

**Canada Foundation for Innovation  
(CFI)**

**Major Science Initiatives Fund (MSIF)  
Data Working Group**

**Report**

**February 5<sup>th</sup>, 2025**

# MSIF Data Working Group: Purpose & Objective

- **Supporting Large-Scale Research Facilities**
  - Recognition that MSIF-funded facilities manage diverse and complex data across disciplines.
  - Data stewardship is critical and linked to specialized personnel recruitment.
- **Building a Strong Data Management Community**
  - CFI aims to foster knowledge sharing and best practices in data lifecycle management.
  - The WG provides strategic guidance on data governance and infrastructure.
- **Developing a Data Lifecycle Model**
  - Identify effective data management practices for large facilities.
  - Define & understand the role of service providers like the Digital Research Alliance of Canada.
- **Enhancing International Collaboration**
  - Address barriers to global data sharing.
  - Align with international data stewardship standards.

# MSIF Data Working Group Mandate

## Advise the CFI on:

- MSIF facilities' data management in relation to their user communities.
- Supporting a data lifecycle model for large facilities.
- Implementation strategies for the data lifecycle model.
- Current and future roles of service providers like the Digital Research Alliance of Canada.
- Enhancing international collaboration by removing barriers to data sharing.

## Engagement & Reporting:

- Work with partners and service providers to understand their role and address MSIF challenges.
- Submit an annual report with key challenges
- Keep the MSIF community informed through workshops and other channels.

# Working Group Members



**Naveed Aziz**

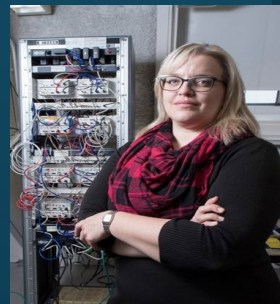
**(Chair)**

*Vice President, Research & Innovation  
Genome Canada*



**Natalie Harrower**

*Executive Director, Canadian  
Research Data Centre Network*



**Kathryn McWilliams**

*Director  
SuperDARN Canada*



**Tanja Niemann**

*Executive Director  
Consortium Érudit (Coalition  
Publica)*



**Lam Pho**

*Chief Information Officer  
Canadian Cancer Trials Group*



**Benoît Pirenne**

*Director, User Engagemen,  
Ocean Networks Canada*



**Sujeevan Ratnasingham**

*Director, Informatics  
Centre for Biodiversity  
Genomics*



**Richard Wintle**

*Associate Director,  
TCAG / CGEn Toronto*

# Information flow

Working Group  
Discussions

MSIF  
Survey

# Needs Identified by Working Group for MSIF Facilities



Implication of  
Data management  
on level of services



Data  
Infrastructure  
and Platforms



Data  
Management and  
Sharing



National  
Coordination &  
Alignment



User Training &  
Open Science  
Standards



Licensing &  
Economies of  
Scale

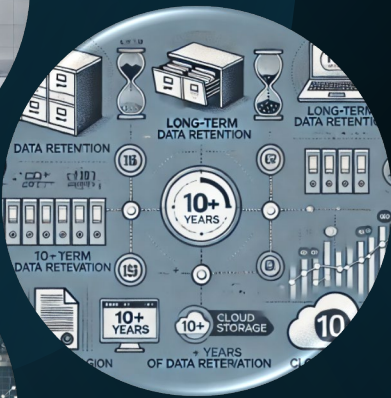


Data Sovereignty  
& National  
Capacity

# MSIF Data Needs Assessment Survey

<b>Level of Service</b>	Corporate-level activities and support	<b>National Coordination and Strategic Alignment</b>	Increased nationwide coordination	
	Guaranteed level of service or special/priority allocation		Data capabilities for specific domains with Canadian focal point	
	Mechanism to apply for data storage as a facility		High-level principles around research data management	
	Service-level agreements		Alignment of the Alliance and MSIF application process	
<b>Data Infrastructure and Platforms</b>	Appropriate data storage platform	<b>User Training and Open Science Standards</b>	Users training	
	On-prem infrastructure		Funder mandates	
	Centralized repository	<b>Licensing and Economies of Scale</b>	Central procurement of licenses and software	
	Commercial lakehouse infrastructure		<b>Data Sovereignty and National Capacity</b>	Sufficient data storage capacity within Canada
	Cybersecurity			
	Data backup services			
<b>Data Management and Sharing</b>	Data curation personnel			
	Back-end reports for data usage			
	Data pipeline			
	Continued provision of FRDR			
	Data Access Committee			

# MSIF Data Working Group: Key Findings



## Critical Needs Identified

- Secure storage, on-prem infrastructure, backup, and cybersecurity.
- Advanced analytics and AI capabilities for large datasets.
- Long-term data retention and archival digitization.

## Challenges Highlighted

- Unmet data lifecycle needs across most facilities.
- Clarity on role of DRAC and CANARIE in terms of accessibility.
- Increasing data volumes outpacing current capacity.

## Emerging Priorities

- Flexible, facility-level service agreements.
- Improved disaster recovery and cybersecurity support.
- Enhanced comms and coordination for national data strategy.



# Next steps...

## Build

- Sustainable
- Scalable and
- Secure data solutions

## Coordinate

- Support multi-scale MSIFs
- Role of CFI and partners in supporting a sustainable data ecosystem

Thank you