NDRIO Call for White Papers on Canada's Future DRI Ecosystem

Submission Response

Strengthening Capacity to Manage, Secure and Protect Sensitive Research Data

12/14/2020

TOPIC: **Research Data Management for Sensitive Data**

CONTRIBUTOR: **Sensitive Data Expert Group, Portage Network** (Lori Walker, Brock University/CAREB Co-Chair; Rachel Zand, University of Toronto/CAREB Co-Chair; Brenda Gagné, Mount Saint Vincent University)

OBJECTIVE: To strengthen Canadian research capacity by identifying key issues related to sensitive data management, and to propose solutions and/or additional support to address these key issues

Background

Sensitive Data refers to data which must be safeguarded against unwarranted access or disclosure.[1] Examples of sensitive data include: personal information, personal health information, responses participants provide to research questions, and institutional records such as educational, criminal, or financial records. Sensitive data are not limited to data about individuals: for example, protected information such as detailed geographic data regarding endangered species are sensitive data; communities, organizations, and other entities may also be the subjects of sensitive data. Sensitive Data also includes Traditional Knowledge (TK), sometimes referred to as "Indigenous Knowledge."

The Open Science movement promotes accessible dissemination of scientific research. However, the benefits of the Open Science movement have not been distributed equally as disciplines which rely on access to sensitive data are limited by the restrictions and complexities involved in the management of sensitive data. The recent pandemic is a timely example of a pressing research problem which can only be addressed through coordinated and appropriate access to sensitive data, such as COVID-19 test results, patient outcomes, reported symptomology, etc.

The Open Science movement, and related initiatives, have provoked a paradigm shift in how sensitive data are conceived. Previously, the typical model for research involving sensitive data would be to publish only the aggregate research results and to destroy the dataset. Contemporarily, there is a push to treat all data as potentially re-usable, and to create policy and data infrastructure to support the re-use of sensitive data. Data publication and re-use allow for research results to be reviewed by peers which strengthens the application of the scientific method. They also allow for researchers to access existing datasets to address research questions, rather than repeatedly studying sample populations which are often already over-studied. Moreover, publication and re-use allow for greater collaboration between researchers and institutions, larger-scale studies which are more

[1] Sensitive Data Toolkit for Researchers Part 1: Glossary of Terms for Sensitive Data used for Research Purposes, https://doi.org/10.5281/zenodo.4088946

geographically distributed, and novel uses of existing datasets. In short, data re-use is the future and its potential benefits are many.

Providing tangible supports to Canadian researchers working with sensitive data will address two aims: 1. improving Canadian capacity to address research questions which rely on sensitive data, such as pandemic response planning, oncology research, community health, etc.; 2. ensuring Canadian research participants' privacy and ethical considerations are respected and supported by robust policies, processes, and best practices. While sensitive data cannot typically be published openly, they can be made discoverable and appropriately accessible. The development of tools and guidance for the management of sensitive data will produce a greater capacity to respect research participant privacy while strengthening and facilitating Canadian research capacity.

The stakeholders involved in developing an ecosystem which supports sensitive data discovery, access, and re-use include research ethics boards, research institutions and their libraries, researchers, privacy regulators and policy-makers, the Tri-Council, data repositories, and NDRIO. A sophisticated and national approach to the management of sensitive data in research will equip Canadian researchers with the tools and supports they need to answer pressing research questions to the betterment of Canada and Canadians.

KEY ISSUES:

Data Custodian Responsibility

In the present regulatory environment, the responsibilities of data owners are not always clear. The custodians or holders of data, in conjunction with the owners of the data, are responsible for determining how data will be stored and who should have access. For example, researchers are often uncertain as to when and how they can share their data with other researchers, and which policies may govern such data sharing. Indigenous and other identifiable communities may own datasets pertaining to their communities, though these datasets may be in the custody of a researcher or institution, and such a custodial arrangement can result in access to the data being inappropriately granted, or access to the data being inappropriately refused. Clearer guidelines for managing these responsibilities are needed to support both the protection of sensitive data and provision of appropriate access to sensitive data.

The custodians, holders, and owners of sensitive data include hospitals, universities, communities, individual researchers, private-sector researchers (such as private medical clinics), libraries, repositories, governments, not-for-profits and NGOs. This list is not exhaustive. Sensitive data are collected in a variety of contexts, and come to researchers with different, often overlapping associated policies. Clarifying and aligning the roles and responsibilities of data custodians, holders, and owners will enable Canadian researchers and the organizations supporting and participating in research, to better manage their sensitive data.

Current Issues

At present, there are few tools or services provided to researchers and research stakeholders to navigate the responsibilities of data ownership or custody. Obligations concerning data management flow from research institutions, research ethics boards, provincial privacy legislation, and funder, repository, and publisher requirements. Researchers are responsible for being aware of the various laws and policies that may govern their work, and for correctly interpreting and applying those regulations, with support typically provided by their home institution. Support is varied and inconsistent across institutions and across jurisdictions.

Future DRI State

A cohesive and national approach to clarifying custodial responsibilities related to data management would enable Canadian researchers to navigate their responsibilities more effectively. NDRIO should support a national approach to sensitive data management to create consistency across jurisdictions and institutions. NDRIO should

also support the development of tools and resources to support researchers and institutions in identifying and addressing data-related responsibilities.

Classification of Data Types – Anonymous vs. Anonymized Data

Data classification types, such as "anonymous data," "anonymized data," and "de-identified data," are used in regulatory guidelines and policies. Data classification types describe the state of the data and may be used to assess the level of risk to privacy associated with the data. Data classification types therefore impact how and when data can be shared, and how it is protected. Both researchers and research participants are often confused about the difference between "anonymous data" and "anonymized data". "Anonymous data" refers to data which were collected with no direct identifiers. At no point in the data collection process is the identity of the respondent known or knowable to the researcher. "Anonymized data" refers to data which were identifiable at the stage of data collection; however, the "study key" or identifying elements of the data have since been destroyed. Terms such as "anonymous," "anonymized," and "de-identified" are not used consistently between jurisdictions, between institutions, and between organizations. For example, a research ethics board may employ the term "anonymized" and a data repository may describe the same dataset as "anonymous." Defining these terms clearly and succinctly in national policies which govern sensitive data collection, such as the TCPS2, would assist researchers and other stakeholders in consistently applying terminology and identifying potential risks associated with privacy. Concise and standardized language for classifying data will allow for consistency between policies, and therefore consistency in research practices.

Current Issues

At present, there are few Canadian resources available to assist researchers in managing sensitive data. The Portage Network Sensitive Data Expert Group has released a bilingual Sensitive Data Toolkit consisting of three parts: a glossary, a risk-matrix, and research data management language for informed consent.[2] The toolkit establishes common definitions of terms, allows researchers to assess and classify sensitivity in terms of risk, and provides language concerning data sharing and deposit for informed consent. Development of other resources is planned and will be influenced by the NDRIO researcher needs assessment.

Future DRI State

Regulatory consistency can be achieved through a national approach. While jurisdictions will be subject to their own local regulations and laws (such as provincial privacy legislation), strong and clear national definitions and standards will enable researchers and institutions to align their practices. NDRIO's support for the development and maintenance of practical tools and resources, such as the Sensitive Data Toolkit, will provide Canadian researchers with effective means of addressing their responsibilities while enabling research.

Secondary use as default

Too often, researchers approach research ethics boards, research libraries, or research repositories with a dataset they intend to share or publish after they have completed their study, only to find that there are obligations or restrictions in place that govern the data which do not allow for sharing or publication. If these issues are addressed at the outset of the research project (for example, at the research ethics application stage), data sharing and deposit can be included in the research design. The language used in consent forms, the requirements of the funding body, the publisher's requirements, the researcher's institutional obligations, and so on, can each be

[2] Part 1: Glossary, English: https://doi.org/10.5281/zenodo.4088946; French: https://doi.org/10.5281/zenodo.4088985
Part 2: Risk-Matrix, English: https://doi.org/10.5281/zenodo.4088954; French: https://doi.org/10.5281/zenodo.4107118
Part 3: Language for Informed Consent, English: https://doi.org/10.5281/zenodo.4060461; French: https://doi.org/10.5281/zenodo.4107185

accounted for to prevent any one of these from blocking the sharing or publication of the data. Experts in research ethics and sensitive data advise researchers to include the expectation of data re-use in their study from the outset. The paradigm shift taking place is a change from a scenario where sensitive data are presumed single-use to one in which sensitive data are presumed re-usable. The specific details of a study will always determine how a particular dataset must be managed, but approaching research with secondary use of data in mind will allow researchers to make small changes at the outset of a study which prevent major hurdles from arising downstream.

Current State

At present, there are no tools or resources to guide researchers in planning for the secondary use of data beyond the definitions and terms set out by the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS2 2018). Individual research ethics boards may have internal practices to assist researchers, but these practices are inconsistent and varied across institutions and jurisdictions.

Future DRI State

NDRIO must support the paradigm shift in sensitive data from single-use to presumed re-usable. Digital infrastructure plays a crucial role in allowing institutions and researchers to protect sensitive data while maximizing its value. Encryption, along with other data security practices, can be employed by institutions and organizations to create digital environments where data access can be appropriately regulated and managed. Such digital infrastructure will require a robust policy infrastructure to support it.

Consistency in research practice, and the risk of a two-tier system

A complex regulatory framework can produce two-tiered (or multi-tiered) requirements for research. Policy differences among funders and institutions can confuse and complicate research processes, and can also contribute to different sets of standards or requirements being applied to relatively similar research. For instance, research projects at any given institution may be subject to different policy frameworks depending on whether or not they receive funding, even if the research is similar in all other respects. Taking measures to prevent a two-tiered policy framework from developing is critical to ensuring that Canadian research participants enjoy the same privacy protections, ethics considerations, and opportunities to contribute to research, regardless of the funding body or specific institutional requirements governing the research study in which they are participating. If we rely on institutions to set their own policies to manage sensitive data, we risk their adopting the bare minimum requirements, which would underserve both researchers at that institution, and research participants recruited by those institutions. While Tri-Council policies apply to all research conducted within a Tri-Council eligible institution, how these standards are applied and enforced sometimes varies. A clear national approach to the management of sensitive data will enable researchers, institutions, and research participants more equal opportunity to benefit from the Open Science movement, and from Canadian research more broadly.

Current State

Research institutions in Canada are developing and adopting Research Data Management (RDM) strategies, in response to the emerging Tri-Agency RDM Policy. Some institutions may opt for a minimal response to RDM requirements, while others may develop more fulsome policies -- resulting in disparities in RDM practice across institutions and jurisdictions. Such disparities challenge multi-site and collaborative research.

Future DRI State

Strong standards set by the Tri-Council can prevent regulatory disparities from growing, and can simplify the complexities involved in multi-site and collaborative research occurring across institutions and jurisdictions. NDRIO is well-positioned to initiate national efforts in this area, and to support the development of tools and resources to educate and assist researchers with regulatory complexity.

Community harms -- to future generations and generations past

In contemporary research ethics there is an additional focus on community risks and harms, and community consent, complementing the continued need to address individual risks and harms, and individual consent. In Canada, we recognize that intergenerational harms and trauma have been created and maintained through certain research practices. We also recognize that some communities have been unduly denied access to the benefits of research, and have been subject to ethically questionable or objectional research practices. Identifying risks and harms as they relate to communities, in addition to individuals, is a critical and meaningful endeavor which must be supported at the federal level. The importance of recognizing such risks and harms in the development of a national approach to the management of sensitive data cannot be understated. Stakeholders from identifiable communities must be consulted and included at every stage in the development of a national approach to the management of sensitive data.

Current State

The most recent update to the TCPS2 encourages researchers to employ the requirements and recommendations of Chapter 9: Research Involving the First Nations, Inuit, and Métis Peoples of Canada to any research involving an identifiable community. All research conducted at Tri-Council eligible institutions must comply with the TCPS2. Following publication of the Truth and Reconciliation Commission Report, the Tri-Council established targeted funding to support research on reconciliation and to support Indigenous research and research training, and the Government of Canada published a strategic plan, "Setting new directions to support Indigenous research and research training in Canada 2019 - 2022." More is needed to identify and address community and intergenerational harms in research, including directed funding to communities to support their research goals and data sovereignty.

Future DRI State

NDRIO is positioned to help address the policy and digital infrastructure gaps involved in Indigenous data sovereignty and capacity-building. Researchers need tools and resources to support the ethical conduct of research involving Indigenous peoples, and communities must be afforded the digital infrastructure and capacity required to engage independently in research. NDRIO must identify relevant stakeholders and provide the funding and support needed to allow for community-driven resources to be developed. For example, "A First Nations Data Governance Strategy" published by the First Nations Information Governance Centre, identifies gaps and requirements in this area.[3] NDRIO must develop a national strategy to enable communities and organizations to address these gaps without sacrificing sovereignty or autonomy in research or research administration.

Contact Person:
Victoria Smith, Policy, Privacy, and Sensitive Data Coordinator for the Portage Network
victoria.smith@carl-abrc.ca

In her role with Portage, Victoria supports the work of the Sensitive Data Expert Group. Please direct any questions concerning this proposal to Victoria and she will facilitate responses from the Expert Group.

This proposal is endorsed by the Canadian Association of Research Libraries (CARL) and the CARL Portage Network.

[3] A First Nations Data Governance Strategy: https://fnigc.ca/wp-content/uploads/2020/09/FNIGC_FNDGS_report_EN_FINAL.pdf