National Data Services Framework Summit 2019
Notes and Resources

Final Version, March 7, 2019
Attendees of the 2019 NDSF Summit

DOI: 10.5281/zenodo.2584260

Introduction

The 2019 NDSF Summit continued the conversations that were initiated in the 2017 Summit, and in various other meetings and workshops. The outcomes of this Summit and related meetings in the same week are summarized in the Kanata Declaration document. The notes presented here were those recorded from the two Breakout sessions.

Presentations

1. 2019 National Data Services Framework Summit : https://doi.org/10.5281/zenodo.2552802
2. Beyond Open Data: How Open Science Aims to Transform Science Partnerships: https://doi.org/10.5281/zenodo.2552804
3. Federated Research Data Repository (FRDR) - Portage and Compute Canada: https://doi.org/10.5281/zenodo.2553612
4. Epigenomic data discovery with the IHEC Data Portal: https://doi.org/10.5281/zenodo.2554255
5. Finding and Accessing "Small" Data through SAGE: https://doi.org/10.5281/zenodo.2554273
6. The CRCDN: Progress, opportunities and challenges: https://doi.org/10.5281/zenodo.2554305
7. Canadian Integrated Ocean Observing System (CIOOS): https://doi.org/10.5281/zenodo.2554389
8. InterOperability: https://doi.org/10.5281/zenodo.2554508
9. SPOR (Strategy for Patient-Oriented Research) National Data Platform: https://doi.org/10.5281/zenodo.2554527
10. DataStream: https://doi.org/10.5281/zenodo.2555189
11. Canadian Health Omics Repository, Distributed (CHORD): https://doi.org/10.5281/zenodo.2555212
12. Dataverse for the Canadian Research Community: Developing reusable and scalable tools for data deposit, curation, and sharing: https://doi.org/10.5281/zenodo.2555322
13. DuraCloud: https://doi.org/10.5281/zenodo.2555344

14. FAIR Repository for Annotations, Corpora and Schemas: https://doi.org/10.5281/zenodo.2555348
15. Federated Geospatial Data Discovery for Canada – Geodisy: https://doi.org/10.5281/zenodo.2555352
16. Making Identifiers Necessary to Track Evolving Data (MINTED) – A Brief Overview: https://doi.org/10.5281/zenodo.2555354
17. Active Research Data Management Tools: Radiam: https://doi.org/10.5281/zenodo.2555356
18. A Research Lifecycle Approach using Islandora 8 Overview: https://doi.org/10.5281/zenodo.2555360
19. Research Portal for Secure Data Discovery, Access and Collaboration: https://doi.org/10.5281/zenodo.2555366
20. World Data System: Trustworthy Data Services for Global Science: https://doi.org/10.5281/zenodo.2555244
21. Building Trust in Scientific Data: Certification & the CoreTrustSeal: https://doi.org/10.5281/zenodo.2555288
22. Open.Canada.ca - Canada's Open Government Portal: https://doi.org/10.5281/zenodo.2555304
23. Data Management and a National DRI Strategy: https://doi.org/10.5281/zenodo.2556688

Breakout 1 Notes

The notes below were transcribed from the two breakout worksheets, and are sorted according to the number of dots placed with each statement. Some grammatical edits were made, and in a few cases multiple statements were combined into a single one where the association between text and dots was not clear.

Statements which received 6 or more dots are reflected in the Declaration draft above. Most of the remaining statements are reflected in the draft through their overlap or support for other statements.

Architecture

1. Common data brokerage system that allows datasets to be aggregated from disparate platforms & locations into new integrated datasets for new analysis, as well as a clear diagrammable agreed architecture showing national and local levels, and showing all platforms/tools/services. (9)
2. Gap between researchers vision of data management and development driven by IT departments. (5)
3. Build on and expand existing and emerging platforms and services; don't forget value of domain-specific repositories and services but aim to coordinate on issues like discovery, minimal metadata standards. (5)
4. Absence of funding and Support for a sustainable federated system. (2)
5. Ensure clearly documented platform APIs. (2)

6. Pathways between active, repository, archival environments for storage, access, reuse. Pathways = protocols, standards, platforms, with maintenance of provenance (e.g. blockchain for traceability). (1)
7. Lack of a clear definition of requirements for an integrated architecture.
8. Long-term storage infrastructure (e.g. health research, 25 years to hold data): we have the infrastructure, but not the political will to enable it (at CC, Compute Ontario, etc.).
9. Centralized vs. distributed/decentralized vs. hybrid architectural structure.
10. Create a common resource of legacy systems (like tape readers) that all Canadian organizations can access for data rescue.
11. Network: ability to transfer large amounts of data.
12. Improve capability and availability of high bandwidth networks, including to remote locations.
13. Web indexing of DOIs from repositories (and other infrastructures), which would help with the incentive model.

Data

1. “Data Review” as part of literature review as an integral part of the research process in understanding existing data. (18)
2. “Data Provenance quality assurance” should be a priority, or part of the process to include the researcher (data depositor) to be involved in documenting the data for reuse. (10)
3. Minimum common denominator for metadata description, including common language for describing “fit for purpose”. (8)
4. Promulgating metadata and exchange standards: more standardization with domains; more harmonization and crosswalks to interoperate across domains; enriched, high-level discovery layer (e.g. DC plus optional subjects/keywords). (6)
5. Addressing societal issues (e.g. environmental, indigenous, health inequalities) can drive interdisciplinary research, enabled by funding agency programs, as a catalyst for developing common standards and tools. (3)
6. Focus on “good enough” metadata standards across disciplines: start somewhere, recognizing it needs to continue to evolve. (3)
7. How do you create incentives (stick/carrot/$$$) for research teams to adopt metadata standards and properly document them. (3)
8. Finding the skills/talent required. (2)
9. Preservation of government data needs to be addressed, and we need solutions for sensitive data; and skills with diverse data types. (2)
10. De-identification awareness, skills, tools, expertise, standards. (1)
11. Curators to enrich the metadata.
12. Highlight quality of data/data usage for rewards and recognition.

Services

1. Data scientists and data librarians to support researchers. (18)
2. Education and training for researchers, ensuring stakeholders have/maintain autonomy in their respective areas/discipline, but at the same have access to, and are aware of best practices, tools, services that support them as “data managers”. Develop use cases for specific pieces and/or the overall framework to be used to develop support for ongoing use and evolution. (15)
3. There are training needs at all levels, and we need to anticipate future data needs (AI, machine learning). (6)
4. Continuous support across the research lifecycle. (4)
5. Acknowledge that researchers need support to get RDM done properly and streamline common services (portal?). (3)
6. Sandboxes for existing services and documentation, as well as metric services (alternatives to Google Analytics) that’s supported by a national service. (3)
7. Funding to institutions to support sustainable services for RDM (built in indirect cost formula funding). (2)
8. Policy interoperability across international as well as national jurisdictions, starting with a Level 2 environmental scan across international jurisdictions, around data and policy management. (2)
9. Lack of training, advice, mentorship for adopting standards and creating metadata: no institutional resources, and we need to ensure institutional support for RDM. (1)
10. Storage and compute co-location.
11. Sustainability is a challenge, as is combining efforts across domains/groups.
12. Education and training about data management.

Access & Interface

1. Need to build a Researcher Dashboard for RDM services, platforms and support. Build off existing open source solutions (e.g. Drupal). (9)
2. Ability to record user prior to downloading a dataset so that consumers can be notified in the future if the data was flagged for invalid data within. (6)
3. Gap between perspective of researcher with regards to security and the classification, platforms, practices provided by IT. (6)
4. Knowing more about all the different repositories and where you can/should deposit. (4)
5. Unified user experience: not “one portal to rule them all”, but a set of APIs that unifies interfaces and services. (3)
6. Have more structured guidelines for deposit. (2)
7. PIDs! Every dataset to have DOI; every researcher an ORCID ID, every Project [RaID], etc. (2)
8. Education of researchers re RDM planning and services, built on initiatives like Portage. (1)

9. Website to help researchers understand what to do and where to go, create communities of expertise and practice, and articulate where the water's edge is for a national data services framework: does it reach into my lab. (1)
10. Standardization of interfaces (API): at a minimum, every API must have working examples and demonstration utilities. (1)
11. Accessibility standards compliance for all.

Rules

1. Encourage participation of emerging (not PI) researchers, and fill the gap in funding for secondary data use. (12)
2. Limited data is audited by 3rd party after archiving to ensure appropriate access. (5)
3. Use stronger language in national policies. (5)
4. Who is accountable for proper research data management (documentation, sharing) researcher/university/funder? What is the role of market leaders (Google etc.) in governance and university research vs. role of librarians relative to IT/OVPR/UL? (3)
5. Ensure mandatory DMPs; and consult with liaison librarians. (3)
6. Government data should be deposited directly into an appropriate repository (NOT departmental).(2)
7. Design rules that leave room for dealing with cultural issues such as fear of reputation damage due to data errors, e.g. data peer review. (1)
8. "AirBnB" type method for establishing trust of data sources and data user (e.g. sensitive data. (1)
9. Don't replicate that which can (and should be) federated/shared.
10. Rules must be human and machine readable.
11. No consistency in data security guidelines across the system.

Governance

1. Inadequate incentive model to encourage participation; citation indexing (inconsistent/inappropriate use); lack of data governance model in institutions (for research and institutional data). (16)
2. Consult with researchers: get more of them in this conversation. (13)
3. There is a culture gap to common understanding and buy-in to what we assume true with respect to open science, open data, etc., by researchers, as well as a education/awareness gap: ensuring our researchers are aware of, educated in, and supported in, their use of and buy-in to the national data services platform, and that there are incentives to support/adoption/adherence to policy. (12)
4. Highlight rewards and recognitions/incentives and engagement, as well as quality of data. (11)
5. Engaging researchers, political and social aspects of engaging researchers, and training in data science, expertise and metadata. (8)

6. Advisory body with representation from all stakeholder groups, and all regions to help prioritize investment and development. (8)
7. Cross-fertilization at a senior level of Board members for RDM organizations.
8. Too much responsibility, too little authority in the RDM ecosystem. (6)
9. Coordinate long-term funding with the big players nationally, and have a representative governance (including librarians) with the big picture view, combined with practical approach to solving problems. (6)
10. Wrap governance and funding around critical services like DOI minters (DataCite Canada). (2)
11. We need overall coordination and sustainability.

Breakout 2 Notes

Recall

1. That we are in the business of scientific development. As such we must organize and structure ourselves as enterprises, focusing on having a system of governance that intends to be sustainable, and economically imparting to the communities we interact with.
2. That the LCDRI has put forward to the Canadian government recommendations regarding the evolution and restructuring of the Canadian DRi ecosystem.
3. The valuable contributions and initiatives of a multitude of actors grows Canada's research data management landscape to support various elements of data management, including data management planning, creation, and preservation.

Reaffirm

1. The reason for RDM is to maximize the science outputs optimizing the initial investment. (6)
2. The need for an organization with a comprehensive national mandate and resources to address the challenges by integrating/leveraging existing and future initiatives, infrastructure and services.(5)
3. The critical role of the federal government and the Canadian RDM stakeholder community to bring together the people, organizations, and resources for concentrated efforts to advance Canadian RDM infrastructure norms and initiatives. (4)

Recognize

1. There is a general lack of awareness within the research community across domains and within the broader Canadian population of the importance and need for good RDM

practice, and the impact of good practice on the Canadian economy, global positioning,a and the quality of life and well being of all Canadians. (15)

2. That research and scholarship are becoming more data intensive and that data is a valuable asset that must be managed. (12)
3. Researcher centric creative engagement, seeing the value of proper data management, both for the researcher and the community at large, with adequate support to participate in open science. (11)
4. That innovation comes from the ground level initiatives, while there needs to be national support to leverage common objectives, approaches, systems, knowledge, and tools. (8)
5. NDSF is not infrastructure, but more about a sustainable process. (7)
6. The public good of RDM in Canadian and international science and society; the imperative to innovate in this space to ensure the future viability, integrity, and stability of Canadian RDM assets; the importance of international engagement and leadership of Canadians as a prerequisite for building capacity and confidence in science. (7)
7. And celebrate that the federal government will take an ongoing funding role at the national layer. (7)
8. The responsibility that all contributors of data must adhere to in regards to: (5)
    a. Sharing datasets/DRI with their peers;
    b. Providing public forums/avenues to knowledge with members in and out of the scientific community;
    c. Developing software platforms that must go beyond the scope of their field, when possible.
9. That researchers' role is to do research, with the necessary RDM supports within the institution. (3)
10. That the adoption of an implementation of a principle that is widely agreed to be good, very much depends on the details (e.g. CCV as an example that most agree is good in principle, but painful in execution.). (1)
11. The federal government's forthcoming DRI strategy and the opportunity for an overarching vision for a national RDM framework and a coordinated and sustainable approach to supporting RDM across Canada.
12. The need for strong governance across the pillars including DRI/RDM/ARC/Software. (1)
13. The critical importance of three categories of stakeholders, Data:
    a. Managers
    b. Stewards
    c. Consumers

Highlight

1. That we should grow:
    a. Capacity for RDM infrastructure and services;
    b. Culture change in researcher adoption of RDM services;
    c. Adoption of RDM services;
    All approximately at the same rate. (9)

2. That data stewardship and management requires specific expertise and often domain-specific knowledge. (2)
3. And recognize that importance of the many pioneering projects that have preceded us, and which we will build upon. (2)
4. That Canada is well positioned to develop a leading-edge model for DRI structures and processes, building upon best international experience and practice. (2)
5. That there are significant existing expertise, resources, tools in the research ecosystem, much of which has emerged through bottom-up collaboration and connections, and that renewed strategic directions should leverage these capabilities and not interrupt the great momentum already built up. (1)
6. That the NDSF needs to be desirable for researchers, efficient for funders, beneficial for research outcomes, and via many motivators.

Resolve

1. To ensure continuous dialogue among the diverse community of scientists, researchers, funders, users and service providers to build trust, understanding and consensus for the Canadian NDSF. (8)
2. To harness the current expertise, services, and infrastructure of research libraries to develop sustainable and enduring RDM ecosystem in Canada. (7)
3. To move forward expeditiously to establish an integrated “federal” DRI ecosystem that sustains, enhances, and advances Canada’s capacity and impact in the global research and innovation community. (6)
4. To protect the essential infrastructure, people and services at the core of responsible RDM: (4)
    a. Physical infrastructure (e.g. data centres, networks);
    b. Interface tools and users;
    c. Security;
    d. Persistence of services.
5. That the NDSF strategy address gaps in awareness of the importance of RDM best practices through better communication of RDM initiatives, strategy, roadmaps, and enhanced engagement of the research community and the general public, and through enhanced metrics embedded within the research lifecycle to better measure the activities and impacts of RDM and research in general. (2)
6. To work together to clearly define needs, priorities and common objectives, and to align activities and resources to further develop existing initiatives and/or implement new activities to achieve the objectives. (1)

Call

1. Researchers to join as partners including the necessary awareness and training to ensure an informed direction forward. (7)

2. For participation from a diverse group of stakeholders, including researchers and community members in strengthening RDM in Canada. (5)

Invite

1. Trusted, established initiatives to join NDSF as anchors, while allowing new initiatives to grow into joining the framework. (3)

Commit

1. To advancing open science through the establishment of a national data repository network, commitment to FAIR Principles, and long-term data stewardship. (15)
2. To providing data management expertise to ensure the proper support/curation/stewardship services are available. (11)
3. To continuing to establish Canada as a leader internationally, working collaboratively with the international community in further development of the NDSF. (6)
4. To the development of a data management framework for all of Canada that is coordinated at all levels and is sustainable. (4)
5. To preserving scientific data in a long-term manner. The data should remain unaltered and free from manipulation of external entities/governments. (3)
6. To providing the full text of RDM resources (ie. infrastructure and HR) required by researchers. (1)