

Institutional Research Data Management Strategy

Author:	Concordia University Research Data Management	Date: February 17, 2023
		Version: 1.0
Conter	nts	
Backgrou	nd	2
Importanc	e of Research Data and Research Data Management	
Oversight	, review and timeline	
Stakehold	lers	5
Definitions	S	6
Current S	upport for Research Data Management at Concordia	
RDM P	roject Team	
Training	g and awareness	
Data Ma	anagement Plans	9
Data re	positories	9
Institutio	onal policies and procedures	

1



Background

In 2016, the Canadian Institutes of Health Research (CIHR), the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Social Sciences and Humanities Research Council of Canada (SSHRC), hereafter known as the Tri-Agency, published a Statement of Principles on Digital Data Management^{1.} which outlines their expectations in achieving excellence in data management practices. The Statement defines the roles and responsibilities of researchers, research communities, institutions, and funders in attaining these expectations through activities such as data management planning, adherence to standards, appropriate collection and storage of research data, metadata creation, preservation, and data citation.

In 2021, the Tri-Agency published a Research Data Management policy with the objective to "support Canadian research excellence by promoting sound [Research Data Management] and data stewardship practices."² The policy is divided into three pillars:

- 1. **Institutional strategy**: The policy requires each institution administering Tri-Agency funds to create an institutional research data management strategy by March 1, 2023, that outlines how institutions will develop awareness of and support for exemplary data management practices.
- 2. **Data management plans**: Refers to the preparation of a formal document outlining how data will be managed throughout the life of a research project and after its conclusion. These plans will be required for an initial set of funding opportunities to be identified by the Tri-Agency in the spring of 2022.
- 3. **Data deposits**: Refers to the transfer of research data, collected as part of a research project, into a research data repository. Data deposits will be phased in only after the Tri-Agency reviews each institution's RDM strategy and assesses the Canadian research community's readiness. At no time will researchers be required to share their data. However, the Tri-Agency anticipates that researchers will provide appropriate access to their research data when ethical, cultural, legal and commercial obligations allow. Note that data should be "as open as possible and as closed as necessary"³: "open" in order to promote data reuse and to advance research, but "closed" to protect the privacy of research participants.

¹ Government of Canada. (2016). *Tri-Agency statement of principles on digital data management*. Retrieved November 9, 2020 from http://www.science.gc.ca/eic/site/063.nsf/eng/h_83F7624E.html

² Government of Canada. (2021). *Tri-Agency Research Data Management Policy*. Retrieved March 15, 2021 from <u>http://www.science.gc.ca/eic/site/063.nsf/eng/h_97610.html</u>

³ European Directorate (2016) *H2020 Programme Guidelines on FAIR Data Management in Horizon 2020.* Retrieved October 19, 2021 from <u>https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf</u>



Importance of Research Data and Research Data Management

Research data are data used as primary sources to support research, scholarship, or creative practices. What counts as relevant research data varies among and across disciplines, areas of research, and modes of inquiry.

Properly managed research data is the cornerstone of high-quality research outputs. It allows researchers to better organize, store, access, reuse and build upon their research. It is at the heart of making data FAIR⁴ (findable, accessible, interoperable and reusable) which is a "key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse"⁵. It allows researchers to effectively comply with the growing number of research data management policies from granting agencies and editors. And crucially, it fosters reciprocity, allowing research data and outputs to be beneficial to both the researcher and their community (peers, participants, and partners).

Therefore, the objective of this document is to help researchers work towards adopting data management best practices and to express Concordia University's commitment towards excellence in this area. A long-term future ideal state for research data management (RDM) at Concordia University is a future where services, staffing resources, and IT infrastructure enable and support best practices in this area.

⁴ Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. <u>https://doi.org/10.1038/sdata.2016.18</u>

⁵ European Commission. (n.d.) Data management. Horizon 2020 Online Manual. <u>https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm#</u>



Oversight, review and timeline

In response to the Tri-Agency RDM policy, the following strategy document has been developed by the Research Data Management (RDM) Project Team. This team reports to the Vice-President, Research and Graduate Studies and the University Librarian, and is comprised of the following members:

- Jared Wiercinski, Associate University Librarian, Research & Graduate Studies (Chair)
- Tarik Alj, Manager, IT Research Support, IITS
- Marie-Pierre Aubé, University Archivist
- Danielle Dennie, Research Data Librarian
- Michael Groenendyk, Digital Scholarship Librarian
- Alex Guindon, GIS & Data Librarian
- Dominique Michaud, Director, Research Development, Office of Research
- Monica Toca, Manager, Research Ethics, Office of Research
- Associate Deans (Research) from the four Faculties (on a consultation basis)

The RDM Project Team is responsible for overseeing the implementation of the strategy. The current strategy document covers a three-year period and will be revised by the RDM Team after this time to assess attainment of deliverables and to reflect changes in the research data management landscape at Concordia, as well as provincially, nationally, and internationally.

Inquiries about this strategy can be directed to the Research Data Management Project Team at lib-research.data@concordia.ca.



Stakeholders

The following groups have a stake in RDM at Concordia and will be made aware of the strategy and associated RDM requirements, resources, and support through actions such as presentations and consultations.

- Vice-President, Research and Graduate Studies
- University Librarian
- Librarians
- Office of Research
- Ethics Unit
- Research Units
- Associate Deans, Research
- Faculty Councils (Arts & Science, Engineering, Fine Arts, JMSB)
- Department chairs
- Researchers
- Indigenous Directions Leadership Council
- Records Management and Archives
- Associate Vice President, Information Systems and CIO
- IITS
- Legal Services



Definitions

Archivematica: A free and open-source digital preservation system that is designed to maintain standards-based, long-term access to collections of digital objects.⁶

Data Management Plan (DMP): A formal statement describing how research data will be managed and documented throughout a research project and the terms regarding the subsequent deposit of the data with a data repository for long-term management and preservation.⁷

Dataverse: An <u>open source data repository</u> originally developed at Harvard University used to share, preserve, cite, explore, and analyze research data. It contains datasets, descriptive metadata and data files.⁸ Note that the Dataverse discussed in this document is not the Microsoft Dataverse software application.

Research Data: Definitions of research data vary greatly, depending on the discipline. CODATA provides the following broad definition of 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.⁹

However, researchers involved in research-creation may define research data differently. Research-creation is an approach to research that involves both creative and scholarly practices. Research-creation data refers to the data that is generated through such critical, practice-based research. This may include traditional creative forms (such as painting, creative writing, music composition, design, architecture, performance art...) but also more conceptual, interdisciplinary and innovative approaches (for example, social and political practices, collaborative, relational and other emerging forms of production). Documentation of this type of research may also involve more traditional forms of data, such as numbers and text. The specific nature of research-creation data will vary depending on the research question and the methods used.

⁶ Archivematica. (n.d.). What is Archivematica. Retrieved December 7, 2020 from <u>https://www.archivematica.org/en/docs/archivematica-</u> <u>1.12/getting-started/overview/intro/#intro</u>

⁷ CODATA. (n.d.). Data management plan. Retrieved March 29, 2023 from <u>https://codata.org/rdm-terminology/data-management-plan/</u>

⁸ Dataverse Project. (n.d.). About. Retrieved December 7, 2020 from <u>https://dataverse.org/about</u>

⁹ CODATA. (n.d.). Research data. Retrieved March 29, 2023 from <u>https://codata.org/rdm-terminology/research-data/</u>



Researchers in the humanities may define research data as: "All materials and assets scholars collect, generate and use during all stages of the research cycle."¹⁰ These materials include whatever the researcher, team and/or collaborators find worthy of their thought and attention. As such, research data in the humanities would include but not be limited to that which they describe, analyze, and/or represent through a variety of sensory means be they musical or otherwise sonic, poetic or literary, experiential or performative, etc.

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 (the proper way to name computer files); 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; notebook protocols (lab or field).¹¹

¹⁰ Harrower, N., Maryl, M., Biro, T., Immenhauser, B., & ALLEA Working Group E-Humanities. (2020) Sustainable and FAIR Data Sharing in the Humanities: Recommendations of the ALLEA Working Group E-Humanities. ALLEA. https://doi.org/10.7486/DRI.tq582c863



Current Support for Research Data Management at Concordia

RDM Project Team

Concordia's RDM Project Team is responsible for creating guidelines, procedures and policies to formalize research data management practices, services and expectations within the University and developing and implementing associated communication plans.

Training and awareness

Currently, different service units provide awareness materials and resources, as well as support and training for the Concordia community. Note that the services listed below are at different maturity levels but are expected to expand as demand for RDM services increases.

Concordia Library

- Provides a <u>Research Data Management Guide</u> on the Library website, which includes a specific section on research with Indigenous communities;
- Offers workshops and consultations to Concordia faculty, students, and staff;
- Trains subject librarians on RDM to inform their work with faculty and students in their assigned departments;
- In coordination with University Communications Services (UCS), communicates with the community through a variety of communication channels;
- In coordination with Data Scientifique, provides data clinics and workshops in the Library on data management, quantitative statistical methods, and data visualization.

Concordia IT Research Support

• Provides <u>consultation services</u> as well as research storage, research server hosting, and research virtualized servers. The support is run by a team that liaises between researchers and IT professionals at Concordia.

Office of Research

- Facilitates workshops on RDM as well as connects researchers with the Library for assistance with issues related to RDM.
- Advises researchers on ethical, legal and commercial issues related to data management, such as data confidentiality and consent forms, that are consistent with requirements expressed in the *Tri-Council Policy Statement: Ethical Conduct for*



Research Involving Humans (2nd edition)¹², the Tri-Agency Framework: Responsible Conduct of Research¹³, as well as other relevant policies.

Records Management and Archives

- Provides workshops on document management, such as retention, preservation, file naming and organization.
- Provides workshops on collecting and protecting personal information.

Data Management Plans

A Data Management Plan (DMP) is a document that helps researchers and funding agencies to understand the type of data that will be produced, how it will be managed and preserved, and how it will be shared.

The DMP Assistant (<u>https://assistant.portagenetwork.ca/</u>), a tool developed by the Digital Research Alliance of Canada (<u>https://alliancecan.ca/</u>) in observance with established best practices, can assist researchers in writing DMPs and provides subject-specific templates and local guidance. Knowledge of this tool is not widespread at Concordia. To increase awareness of the DMP Assistant, a limited number of workshops have been provided since 2020 to librarians and to staff in the Office of Research as well as to researchers. Furthermore, a communication plan¹⁴ was launched in the fall of 2020 in order to convey to researchers and graduate students the importance of DMPs and the availability of the DMP Assistant.

Data repositories

Data repositories allow long-term storage and access to research data. There are a variety of repositories available for researchers. They can be either discipline-specific, general, or institutional.

At the conclusion of a research project, Concordia researchers have different options for storing their data for the long-term. They can choose to deposit their data in subject specific repositories that cater to their specific disciplinary needs. They can also deposit data in institutional repositories like <u>Spectrum</u>, where the data is preserved using the preservation software Archivematica. Concordia

¹² Government of Canada. (2018). *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans – TCPS 2 (2018)*. <u>https://ethics.gc.ca/eng/policy-politique_tcps2-eptc2_2018.html</u>

¹³ Government of Canada. (2016). *Tri-Agency Framework: Responsible Conduct of Research (2016)*. <u>https://rcr.ethics.gc.ca/eng/framework-cadre.html</u>

¹⁴ Concordia University. (2020, November 27). Concordia Library provides grad students and faculty valuable data management resources. Retrieved February 17, 2023 from <u>https://www.concordia.ca/cunews/main/stories/2020/11/27/concordia-library-provides-grad-students-and-faculty-valuable-data-management-resources.html</u>



researchers can also deposit their data in general data repositories like the <u>Concordia instance of Dataverse</u>¹⁵, a Canadian based open-source data repository for small to medium sized datasets. The Library has developed documentation to assist users in the Dataverse deposit process. Datasets in Dataverse are not currently archived with preservation software. Because these two institutional offerings (Spectrum and Dataverse) do not support all types of research data, the <u>Library's RDM guide</u> helps guide researchers towards other repositories which may be better suited to their specific needs.

Institutional policies and procedures

Adopting policies, guidelines and/or procedures helps support institutional awareness of RDM and promote good RDM practices. These policies, guidelines, and/or procedures may address a variety RDM issues, such as:

- Data access and sharing
- Data retention
- Long-term data preservation
- Data management plans
- Privacy, ethical issues, and intellectual property
- Consideration of Indigenous data sovereignty

Many policies at Concordia include, to some extent, subject matter pertaining to research data management. These policies include the Policy for the Responsible Conduct of Research¹⁶, the Policy for the Ethical Review of Research Involving Human Participants¹⁷, the Policy on Intellectual Property¹⁸, the Policy on Contract Research¹⁹, the Policy Concerning the Protection of Personal Information²⁰, the Policy on Data Governance²¹, and the Concordia University Dataverse Policy²². Concordia also passed a Senate Resolution on Open Access²³. There is also an increasing number of journal publishers and funding agencies that have current or upcoming policies related to RDM, and more specifically, pertaining to DMPs and / or data sharing.

¹⁵ <u>https://dataverse.scholarsportal.info/dataverse/concordia</u>

¹⁶ https://www.concordia.ca/content/dam/common/docs/policies/official-policies/VPRGS-12.pdf

¹⁷ https://www.concordia.ca/content/dam/common/docs/policies/official-policies/VPRGS-3.pdf

¹⁸ <u>https://www.concordia.ca/content/dam/common/docs/policies/official-policies/VPRGS-9.pdf</u>

¹⁹ <u>https://www.concordia.ca/content/dam/common/docs/policies/official-policies/VPRGS-1.pdf</u>

²⁰ https://www.concordia.ca/content/dam/common/docs/policies/official-policies/SG-9.pdf

²¹ https://www.concordia.ca/content/dam/common/docs/policies/official-policies/PRVPA-4.pdf

²² https://library.concordia.ca/research/data/Concordia_Dataverse_Policy.pdf

²³ Herland, K. (2010) Concordia opens access to its research output. *Concordia Journal*. http://cjournal.concordia.ca/archives/20100429/concordia opens access to its research output.php



Indigenous Data Considerations

This strategy document does not include details on how Concordia University will approach working with data from or about First Nations, Métis, and Inuit individuals, communities, or nations. Concordia University Indigenous Directions is responsible for guiding the decolonization and Indigenization of the institution through its Action Plan²⁴. The plan recommends seven specific actions with regards to Indigenous research, encouraging Concordia University to commit to "reimagining how ethical, reciprocal and meaningful Indigenous research, in partnership with Indigenous communities, is conducted."²⁵ Specifically, one of the recommended actions is the creation of an Indigenous Research Policy that will take into account the Tri-Agency Research Data Management Policy, among other resources.

²⁴ Concordia University. (2019). The Indigenous Directions Action Plan: Concordia's path towards decolonizing and Indigenizing the University. <u>https://www.concordia.ca/content/dam/concordia/offices/oce/docs/IDLG/indigenous-directions-action-plan.pdf</u>

²⁵ Ibid, 34.



Roadmap towards an ideal state of Research Data Management at Concordia

Many Research Data Management policies, processes, infrastructures, services, and support at Concordia University are not yet formalized or are still under development. Achieving an ideal state of RDM at Concordia will be an incremental process through which the university will guide and support its researchers with the goal of achieving data management best practices. Some elements of an ideal state of RDM would include the following:

- Institutional RDM-related policies, procedures, and guidelines
- Consultation and training in the following areas:
 - RDM and DMPs
 - RDM-related software
 - Data curation
- Availability of active, repository, and archival/preservation storage for both research data and sensitive research data
- Security and risk assessment policies and procedures
- Availability of high-performance computing and file transfer services
- And more...

The following table presents a three-year gradual introduction of RDM related objectives at Concordia. At the end of this three-year period, the objectives will be updated and revised by the RDM Project Team to continue the advancement towards an ideal state of RDM at the University.

1.0 Raising awareness and providing institutional support and training

Communication is key in order to highlight the benefits of and requirements for research data management (RDM) to all Concordia researchers. Outreach activities could include recruiting local champions to promote the value of RDM, engaging with various communities, and developing awareness materials and resources.

Objectives	Strategies	Deliverables	Gaps	Timeline / Priority	Responsibilities
1.1 Recruit local champions	Collaborate with Concordia University Data Science Research Centre on potential data- related training initiatives	One workshop per year through the Library.		December 2023 / Low	Library



	Collaborate with Data Scientifique on RDM- related training initiatives	Continue providing workshops and office hours within the Library	Ongoing / High	Library
	Promote data champions, such as those who deposit in <u>Concordia</u> <u>Dataverse</u>	Communication plan to highlight researchers on Library RDM guide and/or through the Library's social media feeds or news page.	July 2024 / Medium	Library
1.2 Develop awareness materials and resources	Develop brief arguments on the benefits of RDM and use these to tailor different messages for different disciplines or faculties.	Presentation slides on RDM that subject librarians can use for departmental meetings or within graduate workshops.	August 2023 / High	Library
		Promotional material that can be handed out during events such as Open Access Week and graduate and new faculty orientations	August 2023 / Low	Library
	Develop awareness materials surrounding Québec's Law 25, an Act to modernize legislative provisions as regards the protection of personal information	Workshop and / or guide on protecting personal information	December 2024 / High	Records Management & Archives, Office of Research
1.3 Develop training materials	Develop training materials for active data storage	Training capsules on storage of active data and data security, including how to store and share	December 2024 / Medium	IITS training department



	Develop more advanced	sensitive data with collaborators	December	Library
	training materials on research data management	training researchers on anonymization / deidentification methods	2024 / Low	
		List of collaborators for training researchers on data cleaning	December 2024 / Low	Library
1.4 Define need for RDM services	Review the need for increase in staffing to assist with RDM services	Map out time spent on and types of RDM activities performed by the RDM librarian, and level of staffing for different tasks.	June 2024 / Medium	Library
1.5 Enhance current documentation in support of the ethical conduct of research	Create documentation that contains research data management language for informed consent	Consent form templates will include information about data sharing Instructions for writing consent forms will be updated to include information about data sharing	August 2023 / High	Research Ethics Unit (Office of Research)

2.0 Data Management Plans (DMP)

A Data Management Plan (DMP) is a document that helps researchers and funding agencies understand the type of data that will be acquired or produced during a project, how it will be managed, described, analyzed, and stored, and how it will be shared and preserved at the end of the project. Although these issues may be well thought out in advance, formalizing the process in a written document helps to identify potential blind spots or weaknesses in a planned project, and provide a record of the project's intentions.



Objectives	Strategies	Deliverables	Gaps	Timeline / Priority	Responsibilities
2.1 Cultivate awareness and use of DMPs by researchers	Promote the DMP Assistant, especially when DMPs are required by funder, and provide training on how to write a DMP	One or two workshops per year, organized through the Office of Research or targeted to specific faculties, on DMPs and how to use the DMP Assistant		Ongoing / High	Library, Office of Research
	Promote the Library and the Office of Research as places where researchers can get assistance with writing DMPs	Establish a DMP consultation service at the Library	Additional human resources may be required to meet the demand	December 2024 / High	Library, Office of Research
		Train Grant Writing Assistants to provide help in writing DMPs through the Grant Writing Assistant (GWA) Registry		December 2024 / High	Office of Research
		Link to the Library RDM guide from the Office of Research website		July 2023 / Low	Library, Office of Research
		Presentation slides about DMPs that subject librarians can use at departmental meetings or within workshops.		August 2023 / High	Library

3.0 Data repositories and archiving

One of the primary goals of RDM is to store data and its accompanying documentation in a way that allows future access by researchers, including the producers of the data themselves. Data repositories are one way of providing long-term storage and



access to research data. In fact, there are a variety of repositories available for researchers, and whether they are disciplinespecific, general, or institutional, they allow research data to be findable, accessible and reusable.

Objectives	Strategies	Deliverables	Gaps	Timeline / Priority	Responsibilities
3.1 Establish a culture of data deposition and archiving	Promote Concordia Dataverse as a tool for depositing small to medium sized datasets, when appropriate	One or two workshops per year on data repositories, data sharing and how to use Dataverse		Ongoing / High	Records Management & Archives, Library, Office of Research
	Increase the number of datasets deposited by researchers in Concordia Dataverse.	Report for VPRGS and UL that outlines Concordia Dataverse metrics.		Ongoing / Low	Library
	Investigate possibility of preserving datasets through Archivematica	Report to assess the feasibility of archiving research data deposited in Concordia Dataverse either locally or through the creation of a provincial archiving solution. The report would be presented either to the VPRGS and UL or the BCI Groupe de travail sur la Gestion des données de recherche (GT-GDR).	Current lack of archiving solution at the provincial level. Necessary IT hardware and staffing to support archiving several large datasets.	December 2024 / Low	Library, Records Management & Archives
	Collaborate with provincial / national / international organizations for external storage of research data	Training and guidance materials available on the Library RDM guide on using the Federated Research Data Repository (FRDR) for the deposit		Ongoing / Low	Library



	and preservation of large datasets.			
	Representative from Concordia on one of the Aliance Network of Expert groups ²⁶ .		Ongoing / Low	Library
Communicate options to researchers for data sharing according to discipline or types of data produced	Guidance materials available on the Library RDM guide.	Limited options for storing and archiving audio-visual data.	Ongoing / Low	Library

4.0 Institutional policies and procedures

Adopting policies, guidelines and/or procedures helps support institutional awareness of RDM and promote good RDM practices. These policies, guidelines, and/or procedures may address a variety RDM issues, such as: data access and sharing; data retention; long-term data preservation; data management plans; privacy, ethical issues, and intellectual property; and consideration of Indigenous data sovereignty.

Objectives	Strategies	Deliverables	Gaps	Timeline /	Responsibilities
				Priority	
4.1 Develop institutional policies, procedures, or guidelines related to RDM	Develop a data classification standard which outlines how data or information is protected based on its level of sensitivity.	Publish data classification standard and communicate it to researchers		December 2024 / Medium	IITS, Records Management & Archives, Data Governance Steering Committee

²⁶ Digital Research Alliance of Canada. (n.d.). *Network of Experts*. Retrieved February 17, 2023 from <u>https://alliancecan.ca/en/services/research-data-management/network-experts</u>



Develop a cloud directive	Publish institutional cloud	July 2023 /	IITS, Records
which describes the	directive and communicate	High	Management &
protections needed when using cloud services based on the type of data involved and its required security and privacy needs	it to researchers	Ŭ	Archives