



Office of the Chief
Science Advisor of Canada

Bureau du conseiller
scientifique en chef du Canada

Canada



Towards a National Scientific
DATA GOVERNANCE FRAMEWORK

A Report of the Chief Science Advisor of Canada

© His Majesty the King in Right of Canada, as represented by
the Chief Science Advisor of Canada (2025)

Cette publication est aussi disponible en français sous le titre : Vers un cadre national de
gouvernance des données scientifiques

Contact:
data.donnees@ocsa-bcsc.gc.ca

Office of the Chief Science Advisor of Canada
235 Queen Street
Ottawa, Ontario
K1A 0H5 Canada

Cat. no.: lu37-58/2025E-PDF
ISBN: 978-0-660-78294-2

Table of Contents

| | |
|--|-----------|
| A Message from the Chief Science Advisor | 1 |
| Summary of Recommendations | 2 |
| Introduction | 3 |
| Desired Outcome..... | 3 |
| Understanding the Context..... | 4 |
| Principles | 5 |
| A Plan Towards Implementation of a National Scientific Data Governance Framework..... | 6 |
| Recommendations | 6 |
| 1. Establish a Canadian Focal Point | 6 |
| 2. Promote Data Domain Leadership | 7 |
| 3. Adopt Shared Standards and Tools..... | 8 |
| 4. Enhance Security Measures..... | 9 |
| 5. Grow Workforce Capacity and Digital Skills..... | 10 |
| Conclusion | 11 |

A Message from the Chief Science Advisor

Science, technology and innovation have for decades served as a pillar of national security and an engine of economic prosperity. Over the last 20 years, digital technologies have transformed – and continue to transform – those contributions.

The practice of science itself is changed by advanced data generation and analysis tools that enhance collaboration while opening new avenues for research and discovery. Increased access to advanced infrastructure, from instrumentation to large datasets, and computational tools such as artificial intelligence, are changing the pace and conduct of research and allowing innovators and entrepreneurs to accelerate the translation of discovery into socio-economic benefits.

Put plainly: scientific data is more valuable than ever. To maximize its benefits, Canada requires stewardship that ensures data integrity and interoperability and facilitates data reuse within a clear framework.

It is in this context that I convened a group of experts specializing in data governance and management across diverse organizational contexts to inform the development of a scientific data governance framework for Canada. The recommendations of this report reflect their thoughtful deliberations, taking into account the Canadian context and international best practices.

Special thanks are extended to the members of the Advisory Committee, including Mark Daley, Monica Granados, Natalie Harrower, Kevin Kasa, Mark Leggott, Kim McGrail, Benoit Pirenne, Eric Rancourt, Sujeevan Ratnasingham, Kathleen Shearer, Reda Tafirout, Bo Wandschneider, Peter Wilenius, and Lee Wilson. We also acknowledge Amy Buckland and David Castle for their leadership as co-chairs of the Committee, whose work informed this report.

Scientific data governance warrants immediate and serious attention as it is essential for the success of Canada's research initiatives, for supporting innovation and for reaping the full benefits of public investments for sovereignty and economic prosperity.

Mona Nemer, C.M., C.Q., FRSC
Chief Science Advisor of Canada

Summary of Recommendations

Scientific data is a national asset that needs an effective governance framework to maximize its benefits to the research and innovation ecosystem.

A five-point plan is suggested to reach this objective:

1. **Establish a Canadian Focal Point:** Designate an organization to federate actors and foster coherence across scientific data domains and related organizations, ensuring a unified approach to data governance in Canada.



2. **Promote Data Domain Leadership:** Encourage leadership within data domains to mobilize key stakeholders, promote harmonization and synergy within each scientific data domain and achieve benefits from resource-sharing.
3. **Adopt Shared Standards and Tools:** Promote common standards, including standard operating procedures, and tools within and across data domains to advance data stewardship, improve interoperability and support data reuse and AI-enabled analytics.
4. **Enhance Security Measures:** Ensure that data is treated as a national asset by implementing robust security protocols to protect data integrity, safeguard privacy of sensitive information and mitigate risks of data manipulation and theft.
5. **Grow Workforce Capacity and Digital Skills:** Develop a skilled workforce to design and operate the digital systems that underpin science activities, and promote data literacy and analytics among scientific data producers and users.

The report addresses principles and sets out a detailed actionable plan for implementing a national scientific data governance framework.

Introduction

The rapid digitalization of science and the rise of artificial intelligence has led to a significant increase in data production and utilization, requiring scientists and their institutions to address key data management issues. These include safe and secure data storage and use to ensure data integrity as well as access rights and conditions to promote ethical and secure use of data. Further considerations as to data life cycle management such as appropriate safekeeping duration and disposal of data, require harmonized approaches to support multidisciplinary and cross-sector collaboration.

Data domains exist in a context of data production and reuse where common infrastructure and governance frameworks facilitate data pooling, access and sharing. For example, ocean data is produced, collected, shared, and used or reused by a multitude of stakeholders for diverse purposes. They include the study of climate change, marine resource management, planning of transportation routes, ship design and emergency response. Initiatives such as the Ocean Research in Canada Alliance (ORCA) and the Canadian Integrated Ocean Observing System (CIOOS) provide domain leadership for the benefit of end users. Data domains can be based on disciplines, and are most valuable when they build bridges between disciplines in support of national priorities. The development of data domains can be nurtured in strategic fields such as health, agriculture, energy and transportation.

The rise of open science practices introduced specific tools, such as open licenses, and established principles, including FAIR (Findable, Accessible, Interoperable and Reusable), CARE (Collective benefit, Authority to control, Responsibility and Ethics) and TRUST (Transparency, Responsibility, User focus, Sustainability and Technology) to guide data governance and management. In Canada, the introduction of the Roadmap for Open

Science in 2020 has advanced public access to scientific publications resulting from federally supported research. The Roadmap did not yield as visible progress on data. These factors, among others, make it imperative to address scientific data governance.

Maximizing benefits from scientific data at a national level is a complex task that requires coordinated decisions by multiple people and organizations to simultaneously support various goals such as scientific collaboration and protection of sensitive information. The complexity is further compounded by the diversity of data types that have to be considered, such as monitoring data, research data and machine-produced data. This underscores the need for strong scientific data governance frameworks to ensure coherent decision-making for collective benefits.

Data governance frameworks define key roles and responsibilities while providing guidelines to ensure data is accurately managed, stored, shared and protected, maximizing its utility across various sectors. This report emphasizes scientific data governance as a foundational step towards improving current and future Science, Technology and Innovation (STI) performance in Canada, treating data as a national asset that contributes to Canadian prosperity.

Desired Outcome

Canada must adopt a national scientific data governance framework to protect and effectively manage data arising from publicly funded scientific activities and promote its reuse as a cornerstone of a strategy aimed at enhancing discovery and accelerating innovation for economic prosperity. This framework should align with international best practices and adhere to FAIR, CARE and TRUST principles.

Understanding the Context

Digitalization of science and proper governance of scientific data are pivotal to the advancement and well-being of modern societies. As intangible assets, data generated from scientific endeavours significantly contribute to the knowledge economy, influencing everything from policy development to technological innovation. Over the past 15 years, the exponential growth in data and the rise of Artificial Intelligence (AI) have created an urgent need for robust data governance frameworks. Such frameworks ensure data is accurately managed, stored, shared and protected, thereby maximizing its utility across various sectors.

Scientific data refers to three broad categories, each with distinct characteristics that impact their management:

- **Real World Data**, or RWD, comes from sources like clinical care and environmental monitoring. It is messy, and often raises questions of privacy or ownership. Properly managed, however, it can be very valuable on many fronts, from enhancing public services to generating new knowledge and products.
- **Research Data** is produced in controlled environments in the context of specific projects and domains, making it easier to organize and manage. Research data often holds valuable information, beyond its original stated objectives, so facilitating its broader reuse is essential.

- **Sensor or Machine-Generated Data** originates from devices and lab instruments, requiring specialized (often large) infrastructure and international data management regulations.

Concrete examples demonstrate how effective data governance can transform scientific practices. The proper management of health data can enhance patient care and enable rapid responses to public health crises. In the realm of environmental science, well-governed data allows for better weather forecasting and ocean monitoring, enhancing transportation safety while protecting biodiversity and livelihoods. Moreover, inter-domain collaboration, facilitated by shared data standards, can foster social innovation and accelerate technology adoption in many sectors, such as resource management, health care and manufacturing. These examples illustrate how a cohesive approach to data governance can bolster scientific research and accelerate its applications to societal benefit.

The current state of scientific data governance in Canada is marked by organizational fragmentation and inconsistent practices leading to suboptimal data management. The absence of shared standards and tools across domains hinders interoperability and affects data quality. Additionally, security measures at the system level are often insufficient, exposing sensitive information and jeopardizing data integrity, privacy or national security. The recommendations in this report aim to address these gaps by proposing a framework that promotes structural coherence, common standards and improved security. By doing so, the framework seeks to enhance the overall quality of scientific data governance in Canada, thereby effectively supporting research, innovation and the economy.

Principles

In this context, four simple principles should guide actions towards improving scientific data governance in Canada.

- **Reuse, Recycle, Reimagine:** More value will be created from scientific data if data reuse increases within and across scientific domains, including via AI, and through data value chains.
- **Cluster for Success:** Benefits from sharing resources (e.g. infrastructure) and expertise will be maximized if data governance is organized around data domains that attain a critical mass.
- **Lead the Way:** Improvement of practices within and across domains will result from empowering domain-specific leaders to champion common infrastructure and standards.
- **Row in the Same Direction:** Aligning incentives from all science funding organizations and regulators will accelerate the adoption of harmonized scientific data governance.



A Plan Towards Implementation of a National Scientific Data Governance Framework

Current critical gaps in Canadian scientific data governance include organizational and practice fragmentation as well as insufficient security measures. This in turn impacts data quality and accessibility, limiting its use and reuse in an era where access to data holds unlimited potential for innovation and prosperity. These gaps are addressed by the following five recommendations: Establish a Canadian Focal Point, Promote data domain leadership, Develop shared standards and tools, Enhance security measures, Grow workforce capacity and digital skills. Once implemented, these recommendations will increase coherence, synergy and data interoperability, paving the way for a national scientific data governance framework to effectively manage scientific data as the national asset it is.



Recommendations

1. Establish a Canadian Focal Point

Designate an organization to federate actors and foster coherence across scientific data domains and related organizations, ensuring a unified approach to data governance in Canada.

A Focal Point is an organization designated to build national consensus and foster a shared vision across the multiple scientific data domains and between interested organizations such as funders and infrastructure and service providers. In addition to its national function, a Focal Point serves as an international contact point to liaise with other data organizations and provide a forum for cross-domain alignment on standards and best practices.

1.1. Launch the Focal Point with an Initial Set of Pilot Data Domains

The Focal Point could be initiated under the auspices of the Digital Research Alliance of Canada, in collaboration with the Office of the Chief Science Advisor of Canada.

Work could commence using a subset of domains that span all three categories of data. Given their current data domain maturity and their complementarity in data types, the initial Focal Point could include, for instance, ocean data (Canadian Integrated Ocean Observing System (CIOOS)), Arctic data (Canadian Consortium for Arctic Data Interoperability (CCADI)), and social statistics (Canadian Research Data Centre Network (CRDCN)).

Launching the Focal Point with these pilot scientific data domains will help validate the framework, identify areas for improvement and set precedents for future expansions, ensuring a robust implementation.

1.2. Establish a Steering Committee and Identify and Engage Additional Stakeholders

Formation of a Focal Point Steering Committee to ensure light and agile collaborative governance is crucial to the Focal Point's long-term success. This committee may start engaging beyond the initial set of pilot data domains and determine how additional data domains and stakeholders will be brought into the Focal Point ecosystem in support of national priorities.

Collectively, Focal Point members might consider accompanying domains such as health data, biodiversity data and cultural heritage data to foster their consolidation as data domains so that they can rapidly join the Focal Point. Proactive effort should be invested in developing domains that are seen as highly strategic for advancing the notion of data as a national asset for Canada.

1.3. Develop a Scalable and Sustainable Model for the Long-Term Governance of Federally Supported Scientific Data

A formal proposal for the Focal Point's structure and objectives should be developed that will foster alignment with national goals for scientific data governance, contribute to Canada's science and innovation efforts and by extension deliver value to all Canadians.

Securing financial and material resources will be vital for the Focal Point's operational success, enabling it to support overall scientific data domain maturation and scientific data governance effectively. Care should be taken to include an effective accountability mechanism for scientific data governance, which can increase transparency and stakeholder confidence while fostering performance and continuous improvement.

2. Promote Data Domain Leadership

Encourage leadership within data domains to mobilize key stakeholders, promote harmonization and synergy within each scientific data domain and achieve benefits from resource-sharing.

Data domains operate within the context where data is produced and reused and where communities aggregate or cluster the resources and expertise needed for the governance and management of data. Greater benefits are anticipated when critical mass can be attained and shared resources are used efficiently to address similar data domain circumstances and requirements.

Domains should lead the charge on the evolution of practices, working towards the adoption of standards that address practice fragmentation within each scientific data domain. Data domain leadership should help data-producer and user communities prioritize scarce resources and grow data governance maturity.

2.1. Identify Data Domains and Leaders within Each

Data domains that are strategic for Canada's STI performance, sovereignty and economic prosperity should be inventoried and characterized, and then leadership in each of those domains should be identified. Recognizing and supporting committed domain leaders is essential for driving the adoption of standards and good practices within their respective domains, fostering a culture of excellence in data governance. Domain leadership cannot be a peripheral activity.

Data domains bring together participants who share common data culture and practices, where in-group similarities make clustering meaningful. These domains organize around shared context and needs, for example, reliance on common research infrastructure, data access

rules or historical use of particular metadata standards or data ontologies. For each data domain to be sustainable, it must achieve a critical mass and be institutionalized in a way that can support formal governance. This can be achieved through data trusts or formalized strategies and plans to mobilize the community towards domain progress.

2.2. Develop Data Domain Leadership Tools Focused on Data Governance

Through the Focal Point, practical tools and metrics that inform and enable scientific data governance should be established. Common tools and metrics are essential for assessing the capacity and needs within and across data domains as well as for gauging investment performance and sector maturity. For instance, FAIR (Findable Accessible Interoperable and Reusable) data maturity models exist, such as Earth Science Information Partners (ESIP) for AI readiness and Research Data Framework (RDaF) 2.0 for research data management.



3. Adopt Shared Standards and Tools

Promote common standards, including standard operating procedures, and tools within and across data domains to advance data stewardship, improve interoperability and support data reuse and AI-enabled analytics.

Relatively independent decisions are currently being made about scientific data management through its lifecycle by numerous actors across sectors. This excessive diversity of practices poses significant challenges to data quality, curation and access that limits interoperability and effective data reuse. Adoption of data governance principles shared across scientific data domains can encourage system coherence without curtailing the useful diversity of practices between domains.

3.1. Conduct a Review of Existing Standards and Tools within and across Data Domains

Within data domains, and horizontally through the Focal Point, the current standards established by funding and regulating entities as well as those developed by communities of practice should be examined to identify gaps and opportunities for harmonization. This analysis will serve as the foundation for creating a unified set of standards for scientific data interoperability, quality and security in Canada.

3.2. Promote Baseline Standards for Data Interoperability, Quality and Security

Baseline standards and standard operating procedures within and across data domains should be promoted to ensure consistent and high-quality data governance practices that foster seamless data sharing and integration. Toolkits and guidelines will facilitate the adoption of shared standards and make implementation straightforward and efficient

for all data domains, thus improving overall data governance.

Data stewardship standards are available. FAIR data practice and technical standards are increasingly capable of incorporating differences into FAIR Implementation Profiles (FIPs) that can be replicated. While most progress has been made in making quantitative data FAIR, there is also potential to make qualitative data FAIR with appropriate metadata and a focus on machine actionability and reusability.

There is also an opportunity to advance principles and protocols in support of Indigenous self-determination for First Nations, Inuit and Métis, such as the First Nations Principles of Ownership, Control, Access and Possession (OCAP), and with reference to policies such as the Tri-Agency Principles for Digital Data Management. International initiatives are also relevant, for example the recent Australian Framework for Governance of Indigenous Data, the CARE Principles for Indigenous Data Governance, and the Local Contexts initiative.

Furthermore, adoption of baseline standards would enable:

- Canadian interoperability in international research consortia (e.g. CERN, etc.) and other jurisdictions (e.g., compliance with open science commitments under Horizon Europe Pillar 2);
- International commitments (e.g., UNDRIP, UNESCO Recommendation on Open Science);
- Continuous improvement by following the leading global data organizations internationally (ESIP, EOSC, WDS, CODATA, RDA, etc.) and countries with leading national data governance initiatives (UK, FR, GR, AUS, Netherlands, etc.).



4. Enhance Security Measures

Ensure that data is treated as a national asset by implementing robust security protocols to protect data integrity, safeguard the privacy of sensitive information and mitigate risks of data manipulation and theft.

Data security measures are essential for safeguarding data integrity and protecting against the improper or unauthorized use of sensitive information. Digital infrastructure must also be evaluated in the context of data sovereignty and the long term safeguarding of data assets.

Regular monitoring and auditing of data infrastructure and operating practices should be conducted to ensure continuous improvement and adherence to security protocols, and thereby maintain data integrity and trust.

National security considerations require the federal government to work with an array of specialized system actors to achieve integrated coverage of threat vectors related to scientific data. These include the Canadian Security Intelligence Service (CSIS), Communications Security Establishment (CSE), Shared Services Canada, the Digital Research Alliance of Canada (DRAC) and CANARIE. In this respect, the Focal Point can facilitate the important work of these organizations.

5. Grow Workforce Capacity and Digital Skills

Develop a skilled workforce to design and operate the digital systems that underpin science activities, and promote data literacy and analytics among scientific data producers and users.

Effective digital systems are vital for boosting national science, technology and innovation performance and for enhancing productivity and the delivery of public services. Their seamless operation requires a specialized workforce, including data architects who connect people with the information they need, and data engineers who ensure data is available, secure and accessible for data scientists, analysts and other data users.

Making the most of scientific data requires a strong emphasis on data literacy — the ability to read, write, communicate and reason with data. It also includes understanding the implications of data, how to analyze it properly, interpret it correctly, and how to use the findings effectively. Cultivating these essential skills within the scientific community is crucial.

5.1. Target Efforts Towards Developing Skilled Workforce Capacity in High-Need Areas

Given the rapid digitalization of science, a shortage of skilled workers is likely to persist. To address this challenge, periodic assessments through the Focal Point should be conducted, evaluating the available specialized workforce and the evolving needs within each domain across Canada. This information can help domain leaders identify areas where workforce development is most urgently needed.

Where appropriate and desired, opportunities should be created for scientists to transition into specialized roles such as data architects or engineers through additional training, facilitating reskilling and increasing workforce capacity.

5.2. Encourage Data Literacy within the Broad Science Community

Digital technologies and platforms are transforming the practice of science, necessitating new skills for science producers and users. In order to address this evolving landscape, support for training in data science and analytics should be provided through the Focal Point with participation from the domain leadership in collaboration with learning institutions and centres of expertise.

Additionally, the concept of scientific data as a national asset should be integrated into the science curriculum and workplace orientation programs for fostering a comprehensive understanding of its significance. This integration should emphasize potential contribution of scientific data, connections between various data-intensive sciences, technology development, innovation and the real economy.

Conclusion

The digitalization of science and everyday life has unlocked unprecedented advances in data collection, analysis, and utilization. Artificial intelligence and machine learning now allow for rapid data processing, driving innovation across all sectors. As a result, data has become a valuable intangible asset, with its true worth realized through reuse, supported by shared infrastructure and harmonized governance frameworks.

Within this context, data governance and management have evolved into vital elements of the scientific value chain, enabling new technologies and discoveries. Canada, already recognized for its leadership in artificial intelligence research, has the opportunity to set a global standard in data practices by adopting a national scientific data governance framework.

Given data's crucial role in innovation and national security, the Office of the Chief Science Advisor will monitor national progress and international trends. The recommendations in this report aim to strengthen leadership and interoperability among Canada's data domains, promote standardized tools and enhanced security, and boost data literacy throughout the scientific community.