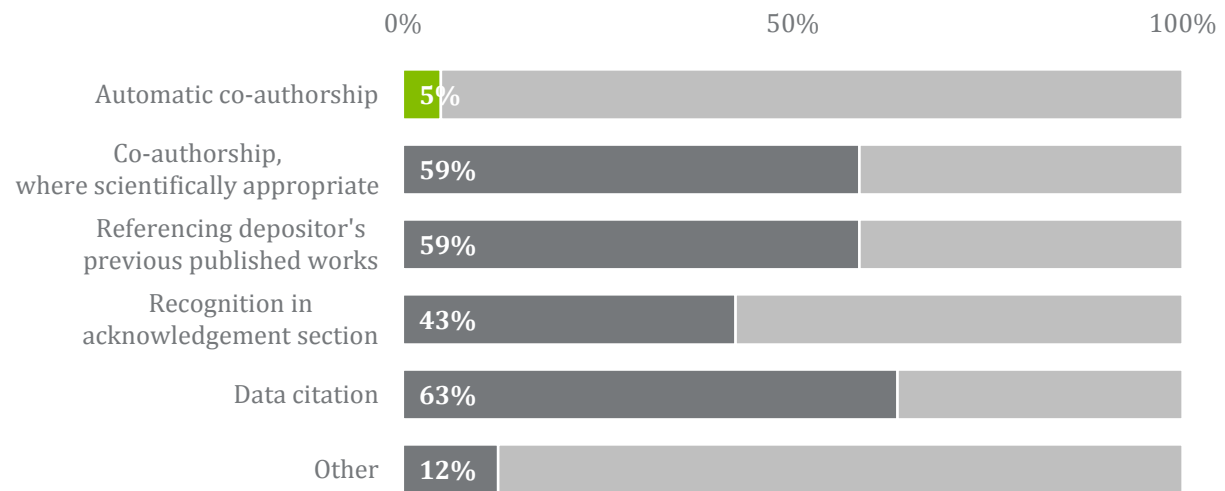
If you were to deposit data with the CDCA, how involved would you want to be in...



2.6. Depositor Recognition

Responding researchers agree with most forms of recognition for depositors, save automatic co-authorship on secondary research. There is also somewhat less agreement that data producers should be formally acknowledged in the publication.

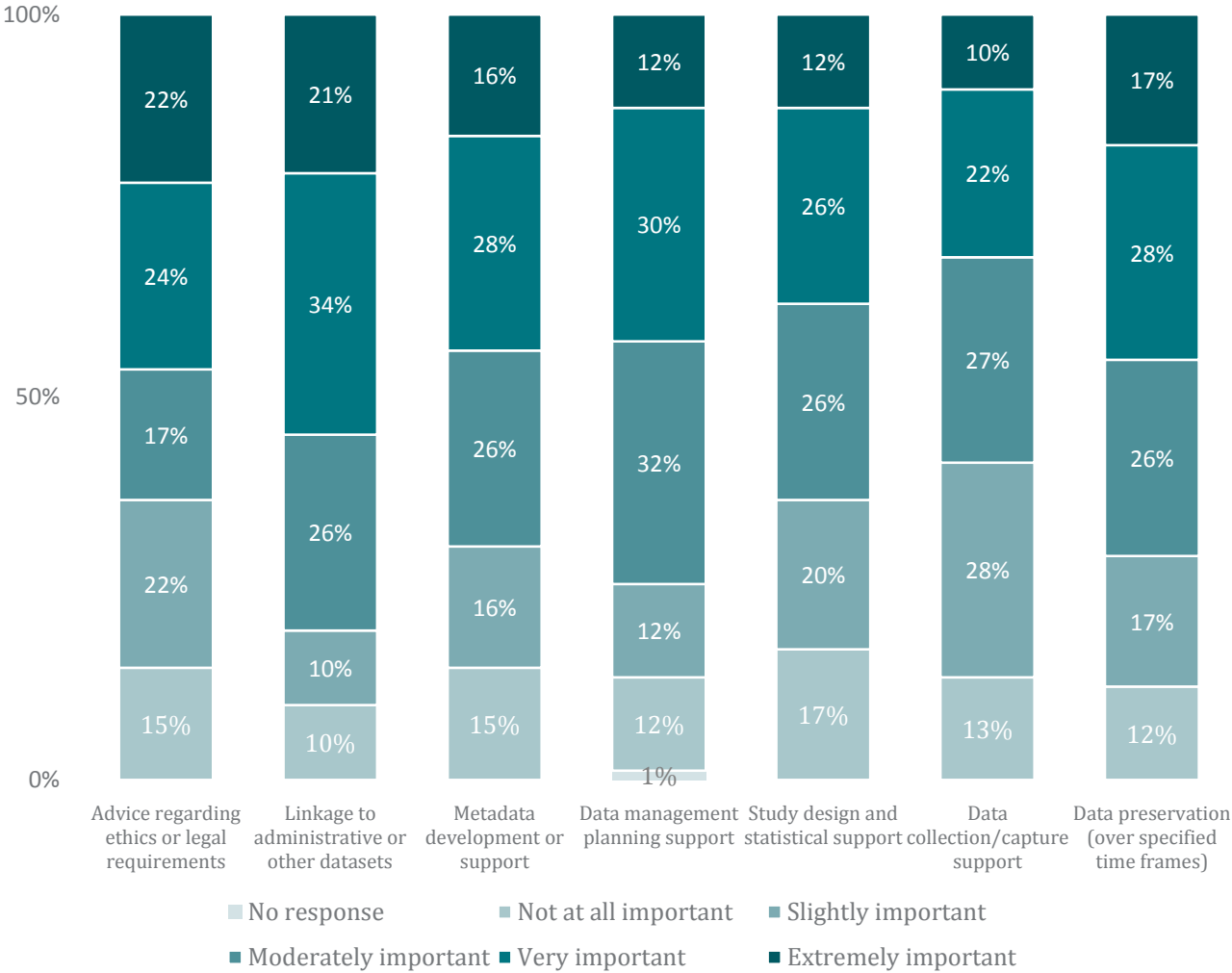
How should the depositor be recognized within the secondary research publication?



2.7. Data Management Needs

For the most part, respondents find all aspects of data management services at least moderately important. The ability to link to administrative and other datasets was somewhat more often rated as very or extremely important, while data collection support was somewhat more often rated as not at all or slightly important. It is worth noting, however, that individual respondents tended to have overarching attitudes towards data management services, either highly valuing such services or not at all.

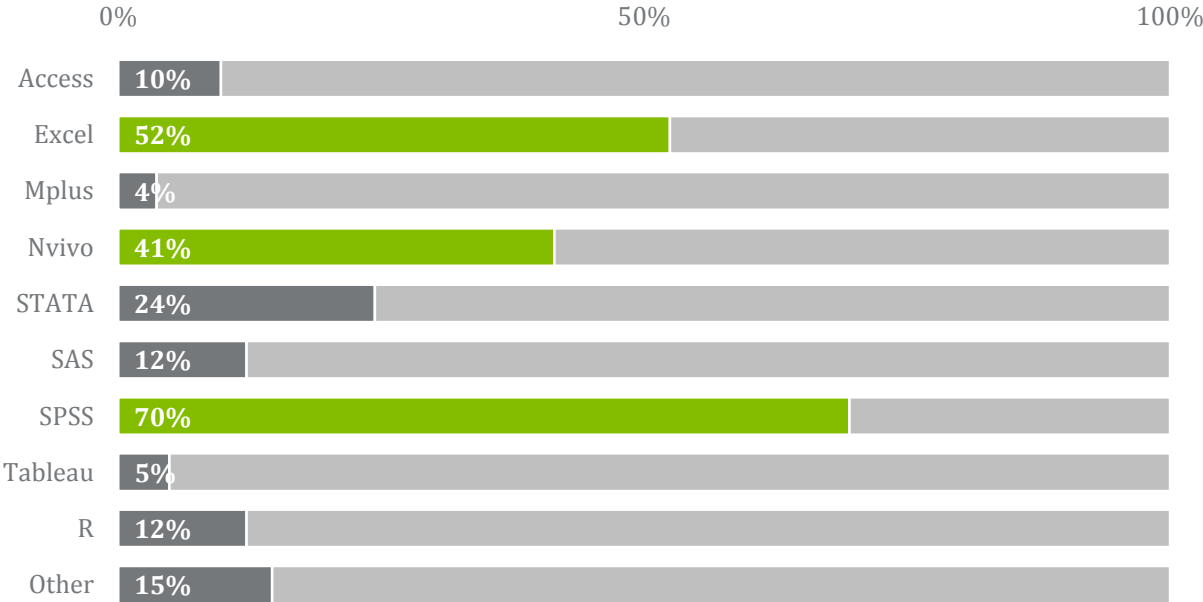
What data management services would meet an unmet need for your research?



2.8. Data-related software

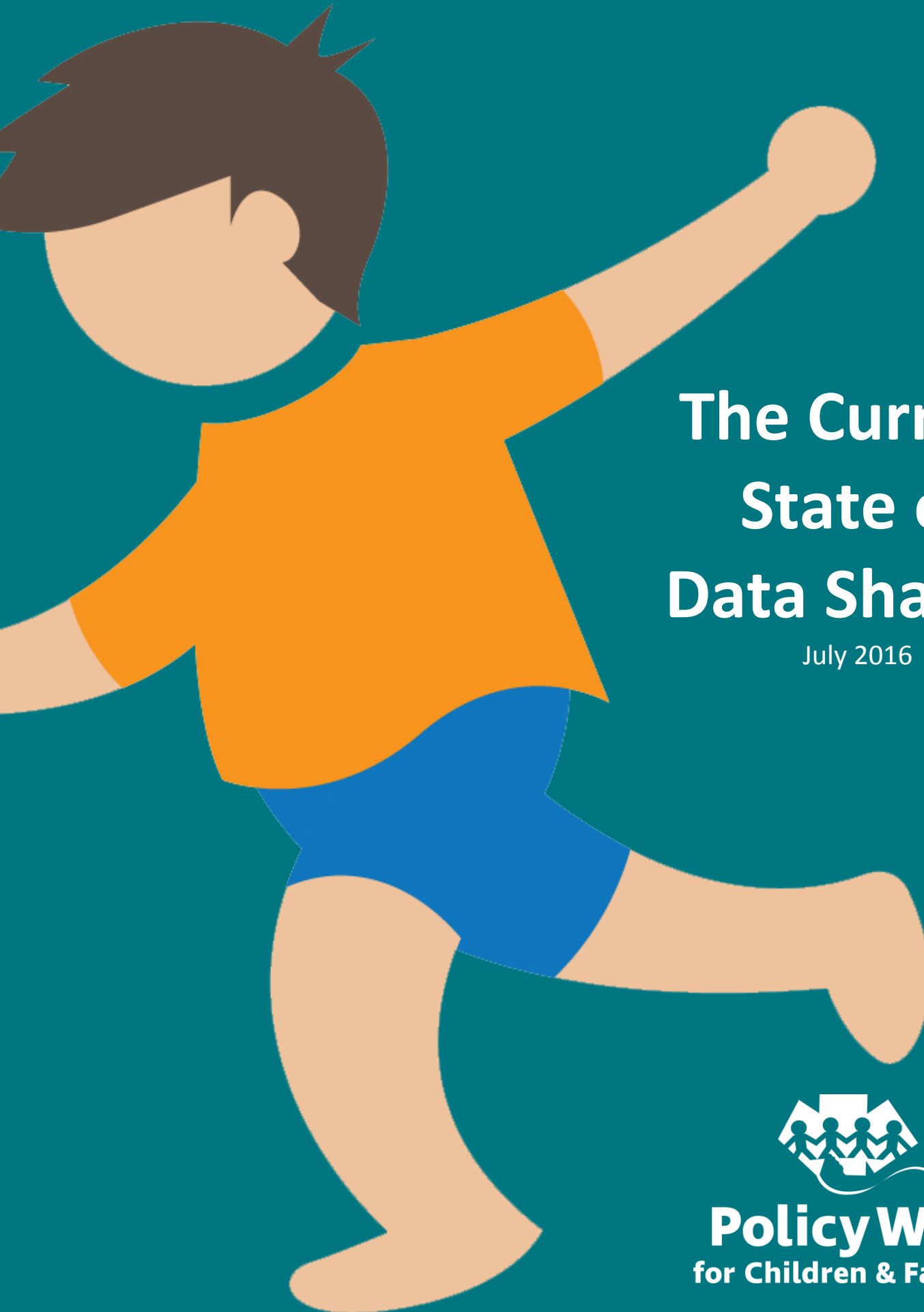
Responding researchers also overwhelmingly use SPSS and NVivo for data manipulation, with Excel also ranked very highly. The design of the study left it unclear whether Excel is used primarily as a data input, visualization or analysis tool for these researchers, but it was listed alone in only 4 cases, which might indicate it is an all-purpose tool rather than used for data analysis.

Which data-related software do you use?



2.9. General Comments

Finally, of note: there were a few comments by researchers regarding the special nature of qualitative data, in particular data collected and shaped both by participants and researchers (community based participatory research, research with aboriginal communities). Because of the unclear or joint ownership of such data, sharing in these cases would require SAGE to develop a relationship not only with the researcher but the community as well, as well as special processes for access. In addition, certain types of “deep” qualitative data are not appropriate for secondary use, as by design such use would take it out of context.



The Current State of Data Sharing

July 2016



PolicyWise
for Children & Families

Table of Contents

Introduction	3
Benefits and Challenges	3
Policy Environment	4
Data Sharing Policy Considerations	5
Overview of Canadian Research Funders Policies	7
Additional Enablers of Data Sharing	8
Conclusion.....	9

Introduction

The global movement toward data sharing is gaining momentum. An increasing number of governments and research communities are developing initiatives and policies to promote data sharing and greater access to data, recognizing their enormous potential for scientific, social, and economic growth. For government, data sharing is the core of their Open Government and Open Data initiatives. Open data is “data that can be freely used, re-used, and redistributed by anyone- subject only, at most, to the requirement that preserve provenance and openness”(1). For the research community, while open data is desirable, various types of data require special considerations (i.e. protection of participant privacy and confidentiality) and cannot be re-used or redistribute as openly. As a result, research data sharing occurs at varying levels of openness. While advancing technologies have made data sharing possible, a number of factors must be considered prior to developing and adopting data sharing policies within the research community. There are multiple stakeholders that can play a role in policy development, each with their own set of responsibilities. This paper focuses on the considerations required at the level of the research funder.

Benefits and Challenges

Data sharing has noteworthy benefits. It allows for scientific replication and verification of research results, enabling greater transparency. In addition, by avoiding research duplication, data sharing accelerates scientific discovery and innovation. New or alternative hypotheses and methods of analysis can be tested, extending data usage beyond the original research objectives and potentially leading to new insights and discoveries (2). Novice researcher training is improved because of greater opportunities to work with complex data and to collaborate with experienced investigators from the original research team. In instances where data linkage is possible, multiple datasets can be combined to answer more complex questions, and potentially engage more interdisciplinary expertise(3). Engagement of and expertise from additional stakeholders, such as policymakers, may expedite knowledge translation of research results beyond the original research audience. Data sharing exponentially increases the long term return on investment from publically funded research.

Despite the benefits and efforts to encourage data sharing, its implementation faces a number of challenges. Researchers, specifically, encounter several obstacles which prevent them from engaging in data sharing. Data sharing requires significant investment of time and resources to ensure well documented, reusable data by others. Detailed information about the data (i.e. metadata) and processes (i.e. data collection, cleaning, etc.) are necessary for secondary users to judge the adequacy and suitability of the data for their research purposes. The original researcher may need to allocate funding to satisfy these requirements, which places increased pressure on limited funds without much added, direct personal benefit. Furthering exacerbating this issue, uncertainty exists around the potential lack of recognition for this investment. There is also uncertainty around the operationalization and governance of data sharing. Researchers may worry that the lack of governance and control leaves them vulnerable to have their ideas “scooped” by other researchers. Finally,

historical research practices have dis-incentivized data sharing. Academic recognition and advancement criteria are largely contingent on publishing manuscripts. Researchers who “hoard” their data become more likely to be rewarded than those that share their data.

There is a lack of clarity around necessary legal and ethics processes plague those researchers that share data. Responsible data sharing requires that there are no violations of privacy, confidentiality, or intellectual property rights. These considerations become complicated due to the different types of research data, and their respective governing jurisdictions (3). In cases where participant consent does not explicitly permit sharing of data outside the research team, research ethics boards may also require participants re-consent prior to sharing. These additional considerations may dissuade researchers from data sharing.

Policy Environment

Data sharing initiatives and policies have been adopted in various countries since early 2000. The G8 Ministers’ Statement of 2013 highlights an international effort “to promote policies that increase the access to the results of publicly funded research” (5) and build upon existing policy work. National research bodies (i.e. the Tri-Council in Canada and the Research Councils of UK (RCUK)) often support data sharing policies, while research funding agencies have taken on the responsibility of the developing and implementing these policies (4).

Internationally, the United Kingdom has become a leader in data sharing policy development. In 2011, RCUK issued the “Common Principles on Data Policy” (6) which set the expectation for management and sharing of research data. The principles have been adopted by each of the seven councils of the RCUK as well as Wellcome Trust, a large charitable foundation that funds research. For the Wellcome Trust, data management and sharing plans are considered an integral part of funding decisions (7). In addition to fostering best data sharing practices, the Wellcome Trust has committed to supporting grants holders on an ongoing basis to maximize the long-term value of research data.

The United States has developed comprehensive data sharing policies. In 2013, the White House’s Office of Science and Technology Policy (OSTP) issued a memorandum to its federal scientific funding agencies, which have over \$100 million in research and development expenditures, to develop plans that support increased public access to the research publications and data (8). This memorandum highlights the early efforts of the National Institutes of Health (NIH) and the National Science Foundation (NSF), who had already established mandatory data sharing policies for their grant recipients prior to the OSTP directive. For grants seeking \$500,000 or more, the NIH requires “a plan for sharing final research data for research purposes” or an explanation as to why data sharing is not possible (9). The NSF, on the other hand, requires data management plans from all grant recipients outlining how investigators will share primary data, samples, physical collections and other support materials within a reasonable time (10).

In 2016, the Tri-Agencies in Canada, which include the Canadian Institute of Health Research (CIHR), the Natural Sciences and Engineering Research Council (NSERC), and Social Sciences and Humanities Research Council (SSHRC), outlined their harmonized support for research data management and sharing in the Tri-Council Statement on Principles of Digital Data Management. Similar to the principles issued by the Research Council UK, the Tri-Council statement provides a guideline on data management. In regards to data sharing, it states that agency funded research should be “preserved in a publicly assessable, secure and curated repository or other platform for discovery and reuse by others” (11).

Prior to the over-arching Tri-Council statement, each Agency had independently addressed data sharing in accordance to the Tri-Agency Open Access Policy (2012), albeit with varying levels of readiness to implement policies. While NSERC is considering policy options that will align with the other Agencies, they currently do not have a data sharing policy. Instead, NSERC has chosen to raise awareness for, promote, and support open access on a global scale (12). CIHR currently requires the deposit of only bioinformatics, atomic, and molecular coordinator data (i.e. data related to genes, nucleic acid, proteomics, etc.) into appropriate public databases immediately upon publication of research results (13). However, CIHR is exploring the possibility of including the sharing of other research data within their policy. Finally, SSHRC has been an early adopter of data sharing, as demonstrated by their 1990 Research Data Archiving Policy. The policy states that all SSHRC funded research data “must be preserved and made available for use by others within a reasonable period of time” (3). Furthermore, data sharing costs, more specifically those associated with preparing data for deposit, are eligible expenses in SSHRC research grant programs (3).

Despite the Tri-Agencies’ data sharing policies, only a few other Canadian research funders have adopted similar policies. As a result, funders at all levels are in a unique position to be leaders in encouraging and enabling data sharing.

Data Sharing Policy Considerations

Data sharing policies are often included in broader data management policies. While data management policies have been adopted around the world, policies differ across jurisdictions. In addition, the focus of these policies will differ depending on the objectives and guiding principles with which they are based (14). For example, a policy that focuses on data sharing rather than data stewardship may have a greater emphasis on activities that facilitate data access (14).

For funders looking to develop data sharing policies, there are a number of common elements that constitute a policy. These elements are summarized in the following table:

TABLE 1

Typical Data Management Policy

Adapted from (4) and (14)

Policy Elements

Data management plan	Data management plans are formal, overarching guides or protocols outlining how data will be collected, organized, standardized, preserved, and shared across the research lifecycle. Policies should encourage investigators to submit data management plans along with their proposal to demonstrate that considerations were given to that data management and sharing during study inception.
Data quality and standards	<p>International standards exist for many types of data. Adherence to such standards ensure data quality control and comprehensive documentation, both of which enable data accessibility and re-use</p> <p>Use of common and open-source formats (for example, open source data file formats like CSV or metadata standards such as DDI or CDISC) also promote interoperability.</p>
Data sharing	<p>Data should be deposited in relevant subject, institutional, or general repositories. Policies may recommend repositories depending on the type of data. At minimum, if no suitable repository exists, investigators should retain data and follow standards to ensure that quality data are available upon request by other researchers.</p> <p>Depending on the sensitivity of the data, some repositories may impose different access conditions or permit differing levels of data access for reuse.</p> <p>Data sharing agreements may be developed to protect the investigators and study participants.</p> <p>Investigators should make data accessible in a timely manner. However, data may be embargoed for specified, reasonable amounts of time until analyses are completed by the original investigator.</p>
Data retention and preservation	<p>Data should be retained for a minimum time period. Retention times tend to vary across disciplines, research purpose, as well as kind of data collected.</p> <p>Also, where available, data should be deposited in a long-term repository or archive to ensure the preservation of data.</p>
Compliance and monitoring	<p>Policies can be either mandatory or voluntary. While voluntary policies are attractive, they may have lower rates of compliance.</p> <p>The methods with which data management and sharing will be monitored or enforced should also be outlined in the policy.</p>

Other Considerations

Privacy and Confidentiality	For research involving human participants, the privacy and confidentiality of participants should always be protected. Safeguards such as anonymization or de-identification of data should be undertaken to reduce the risk of identity disclosure. The safeguard should be proportionate to the sensitivity of the data and associated risk.
Traditional knowledge	<p>Traditional knowledge includes the beliefs, knowledge, practices, innovations, arts, spirituality, and other forms of culture experience and expression that belong to indigenous communities. (15).</p> <p>The rights of traditional knowledge holders shall not be compromised. Different sharing practices and permissions are required; in particular, appropriate engagement of the people, community, or organization is required through the entire data lifecycle (14).</p>
Intellectual property/data ownership	A delay in data sharing may be permitted in instances where investigators or their institutions are developing or submitting applications.

While adopting comprehensive and clear policies are vital for promoting data sharing, funders must also consider what type of support will be provided to researchers to enable compliance (for instance, permitting/requiring budget items for data management and sharing costs in funding applications).

Overview of Canadian Research Funders Policies

The Digital Curation Centre (DCC) is an internationally recognized centre from the United Kingdom that provides expertise and support for research organizations that want to store, management, and share research data. In 2009, the DCC developed a rubric to review the coverage of United Kingdom research funders' data management policies.

Using the same rubric, the following table evaluates the coverage of the Tri-Agencies data management policies and supports that are offered.

TABLE 2
Policy Elements covered by Canadian Research Funders
Adapted from (16)

● Required ● Suggested ○ Omitted

		NSERC	CIHR *	SSHRC
Policy Elements	Data Management Plan	○	○	○
	Data quality and standards	○	○	○
	Data Sharing	○	●	●
	Data Preservation and Retention	○	●	●
	Compliance and Monitoring	○	○	○
Support	Guidance	○	○	○
	Recommended Repository	○	●	○
	Costs	○	○	●

*Data sharing is only required for certain types of data

The overview suggests there are clear gaps within each policy. It should be noted that these policies will likely require revision in order to align with the recently issued Tri-Council Statement on Principles of Digital Data Management.

Additional Enablers of Data Sharing

Until new policies can be implemented, there are a number of ways that funders can promote data sharing. These include:

1. Providing guidance and suggesting support service to encourage best data management practices and increase data accessibility
2. Work collaboratively with researchers and recognize any challenges presented by data management and sharing for different projects
3. Recognize the costs of data cleaning, documentation and sharing as valid expenses and valuable activities during funding competitions
4. Recognize and encourage the production and sharing of high-impact data by including data sharing as a formal criterion for assessing a researcher’s track record and achievements
5. Designate funds for research using secondary data
6. Support data sharing infrastructures (i.e. data repositories)

These measures will enhance a funders' ability to produce effective policies as well as embed sharing practices into the research community.

Conclusion

Data sharing is increasingly supported by government and national research bodies because of the financial and human benefits of access and reuse (14). In Canada, the Tri-Agencies have developed data sharing policies; however, funders at all levels are in a unique position to be leaders in shift towards data sharing. Understanding and addressing issues within the research community that inhibit data sharing will be important during the development and implementation of data sharing policies. Funder initiated policies will prove to be powerful tools to raise awareness, promote, incentivize and support data sharing practices. Although policy elements may differ, a coordinated approach to data sharing policies and their implementation will facilitate uptake by researchers. Funders willingness to promote, enforce, and support data sharing will impact our abilities to create leading-edge research and optimize use of scarce resources. Data sharing will amplify the new knowledge generated from extensive public investments in current and future research (17).

References

1. Open data. (n.d.). In *Research Data Canada – Glossary*. Retrieved from <http://www.rdc-drc.ca/glossary/>
2. Organisation for Economic Co-operation and Development. (2007). *OECD Principles and Guidelines for Access to Research Data from Public Funding*. Retrieved from <http://www.oecd.org/science/scitech/38500813.pdf>
3. Social Sciences and Humanities Research Council. (1990). *Research Data Archiving Policy*. Retrieved from http://www.sshrc-crsh.gc.ca/about-au_sujet/policies-politiques/statements-enonces/edata-donnees_electroniques-eng.aspx
4. Government of Canada. (2011). *Research Data*. Retrieved from http://www.science.gc.ca/default.asp?lang=en&n=2BBD98C5-1#_ftn7
5. G8 Science Ministers. 2013. *G8 Science Ministers Statement*. Retrieved from <https://www.gov.uk/government/publications/g8-science-ministers-statement-london-12-june-2013>
6. Research Council UK. 2011. *RCUK Common Principles on Data Policy*. Retrieved from <http://www.rcuk.ac.uk/research/datapolicy/>
7. Wellcome Trust. (n.d.). *Policy on Data Management and Sharing*. Retrieved from <https://wellcome.ac.uk/funding/managing-grant/policy-data-management-and-sharing>
8. Executive Office of the President Office of Science and Technology Policy. (2013, February 22). *Memorandum for the Heads of Executive Departments and Agencies: Increasing Access to the Results of Federally Funded Scientific Research*. Retrieved from https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf
9. National Institutes of Health. (2013, March 5). *NIH Data Sharing Policy and Implementation Guidance*. Retrieved from http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm
10. National Science Foundation. (n.d.) *Dissemination and Sharing of Research Results*. Retrieved from <http://www.nsf.gov/bfa/dias/policy/dmp.jsp>
11. Government of Canada. (2016). *Tri-Agency Statement of Principles on Digital Data Management*. Retrieved from <http://www.science.gc.ca/default.asp?lang=En&n=83F7624E-1>
12. Natural Sciences and Engineering Research Council of Canada. (2014). *Policies and Guidelines: Draft Tri-Agency Open Access Policy Frequently Asked Questions*. Retrieved from http://www.nserc-crsng.gc.ca/NSERC-CRSNG/policies-politiques/OpenAccessFAQ-LibreAccesFAQ_eng.asp#a4
13. Canadian Institutes of Health Research. (2013). *CIHR Open Access Policy*. Retrieved from <http://www.cihr-irsc.gc.ca/e/46068.html>
14. Shearer, Kathleen. (2015). *Government of Canada: Comprehensive Brief on Research Data Management Policies*. Retrieved from <http://www.science.gc.ca/default.asp?lang=En&n=1E116DB8-1>
15. Simeone, Tonina. (2004). *Parliament of Canada: Indigenous Traditional Knowledge and Intellectual Property Rights*. Retrieved from <http://www.lopparl.gc.ca/content/lopparl/researchpublications/prb0338-e.htm>
16. Digital Curation Centre. (n.d.) *Overview of funders' data policies*. Retrieved from <http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies>

17. Government of Canada. (2013). *Capitalizing on Big Data: Toward a Policy Framework for Advancing Digital Scholarship in Canada*. Retrieved from http://www.sshrc-crsh.gc.ca/about-au_sujet/publications/digital_scholarship_consultation_e.pdf