

 <p style="text-align: center;"><b>Institutional Research Data Management Strategy</b></p>	<b>Sponsor:</b> Vice-President Research & Innovation	<b>Approval Date:</b> February 14 2023
	<b>Responsible Unit:</b> Research Administration	<b>Revisions:</b> 3

## A. Purpose

The purpose of the *Institutional Research Data Management (RDM) Strategy* is to foster a culture and develop capacity that supports researchers in adopting responsible RDM practices, following the FAIR Principles<sup>1</sup> to make research data findable, accessible, interoperable and reusable. The FAIR principles are balanced by the CARE Principles for Indigenous Data Governance<sup>2</sup> with considerations for collective benefit, authority to control, responsibility, and ethics.

In 2018 the Canadian Tri-Agencies released a draft of the *Tri-Agency Research Data Management Policy for Consultation*, which will require institutions to create an institutional RDM strategy. This strategy complies with that requirement.

In 2017, an RDM Advisory Committee was formed by the Toronto Area Health Sciences Network (TAHSN) hospitals affiliated with the University of Toronto and have encouraged hospitals to develop their own strategies while coordinating with the University of Toronto libraries.

This document will be used by the Sunnybrook Research Institute (SRI)'s Research Executive committee and Research IT Advisory committee, Research Ethics Board and Sunnybrook leadership to better inform future policy development and operational policies and procedures.

### Definitions

Abbreviations and terms used in this strategy are listed below.

- ❖ **UofT** – University of Toronto
- ❖ **TAHSNr** – Toronto Area Health Sciences Network Research committee membership is mostly Vice Presidents of Research and Innovation.
- ❖ **DMP** – Data management plan
- ❖ **REC** – SRI Sunnybrook Research Executive Committee
- ❖ **RIA** – SRI Research IT Advisory Committee

- ❖ **Research Data** – 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. All other digital and non-digital content have the potential of becoming research data. Research data may be experimental data, observational data, operational data, third party data, public sector data, monitoring data, processed data, or repurposed data<sup>7</sup>.
  
- ❖ **Research Data Management (RDM)** – Data management refers to the storage, access and preservation of data produced from a given investigation. Data management practices cover the entire lifecycle of the data, from planning the investigation to conducting it, and from backing up data as it is created and used to long-term preservation of data deliverables after the research investigation has concluded. Specific activities and issues that fall within the category of data management include file naming conventions; data quality control and quality assurance; data access; data documentation (including levels of uncertainty); metadata creation and controlled vocabularies; data storage, data archiving and preservation; data sharing and reuse; data integrity; data security; data privacy; data rights; and notebook protocols (lab or field)<sup>8</sup>.

## B. Strategy and Roadmap

### 1. Assess Institutional Readiness and ideal state for RDM

An institutional RDM strategy includes an assessment of institutional readiness. This includes identifying and reviewing the data landscape at Sunnybrook Research Institute and assessing existing capacity for RDM. It is also important to define the ideal state for RDM at SRI. An ideal state for RDM at SRI is defined as: researchers will be aware of the supports available to assist them in best data management practices and planning. Researchers will be aware and know how to access appropriate data management planning tools, repositories and resources.

### 2. Support the use of Data management Planners

Identifying stakeholders and providing appropriate resources and materials is essential in delivering effective RDM services. Researchers and staff, are instrumental in helping to promote the value of RDM and use of DMPs and can assist in outreach and communication.

The Research Executive Committee and Research IT Advisory Committee have approved the use of the data management planner guide, online tools and resources currently provided with links to Sunnybrook and University of Toronto resources.

<b>Assess Institutional Readiness</b>				
<b>Objectives</b>	<b>Current Practices &amp; Supports</b>	<b>Gaps &amp; Resources Required</b>	<b>Timeline</b>	<b>Responsibility</b>
2.1 Undertake a survey of institutional data assets and data management practices on campus	Data management practices vary widely across campus. Institutional data assets include: Sunnybrook and SRI infrastructure, external and cloud based repositories,	Data Management practices are highly varied within SRI. Inventory of all data, data types, data owners is challenging.	Ongoing	Research Administration
2.2 Complete a needs assessment of indigenous data collection and management practices	In support of Indigenous persons, Sunnybrook will be assessing its governance and leadership capacity with regards to Indigenous data collection practices.	Sunnybrook requires dedicated resources in support of developing Indigenous data management governance framework	Ongoing	Sunnybrook leadership
<b>Objectives</b>	<b>Current Practices &amp; Supports</b>	<b>Gaps &amp; Resources Required</b>	<b>Timeline</b>	<b>Responsibility</b>
<b>Evaluate existing RDM services</b>				
2.2.1 Data Management Plans	SRI has adopted Portage's DMP Assistant tool. SRI has also created a guide to support researchers and will be updating as policies/technologies change.	DMPs are not currently required by all competitions within Tri-council. Preliminary feedback from Scientists is that the DMPs should include more "boilerplate" information.	Ongoing	Research Administration
2.2.2 Institutional support and training	SRI will establish an internal support group for scientists comprising platform/program resources and IT support.  It will also coordinate with University of Toronto to provide training to scientists on RDM.	More human resources are required to continue to develop training materials and participate in regional and national organizations. Financial resources may be required to hire and/or train in the areas of metadata and data anonymization.	Ongoing	Research Administration

2.2.3 Data repositories and archiving	<p>Currently SRI does not have the capacity to centralize storage and archival for all the scientists' data.</p> <p>Select groups of SRI scientists use the on-premise corporate research storage array but the majority of laboratories collect and store their data within the laboratory.</p> <p>SRI currently does not have a comprehensive data repository of all scientists' data and their primary storage location.</p>	<p>Human resources needed to create and support data repository registry.</p> <p>SRI has committed to enhancing its data storage capacity and backup to include centralization of research data capture and archiving.</p> <p>SRI is exploring the use of University of Toronto's Borealis (Dataverse) repository for data deposit.</p>	Medium term (1–2 years)	Research Computing
2.2.4 Institutional policies, guidelines and/or procedures	<p>SRI has drafted data deposit policy, scientist departure policy and is working on clinical research data access policies. Other RDM policies include <i>Retention and Destruction of Records Policy and Schedule, Access to and Disclosure of Personal Health Information, Ethical Conduct of Research Involving Humans Policy, etc</i></p>	<p>.Human resources required for consultation process with Indigenous researchers and the Indigenous Community, and with the Indigenous Advisory Council specifically.</p>	Long term (2–3 years or more)	Research Executive Committee and hospital senior leadership committee

### **3. Formalize RDM practices and procedures**

The *Tri-Agency Research Data Management Policy* will be formalized shortly. RDM is required by some journals and funding agencies. In addition, Research Ethics applications require some components of RDM. SRI/Sunnybrook has drafted data deposit and scientist departure policies and procedures to better inform the research hospital community.

Other policies are related to the data repository and include data quality and standards, data retention, and long-term data preservation. A clinical research data governance policy, while not specific to RDM, has been drafted and will also inform RDM practices.

SRI will be adopting the online Data management plan system and incorporate the workflow into its pre-awards and grant application submission process to better support the researchers in this process.

In addition, SRI may wish to develop policies related to unfunded research and student research, including undergraduate and graduate.

Some gaps in policy can be filled by continuing to work with the TAHSNr working groups who are collaboratively developing resources, services, infrastructure and policy related to RDM, and with those leading Indigenous Data Governance initiatives. Policy development is a long-term goal that will be led by the Research IT Advisory Committee.

#### Technology Roadmap

Sunnybrook currently is sourcing an enterprise-wide data storage and archival solution for clinical and non-clinical research data. We are also looking at cloud based data management solutions and will finalize our technology roadmap in Spring of 2023.

SRI is also researching additional tools for public data deposit, data inventory tools de-identification tools.